7[1]

# Intuitions and the Theory of Reference

Jennifer Nado and Michael Johnson[2]

Do philosophers rely on intuitions as their primary evidence for philosophical theories? Many philosophers have long simply assumed that they do. Some have gone even further and asserted that intuitions are the data of philosophy and that the goal of a philosophical theory is to "capture" our intuitions. Recently, however, this picture of philosophy has come under question. Authors like Timothy Williamson, Max Deutsch, and Herman Cappelen have challenged the thesis that Cappelen calls "Centrality": the thesis that philosophical evidence consists chiefly of intuitions. Further, these philosophers have suggested that the falsity of Centrality undermines the primary project of experimental philosophy. Experimental philosophy, according to experimental philosophers themselves, mainly involves conducting surveys of folk intuitions; if intuitions are not a source of evidence in philosophy, or only a very peripheral source, then it wouldn't appear very fruitful to spend lots of resources empirically studying them.

Although Williamson, Deutsch, and Cappelen have set their sights on philosophical methodology as a whole, other philosophers have leveled strikingly similar challenges at one particular experimental study. In that study, Edouard Machery, Ron Mallon, Shaun Nichols, and Stephen Stich (hereafter, MMNS) uncovered cross-cultural differences in responses to one thought experiment drawn from Saul Kripke's *Naming and Necessity*, and suggested that these differences undermined Kripke's methodology and conclusions. Subsequently, the relevance of MMNS' finding to Kripke's project has been vigorously challenged by Michael Devitt, Genoveva Martí, and others, on the grounds that MMNS had over-exaggerated the role intuitions about reference play in Kripke's arguments and in theories of reference in general.

Interestingly, points of contact between these two debates have been few.[3] At first glance, one might expect that the arguments presented by Devitt, Martí, and others, could be taken by fans of the "anti-Centrality" position to be just more fuel for their fire. But, as we shall argue, the relationship between the two debates is more complex. Philosophy is a diverse field, and what is true for the theory of reference may not be true for other lines of philosophical inquiry. Indeed, what is true for Kripke's particular theory of reference and the arguments he gives for it may not be true more generally for other theories of reference. Thus we think the question raised by Williamson, Deutsch, and Cappelen—whether philosophers rely centrally on intuitions as evidence—is unlikely to have an unequivocal answer across all philosophical disciplines.

In this paper, we will examine the role that intuitions and responses to thought experiments play in confirming or disconfirming theories of reference, using insights from both debates as our starting point. Our view is that experimental evidence of the type elicited by MMNS does play a central role in the construction of theories of reference. This, however, is not because such theory construction is accurately characterized by "the method of cases." First, experimental philosophy does not directly collect data about intuitions, but rather about people's responses to thought experiments, which may reflect their intuitions but may well not. Second, unusually in the case of the theory of reference, experimental prompts involve elicitation of the phenomenon under investigation—that is, *referring*—and it is the reference facts rather than the intuitions that a theory of reference should capture. Finally, "best fit" models like the method of cases are inconsistent with the actual practice of semantic theorists, who appeal to general principles of theory construction (e.g. beauty and simplicity) as well as considerations native to the theory of reference (such as Grice's razor), often in the service of rejecting intuitions. Indeed, Kripke himself viewed several claims of his theory as unintuitive, but felt no pressure to alter them on that account.

These facts, we'll argue, suggest that the relevance of experimental methods in the theory of reference cannot be straightforwardly extended to other fields: an experimental prompt containing Newton's two rocks tied together in empty space may elicit referring, and may also elicit intuitions, but it doesn't also elicit the phenomenon under investigation (rocks rotating in empty space). Furthermore, the considerations native to the theory of reference are not part of physics, which has its own distinct set of principles and considerations. This suggests that an evaluation of the centrality of intuitions/ responses to thought experiments in philosophical methodology must proceed in a piecemeal fashion. We conclude that the potential contributions of experimental philosophy will take different forms in different sub-fields of philosophical inquiry.

## 1. Intuitions in philosophy: Two debates

As we mentioned, there are two separate challenges that have been leveled at experimental philosophy's focus on intuition. The broader challenge, which, following Cappelen, we'll call the "anti-Centrality" challenge, questions the role of intuition in philosophy as a whole.[4] The narrower challenge has focused on the role of intuitions about reference in Kripke's arguments against descriptivist theories, and the role of intuitions in confirming or disconfirming theories of reference more generally. While the two challenges are superficially similar, they differ in the specific arguments they advance. We start first with a discussion of the broader challenge.

A prominent theme in Cappelen's and Williamson's anti-Centrality arguments—though not Deutsch's—is a general skepticism regarding the very concept of an intuition. For Cappelen, philosophers' usage of the term "intuition" fails to reflect a clear, commonly agreed-upon definition. Some philosophers insist that intuitions must arise from conceptual competence; others do not. Some insist that intuition has a distinctive phenomenology; others do not. And so on. Similarly, Williamson notes that the label "intuitive" can be applied to nearly any claim—even ones that are straightforwardly perceptual. There is simply no clear dividing line between philosophical intuitions and such everyday judgments as mundane cases of concept application. Williamson goes so far as to claim that "philosophers might be better off not using the word 'intuition' and its cognates" (Williamson 2007: 200); Cappelen would presumably agree.[5]

Of course, if basically everything counts as an "intuition" or if "intuition" fails to pick out a clear referent, it becomes difficult even to make sense of the claim that philosophy centrally relies on intuitions as evidence. This cuts both ways, however: it becomes difficult to make sense of anti-Centrality claims as well. If there aren't any intuitions, because the notion is ill-defined, then one doesn't really need a case study to show that philosophers don't appeal to them.

According to the anti-Centrality camp, how then do philosophers arrive at evidence for or against philosophical theories, if not by way of an evidential appeal to intuition? The answer given is multi-faceted. Both Deutsch and Williamson, for instance, reject the view that our evidence consists of intuition facts like "I intuit that P" in favor of the view that our evidence consists of facts like p itself. In Williamson's terms, the idea that philosophical evidence consists of intuition-facts involves an illegitimate "psychologization" of our evidence: it puts us in the undesirable position of having to argue from psychological premises to conclusions about the non-psychological (metaphysics, epistemology, etc.). Williamson holds that such a gap is "not easily bridged" (Williamson 2007: 211). However, denying that it's appropriate to psychologize our evidence is compatible with an evidential role for intuition. Compare vision: if we assume that when we see that P, our evidence is p and not merely that we see that P, it's still true that the reason we're justified in believing p is that we'd seen it.

Cappelen claims that what may look like an appeal to "intuition" is often really just an appeal to what is in the common ground—that is, to a proposition that needs no further support in the current dialectical context. Such common ground claims may be based on any evidential source whatsoever, and they need not be *a priori*. Furthermore, Cappelen and Deutsch both point out that philosophers who employ thought experiments frequently engage in careful argumentation, rather than brute appeal to intuitions about the target case. Deutsch, for instance, notes that Gettier discusses the disconnect in his counterexamples between what makes the proposition true and what makes it believed. Cappelen, meanwhile, claims that close examination of paradigm thought experiments reveals no mention of reliance on states with the sorts of features associated with intuition, such as special phenomenology or a "rock-bottom" evidential status.

We agree with many of these points. Like Williamson and Cappelen, we have doubts about the category of "intuition." As one of us has previously argued (Nado 2014), the states covered by the term "intuition" aren't likely to form a natural kind at the psychological level: in different domains like moral judgment or mathematical judgment, distinct mechanisms with distinct assumptions and inferential procedures will produce the spontaneous, non-consciously-inferred judgments we call "intuitions." However, beyond the fact that such judgments are spontaneous and not consciously inferred, there will be almost nothing in general to say about them—for example, about how reliable they are or how robust they are in unusual environments. This is part of our reason for believing that the Centrality question—and the more general question of the viability of experimental philosophy—can have no simple answer. We shall return to this issue at a later point in the paper.

We also agree that, when (good) philosophers present their judgments regarding thought experiments, they also provide substantial argumentation to support those judgments. For us, however, this merely shows that the "method of cases" is at best a caricature of actual philosophical practice. Philosophers present an array of evidence for their claims. A particular philosopher may be justified in believing p because a particularly reliable mental mechanism caused her to spontaneously judge p in response to a thought experiment (without needing to argue from the latter fact to the former). She might present the thought experiment and her judgment, expecting her audience to perhaps spontaneously judge as she does. It doesn't follow that we expect philosophers to explicitly appeal to intuitions in arguing for P. First, if they expect their readers to have the same intuition, appealing to their intuitions is redundant. Compare: if A and B both see that P, A doesn't need to tell B that A sees that p in order to get B to believe P. Second, if they expect their readers to have different intuitions, independent argumentation is necessary. Thus, we find it no wonder that Cappelen's case studies uncovered that, in a number of classic thought experiments, a judgment was presented, and then independent argumentation was given for that judgment. That's how one covers one's bases. It doesn't show that intuitions aren't evidence in philosophy.

Let's turn now to the narrower of the two debates, which focuses on MMNS's cross-cultural findings on subjects' responses to thought experiments concerning reference. MMNS's study employed two well-known thought experiments used by Kripke in his arguments against the descriptivist theory of reference—the "Gödel" case and the "Jonah" case. It will be helpful to have

9781474223218_txt_print.indd 129 28/09/2015 13:27

Kripke's versions of these thought experiments fresh in mind, so we'll reprint them here:

> Suppose that Gödel was not in fact the author of [Gödel's incompleteness theorem for arithmetic]. A man called "Schmidt," whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the view in question, then, when our ordinary man uses the name "Gödel," he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description "the man who discovered the incompleteness of arithmetic"… when we talk about "Gödel," we are in fact always referring to Schmidt. But it seems to me that we are not. We simply are not. (Kripke 1980: 83–4)

> Suppose that someone says that no prophet ever was swallowed by a big fish or a whale. Does it follow, on that basis, that Jonah did not exist? There still seems to be the question whether the Biblical account is a legendary account of no person or a legendary account built on a real person. In the latter case, it's only natural to say that, though Jonah did exist, no one did the things commonly related to him. (Kripke 1980: 67)

MMNS presented simplified versions of these cases to subjects in the United States and in Hong Kong (the prompts were given in English in both groups). Their Gödel prompt, for instance, describes a man named John who knows of Gödel only that he proved incompleteness; the prompt then tells of Gödel's "deception," and asks subjects to judge who John refers to when he uses the name "Gödel." They found that there was a statistically significant difference in the responses between the US subjects and the Hong Kong subjects, and that the Hong Kong subjects were more likely to report the more "descriptivist" judgment that John refers to Schmidt. They found no significant difference in the groups' responses to the Jonah case.

It's worth remarking at the outset that while MMNS claim that these findings show that there are cross-cultural differences in *intuitions* regarding the Gödel case, what they really reveal is a cross-cultural difference in the *judgments subjects report* about the thought experiment. Even on the thinnest conception of intuitions, where intuitions are spontaneous judgments that are not the conclusions of conscious reasoning, it's questionable whether MMNS received many intuition-reports. Their subjects were university students in an academic setting. Students have a proclivity to spend at least some non-zero amount of time consciously reasoning about which answers they are to give in an academic setting.

Be that as it may, MMNS *do* take their results to show a cross-cultural difference in intuitions. Furthermore, in the paper where these findings were reported, MMNS make the following claims about the methodology underlying Kripke's work (and theories of reference generally):

> There is widespread agreement among philosophers on the methodology for developing an adequate theory of reference. The project is to construct theories of reference that are consistent with our intuitions about the correct application of terms in fictional (and non-fictional) situations. Indeed, Kripke's masterstroke was to propose some cases that elicited widely shared intuitions that were inconsistent with traditional descriptivist theories. … Even contemporary descriptivists allow that these intuitions have falsified traditional forms of descriptivism. (Machery et al. 2004: B3)

In a later paper, MMNS clarify that the methodology they have in mind is what is often known as the "method of cases." There, they characterize the method as follows: "The correct theory of reference for a class of terms T is the theory which is best supported by the intuitions competent users of T have about the reference of members of T across actual and possible cases" (Mallon et al. 2009: 338).[6]

If we assume the method of cases (as characterized above) is the method philosophers employ for evaluating theories of reference, and we assume that MMNS have found a cross-cultural difference in intuitions about a particular possible case, then it's fairly clear how those findings have problematic consequences for that project. If different demographic groups differ in their intuitions, whose intuitions are to be captured by the theory of reference? What reason could possibly be given to think that Western subjects, say, have intuitions that track the true theory of reference, whereas East Asian subjects, on the other hand, do not? It's possible, of course, to take the line that, in Hong Kong, names are used descriptively, whereas in the United States, they are not. But even a substantial minority of the Western subjects reported descriptivist judgments. Surely students from the same culture, on the same campus, in the same classroom, speaking the same language, aren't governed by different theories of reference.

MMNS's study has been challenged on several grounds (see for instance Ludwig 2007; Deutsch 2009; Lam 2010), but our particular focus will be on challenges that have been made to MMNS's implicit and explicit claims about the methods of research in the theory of reference. MMNS asked subjects to report which person, Gödel or Schmidt, a hypothetical speaker, John, would be "talking about" when he used the name "Gödel." Genoveva Martí (2009) claims that the intuitions subjects have about this question reflect only their own *theories* about the reference of proper names, but not how they in fact *use* proper names. Intuitions about reference, such as those in response to MMNS's version of the Gödel case, provide little evidence for or against theories of reference, according to Martí: subjects might easily hold mistaken theories that don't accurately reflect their own usage. Nonetheless, Martí does think that some types of intuitions provide evidence about actual usage, and may therefore provide evidence for theories of reference. Instead of asking subjects what a name refers to, or what some person is talking about when they use a name, Martí claims that we should elicit their reactions to dialogues in which a name is used in certain ways. Here is her example:

> In order to determine whether users of names in the two experimental groups use or don't use 'Gödel' according to what is predicted by the causal-historical picture, it would be best if the end of the story, and the question asked, went along the following lines: One day, the fraud is exposed, and John exclaims: 'Today is a sad day: we have found out that Gödel was a thief and a liar'. What do you think about John's reaction? (Martí 2009: 47)

This line of questioning doesn't employ semantic terminology like "reference," and does not aim to produce judgments about reference. Yet it can be used to get at the referential facts, according to Martí, since a subject for whom names refer non-descriptively should report finding John's reaction deeply strange.

Meanwhile, Michael Devitt (2011), and Jonathan Ichikawa, Ishani Maitra, and Brian Weatherson (2012) argue that the Gödel and Jonah cases play a minimal role at best in Kripke's arguments against descriptivism. Both Michael Devitt and Ichikawa et al. note that Kripke targets two distinct views in *Naming and Necessity*.[7] Devitt distinguishes between descriptivism as a theory of meaning vs. as a theory of reference determination; Ichikawa et al. use "strong descriptivism" and "weak descriptivism" for the same distinction.

The strong "theory of meaning" version of descriptivism tells you what the meaning of a name is—it says that the meaning of a proper name is the same as the meaning of some description. The weaker "theory of reference" version of descriptivism tells you why a name refers to what it does—it says that for each name N, there is a description that speakers associate with N, and whatever uniquely satisfies that description, if anything, is the referent of N.

Kripke gives arguments against both varieties of descriptivism. He notes that, since the strong thesis says that the name "Aristotle" is synonymous with some description "the D," this predicts that "Aristotle might not have been the D" is synonymous with "the D might not have been the D." But Kripke argues that while the latter has a necessarily false reading, the former does not, at least for any of the ordinary descriptions purported to be synonymous with "Aristotle." For instance, it's certainly possible that Aristotle might not have been the teacher of Plato, or the last great philosopher of antiquity, or the man named "Aristotle." The two phrases cannot therefore be synonymous. Devitt calls this argument "Unwanted Necessity," and, according to him[8], it does appeal to an intuition—not an intuition about reference, but, rather, a modal intuition.

This modal claim does not, however, undermine the weaker version of descriptivism. Weak descriptivism says not that names are *synonymous* with descriptions but that descriptions *fix referents*: for each name N, there is a description that speakers associate with N, and whatever uniquely satisfies that description, if anything, is the referent of N. Against this view, Kripke presents an argument largely based on a variety of actual cases. First, he argues that often speakers lack enough information to uniquely identify individuals that the names they use nevertheless refer to. For example, speakers who know of Feynman only that he was a famous physicist still succeed in referring to Feynman when they use the name "Feynman;" speakers who know of Cicero only that he was a famous orator still succeed in referring to Cicero when they use the name "Cicero." But according to weak descriptivism, Feynman and Cicero cannot be the referents of these names for these speakers, since neither Feynman nor Cicero uniquely satisfies the descriptions the speakers associate with "Feynman" and "Cicero," respectively. Devitt calls this the "Argument from Ignorance."

Moreover, Kripke argues, many speakers associate descriptions with names that are false of those names' referents. Some people only "know" of Columbus that he was the first European to discover the Americas; others only "know" of Einstein that he invented the atomic bomb. Weak descriptivism predicts that the former speak of Leif Ericson when they say things like "Columbus was European." Kripke is right to point out that, nevertheless, they are speaking of Columbus. Devitt calls this the "Argument from Error."

In addition to the actual cases Kripke employs in service of the Ignorance and Error arguments, he also presents several non-actual cases—including the Gödel case. *Contra* MMNS, however, Devitt argues that the Gödel case is not central to Kripke's arguments against descriptivism, for the evidence it provides is fairly weak when compared to the evidence provided by his other cases. The Gödel case is an unusual fictional scenario; Devitt claims that our intuitions about unusual fictional scenarios are generally not as strong as they are on actual cases, and that we have more reason to expect error. This is especially true of the intuitions of the folk, who may not be used to thinking about such cases. Devitt holds that the experimental method is appropriate, but that it is better to conduct surveys about "humdrum" cases, not intuitions on cases that are "counterfactual, hypothetical, or fictional" (p. 421). Here, Devitt emphasizes that in the actual-world case MMNS surveyed (the Jonah case), MMNS found no cross-cultural difference in subjects' reported judgments. Devitt's considered view is that, on balance, the evidence we have still supports Kripke's rejection of descriptivism.

Ichikawa et al. (2012) are largely in agreement with Devitt that the Gödel case plays a minimal role in Kripke's overall argument against descriptivism, but for a different reason. Devitt expresses skepticism about the evidential value of non-expert intuitions on the Gödel case, due to its unfamiliar, counterfactual character. Ichikawa et al., however, see no reason to doubt even naïve intuitions on the case. Instead, they take the Gödel case to play a minimal role simply because its primary use is to argue against an ultra-weak version of descriptivism according to which descriptions determine reference only in cases where the speaker possesses an individuating description. They argue that, since it is largely agreed that certain names like "Jack the Ripper" do function descriptively, then if the intuitions go against Kripke on the Gödel case, "that would at most show that there are a few more descriptive names than we thought there were" (Ichikawa et al. 2012: 6).

Note that Martí, Ichikawa et al., and Devitt—unlike Cappelen and Williamson—are more than happy making use of the notion of an "intuition."

9781474223218_txt_print.indd 134 28/09/2015 13:27

Indeed they each appear to accept that intuitions play an evidential role in the project of theorizing about reference. Martí writes that the primary flaw in MMNS's study is not that they tested intuitions, but that they "did not test the right kind of intuitions" (2009: 44). Ichikawa et al. claim that "most of our intuitions in this field are surely correct" (2012: 7) And Devitt goes so far as to write that "Machery et al. are surely right in claiming that theories of reference are assessed by consulting intuitions: this practice does seem to be the method of semantics" (2011: 419). Elsewhere, Devitt even expresses bewilderment at how "strangely critical" Deutsch is of the standard characterization of philosophical methodology (2011b: fn. 5).[9] It might seem then that these authors are rather pro-Centrality[10], despite their view that the particular results of MMNS's surveys are of little evidential significance. However, when we look a little closer, it turns out that not all the participants in the two debates are agreed on what counts as an intuition, and thus what the method of cases is, or what Centrality really amounts to. In the end, there is more agreement between the participants in both debates than initially appears on the surface.

## 2. Centrality and theories of reference

Centrality is a slippery thesis. It says that the primary evidence for philosophical theories consists of intuitions. But how one understands this thesis depends on how one understands "intuition," and how one understands what role something has to play to count as evidence. On different interpretations of the thesis, the authors in the MMNS debate can be viewed either as advocates or as opponents of Centrality. Below, we'll give our take, and discuss whether Centrality holds for the theory of reference given the arguments Devitt, Ichikawa et al., and Martí have offered.

### 2.1 What is an intuition?

As we noted, several statements Devitt gives in his discussion of MMNS suggest a pro-Centrality view. But Devitt's understanding of "intuition" complicates the picture. For Devitt, intuitions are simply a type of judgment. What makes a judgment an intuition is the process by which you arrive at it:

> [I]ntuitive judgments are empirical theory-laden central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning. (Devitt 2006: 10)

Many non-philosophical sorts of judgments, even perceptual classification judgments like *the cat is on the mat*, plausibly turn out to be intuitions on this account. Thus Devitt's view of intuition is in fact fairly Williamsonian, in that it countenances no sharp line between categorizing mundane objects and kinds in the world and categorizing philosophical phenomena in thought experiments. It's also worth noting that Devitt's definition of an intuition does not assign to it any of the features which Cappelen suggests that Centrality proponents ascribe to intuition (namely, special phenomenology, "rock-bottom" evidential status, and an etiology based in conceptual competence).[11] So, in a sense, Devitt likely agrees with Cappelen that philosophers make no use of "intuitions" in the sense Cappelen has in mind.

Nevertheless, Devitt differs from Williamson and Cappelen in that he explicitly holds that philosophical methodology does involve an appeal to facts like *X intuits that p* as evidence for p itself, and that such an appeal is legitimate, when intuition is understood in Devitt's sense. The role Devitt ascribes to intuition, however, is much more limited than a traditionalist might hope. On Devitt's picture, intuition provides a starting point for more systematic inquiry. Various individuals in the community are more or less masters of different terms, like "grass" or "rock" or "elm" or "quartz." Depending on the quality of their underlying theories (recall that for Devitt, intuition is theory-laden), these different individuals are better or worse sorters of objects into these categories. Insofar as they are fairly good sorters of a category F, their intuitions will be a good preliminary guide to identifying Fs, so we may inquire further into the nature of the Fs. A deep understanding of the category F, however, involves knowing what is "common and peculiar" to the Fs. In the absence of developed scientific theory, our grip on the boundaries of F-hood is largely beholden to intuition; and the intuition of experts is given greater weight, due to the superiority of their theories. However, even expert intuitions are not decisive with regard to what is and is not an F. Sometimes our science is sufficiently advanced for us to move beyond intuition—we then reason consciously from our scientific theories and our initial intuitive sorting to discern the underlying nature of the Fs. In such cases, scientific theory trumps even expert intuition. Devitt therefore rejects the "method of cases" as MMNS characterize it.

We're attracted to a very "thin" conception of intuition, as Devitt is. Devitt (forthcoming) makes a good case that the common, core meaning of "intuition" is "immediate judgment that's not the result of conscious reasoning." Given our commitment to the heterogeneity of the psychological processes that produce such immediate judgments, we think such a "thin" account is the best account, and should be adopted[12]. Devitt additionally holds intuitions to be theory-laden and driven by central (as opposed to modular) processing. We, on the other hand, find it likely that some intuitions are innate, largely encapsulated, and "theory free" (cognitively impenetrable)—for instance, the effortless and instantaneous numerosity judgments we make in response to being shown a collection consisting of four or fewer objects. Other intuitions are, as Devitt says, theory-driven and acquired through experience.

On our liberal understanding of "intuition," it's almost inevitable that philosophers are constantly relying on intuitions during their theorizing, both in the theory of reference and in other fields. After all, they surely rely on, say, immediate concept-application judgments—it's hard to see how anyone could reason without doing so. Indeed, we're essentially including much of Cappelen's "common ground" in the category of "intuition." It would be quite cheap to use such a definition in service of a rejection of Cappelen's arguments, and we don't intend to do so. We simply hold, with Williamson, that the states philosophers are inclined to call "intuitions" are so heterogeneous that no thicker definition will cover them all. Inevitably, our ultra-thin definition lets in quite a lot of ordinary judgment, and therefore makes it almost trivially true to say that philosophers use intuition in their theorizing. There remains the more interesting question of whether theories of reference rest on intuitions as their *primary*, or even their exclusive, source of support. We'll deal with this in section 2.2.

Of course, Williamson, Cappelen and Deutsch would note that, even if philosophers *rely on* intuitions in their theorizing (in the sense that intuitions are causally responsible for many of the judgments philosophers make), this doesn't show that philosophers *appeal to intuitions as evidence* (in the sense that intuition-facts serve as premises in their argumentation). And this is fair enough; we'll return to this issue in section 2.3. But for now, it's worth noting that we think that the former claim is enough to justify the project of experimental philosophy. If a philosopher holds premise p because she has had an intuition that p, and if there is empirical evidence that intuitions about p tend to be subject to biases, cultural variation, etc., then we have *prima facie* reason to be concerned that those biases may have affected the philosopher's judgment in a problematic way. We say *prima facie*, for of course the judgment may well still be true and the philosopher may still have good (intuition-independent) justification for judging as she does. We don't, then, think that experimental findings provide sufficient reason to immediately cease relying on intuitive judgments—they merely warrant concern, and ought to prompt further investigation and reflection on our methodology. Further, since we take intuition to be utterly heterogeneous, we don't think that any particular experiment is ever likely to cast doubt on intuition *as a whole*—some types of intuitions are likely to be largely bias-free, while others may be riddled with epistemological flaws. But all this means is that experimental philosophers have more work on their hands than they may have hoped.

## 2.2 Intuitions and the method of cases

Ichikawa et al. do not provide much by way of explicit views on the overall role of intuition in the paper we've been discussing. However, at least two of the authors (Ichikawa and Weatherson) have written independently about intuitions and their evidential role. Here we will focus on a particular aspect of the view Weatherson advocates in his (2003). There, he puts forth a moderate view on intuition, according to which theoretical considerations often ought to trump intuition. Indeed, he argues that theoretical considerations should lead us to reject the Gettier intuition in favor of the classical JTB account of knowledge, on the grounds that JTB possesses greater simplicity and naturalness. The correct account of knowledge, Weatherson holds, is one that best *balances* our intuitions with various important theoretical constraints. Contrast this with the "method of cases" as characterized by MMNS, according to which the best theory is simply the one that captures the most intuitions.

We take this contrast to be important, because one of the principal arguments offered against Centrality is that, in the contexts where philosophers have been alleged to appeal to intuitions (e.g. in Kripke's Gödel/Schmidt case or in Gettier's ten coins case), the relevant philosophers present independent arguments for their conclusions. Cappelen goes so far as to claim that if a philosopher gives significant argumentation for a proposition P, that is evidence that the philosopher does *not* rely on intuition as evidence for P. Cappelen characterizes proponents of Centrality as ascribing to intuition a "Rock" evidential status—intuitions justify, but themselves need no further justification. Since (good) philosophy centered around thought experiments typically involves arguments, Cappelen holds that the Centrality proponent's picture of philosophical methodology is false.

We agree that sometimes, some experimental philosophers can make it seem as though (for example) Kripke and Gettier simply presented thought experiments, had some intuitions about them, and went home without providing any reasons beyond their intuitions for their conclusions. Such a picture obviously bears no resemblance to the texts in question. But, to us, this fact merely supports the claim that philosophers don't use the method of cases; it doesn't support the claim that intuitions aren't evidence in philosophy. We're inclined to think that, in actual practice, things more often approximate the ideal Weatherson suggests—that is to say, intuitions are weighed against competing theoretical considerations, and often rejected on those grounds. A good argument can lead us to reject an intuition; but, if that is so, then it stands to reason that a good argument can also be used in support of an intuition. Perhaps this undermines Centrality insofar as we understand that thesis as requiring that intuitions be the *primary* source of evidence for philosophy. But if experimental philosophy's importance hinges on any version of Centrality, it certainly doesn't hinge on that version.

The actual extent to which philosophical practice approaches Weatherson's ideal is likely to vary from field to field. In some fields, intuitions may be given heavy weight, while in others they may be easily rejected (indeed, Weatherson suggests as much himself). So let's consider how theoretical considerations and intuitions interact when it comes to the theory of reference.

Consider the thesis that semantic reference is speaker reference—that a name N refers to object O in virtue of the fact that a speaker intends to refer to O by his use of N. This view is likely to line up quite well with many people's judgments about particular cases, since even relative experts (like undergraduates who've completed a philosophy of language course) still struggle making the distinction. So if we surveyed a wide array of native speaker judgments, and we had good reason to think those judgments corresponded to their intuitions, and we used the method of cases, we might well arrive at the view that semantic reference is after all speaker reference.

There are, however, strong reasons—specific to semantic theory—that lead us to reject the theory that best fits speakers' intuitions. Suppose that, up until now, S has used "Bob" to refer to some individual X. On one occasion, S speaker-refers to some $Y \neq X$ using "Bob." Accepting the identification of speaker reference with semantic reference would lead us to posit that S has two homophonic names "Bob"—one of which semantically refers to X, and the other of which refers to Y. This might be seen as at odds with Grice's razor: don't postulate senses beyond necessity. Additionally, it might be seen as at odds with certain Fregean principles, e.g. that semantic content should be preservable and communicable; for instance, S's morning diary entry about "Bob" won't be about the same person as his evening diary entry about "Bob." Finally, the proposal will introduce massive indeterminacy into our theories. Suppose S wonders aloud "What's Bob doing now?," without any additional commentary and without any associated mental imagery. Which "Bob" is he wondering about, and, in virtue of what is it, say, X rather than Y[13]?

Semantic theorists invoke a plethora of principles—Grice's razor, semantic innocence, charity, compositionality, theoretical simplicity, and so on—both in arguing for and against certain theories, and in deciding what to say about different cases. It seems false, then, that the accepted method of theory construction among semanticists is anything like the "method of cases" whereby theorists simply try to fit the theory to the intuitions. Theorists of reference are often quite happy to reject intuitions in favor of theoretical considerations like those just mentioned.

Consider, for instance, Kripke on unicorns. Kripke writes that "Some of the views that I have are views which may at first glance strike some as obviously wrong," and that his "favorite" such view is his denial of the claim that "under certain circumstances, there *would* have been unicorns" (pp. 22–3). This wouldn't be coherent if Kripke were just trying to capture the intuitions.

Kripke himself would probably predict that, if you did an experiment and tested people's reactions to unicorn cases, the intuitions would not support him. And yet he does not on this basis abandon his theory or modify it so that it does not have this consequence. Kripke never says why he ignores the intuitions in this case, but there are several responses available to him: he could say, for instance, that his theory has enough intuitive support, and that modifying it to handle specific cases would violate various principles of theoretical beauty and simplicity. Alternatively, he could say that, as reliable as intuitions are in central cases, they aren't of any (or much) evidential weight in arcane cases. Such moves are inconsistent with the method of cases as MMNS characterize it, but they aren't in our view inconsistent with intuitions playing a role in semantic theory-building.

### 2.3 Intuition facts and language facts

Martí's critique of MMNS does not provide an explicit picture of the role of intuition in the theory of reference, but we can extract a bit from what she does say. Consider Martí's suggestion that we employ experiments to elicit reactions to uses of the name "Gödel," rather than intuitions about what the name "Gödel" refers to. On one reading, Martí is suggesting that we simply elicit a different type of intuition—and indeed, this seems to be supported by certain lines in the original text: "[MMNS's prompt] does not test the right kind of intuitions. It does not test the intuitions that could allow us to tell whether or not the participants in the experiment use names descriptively" (Martí 2009: 44). In fact, this is how Machery, Olivola and De Blanc (MOD) interpret her in a paper responding to Martí's criticisms. They characterize Martí as rejecting "metalinguistic" intuitions (intuitions about semantic properties such as reference) in favor of "linguistic" intuitions (intuitions about individuals and their properties). In other words, they characterize Martí as rejecting intuitions about "Gödel" in favor of intuitions about Gödel. In this vein, they performed a follow-up experiment in which a speaker in the vignette makes an utterance using the target name—e.g. "Gödel was a great mathematician."[14] Rather than asking who the speaker refers to, MOD instead asked subjects whether the speaker's utterance was true or false. MOD found no significant difference between responses to the "linguistic" vignettes and the responses to the corresponding "metalinguistic" vignette in the original prompt.

Martí (2012) objects that MOD's case asks subjects to reflect on another person's use of a name; it does not prompt them to *use* the name themselves. What is really needed, Martí claims, is data about the *subjects' own uses* of the name in question. Insofar as this is what Martí has in mind, this perspective fits quite closely with Deutsch's and Williamson's suggestions that our evidence consists of facts about P, rather than intuitions about P. Arguments in support of a given theory of reference should invoke premises concerning reference-facts, not intuition-facts; thus the appropriate data for a theory of reference consists of cases of referring, not intuitions about cases of referring. And here we are fully in agreement. When one is embarked upon construction of a theory of, say, porcupines, it is obviously preferable to *observe porcupines* rather than merely asking folks how many spines something must have to count as a porcupine. Similarly for a theory of reference; ideally, one gathers data by observing cases of referring.

Yet, Martí also suggests that (the right kinds of) intuitions are evidence; and this at least *prima facie* conflicts with the position just outlined. Here, though, we should note that the case of reference is somewhat unique, in that most cases of referring also happen to be cases of humans expressing judgments[15]—often, spontaneous judgments that we would classify as intuitions. We might, then, make sense of Martí's suggestion by saying that intuitions do have an evidential role to play in theorizing about reference—not *qua* intuitions, but *qua* instances of the phenomenon under observation. But in fact, we still think this isn't *quite* right. The data in question aren't intuitions themselves, but *verbal expressions* of intuitions. What we're really interested in is the linguistic behavior; it's neither here nor there whether the mental state which caused the linguistic behavior has whatever features it must have to count as an intuition.

Thus, while we think there's something very right about Martí's suggestion, we also think such surveys involve an "evidential role" for intuition only in a fairly minimal, roundabout way. From the perspective of the theory of reference, it doesn't matter whether surveys elicit intuitions or reasoned judgments. If you've gotten your subjects to use the name in question, you've got perfectly good evidence in the form of *speakers applying the name to a thing*. That's the core data for the theory of reference. The method is called elicited production, and it's a tool that linguists frequently make use of.

It is important to keep in mind, too, that straightforwardly applying a type of "method of cases" to such survey data would be laughably bad methodology. We can't just collect a bunch of instances of speakers applying names to things and find the theory that best fits those applications. This is because, patently, speakers incorrectly apply names to things, and they do so sometimes systematically. Sometimes they misapply names because they have inferred who their bearers are from false principles; other times they misapply names because they spontaneously judge the name applies, but would realize it did not, had they had more information or spent some time thinking about the case. As always, we take our data and our theoretical considerations, and then find the best set of theories that explain some of the data, explain away other of the data, and satisfy best our theoretical considerations. It's hard work.

So what of the philosopher in her armchair—must she refrain from claims about reference facts until the surveys have come in? Not necessarily. It's not like the armchair philosopher has no evidence whatsoever; in fact, more so than in the survey case, intuitions really are a form of evidence here—though perhaps not ideal evidence. Suppose the philosopher intuits in response to a case that she would say thus-and-so. That's very good evidence that she would indeed say thus-and-so. She could even say it aloud to the empty room and have an instance of her applying a name to a (hypothetical) thing. This is evidence that, in her idiolect, the name applies to the thing. Of course, there may be theoretical reasons for rejecting this intuition-caused bit of verbal behavior. But, being a theorist of reference, she is in an excellent position to reflect on this data point's fit with various theories, and those theories' fit with various theoretical considerations, and to reach a non-intuitive, reasoned conclusion about the case. And this is even better evidence concerning what refers to what. Is it sufficiently strong evidence to wholly obviate the need for systematic survey data? It seems to us not, but there are complicated epistemological considerations that will hinge on the details of the case and the centrality of the case and the extent of the disagreement and the available methods for explaining away certain data points.

To sum up: Martí wants empirical studies that are designed to elicit speakers' uses of names in response to vignettes. This is a wholly appropriate methodology for the theory of reference, called elicited production. It's appropriate not because it elicits the "right" type of intuitions, but because it elicits the central data for the theory of reference: speakers applying terms to things.

## 3. Making room for thought experiments

So far, we have endorsed what is, from a traditionalist point of view, a fairly Centrality-unfriendly position. We've agreed that the term "intuition" is a bit of a mess, and we've agreed that there's no obvious way to carve off "philosophical intuitions" from the swaths of mundane categorization judgments that comprise much of our ordinary cognition. We've agreed that actual philosophical practice is a far cry from MMNS's characterization of the "method of cases," involving as it does a great deal of argumentation and weighing of various theoretical considerations and constraints. And we've agreed that when studying some phenomenon P, the most immediately relevant data consists of facts about P, not facts about intuitions about P.

Nonetheless, we do have points of disagreement with the opponents of Centrality. Unlike Cappelen, we hold that intuitions are best defined as spontaneous, unreflective judgments. And unlike all opponents of Centrality, we hold that such judgments are frequently relied upon (if not *appealed to as evidence*) by philosophers when they defend their views. Finally, we hold that empirical evidence about such judgments can at least in principle have important consequences for assessments of our current methods.

With regard to theories of reference in particular, we have argued that survey methods are in fact the best possible source of evidence, with the method of elicited production being the appropriate model for such data collection. We've also (*contra* both opponents of Centrality and many experimentalists) argued that armchair appeal to intuition is in fact a viable source of evidence for theories of reference, though elicited-production evidence is generally superior when available. We'd hazard a guess that both Martí and Devitt, at least, would be more or less amenable to these conclusions; in fact, our considered position on the role of intuition in theories of reference comes closest, we think, to Devitt's. However, *contra* Devitt, and *contra* at least some proponents of negative experimental philosophy (e.g. Weinberg 2007), we want to argue in favor of the value of eliciting intuitions about unusual, far-fetched, non-"humdrum" hypothetical scenarios.

Skepticism about cases that are merely counterfactual, hypothetical, or fictional seems, at first glance, a tad strange. The Peano/Dedekind case is none of these things, yet it's hard to see how that *alone* makes intuitions about it evidentially superior to those elicited by the Gödel case. In a fictional case you make up characters (like Schmidt) and sometimes ascribe counterfactual properties to other characters (e.g. Gödel being a thief). But most English speakers haven't heard of Peano, Peano's axioms, Dedekind, or Dedekind's original formulation of Peano's axioms. *For them, those things might as well be made up*. The difference between the two cases is really the word "suppose" out front, and, for the vast majority of people you might hand a prompt to, they wouldn't notice if that word went missing.

Of course, presenting cases to naïve subjects wouldn't be Devitt's preferred way of attempting to resolve metasemantic questions. Instead, he would presumably appeal to the simple fact that people who learn first about Peano's axioms, then later discover that Dedekind originally formulated them, don't spontaneously blurt out things like, "Peano was German!" The observed absence of such utterances in ordinary conversation is more relevant than the linguistic intuitions of confused neophytes filling out philosophy surveys. Here we have the linguistic reality itself, available for direct investigation. And so it might be thought that people's utterances about the actual world and actual objects and events within it are of central importance, and people's judgments about hypothetical cases more peripheral.

However, hypothetical and counterfactual cases or, at least, cases involving telling people a lot of stuff they might currently believe the opposite of, are arguably central to adjudicating questions in the theory of reference—and, if this is so, we may need to crucially rely on judgments in response to cases in addition to simple observation of linguistic behavior "in the wild." A brief survey of the cases Kripke and Putnam adduce against descriptivism for natural kind terms reveals bizarre scenarios like duplicate planets full of chemically impossible substances, pencils that are secretly animals, atmospheric disturbances that make blue gold appear yellow, or 3-legged tigers appear 4-legged, and so on. This isn't philosophers taking shortcuts with convenient thought experiments rather than investigating real world cases. We can't just look at how people use words in actual scenarios.

Why not? Consider the case of fish. Let's suppose that English speakers use "fish" to refer to whatsoever kind, in the historical chain of usage of that word, samples of which were baptized "fish." Actual usage of "fish," however, is going to be wildly divergent, and not just because sometimes a boot on your line might tug like a fish. Some individuals are going to call whales "fish," because they are under the impression that whales and fish form a natural kind that excludes things like pigs and lemurs. Others are going to not call whales "fish," because they think whales are of a kind with pigs and lemurs, but not rays and lungfish. Still others might call whales "fish," because they think any kind including rays and lungfish must include whales, pigs, and lemurs (e.g. if they think biological kinds are clades). Without knowledge of the beliefs that drive a person's applications of "fish," the data they produce—those applications—are at best murky with regards to their relevance.

Hypothetical scenarios are a means of *controlling for* what people know or believe about a case. If you make up a case and then stipulate all the facts of it, then you know precisely what information is driving subjects' applications of terms in the case. You might have stipulated impossible things, or you might have failed to stipulate enough things to decide the questions you'll be asking your subjects. But this form of inquiry—asking people to respond to hypothetical scenarios—is of paramount importance in the theory of reference.[16]

This is not to say that hypothetical cases trump "the linguistic reality itself," or that people's actual applications are irrelevant. *Both* matter. What a word does refer to is different from what people think it would refer to, were they to know all the relevant facts. People might think that they'd call mammals "fish" were cladism to be true, but then fail to do so, even after accepting cladism—in such a case their intuitions would clash with the observed linguistic reality. In such cases linguistic reality ought to trump intuition; however, for many scenarios of interest little to no linguistic reality exists to be observed, and thus we must fall back on the use of hypothetical cases. In other words, Devitt is right about the decisiveness of various humdrum cases; the Peano case is so compelling because it's actual, and because it's unlikely that the widespread use of "Peano" to refer to Peano is the result of some misconception among users of that name. But especially for natural kind terms, how people actually use the terms can be highly equivocal and extremely misleading. The theory of reference requires a balance between real cases and hypothetical cases (which is precisely what Kripke provides us).

Judgments about what refers to what, or what applies to what, or what is true, depend on a wide range of ordinary factual beliefs, like (possibly) who wrote what theorem when, or whether whales are more closely related to lungfish than the latter are to rays. In syntax, you only have to look at what sentences people produce. You assume there's some noise in the channel, but you don't assume that the quality of the evidence varies greatly from time to time, circumstance to circumstance, and person to person, as the knowledge bases of those people in those circumstances at those times vary. That a sentence was uttered is basically all you need to know about it for purposes of syntactic theorizing—not who uttered when and what they believed at that time.

This is patently not true for semantics. Whenever traits are hidden at all behind the appearances, we have to fallibly infer from the appearances to the traits, and such inferences are infected with our beliefs, which vary wildly from person to person, time to time, and circumstance to circumstance. At one time, all whales were called "fish." In no sense does it follow from this that at that time, "fish" applied to whales. Compare syntax: all English sentences have subjects. This is very good evidence that English isn't a pro-drop language. No one seriously considers the possibility that English is really like Spanish, but, due to some false belief, English speakers put in subjects when they don't have to. It is a live option, however, that, due to some false belief, *everyone* misapplies "fish." This doesn't have anything particular to do with natural kind terms either. Everyone quite frequently misapplies "smart" and "friendly" and "Tuesday" as well. It's not always obvious whether it's Tuesday or not.

As we've noted, devising theories of reference and application requires controlling for what people believe. It's not requisite that people have *true* beliefs. You can test whether Kripke and Putnam's judgments about natural kinds are right without completing science and then explaining it all to your undergraduate subjects. All you need to do is stipulate, in a particular case, which things are the kinds, and which things satisfy which descriptions, and then see how people apply terms to things so-described.

In sum, MMNS are right that getting people to report their judgments about cases—sometimes strange hypothetical cases—is the primary methodology in the theory of reference. What they're wrong about is their claim that we have to take these data at face value (i.e. we have to employ the method of cases). *Improperly controlled* thought experiments aren't particularly telling. And, as Martí points out, MMNS's prompts are improperly controlled, as they don't control for which theory of reference (if any) the subjects explicitly hold. Data gathering for the theory of reference from naïve subjects is extremely tricky. Whatever details you don't supply to subjects are ones they might fill in in different ways.

To be clear, this case is tricky. You don't control for differing explicitly held theories of reference by stipulating that one of them is true in the thought experiment. If you did that, you'd only get out of your surveys what you put into them. But it's not better to get out of them what your subjects willy-nilly put into them. What you want is to instruct the subjects to give you a judgment on the case without deducing it from their explicitly held theory of reference (if any), and that's a hard instruction to convey to naïve subjects. As we mentioned earlier, Martí seemed to think that one could obviate the need for such an instruction by getting subjects to use the name, but we don't see why that would be. This is one positive of armchair theorizing about reference: philosophers at least understand that, when considering these cases, they ought not to be deducing what refers to what from their prior commitments in the theory of reference. Not if what they're trying to do is to determine whether those commitments are correct.

## 4. Conclusion

On our view, verbal reports of judgments about appropriately elaborate hypothetical cases are the central data for the theory of reference, for they are instances of speakers applying terms to things that have been generated under controlled conditions to test the predictions of different theories. Philosophers frequently substitute their own judgments for the more general "what we would say" about the case. This has some virtues, as we've argued, for philosophers have a good grasp on what information can and can't be brought to bear

9781474223218_txt_print.indd 148 28/09/2015 13:27

in this activity, but it does run the risk that philosophers are a poor barometer of "what we would say," or perhaps that different groups would say different things. Additionally, verbal responses to hypothetical cases are not the only relevant data—all cases of referring count, whether or not they were prompted by intuition; and responses to actual cases are in some respects better than hypothetical cases (although in some respects worse, as well). Finally, the theory that best fits that data is not necessarily the best theory overall, for there are various theoretical considerations that may necessitate rejection of certain judgments as unreliable. Given that verbal reports of judgments about hypothetical cases are undoubtedly the sort of thing that experimental philosophers study, experimental studies are relevant to evaluating theories of reference, and may indeed be necessary, if it turns out that different demographics would say different things about different cases, even under the appropriately controlled conditions. Nevertheless, because theoretical constraints play a role, and because of the difficulty in adequately controlling naïve subjects' beliefs, armchair argumentation and seasoned philosophical judgment are also necessary components of theory construction.

Thus we have provided a limited defense, with respect to the theory of reference, of (a) relying on armchair judgments about cases and (b) using experimental methods. But this defense does not straightforwardly extend to other domains of philosophy. As we noted earlier, the primary data for a theory of reference consists of cases of people applying terms to things—and such applications are standardly caused by judgments, many of which are (or perhaps are themselves caused by) intuitions. It makes sense, then, that intuitions would provide some evidence about the nature of reference—since, when one intuits that a term T refers to P, one is also highly likely to apply the term to P. But this is a peculiarity of studying human language; most other philosophical fields plausibly lack such a straightforward link between intuition and their target phenomena. Intuitions frequently cause applications of terms to things, but they do not frequently cause instances of other philosophically relevant phenomena like time, consciousness, scientific progress, or God.

Nonetheless, there is a complication here. Consider epistemology: specifically, suppose someone reads a Gettier case and intuits that it is a case of knowledge. On our account this is at least some evidence for the theory of reference, supporting the claim that the individual would likely apply "knowledge" to that case, and thereby the claim that "knowledge" *in fact* applies to that case. But then it follows that the intuition is also *equally* good evidence that that case *is* knowledge, because anything that "knowledge" applies to is knowledge. So doesn't the evidential link between intuition and referring have to be exactly as strong as the evidential link between intuition and literally *anything* else?

In a sense yes, and in a sense no. As we've noted, though instances of persons applying terms to things are the primary data for a theory of reference, such data have a greater potential to mislead than, e.g. the data for a theory of syntax. A given piece of linguistic data can of course fail to reflect the grammatical facts due to performance error; but there is no possibility that English is in fact a subject-object-verb (SOV) language despite the fact that every native English speaker employs subject-verb-object (SVO) word order. By contrast, "fish" fails to apply to whales despite the fact that there was a time when (more or less) every native English speaker applied "fish" to whales. The facts about what refers to what are hostage to certain extra-linguistic facts in the way that the facts about what is grammatical are not. Importantly for our purposes, our access to these extra-linguistic facts varies dramatically.

Contrast, for instance, "water" with "Saul Kripke." Before anyone could determine the correct application of the term "water," quite a lot of science had to be done—not so for "Saul Kripke." Note that this is not because it's an analytic, a priori truth that "Saul Kripke" refers to Saul Kripke, or anything like that; it's simply that the relevant empirical facts are (assuming no Gödel-type case holds for Professor Kripke) more readily available. Note also that, prior to doing the actual science, probing people's intuitions with hypothetical twin-earth scenarios *would not have told us what water in fact refers to*. It would at best have told us what water *would* refer to if the scientific facts happen to turn out to be such-and-so.

There continue to be a large number of terms—for instance, many terms from the sciences—for which we simply cannot yet determine their correct application. But we don't need to determine the correct application of every single term in order to develop a successful theory of reference. We can plausibly develop our theory using only the cases (like that of "Saul Kripke") where our access is reasonably good, and our confidence that we've gotten things right is therefore reasonably high. For the theory of reference, we (quite fortunately) probably have enough non-misleading data to make progress. Notice that we're not just making the general claim "use uncontroversial, central cases to develop a theory" here—prior to modern chemistry we had a plethora of uncontroversially correct judgments about water, but this failed to be sufficient for determining the nature of water.

What, then, does any of this imply about the use of intuitions in other fields of philosophy? Our intuitions will fail to be a good guide to the philosophical phenomena in cases where the correct application for the relevant term is dependent on extra-linguistic facts to which we have poor access—be these facts about the direction of future science, facts about the linguistic usage of others in our community, or even facts about undesirable consequences of a given theory that no philosopher has managed to suss out yet. It's quite likely, then, that different areas of philosophy will vary in the degree to which intuition is a good source of evidence. We'd be willing to bet, for instance, that the correct application of "time" and "consciousness" are quite heavily dependent on highly non-accessible extra-linguistic facts—and that intuition is therefore of little evidential use in the study of time and consciousness.

All of this also implies that the relevance of at least *some* types of experimental philosophy will differ from field to field. We've argued that experimental work of broadly the sort that MMNS embarked upon is crucial to a successful theory of reference; by contrast, recent studies of folk notions of consciousness (e.g. Knobe and Prinz 2008; Sytsma and Machery 2010) tell us essentially nothing about consciousness. Such studies do have value, but their value largely consists in contributing to the inherently interesting project of characterizing our everyday cognition about philosophical phenomena—which in turn may (or may not) reveal something interesting about how philosophers have come to the theories they hold.

Of course, much of the controversy over experimental philosophy concerns not "positive" projects like the one just characterized, but the "negative" project of casting doubt on intuition. On our view, such projects are more or less generally legitimate—but they must be qualified in certain ways, and the conclusions they warrant will likely differ for different fields. As we've already noted, casting doubt on "intuition" in our ultra-thin sense is a hopelessly over-broad goal. But there's quite a lot of sense in, say, studying the psychological

9781474223218_txt_print.indd 151 28/09/2015 13:27

processes underlying intuitive moral judgments and using those findings to critique the practices of ethicists; and *mutatis mutandis* for other fields or sub-fields. The arguments of the anti-Centrality folks and of the critics of MMNS show, however, that the arguments here must be careful—evidence that intuitive moral judgments are affected by emotional state, e.g. don't imply that the entire project of ethics is doomed to failure. Yet insofar as such judgments causally influence philosophers' arguments—and we think it is clear that they do, at least to some degree—there's cause for concern. Our painfully unexciting prediction: for *some* areas of philosophy, *some* amount of modification to *some* of our methods will likely be needed. Given the complexity, subtlety, and diversity of the methods and tools of philosophy, no broader conclusions are likely warranted.

## Notes

1 Thanks to Max Deutsch and Michael Devitt for helpful comments on an earlier draft.

2 The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. LU 359613).

3 Though Deutsch has also argued directly against MMNS, he was writing before the publication of, e.g. Devitt's and Martí's arguments and thus does not engage with them. We have categorized him as a participant in the former debate due to the broad scope of his argumentation.

4 For the central works here, see Williamson (2007), Deutsch (2009, 2010), and Cappelen (2012).

5 Deutsch does not—he is perfectly happy to speak of intuitions, he simply thinks they do not play an evidential role.

6 This is not the only possible characterization of the "method of cases." Cappelen's characterization is as follows: "T is a good theory of X only if it correctly predicts our intuitions about X-relevant cases." This imposes only a necessary condition, and makes no claim that the "best fit" theory is automatically the winner.

7 Deutsch (2009: 446) also makes this point.

8 Of course, Deutsch and Cappelen would disagree that Kripke's argument here relies on intuition. However, in this section our goal is merely to articulate

various positions in the debate over the Gödel case, and not to argue for or against the claim that Kripke relies on intuitions as evidence.

9 See also Devitt (forthcoming) for an extended critique of Cappelen and Deutsch.

10 At least with regard to the theory of reference.

11 To clarify the dialectic—Cappelen's position is that "intuition," as used by non-philosophers, does not serve to pick out a special kind of mental state. It is instead used, e.g. for hedging one's claims. Insofar as "intuition" is used in a technical sense in philosophy, it is a defective technical term lacking a well-agreed upon definition. Of course, regardless of the status of the term 'intuition,' it may well be that philosophers make evidential use of states possessing some of the features commonly ascribed to intuition. To argue against this possibility, Cappelen outlines the above-mentioned diagnostic criteria for intuition and argues that philosophical practice shows no reliance on states possessing those features.

12 The account might be refined a little: an intuition is an unreflective judgment that is not produced by memory, not produced by vision, not produced by proprioception …

13 This isn't merely the result of the existence of homophonic names. In standard cases of homophones, there's a perfectly good way to say which of them is being used on each occasion—look at what the speaker intended. But in this case, S doesn't realize he has two names 'Bob' referring to X and Y. So nothing about his mental states can tell us who he's referring to on any occasion.

14 The actual prompts used by MOD concerned the Chinese astronomer Tsu Ch'ung Chih rather than Gödel—this was an alternate case used in MMNS's original experiment.

15 Some are not—instances of referring also occur in, e.g. questions or exclamations.

16 One could get at the beliefs driving a person's applications more directly—say, by asking them whether they believe in cladism, etc. But use of hypothetical scenarios is a more efficient way of getting at the data we need, for it has the same virtues of controlled experimentation over mere observation. Suppose we are interested in the effect of caffeine on concentration; we could go search for people who have had exactly three cups of coffee in the past eight hours and measure their concentration, or we could take matters into our own hands and give subjects coffee. Similarly, by presenting a hypothetical scenario we are asking subjects to adopt certain beliefs (about the hypothetical scenario), rather than searching for subjects who already have those beliefs (about reality).

## References


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

Deutsch, M. (2009), "Experimental Philosophy and the Theory of Reference," *Mind and Language* 24.4: 445–66.

Deutsch, M. (2010), "Intuitions, Counter-Examples, and Experimental Philosophy," *Review of Philosophy and Psychology* 1.3: 447–60.

Devitt, M. (2006), "Intuitions in Linguistics," *The British Journal for the Philosophy of Science* 57.3: 481–513.

Devitt, M. (2011a), "Experimental Semantics," *Philosophical and Phenomenological Research* 82.2: 418–35.

Devitt, M. (2011b), "Whither Experimental Semantics?," *Theoria* 72: 5–36.

Devitt, M. (forthcoming), "Relying on Intuitions: Where Cappelen and Deutsch Go Wrong," *Inquiry*.

Ichikawa, J., Maitra, I., and Weatherson, B (2012), "In Defense of a Kripkean Dogma," *Philosophical and Phenomenological Research* 85.1: 56–68.

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

Lam, B. (2010), "Are Cantonese Speakers Really Descriptivists? Revisiting Cross-Cultural Semantics," *Cognition* 115.2: 320–9.

Ludwig, K. (2007), "The Epistemology of Thought Experiments: First Person versus Third Person Approaches," *Midwest Studies in Philosophy* 31.1: 128–59.

Machery, E., Mallon, R., Nichols, S., and Stich, S. (2004), "Semantics, Cross-Cultural Style," *Cognition* 92.3: B1–12.

Machery, E. Olivola, C. and De Blanc, M. (2009), "Linguistic and Metalinguistic Intuitions in the Philosophy of Language," *Analysis* 69.4: 689–94.

Martí, G. (2009), "Against Semantic Multi-Culturalism," *Analysis* 69.1: 42–8.

Nado, J. (2014), "Why Intuition?," *Philosophical and Phenomenological Research* 89.1: 15–41.

Weatherson, B. (2003), "What Good are Counterexamples?," *Philosophical Studies* 115.1: 1–31.

Williamson, T. (2007), *The Philosophy of Philosophy*, Hoboken, NJ: John Wiley & Sons.