January 2012

Experimental Philosophy and Philosophical Disputes

Justin Sytsma
East Tennessee State University

Jonathan Livengood
University of Illinois Urbana-Champaign

Experimental Philosophy and Philosophical Disputes

Justin Sytsma, East Tennessee State University
Jonathan Livengood, University of Illinois Urbana-Champaign

Abstract

One view of philosophy that is sometimes expressed, especially by scientists, is that while philosophers are good at asking questions, they are poor at producing convincing answers. And the perceived divide between philosophical and scientific methods is often pointed to as the major culprit behind this lack of progress. Looking back at the history of philosophy, however, we find that this methodological divide is a relatively recent invention. Further, it is one that has been challenged over the past decade by the modern incarnation of experimental philosophy. How might the reincorporation of empirical methods into philosophy aid the process of making philosophical progress? Building off of the work of Sytsma (2010), we argue that one way it does so is by offering a means of resolving some disputes that arise in philosophy. We illustrate how philosophical disputes may sometimes be resolved empirically by looking at the recent experimental literature on intuitions about reference.

1. Good Questions, Poor Technique

One common view of philosophy from other parts of the academy is that while philosophers are great at posing questions, they are not so good at answering them. The criticism is not that philosophers do a poor job at coming up with *possible* answers, but that they lack a means of determining which amongst the possible answers are the right ones. For example, Nobel laureate and co-discover of DNA, Sir Francis Crick, offered this blunt assessment:

> Essentially philosophers often ask good questions, but they have no techniques for getting the answers. Therefore you should not pay too much attention to their discussions. And we can ask what progress they have made. A lot of problems which were once regarded as philosophical, such as what is an atom, are now regarded as part of physics. Some people have argued that the main purpose of a philosopher is to deal with the unsolved problems, but the problems eventually get solved, and they get solved in a

---

Corresponding Author: J. Sytsma
East Tennessee State University
email – jmsytsma@gmail.com

scientific way. If you ask how many cases in the past has a philosopher been successful at solving a problem, as far as we can say there are no such cases. (Blackmore, 2005, 75)

In this passage Crick suggests that there is a sharp methodological divide between philosophy and the sciences. With the development of experimental philosophy over the past decade, however, an increasing number of philosophers are crossing this divide. One motivation for doing so is the reason that Crick suggests: Experimentation, and empirical investigation more generally, has shown itself to be an indispensable tool in making progress with regard to many important questions about our world.

Of course, historically speaking, the divide between philosophy and the sciences has not always been so sharp as Crick suggests. Not only was it once philosophers who were primarily concerned with questions about the nature of matter, for example, but those philosophers made significant progress toward answering them. In fact, the first people to call themselves experimental philosophers—Robert Boyle, Isaac Newton, and some of their contemporaries—were working on problems in what we would now call physics or chemistry.[1]

Boyle's experimental work was specifically directed at questions about the nature of matter—especially the nature of air. In the course of his inquiry into the spring of the air and the possibility of a vacuum, Boyle investigated the relationship between air and sound. In one beautiful experiment, he suspended a mechanical watch by a thin thread in a bell jar connected to an air pump. Boyle then listened to the sound of the watch as the pump removed the air. He noted that the sound of the watch steadily decreased in volume as the air was removed until it could not be heard at all. Thus, Boyle convincingly demonstrated that sound is transmitted through the air but not through a vacuum.

Newton's experimental work was primary directed at questions about the nature of space, time, gravitation, and light, although in the second half of his career he inquired into the nature of matter as well. In fact, Newton's famous dictum against hypotheses, *hypotheses non fingo*, was explicitly a statement about the proper methodology of experimental philosophy (as opposed to speculative philosophy). Newton said that he framed no hypotheses because "hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, *have no place in experimental philosophy*" (1946 [1713], 314, emphasis added).

By the nineteenth century, experimental philosophy appears in the titles of textbooks and handbooks for students of what we would now call the physical sciences.[2] Even Michael Faraday seems to have regarded his work as philosophical. In public lectures published in 1861 as *The Chemical History of a Candle*, for example, Faraday remarked: "We come here to be philosophers, and I hope you will always remember that whenever a result happens, especially if it be new, you should say, 'What is the cause? Why does it occur?' and you will, in the course of time, find out the reason" (2008, 9).[3] Of course, Faraday

tracked down the causes of the unusual behaviors that he observed (in candles and elsewhere) through sequences of experiments.

It should be noted that Crick's concern in the above quotation was not really with the questions about the nature of matter, however, but with questions concerning the working of the human mind, and specifically the nature of consciousness.[4] But again we find that canonical philosophers—such as Rene Descartes, John Locke, George Berkeley, David Hume, Immanuel Kant, John Stuart Mill, Charles Peirce, and William James—either suggested or performed experiments to learn things about human cognition, perception, and emotion.

Looking across the history of philosophy since the early modern period, what we see is a process of specialization, with the sciences slowly breaking away from the mother discipline of philosophy. And, as this has occurred, philosophy has increasingly become focused on theoretical speculation over empirical investigation. In fact, many philosophers came to embrace methods that have been thought to be independent of empirical investigation, such as conceptual analysis. One problem with such an anti-experimental turn, however, is that it leaves philosophers without the means to resolve some of the disputes that arise in employing these methods, as will be illustrated in the following section.

2. Resolving Disputes

In previous work (Sytsma, 2010), one of us has argued that one problem with the standard practice of conceptual analysis in philosophy is that it lacks a means for resolving conceptual disputes when they arise. The claim is not that analyzing concepts is uninteresting, or that conceptual analysis has no role to play in modern philosophy, but that the practice should be expanded to make use of empirical methods. The primary reason is that empirical investigation provides a needed means for resolving philosophical disputes that arise in the practice of conceptual analysis.

This was illustrated by considering one such dispute: Max Bennett and Peter Hacker (2003) have raised concerns about the (supposed) misuse of terms for ordinary, non-technical concepts in the brain sciences. Specifically, they have claimed that ordinary use is such that psychological predicates only logically apply to whole living animals. Of course, it follows that many brain scientists disagree—at least insofar as they take themselves to be employing these terms in a standard way—as they use them far more broadly than Bennett and Hacker would allow. Which group is right about the ordinary use of these psychological predicates? To answer this question we need empirical evidence. And Sytsma discussed a range of empirical studies that indicate that ordinary

people are perfectly willing to apply psychological predicates to non-animals in a variety of cases.

More important than the specific example is the lesson that experimental research can help to resolve conceptual disputes that might otherwise result in an impasse. Further, the lesson that experimental research can help resolve conceptual disputes is an instance of a larger point: In providing a way of resolving some of the disputes that arise in the process of philosophical research, empirical investigation leads to progress in philosophy. The larger point is illustrated in the following section by considering one line of research in the recent literature—starting with Edouard Machery, Ron Mallon, Shaun Nichols, and Stephen Stich's (2004) work on intuitions about reference—and showing how some of the objections to that work that might have resulted in an impasse have been resolved through further empirical investigation.

3. Resolving Disputes about Semantic Intuitions

In recent years the study of linguistics has undergone a methodological shift. Generative linguists, in particular, have moved beyond a reliance on their own intuitions about the grammaticality of bits of language and have begun to incorporate systematic empirical studies of the intuitions of ordinary people into their work. Carson Schütze concisely captures the motivation for this shift when he writes:

> Linguistics has much to gain and nothing to lose by taking data collection, and particularly judgment collection, much more seriously, both with regard to the insights that will be gained and the theoretical issues that will be clarified, and with regard to the standing of the field as a scientific endeavor in the larger academic setting. The realization seems to be growing that the psychology of grammaticality judgments can no longer be ignored. (1996, 212)

The same motivations hold for philosophy of language, prompting Edouard Machery and Stephen Stich to urge that "it is time for philosophers of language to emulate linguists in taking an experimental turn" (forthcoming, 28).

In fact, much of the work that has been done in experimental philosophy of language to date has been directed toward making the case for such an experimental turn. Most prominently, it has been argued that the intuitions of philosophers of language about a prominent thought experiment in the literature—Saul Kripke's (1972) Gödel example— do not reflect the diversity of intuitions found in the larger population; it is then contended that this disparity raises serious questions about the use of philosophers' intuitions as evidence in their theorizing about reference.

Kripke's concern in giving the Gödel example was to determine the correct theory of reference: To understand how our words refer to things (entities, events, properties); for example, how the proper name "Gödel" picks out a specific person. The underlying concern for recent experimental work on the Gödel example, however, is not so much with what the correct theory of reference is, but with what should count as compelling evidence for a theory of reference. Specifically, should the intuitions of philosophers be taken to be compelling evidence?

Kripke treated his intuitions about the Gödel example as evidence against a *descriptivist* account of reference, and evidence for a *causal-historical* account. At the time, the dominant view was descriptivist. It was held that proper names, for example, refer to those individuals that correspond best with the descriptions associated with the names. Kripke constructed the Gödel example to put pressure on this type of account. The basic idea is simple: The case is constructed so that the description generally associated with the name "Gödel" best corresponds with someone other than the man given that name at birth. Kripke writes (1972, 83-84):

> Let's take a simple case. In the case of Gödel that's practically the only thing many people have heard about him—that he discovered the incompleteness of arithmetic. Does it follow that whoever discovered the incompleteness of arithmetic is the referent of "Gödel"? Imagine the following blatantly fictional situation. (I hope Professor Gödel is not present.) Suppose that Gödel was not in fact the author of this theorem. A man named "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".… So, since the man who discovered the incompleteness of arithmetic is in fact Schmidt, we, when we talk about "Gödel," are in fact always referring to Schmidt. But it seems to me that we are not. We simply are not.

Thus, Kripke's intuition about this case—which he assumes that *we* will all share—is that despite the fact that someone else better fits the description associated with the name "Gödel," the name nonetheless refers to Gödel. And most philosophers of language have agreed.

While Kripke's intuition about the Gödel case runs counter to the standard descriptivist account, it is in accord with the causal-historical account that he puts forward. With

regard to proper names, his account contends that they refer to the individuals that were dubbed with those names—so long as there is an appropriate causal connection between that dubbing and the current use. Following Kripke, many have taken their intuitions about the Gödel example to provide evidence in favor of causal-historical theories of reference. More generally, this can be seen as an instance of a standard methodology in philosophy. Joshua Alexander and Jonathan Weinberg articulate this methodology as follows:

> Going back arguably at least to Frege (and, in some sense, all the way back to Socrates), it has been a standard practice in analytic philosophy to employ intuitions generated in response to thought-experiments as evidence in the evaluation of philosophical claims. A philosopher, wishing to either establish or prosecute some philosophical claim proposes a thought-experiment intended to generate an intuition relevant to evaluating the philosophical claim. According to standard philosophical practice, the generated intuition provides evidence for the acceptance or rejection of the philosophical claim: the philosophical claim is *prima facie* good to the extent that it accords with the generated intuition, *prima facie* bad to the extent that it fails to accord with the generated intuition. (2007, 56)

Applied to work on reference, this standard methodology holds that the correct theory of reference will be the one that best accords with intuitions about various relevant cases. In other words, intuitions about cases like Kripke's Gödel example are thought to place constraints on theories of reference.

Of course, the intuitions that have been considered in practice are those of a small number of philosophers. This raises a rather obvious question: Why focus on those intuitions? Unfortunately, practitioners of the standard methodology have not been as explicit about their reasons as one might like. A charitable explanation for this is that Kripke, for example, did not actually take himself to be focusing on the intuitions of philosophers; rather, he took his intuition about the Gödel example to reflect something utterly obvious about the case—so obvious that just about everybody would share his intuition. In other words, on this reading Kripke made what we have called the *uniformity conjecture* (Sytsma and Livengood, 2011): He assumed that intuitions about the Gödel case are essentially uniform. However, recent experimental work has called this uniformity conjecture into doubt.

3.1. Cross-cultural Semantics

In one of the foundational articles in the recent incarnation of experimental philosophy, Edouard Machery, Ron Mallon, Shaun Nichols, and Stephen Stich (2004) challenged the

uniformity conjecture suggested by Kripke—and through it the standard methodology applied to work on reference. Machery et al. presented evidence that there is significant variation in intuitions about the Gödel case, both within and across cultures. In other words, their data suggests that intuitions about this example are far from uniform.

Machery et al. presented two groups of English-speaking undergraduates—one group from Rutgers University and one group from the University of Hong Kong—with two probes modeled on Kripke's Gödel example. The probes were presented in English and differed in whether Western or Chinese names were used. Each participant was given both probes. The Western-name probe reads as follows:

> Suppose that John has learned in college that Gödel is the man who proved an important mathematical theorem, called the incompleteness of arithmetic. John is quite good at mathematics and he can give an accurate statement of the incompleteness theorem, which he attributes to Gödel as the discoverer. But this is the only thing that he has heard about Gödel. Now suppose that Gödel was not the author of this theorem. 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 claimed credit for the work, which was thereafter attributed to Gödel. Thus, he has been known as the man who proved the incompleteness of arithmetic. Most people who have heard the name "Gödel" are like John; the claim that Gödel discovered the incompleteness theorem is the only thing they have ever heard about Gödel. When John uses the name "Gödel," is he talking about:
>
> (A) the person who really discovered the incompleteness of arithmetic? or
> (B) the person who got hold of the manuscript and claimed credit for the work?

Machery et al. scored answers as either 0 or 1, where 0 corresponds with an (A) answer and 1 corresponds with a (B) answer. They then added the scores for the two probes together, resulting in a scale running from 0 to 2 for each participant. Using this scale, Machery et al. found a mean of 1.13 for American participants compared to a mean of 0.63 for Chinese participants. Converting the mean scores to percentages of (B) answers, they found that 57% of American participants answered (B) compared to 32% of Chinese participants. The difference between the responses for the two groups is statistically significant.

On the basis of this study, Machery et al. tentatively concluded that there is cross-cultural variation in intuitions about the Gödel case. Further, they pointed out that there is also

significant variation within each sample—neither the American nor the Chinese participants were uniform in their responses. These results suggest that the uniformity conjecture does not hold, at least with regard to Kripke's Gödel example.

A number of different responses have been made to Machery et al.'s work, often with slightly different targets. Some critiques have raised concerns about their interpretation of their empirical results, while others have instead focused on the conclusion that Machery et al. draw from those results. In each case, the major criticisms that have been raised are susceptible to further experimental testing, as has often occurred in the subsequent literature. In other words, the literature surrounding Machery et al.'s work provides a further example of how empirical investigation can help to resolve some philosophical disputes.

3.2. The Expertise Objection

Michael Devitt (forthcoming) accepts Machery et al.'s interpretation of their results, while challenging the conclusion they draw from those results. He argues that owing to their greater expertise, the semantic intuitions of philosophers are more reliable than those of non-philosophers. As such, their intuitions should be accorded greater evidential weight in assessing theories of reference.

Machery et al. briefly considered this *expertise objection* in their original article. They noted that the focus on the intuitions of philosophers that is typical of work using the standard methodology could be justified in another way. Rather than relying on the uniformity conjecture, advocates of the standard methodology could grant that the intuitions of philosophers are not representative of the intuitions of the broader population, and instead argue that the intuitions of philosophers are simply better.

Machery et al.'s response takes the form of a challenge. Accepting that the evidence suggests that there is a good deal of variation in intuitions about reference for the Gödel case, further justification is needed for the assumption that the intuitions of philosophers are more reliable than the intuitions of non-philosophers:

> We find it *wildly* implausible that the semantic intuitions of the narrow cross-section of humanity who are Western academic philosophers are a more reliable indicator of the correct theory of reference (if there is such a thing…) than the differing semantic intuitions of other cultural or linguistic groups. Indeed, given the intense training and selection that undergraduate and graduate students in philosophy have to go through, there is good reason to suspect that the alleged *reflective* intuitions may be *reinforced* intuitions. In the absence of a principles argument about hwy

philosophers' intuitions are superior, this project smacks of narcissism in the extreme (2004, B9)

Devitt takes up Machery et al.'s challenge, claiming to offer a principled argument for the superiority of the intuitions of philosophers. The principle he calls on is expertise. In essence, Devitt responds by claiming that semanticists (prominently including the relevant philosophers) have greater expertise concerning semantics than do the folk. If he is correct, then the intuitions of the folk are relatively unreliable (compared to the intuitions of philosophers), and this could be called on to explain away the variation in intuitions that Machery et al. found.

Is Devitt's claim accurate? This strikes us as an empirical question. And absent empirical evidence about the relative reliability of the intuitions of semanticists and the folk, it is unclear how we should answer it. Clearly Devitt takes it to be obvious that the intuitions of semanticists are more reliable; but, equally clearly, Machery et al. do not agree. Absent empirical evidence, we have a clash of opinions, and it is unclear how this dispute could be resolved. Fortunately, Edouard Machery (forthcoming) has presented new empirical evidence that helps to resolve the dispute.

To test the claim that the semantic intuitions of experts are more reliable than non-experts, Machery used a snowball sampling method to survey both philosophers of language and linguists. For comparison, lay people (people without training in philosophy or linguistics) were also sampled and matched to the experts for general level of education. Participants were given a variation[5] on the Chinese-name version of Machery et al.'s original Gödel probe:

> Ivy is a high school student in Hong Kong. In her astronomy class, she was taught that Tsu Ch'ung Chih was the man who first determined the precise time of the summer and winter solstices. But, like all her classmates, this is the only thing she has heard about Tsu Ch'ung Chih. Now suppose that Tsu Ch'ung Chih did not really make this discovery. He stole it from an astronomer who died soon after making the discovery. But the theft remained entirely undetected and Tsu Ch'ung Chih became famous for the discovery of the precise times of the solstices. Everybody is like Ivy in this respect; the claim that Tsu Ch'ung Chih determined the solstice times is the only thing people have heard about him.

Having read the above story and accepting that it is true, when Ivy uses the name "Tsu Ch'ung Chih," who do you think she is actually talking about:

> (A) The person who (unbeknownst to Ivy) really determined the solstice times?

> (B) The person who is widely believed to have discovered the solstice times, but actually stole this discovery and claimed credit for it?

The linguists surveyed were sorted into eight groups on the basis of their area of specialization. Including philosophers of language and lay people, this gave ten groups total. Machery found that the majority of the participants in each group reported the causal-historical intuition, giving the (B) answer. At the same time, Machery found variation in the responses within each group, with the proportion varying from 67% (B) answers to 89% (B) answers. Further, sorting philosophers of language and linguists into two groups on the basis of how likely they were to have read Kripke's *Naming and Necessity*, Machery found that philosophers and semanticists were significantly more likely to report having the causal-historical intuition than were linguists working in discourse analysis, historical linguistics, and sociolinguistics.

On the basis of these findings, Machery argues that expertise about language does not appear to increase the reliability of people's semantic intuitions. He reasons that if such expertise improves reliability, then the responses of different groups of experts should not be significantly different. This is not what he found, however: The group of experts who were more likely to have read Kripke were more likely to report having the Kripkean intuition than were lay people, while the group of experts who were less likely to have read Kripke were less likely to report having the Kripkean intuition than were lay people.

3.3. The Speaker's Reference Objection

We have just considered one prominent objection to Machery et al.'s work on the Gödel example—the expertise objection—and seen how further experimental work has helped to resolve the issue. This objection did not take issue with their interpretation of their experimental results, but instead targeted the argument that made use of those results. Other responses in the literature have challenged Machery et al.'s interpretation of their results, noting potential confounds with their study. Perhaps the most prominent such response argues that the question in their original Gödel probe was ambiguous with regard to Kripke's distinction between speaker's reference and semantic reference.

We will focus on the version of the *speaker's reference objection* put forward by Kirk Ludwig (2007). He writes:

> For anyone at all familiar with work in philosophy of language, it is immediately evident that the question [in Machery et al.'s Gödel probe] does not clearly distinguish between two things: Whom John intends to be talking about (or speaker's reference) and who the name John uses refers

to, taken literally in the language he intends to be speaking (semantic reference). Experts may well negotiate this infelicity in the formulation of the question without much difficulty, but that is because they have some relevant expertise about hard-won distinctions developed in the field and will likely understand what is intended. (150)

The underlying contention for Ludwig's objection is that in the Gödel story John lacks some knowledge that is supplied to the reader by the narrator. Specifically, John does not know about Schmidt, and thus does not know that the person who discovered the incompleteness of arithmetic is not the person who is commonly credited with that discovery. As such, there is the potential for confusion: It is possible that participants held that in using the name "Gödel" John meant to be talking about the person who discovered the theorem, even though that name literally refers to the person commonly credited with the discovery. And if participants then focused on who John meant to be talking about when they answered the probe question, their answers would not correspond with their semantic intuitions.

In the absence of empirical evidence that people's answers reflected their semantic intuitions, it is possible that Machery et al.'s results reflect the ambiguity between speaker's reference and semantic reference. Unfortunately, Ludwig offers no evidence that this is actually the case. In other words, Ludwig presents a *theoretical response*, but does little to establish that the potential confound at issue is likely to have driven the results. This is an issue because in the absence of empirical evidence, we are left at an impasse: Ludwig takes it to be obvious that this potential confound drove Machery et al.'s results, while they are inclined to see it as at most a relatively minor factor.[6] Fortunately, we find that once again, further experimentation can help to resolve this dispute.

In fact, Edouard Machery, Max Deutsch, and Justin Sytsma (forthcoming) have run a series of experiments to assess the speaker's reference objection. Focusing on their first study, they clarified the question in Machery et al.'s original Gödel probe, aiming to remove the ambiguity between speaker's reference and semantic reference. They asked English-speaking participants in both United States and China the following question:

> When John uses the name "Gödel," *regardless of who he might intend to be talking about*, he is *actually* talking about:
>
> (A) the person who really discovered the incompleteness of arithmetic.
> (B) the person who got hold of the manuscript and claimed credit for the work.

e𝑝

Machery, Deutsch, and Sytsma found that while 59% of the American participants gave the causal-historical answer (B), only 39% of the Chinese participants gave that answer. Again we see significant variation in participants' intuitions about the Gödel case, even after the probe question has been clarified.

3.4. The Linguistic Competence Objection

Barry Lam (2011) offers another objection to Machery et al.'s work that is focused on interpretation. Lam argues that the cross-cultural variation that they found between the intuitions of Westerners and East Asians might not reflect differences in their intuitions, but differences in linguistic competence. Thus, Lam notes that while the American participants responded to questions in their native tongue, this was a second language for the Chinese participants.

Interestingly, Lam did not simply put forward the linguistic competence objection, but sought to empirically test it. To do so, he presented both native speakers of Cantonese and native speakers of English the following probe in their respective languages:

> Suppose there is a group of people who do not know anything of the English author Shakespeare except the name and that he is the author of "Romeo and Juliet." Unbeknownst to this group of people, Shakespeare did not in fact write the play "Romeo and Juliet"; in fact, a German man named "Spencer" wrote the play, but Spencer was an obscure writer who died before the play was published. Shakespeare in fact found the play and published it as his own. Nobody knows this. This group of people otherwise use the name "Shakespeare" and can use it in conversation, for instance, they may ask each other, "I wonder whether Shakespeare was English or German?"

> When these people use the name "Shakespeare" in a conversation, is their use of the name to talk about:

> (A) Shakespeare
> (B) Spencer

Lam found that a significantly higher proportion of the participants who were native Cantonese-speakers gave the causal-historical answer "Shakespeare" (87%) than did the participants who were native English-speakers (65%). Thus, while Lam's results indicate some cross-cultural variation in intuitions about reference, this variation runs in the opposite direction to that found by Machery et al., putting pressure on their conclusion.

Edouard Machery, Max Deutsch, Ron Mallon, Shaun Nichols, Justin Sytsma, and Stephen Stich (2011) have responded to Lam's objection, pointing out that there are two key differences between his Shakespeare vignette and the original Gödel vignette. First, while the Gödel vignette states that the only thing that *most people* who have heard the name "Gödel" know about him is that he discovered the incompleteness of arithmetic, the Shakespeare vignette speaks of *a group of people* who know nothing about Shakespeare except that he is the author of *Romeo and Juliet*. The latter claim, however, is compatible with *most competent speakers* knowing far more about Shakespeare than this. Second, while the answer choices in the Gödel probe were definite descriptions, the answer choices in the Shakespeare probe were proper names. These differences introduce a potential confound: It is unclear whether Lam's results primarily reflect that Chinese participants received the probe in their native language, or the other differences that Lam introduced. And, once again, further experimental work can help to resolve this dispute.

The confounds that Machery and colleagues note can be readily avoided by repeating Lam's experiment using a Chinese translation of the original Gödel probe, rather than the Shakespeare probe. And this is exactly what they did, finding that only 39% of Chinese participants gave the causal-historical answer to the translated probe. This is consistent with Machery et al.'s original results, suggesting that the cross-cultural variation that they found cannot be explained away by the probe having been presented in English rather than in Chinese.

3.5. The Perspectival Ambiguity Objection

In Sytsma and Livengood (2011) we raise a further objection to Machery et al.'s interpretation of their results. Specifically, we argued that there is a *perspectival ambiguity* in the question that they asked. This ambiguity concerns the epistemic perspective that the participant should adopt in reading the definite descriptions given as the answer choices. Thus, we noted that there is an asymmetry between *what John knows* in the Gödel story and *what the narrator knows* in telling us the story. As such, from the narrator's perspective, Schmidt is the person who discovered the incompleteness of arithmetic, while Gödel is merely the person who got hold of the manuscript and claimed credit for the work. As far as John knows, however, it is Gödel who discovered the theorem.

While Machery et al. assumed that participants would adopt the narrator's perspective in responding to the Gödel probe, the question asks who *John* is talking about, which might lead participants to instead adopt John's perspective in answering the question. If they did so, however, their responses would not necessarily correspond with their semantic intuitions: They might give the "descriptivist" response, despite having causal-historical

intuitions, simply because they read "the person who really discovered the incompleteness of arithmetic" from John's perspective as denoting Gödel.

To test whether the perspectival ambiguity we identify actually does impact the results for the Gödel probe, we ran a series of studies in which we changed the question in the Gödel probe to indicate that participants should either answer from John's perspective or the narrator's perspective. To emphasize John's perspective, we asked participants who John thinks he is talking about. To emphasize the narrator's perspective, we asked participants who they take John to actually be talking about, in addition to clarifying the definite descriptions in the answer choices:

> (A) the person who (unbeknownst to John) really discovered the incompleteness of arithmetic?

> (B) the person who is widely believed to have discovered the incompleteness of arithmetic, but actually got hold of the manuscript and claimed credit for the work?

Responses were collected from native English-speakers. We found that while 39% of the participants receiving Machery et al.'s original question gave the "causal-historical" answer (B), only 22% of those receiving the question emphasizing John's perspective gave that answer; further, 74% of the participants receiving the probe emphasizing the narrator's perspective gave the "causal-historical" answer.

We argued that these results provide evidence that Machery et al.'s original Gödel probe is ambiguous. The perspectival ambiguity we identified significantly affects the answers given by native English-speakers. This raises significant doubts about whether their results reflect participants' semantic intuitions. We went on to argue that although we only tested native English-speakers, our results nonetheless challenge Machery et al.'s claim to have shown significant cross-cultural variation between Westerners and East Asians, as well as significant variation within East Asians. In brief, having evidence that the original Gödel probe does not reliably track semantic intuitions in Westerners, it should not be assumed to do so for East Asians.

While our studies cast doubt on Machery et al.'s evidence for the claim that there is significant variation in intuitions about reference, including cross-cultural variation between the intuitions of Westerners and East Asians, they do not provide evidence either for or against Machery et al.'s claim itself. To determine whether or not there is cross-cultural variation between the intuitions of Westerners and East Asians, for example, cross-cultural data is needed using probes that control for the perspectival ambiguity we identified.

We have recently conducted a series of cross-cultural studies with Ryoji Sato and Oguchi Mineki using such probes. In our new experiments, we first presented the suite of Gödel probes used in Sytsma and Livengood (2011) to Western participants in English to get a baseline. We then had those probes translated into Japanese and presented them to East Asian participants at five different Japanese universities. Finally, we had the translated probes reverse-translated back into English, by a different translator, and presented them to a new set of Western participants to check for potential problems with the translation.

What we found is that there was much smaller variation for the Japanese translations when compared with the other sources—especially between the percentages for the John's perspective and narrator's perspective probes. These new results suggest that despite the effect of the perspectival ambiguity on responses to Machery et al.'s Gödel probe for Westerners, their original claim is nonetheless correct: East Asians appear to be significantly more likely to have descriptivist semantic intuitions than do Westerners.

4. Conclusion

One view of philosophy is that while philosophers often ask good questions, they have a poor track record with regard to producing the right answers. We have argued that this view is somewhat unfair, historically speaking. Nonetheless, it is fair to suggest that as philosophy has moved away from the use of empirical methods, its ability to make progress has diminished. In this paper we have suggested one way in which this can occur: Empirical investigation can provide a means for resolving some of the disputes that arise in philosophy, as we illustrated with a survey of recent experimental literature on intuitions about reference.

References

Alexander, Joshua and Jonathan Weinberg (2007). "Analytic Epistemology and Experimental Philosophy." *Philosophy Compass*, 2(1): 56–80.

Bennett, M. R., and Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Blackmore, Susan (2005). *Conversations on Consciousness*. Oxford: Oxford University Press.

Crick, Francis (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Touchstone.

Devitt, Michael (forthcoming). "Experimental Semantics." *Philosophy and Phenomenological Research*.

Faraday, Michael (2008 [1861]). *The Chemical History of a Candle*. Echo Library.

Ludwig, Kirk (2007). "The Epistemology of Thought Experiments: First Person versus Third Person Approaches." *Midwest Studies in Philosophy*, 31(1): 128–159.

Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen Stich (2004). "Semantics, Cross-cultural Style." *Cognition*, 92(3): B1–B12.

Machery, Edouard and Stephen Stich (forthcoming). "Experimental philosophy of language." In G. Russell and D. G. Fara (Eds.), *The Routledge Companion to the Philosophy of Language*. Routledge.

Machery, Edouard, Max Deutsch, and Justin Sytsma (forthcoming). "Speaker's Reference and Cross-cultural Semantics." In A. Bianchi (Ed.), *On Reference*. Oxford: Oxford University Press.

Machery, Edouard, Max Deutsch, Ron Mallon, Shaun Nichols, Justin Sytsma, and Stephen Stich (2011). "Semantic Intuitions: Reply to Lam." *Cognition*, 117: 363–366.

Newton, Isaac (1946 [1713]). *Mathematical Principles of Natural Philosophy*. Translated by Andrew Motte (1729), revised by Florian Cajori (1946). Berkeley: University of California Press.

Schütze, Carson (1996). *The Empirical Base of Linguistics*. Chicago: University of Chicago Press.

Sytsma, Justin (2010). "The Proper Province of Philosophy: Conceptual Analysis and Empirical Investigation." *Review of Philosophy and Psychology*, 1(3): 427–445.

Sytsma, Justin and Jonathan Livengood (2011). "A New Perspective Concerning Experiments on Semantic Intuitions." *Australasian Journal of Philosophy*, 89(2): 315–332.

---

[1] A few university chairs of experimental philosophy remain in Britain from the period immediately following the initial rise of experimental philosophy. For example, the Erasmus Smith Professorship of Natural and Experimental Philosophy was established at the University of Dublin in 1724, and the Plumian Professorship of Astronomy and Experimental Philosophy was established at Cambridge University in 1704. One of Newton's early students, Roger Cotes, was the first Plumian professor (from 1707 to 1716), and several notable people have held the professorship, including George Darwin (from 1883 to 1912),

Arthur Eddington (from 1913 to 1944), Harold Jeffreys (from 1946 to 1958), and Fred Hoyle (from 1958 to 1972).

[2] For example, George Gregory's 1808 *Lectures on Experimental Philosophy, Astronomy, and Chemistry; Intended Chiefly for the use of Students and Young Persons*, or Richard Green Parker's *School Compendium of Natural and Experimental Philosophy* published in 1852, which covered topics ranging from mechanics, acoustics, and optics to electricity, magnetism, and astronomy. Also Cassell, Petter, and Galpin's 1861 *Hand-Book of Natural and Experimental Philosophy* and an 1826 *Grammar of Natural and Experimental Philosophy* by the reverend David Blair.

[3] This is not an isolated instance. For example, later on in the same volume, Faraday writes: "Now we philosophers—I hope that I may class you and myself together in this case—speak of water as water, whether it be in its solid or liquid or gaseous state—we speak of it chemically as water." (2008, 23).

[4] As Crick wrote a decade earlier: "The… thing to stress is that the study of consciousness is a scientific problem. Science is not separated from it by some insurmountable barrier. If there is any lesson to be learned from this book it is that we can now see ways of approaching the problem experimentally. There is no justification for the view that only philosophers can deal with it. Philosophers have had such a poor record over the last two thousand years that they would do better to show a certain modesty rather than the lofty superiority that they usually display." (1994, 257-258).

[5] In this variation, the probe question was clarified following the work of Sytsma and Livengood (2011) discussed below.

[6] One theoretical reason to see this potential confound as at most a relatively minor factor is that for it to explain the variation in responses that Machery et al. found for their Gödel probe, participants who selected the "descriptivist" answer (A) would not only need to be answering with regards to the speaker's reference of John's utterances of "Gödel," but would have needed to assume that John meant to be talking about the person who discovered the incompleteness of arithmetic. It is not clear, however, why participants would think this. After all, John's intentions would be relative to particular statements involving the name "Gödel," but Machery et al.'s probe did not include any such statements. As such, participants would first need to imagine a particular statement. And, indeed, for some such statements, the most charitable reading would be to treat John as intending to talk about Schmidt ("Boy, Gödel sure was a mathematical genius!"). The problem is that for other statements the most charitable reading would be to treat John as intending to talk about Gödel ("Boy, Gödel sure had a funny name!").