Reference in the Land of the Rising Sun:
A Cross-cultural Study on the Reference of Proper Names

Justin Sytsma, Jonathan Livengood, Ryoji Sato, and Mineki Oguchi

A standard methodology in philosophy of language is to use intuitions as evidence. Machery, Mallon, Nichols, and Stich (2004) challenged this methodology with respect to theories of reference by presenting empirical evidence that intuitions about one prominent example from the literature on the reference of proper names (Kripke's Gödel case) vary between Westerners and East Asians. In response, Sytsma and Livengood (2011) conducted experiments to show that the questions Machery and colleagues asked participants in their study were ambiguous, and that this ambiguity affected the responses given by Westerners. Sytsma and Livengood took their results to cast doubt on the claim that the current evidence indicates that there is cross-cultural variation in intuitions about the Gödel case. In this paper we report on a new cross-cultural study showing that variation in intuitions remains even after controlling for the ambiguity noted by Sytsma and Livengood.

In a widely discussed article, Machery, Mallon, Nichols, and Stich (2004, MMNS) argued that the common practice of appealing to one's own intuitions about cases as evidence for or against philosophical theories of reference is suspect. Specifically, MMNS targeted a standard justification for this practice. They noted that philosophers of language often assume that the relevant semantic intuitions are widely held and do not systematically differ across theoretically salient groups, taking such widespread agreement to indicate that the intuitions are reliable. Following Sytsma and Livengood (2011), we will refer to this assumption as the *uniformity conjecture*. MMNS presented empirical evidence suggesting that the uniformity conjecture is false: Intuitions about the semantic reference of a proper name in one prominent case from the literature (Kripke's Gödel case) differ in a statistically significant way between Western and East Asian cultures.

In response, Sytsma and Livengood (2011) presented experimental evidence indicating that a *perspectival ambiguity* in the questions that MMNS asked their participants affected the responses of one of their target groups (Westerners). Sytsma and Livengood argued that their results cast doubt on MMNS's assumption that their survey instrument reliably indicates the semantic intuitions of Westerners. On this basis, Sytsma and Livengood charged that MMNS's study does not allow us to conclude that semantic intuitions differ across cultures. In other words, they claimed that MMNS's "results do not provide the basis for a compelling case against the uniformity conjecture" (316). In doing so, however, Sytsma and Livengood did *not* claim that the uniformity conjecture is true with respect to Kripke's Gödel case. Their point was rather that MMNS's study did not show that the uniformity conjecture was false. As such, the status of the uniformity conjecture with respect to intuitions about reference remains an open empirical question.

In this paper, we present new evidence that the uniformity conjecture does not hold for Kripke's Gödel case. We report the results of a study comparing the responses of one group of Westerners and one group of East Asians for the suite of probes used by Sytsma and Livengood. What we found is that while the Western participants were quite sensitive to the perspectival ambiguity, the East Asian participants were not. Further, while the Western participants tended to give answers consistent with Kripke's intuition about the Gödel case once the perspectival ambiguity was clarified, a majority of the East Asian participants gave the opposite answer for each probe. Assuming that the clarified Gödel probe tracks semantic intuitions, the semantic intuitions of the Western participants were statistically significantly different from those of the East Asian participants, suggesting that the uniformity conjecture is in fact false for Kripke's Gödel case.

1. The Gödel Case and the Uniformity Conjecture

Machery, Mallon, Nichols, and Stich (2004) challenge a standard methodology used in recent philosophy of language. They note that "philosophers agree that theories of reference for names have to be consistent with our intuitions regarding who or what the names refer to" (B2). This is drawn out by considering a prominent example from the literature— Kripke's (1972) Gödel case.

When Kripke developed this case, he was replying to theories of reference that followed Russell (1905, 1919) in taking an associated definite description to express the meaning of a proper name. The Gödel case purports to show that descriptivist accounts of proper names cannot handle cases where ignorance leads a speaker to associate some description with the wrong individual. In order to illustrate this problem, Kripke sets up his case so that the description generally associated with the name "Gödel" best corresponds with someone other than the man given that name at birth (a man named "Schmidt"). And, as MMNS note, this story has elicited widely shared intuitions amongst philosophers that run counter to traditional descriptivist accounts. Further, many philosophers have held that theories of reference need to accommodate these intuitions.

MMNS challenge this standard methodology by raising doubts about a plausible justification for the practice. With regard to the Gödel case, they write:

> Philosophers typically share the Kripkean intuitions and expect theories of reference to accommodate them.... [We] suspect that most philosophers exploring the nature of reference assume that the Kripkean intuitions are universal. Suppose that semantic intuitions exhibit systematic differences between groups or individuals. This would raise questions about whose intuitions are going to count, putting in jeopardy philosophers' methodology. (B4)

Put another way, MMNS's challenge focuses on the *uniformity conjecture*—the assumption that intuitions about key cases will be (near) universally shared in the general population. MMNS argue that the truth of the uniformity conjecture with regard to intuitions about Kripke's Gödel

case would support the assumption that theories of reference should be consistent with these intuitions. And, conversely, they note that the falsity of the uniformity conjecture would raise significant doubts about this assumption.

Whether or not the uniformity conjecture holds for the Gödel case, however, is rather clearly an empirical question. If it were to turn out that there is variation in intuitions about the case within or across theoretically interesting populations, then advocates of the standard methodology would be under pressure to provide a compelling alternative justification for the practice. Failure of the uniformity conjecture would seriously undermine common practice in philosophy of language. Hence, the burning question is whether the uniformity conjecture actually holds.

Based on work in cultural psychology indicating that there are systematic cognitive differences between Westerners and East Asians (see Nisbett et al. 2001), MMNS predicted that East Asians would be more likely to have descriptivist intuitions than Westerners, and the results of their empirical studies were in line with their prediction. MMNS presented two groups of English-speaking undergraduates—one group from Rutgers University (Westerners) and one group from the University of Hong Kong (East Asians)—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?

MMNS 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, MMNS 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 56.5% of American participants answered (B) compared to 31.5% of Chinese participants.

The difference between the responses for the two groups is statistically significant.[1] From this, MMNS concluded that there is cross-cultural variation in intuitions about the Gödel case. Further, they pointed out that there is significant variation within each sample—neither the Western nor the East Asian participants were uniform in their responses. These results suggest that the uniformity conjecture does not hold, which in turn puts pressure on practitioners of the standard methodology to offer an alternative justification from the uniformity conjecture for the practice of treating their own intuitions about such cases as evidence in constructing theories of reference.

Several objections have been raised against MMNS's work and subsequently been answered in the literature (for criticisms, see Ludwig 2007, Deutsch 2009, Martí 2009, Cullen 2010, Lam 2010, Ichikawa et al. 2012, Devitt 2011, Devitt 2012a, and Devitt 2012b; for replies,

[1] One might worry that although the difference is statistically significant, it is not very *large*. Indeed, according to Cohen's h, the effect size is given by $h = \arcsin(0.565) - \arcsin(0.315) = 0.280$, which would ordinarily be classified as a small effect. However, it is not especially clear that the usual interpretation of effect size is appropriate in this context. Whether an effect of this size is philosophically interesting depends on the details of the arguments launched from its foundation.

see Machery et al. 2009, Machery et al. 2010, Machery 2012a, Machery 2012b, and Machery et al. forthcoming). One objection that has not yet been addressed in the current literature was put forward by Sytsma and Livengood (2011). Focusing on the Western name probe seen above, Sytsma and Livengood argued that there is a perspectival ambiguity in the question that MMNS asked participants. This perspectival ambiguity concerns the epistemic perspective that a participant adopts in reading the definite descriptions given as the answer choices. Sytsma and Livengood noted that there is an asymmetry between what John knows in the Gödel story and what the narrator knows in telling the story. From the narrator's perspective, Schmidt discovered the incompleteness of arithmetic, while Gödel is merely the person who got hold of the manuscript and claimed credit for the work, but as far as John knows, Gödel discovered the theorem. While MMNS assumed that participants would adopt the narrator's perspective in responding to the Gödel probe, participants might have adopted John's perspective instead. But if participants adopted John's perspective, then their responses might not correspond with their semantic intuitions: They might give the "descriptivist" response despite having causal-historical intuitions, since John does not know anything about the theft.

To test whether perspectival ambiguity actually affected responses to MMNS's Gödel probe, Sytsma and Livengood ran a series of studies testing different versions of the probe question. In addition to MMNS's original question, they designed variations intended to implicate that participants should answer either from John's perspective (one variation) or from the narrator's perspective (two variations). The result was a suite of four probes, with each using the Western-name vignette developed by MMNS and only the probe question being varied:

Original: 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?

John's Perspective: When John uses the name "Gödel," does John think he is talking about: (A) the person who the story says really discovered the incompleteness of arithmetic? Or, (B) the person who the story says got hold of the manuscript and claimed credit for the work?

Narrator's Perspective: When John uses the name "Gödel," is he actually talking about: (A) the person who the story says really discovered the incompleteness of arithmetic? Or, (B) the person who the story says got hold of the manuscript and claimed credit for the work?

Clarified Narrator's Perspective: Having read the above story and accepting that it is true, when John uses the name "Gödel," would you take him to actually be talking about: (A) the person who (unbeknownst to John) really discovered the incompleteness of arithmetic? Or, (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?

These four probes were given to Western participants (native English-speakers), using both between-subjects (Studies 1 and 2) and within-subjects designs (Study 3). The overall results were similar in both cases (see Figure 1). Compared to the original probe, participants were significantly less likely to select the "causal-historical" answer (B) when given the John's perspective probe, and significantly more likely to select the "causal-historical" answer when given either of the narrator's perspective probes.

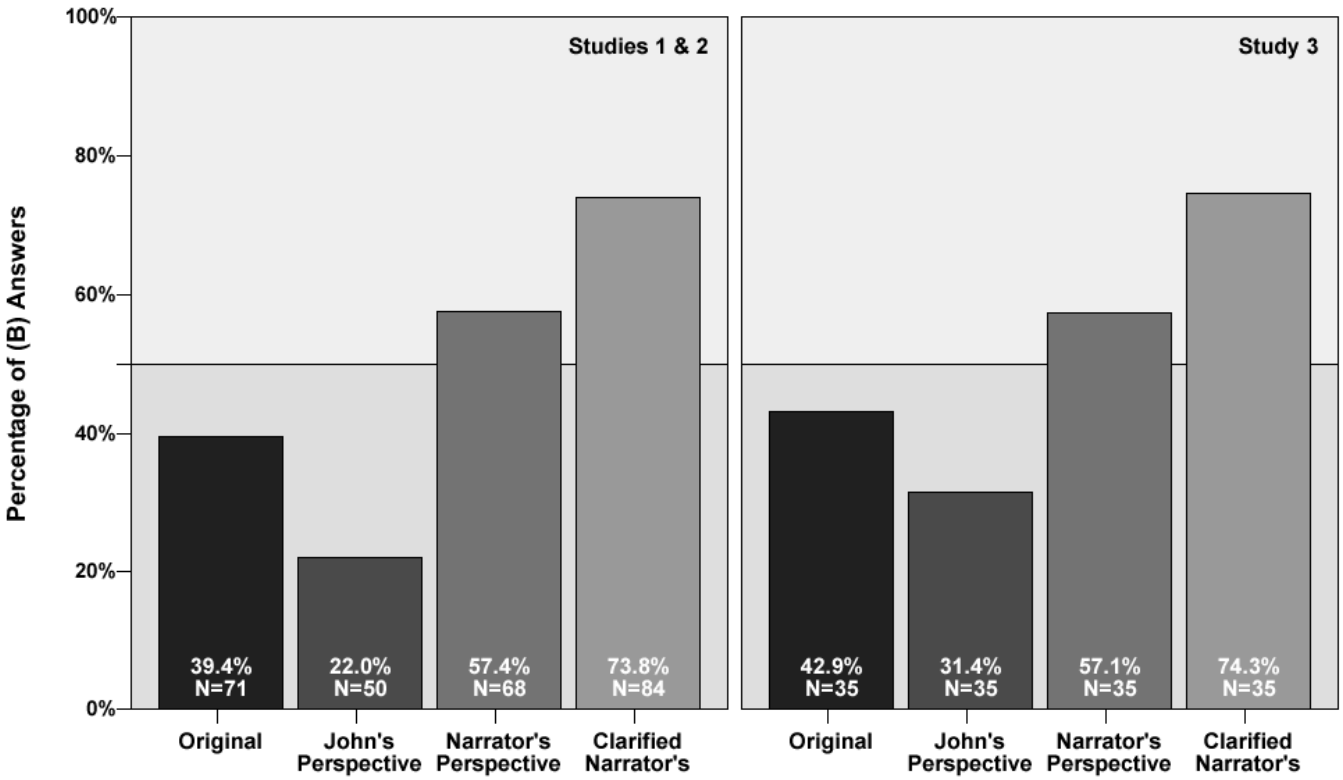

Figure 1: Results of the first three studies from Sytsma and Livengood (2011).

On the basis of these studies, Sytsma and Livengood claimed that MMNS's original Gödel probe is ambiguous for Western participants. They argued that their studies raise significant doubts about whether MMNS's results actually reflect participants' semantic intuitions about the Gödel case, as opposed to the epistemic perspective adopted in reading the probe question. Sytsma and Livengood went on to argue that although they only tested Westerners, their results nonetheless challenge MMNS'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, they argued that it should not be assumed to do so for East Asians.

It is important to note that Sytsma and Livengood did not provide positive evidence for the uniformity conjecture, nor did they argue that MMNS's claim that the uniformity conjecture does not hold for the Gödel case was false. Rather, they questioned the reliability of the evidence that MMNS provided for that claim. In other words, given that the perspectival ambiguity noted had a significant effect on the responses of Western participants, they argued that whether or not the uniformity conjecture holds remains an open question. As such, further experimental work is needed to answer that question, and such work should use probes that control for the perspectival ambiguity identified. In the following section we present a new cross-cultural study that does this. The results suggest that MMNS were right after all: The uniformity conjecture is false.

2. New Evidence against the Uniformity Conjecture

The cross-cultural study we conducted had three primary goals. First, we wanted to test whether or not the perspectival ambiguity identified by Sytsma and Livengood significantly affects the responses of East Asians to MMNS's Gödel probe, as it appears to do for Westerners. Second, we wanted to test whether the cross-cultural variation found by MMNS remained after the probe question was rewritten to control for the perspectival ambiguity. Third, we wanted to test Lam's (2010) charge that the cross-cultural variation found by MMNS reflects that Western participants received the probe in their native language, while East Asian participants received it in a second language. Since both groups received the probes in English, and since it is plausible that native English-speakers have greater competency with the language than those for whom English is a second language, it is possible that MMNS's results primarily reflect differences in linguistic competency rather than semantic intuitions.

The study we developed consists of three stages. In the first stage we presented Sytsma and Livengood's suite of Gödel probes to American participants in English to get a baseline response rate for each of the four probes. We then translated these probes into Japanese and, in the second stage of the study, presented the translated probes to participants at five different Japanese universities. Finally, in the third stage of the study, we had the Japanese-language versions reverse-translated back into English, and we presented the reverse-translated probes to a new set of American participants to check for potential problems with the Japanese translation. The full set of probes is given in the appendix.

For each stage we used a between-subjects design, with each participant receiving only one of the four types of probes. Following Sytsma and Livengood, (A) and (B) answer choices were counterbalanced for order, although we will report the results using the ordering given in the previous section: "(A)" for the descriptivist answer choice, "(B)" for the causal-historical answer choice. In addition, participants were asked for demographic information and about any training they had in philosophy. The first and third stages of the study were conducted online, while the second stage was conducted in person in university classrooms.[2] Participants were excluded if they were under 18 years of age, did not complete the survey, had previously participated, or had more than minimal training in philosophy.[3] In addition, participants in Stages 1 and 3 were excluded if they were not native speakers of English born in and currently residing in the United States. Participants in Stage 2 were excluded if they were not native speakers of Japanese born in and currently residing in Japan. Responses were collected from 596 participants

[2] Online responses were collected through the Philosophical Personality website (philosophicalpersonality.com).
[3] Participants were counted as having more than minimal training in philosophy if they were philosophy majors, had completed a degree with a major in philosophy, or had taken graduate-level courses in philosophy.

in Stage 1, from 221 participants in Stage 2, and from 583 participants in Stage 3.[4] The results are shown in Table 1 below, which gives the proportion of (B) responses for each probe by stage.

	Stage 1	**Stage 2**	**Stage 3**
	(English Baseline)	**(Japanese Translation)**	**(Reverse-Translation)**
Original:	54.7% (N=161)	29.9% (N=67)	60.0% (N=140)
John's Perspective:	31.3% (N=144)	30.8% (N=52)	20.0% (N=145)
Narrator's Perspective:	57.4% (N=148)	39.2% (N=51)	67.3% (N=150)
Clarified Narrator:	68.5% (N=143)	41.2% (N=51)	56.8% (N=148)

Table 1: Proportion of (B) answers by stage.

The results for American participants in Stage 1 replicate the pattern of variation across probes found by Sytsma and Livengood, although it should be noted that the percentage of (B) answers for the original probe is more in keeping with that originally reported by MMNS. As predicted, the percentage of (B) answers is significantly lower for the John's perspective probe than for the original probe, and is significantly higher for the clarified narrator's perspective probe than for the original probe.[5] Hence, our Stage 1 results provide further support for the

[4] Participants in Stage 1 were 73.8% female, with an average age of 35.2 years, and ranging in age from 18–79 years old. Participants in Stage 2 were 45.7% female, with an average age of 20.1 years, and ranging in age from 18–39 years old. Participants in Stage 3 were 75.1% female, with an average age of 36.9 years, and ranging in age from 18–79 years old.

[5] A χ^2 test of independence indicates that Probe and Response are associated for Americans (χ^2=42.54, df=3, p=3.08e-9). Fisher's exact test yields a p-value with the same order of magnitude (p=2.25e-9). We tested the expectation—derived from Sytsma and Livengood's results—that the four probes would be ordered by the proportion of (B) answers from least to greatest like John's Perspective, Original, Narrator's Perspective, and Clarified Narrator's Perspective. Hence, we conducted three one-sided tests of proportions. We do not correct for multiple comparisons here, since some authors argue that such corrections are unnecessary when comparisons are planned in advance (see for example, Chapter 22 in Motulsky 2010). The reader may make whatever corrections seem prudent. We found that the proportion of (B) answers in the John's Perspective probe was statistically significantly smaller than the proportion of (B) answers in the Original probe (χ^2=15.99, df=1, p=3.17e-5) and that the proportion of (B) answers in the Original probe was statistically significantly smaller than the proportion of (B) answers in the Clarified Narrator's Perspective probe (χ^2=3.38, df=1, p=0.033). However, we could not reject the hypothesis that the proportion of (B) answers in the Narrator's Perspective probe was equal to the proportion of (B) answers in the Clarified Narrator's Perspective probe (χ^2=0.141, df=1, p=0.353). At the request of an anonymous referee, we also calculated 95% confidence intervals for the proportion of (B) answers for each probe. The intervals are (0.466, 0.624), (0.239, 0.396), (0.490, 0.654), and (0.601, 0.759) for the original, John's perspective, narrator's perspective, and clarified narrator's perspective probes, respectively. Using a uniform prior, we also calculated 95% highest-density credible intervals for the proportions to be (0.470, 0.622), (0.241, 0.391), (0.494, 0.652), and (0.607,

claim that the perspectival ambiguity that Sytsma and Livengood describe significantly affects the answers given by Westerners.

In contrast, the variation across the four probes for Japanese participants in Stage 2 was minimal. We found that in each case the percentage of (B) answers was within 10 percentage points of the 31.5% that MMNS reported for East Asian participants on the original probe, with only a minority of participants giving the (B) answer in each case. In line with the findings of MMNS, this suggests that *contra* Lam the language of presentation (native language versus second language) does not explain the cross-cultural variation found by MMNS. Further, emphasizing either John's perspective or the narrator's perspective had minimal impact on the responses. For example, while the percentage of (B) answers for American participants was 37.2 percentage points higher for the clarified narrator's perspective probe than it was for the John's perspective probe, for Japanese participants the percentage of (B) answers was only 10.4 points higher for the clarified narrator's perspective probe than it was for the John's perspective probe. This indicates that the perspectival ambiguity identified by Sytsma and Livengood does not have a major impact on the responses of at least one group of East Asians.

Most importantly, the results of the first two stages of our study suggest that significant cross-cultural variation remains even after controlling for the perspectival ambiguity and the language of presentation. Whereas the probe a participant saw made a significant difference to the responses of American participants, the probe a participant saw made no statistically significant difference to the responses of Japanese participants.[6] Moreover, consistent with

---

0.757). Finally, using a prior informed by the results reported in Sytsma and Livengood (2011), we calculated 95% HDIs for the proportions to be (0.431, 0.550), (0.236, 0.353), (0.512, 0.634), and (0.653, 0.763).

[6] A χ^2 test of independence failed to detect any association between Probe and Response for Japanese speakers (χ^2=2.45, df=3, p=0.485). Fisher's exact test yields a similar p-value (p=0.488). We used simulations based on the proportions observed among English speakers to estimate the power to detect a similar effect at the 0.05 significance level with our Japanese sample sizes. Probe and Response were associated according to a χ^2 test at the 0.05 significance level in 928 out of 1000 simulations. They were associated according to Fisher's exact test at the 0.05

MMNS's findings, Japanese participants were less likely than American participants to choose the (B) answer in response to the Original probe, the Narrator's Perspective probe, and the Clarified Narrator's Perspective probe.[7] In fact, we find that while a majority of the American participants give that response in each of those cases, only a minority of the Japanese participants do. Based on these results, we tentatively conclude that MMNS were correct: The uniformity conjecture does not hold for the Gödel case.[8]

---

significance level in 933 out of 1000 simulations. We calculated 95% confidence intervals for the original, John's perspective, narrator's perspective, and clarified narrator's perspective probes to be (0.196, 0.424), (0.191, 0.453), (0.262, 0.539), and (0.279, 0.558), respectively. Using a uniform prior, we calculated analogous 95% highest density credible intervals to be (0.199, 0.413), (0.195, 0.439), (0.268, 0.527), and (0.285, 0.547).

[7] One-sided χ^2 tests of proportions showed that the proportion of (B) answers in the Original probe was statistically significantly smaller for Japanese than for Americans (χ^2=10.71, df=1, p=0.00053), the proportion of (B) answers in the Narrator's Perspective probe was statistically significantly smaller for Japanese than for Americans (χ^2=4.35, df=1, p=0.0186), and the proportion of (B) answers in the Clarified Narrator's Perspective probe was statistically significantly smaller for Japanese than for Americans (χ^2=10.74, df=1, p=0.00053). We could not reject the hypothesis that the proportion of (B) answers for the John's Perspective probe was the same for Americans and Japanese (χ^2=0, df=1, p=0.5). An anonymous referee worried that the differences we would like to attribute to culture might actually be due to systematic differences in the gender and age composition of the samples. The Americans we sampled were more likely to be female than the Japanese we sampled, and they were more likely to be older. In order to control for the gender disparity, we did the following 1,000 times: We randomly selected a sub-sample from our American female participants so that we had a new, smaller sample that was 50% male and 50% female—basically in line with our Japanese sample; we then made the same comparisons as above and recorded how often we got a significant result at the 0.05 and at the 0.001 levels in each case. The results were as follows: In all 1,000 repetitions, the percentage of (B) answers to the Original probe among the Japanese was statistically significantly smaller than the percentage among the Americans at the 0.05 level, and in 424 of those, the Japanese percentage was statistically significantly smaller at the 0.001 level; in no repetitions was the percentage of (B) answers to the John's Perspective probe statistically significantly smaller among the Japanese at either the 0.05 or 0.001 level; in 776 repetitions, the percentage of (B) answers to the Narrator's Perspective probe, the Japanese percentage was statistically significantly smaller at the 0.05 level, and in 2 of those the Japanese percentage was also smaller at the 0.001 level; finally, in all 1,000 repetitions, the percentage of (B) answers to the Clarified Narrator's probe among the Japanese was statistically significantly smaller at the 0.05 level, and in 715 of those the Japanese percentage was also smaller at the 0.001 significance level. We found it practically impossible to directly control for the differences in age composition without serious loss of power. However, we controlled indirectly by fitting a logistic regression model predicting responses based on age using our Western data. If the difference in age distributions mattered, we would expect to see a significant coefficient for age in the model. But age was not a statistically significant predictor of participant responses for any of the four probes (with p-values of 0.790, 0.802, 0.940, and 0.468). Hence, we have reason to doubt that differences in the age compositions of the two samples accounts for the differences in responses.

[8] It is worth noting that our study did not specifically control for a third potential confound that has been pressed in the literature—that MMNS's original Gödel probe question is ambiguous with regard to asking about the speaker's reference or semantic reference of John's use of the name "Gödel" (Ludwig 2007, Deutsch 2009). Nonetheless, while this ambiguity is distinct from the perspectival ambiguity, it is plausible that the two narrator's perspective probes clarify it as well, as discussed by Sytsma and Livengood (2011). Further, Machery, Sytsma, and Deutsch (forthcoming) present independent evidence that the speaker's reference/semantic reference ambiguity does not explain away the cross-cultural variation found by MMNS.

One potential issue with the comparison between the responses of American and Japanese participants in Stages 1 and 2 of our study, however, is that they might reflect the translation used in Stage 2 rather than differences in semantic intuitions. The third stage of our study was designed to control for problems with the translation. If the difference between American and Japanese responses observed in Stages 1 and 2 is due to something in the translation rather than to semantic intuitions, then we would expect the responses of Americans to the reverse-translated probes used in Stage 3 to be different from those received in Stage 1. Specifically, we would expect the responses of Americans to the reverse-translated probes to be more uniform across the four conditions. But that is not at all what we found. The patterns of response in Stages 1 and 3 were statistically very similar to each other, and they were statistically very different from the pattern in Stage 2.[9] Moreover, with the exception of the Clarified Narrator's Perspective probe, which we discuss in greater detail below, the variation in American responses to the reverse-translated probes was *more pronounced* than the variation in American responses to the baseline (un-translated) probes: the exact opposite of the prediction made on the assumption that the differences between Stages 1 and 2 were due to translation.

Earlier, we compared the results of Stages 1 and 2, and we found that Japanese participants were less likely than American participants to choose the (B) answer in response to the Original probe, the Narrator's Perspective probe, and the Clarified Narrator's Perspective probe. Comparing the results of Stages 2 and 3, we again found that Japanese participants were less likely than American participants to choose the (B) answer in response to the Original probe,

[9] We conducted a Cochran-Mantel-Haenszel test for conditional independence to compare the responses in Stages 1 and 3. Specifically, we tested the null hypothesis that Stage (either 1 or 3) and Response (either A or B) were independent conditional on Probe (1-4). We could not reject the null hypothesis (M^2=0.347, df=1, p=0.556). By contrast, the same test using data from Stages 1 and 2 *did* reject the null hypothesis (M^2=21.24, df=1, p=4.05e-6). And the null hypothesis was rejected when the data from all three stages were analyzed together (M^2=23.59, df=2, p=7.55e-6).

the Narrator's Perspective probe, and the Clarified Narrator's Perspective probe.[10] However, we had expected that in Stage 3, as in Stage 1, the proportion of (B) answers for the Clarified Narrator's Perspective probe would be greater than the proportion of (B) answers for the Narrator's Perspective probe—an expectation that was not fulfilled.[11] The unexpectedly low proportion of (B) answers for the reverse-translated Clarified Narrator's Perspective probe does not raise any issues for the other probes (for the reasons already given above), but it does call for some explanation. In the remainder of this section, we will give reasons for thinking that the translation problem arose in the reverse-translation step.

Recall that the baseline English version of the question for the Clarified Narrator's Perspective probe reads as follows:

> Having read the above story and accepting that it is true, when John uses the name "Gödel," would you take him to actually be talking about: (A) the person who (unbeknownst to John) really discovered the incompleteness of arithmetic? Or, (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?

Translated into Japanese, the question reads:

---

[10] One-sided χ^2 tests of proportions showed that the proportion of (B) answers in the Original probe was statistically significantly smaller for Japanese participants than for American participants (χ^2=15.29, df=1, p=4.61e-5), the proportion of (B) answers in the Narrator's Perspective probe was statistically significantly smaller for Japanese participants than for American participants (χ^2=11.41, df=1, p=3.65e-4), and the proportion of (B) answers in the Clarified Narrator's Perspective probe was statistically significantly smaller for Japanese participants than for American participants (χ^2=3.10, df=1, p=0.0393). As before (see Footnote 5), we did not correct for multiple comparisons. We calculated 95% confidence intervals for the reverse-translated probes in Stage 3 to be (0.514, 0.681), (0.140, 0.276), (0.591, 0.746), and (0.484, 0.648). Similarly, assuming a uniform prior, 95% highest density credible intervals for the reverse-translated probes are (0.518, 0.678), (0.141, 0.270), (0.596, 0.745), and (0.488, 0.645).

[11] Parallel to what we did in Stage 1, we conducted three one-sided tests of proportions in order to test an expected ordering for the proportion of (B) answers in the reverse-translated probes. As before, we do not correct for multiple comparisons. We found that the proportion of (B) answers in the John's Perspective probe was statistically significantly smaller than the proportion of (B) answers in the Original probe (χ^2=45.97, df=1, p=6.00e-12). The proportion of (B) answers in the Original probe was smaller than the proportion of (B) answers in the Narrator's Perspective probe, but the difference was not statistically significant (χ^2=1.38, df=1, p=0.120). And the difference between the Narrator's Perspective probe and the Clarified Narrator's Perspective probe did not even run in the right direction (χ^2=3.105, df=1, p=0.961).

上のストーリーが実話だったとしよう。ジョンが「ゲーデル」という名前を使うとき、あなたは彼が語っているのは実際のところ誰についてだと思うか？
（A）算術の不完全性を本当に発見した人物（ジョンはこのことを知らない）
（B）算術の不完全性を発見したと広く信じられているが、実際には、手稿を手に入れ、その業績を自分のものだと主張した人物

When reverse-translated into English, the question reads like this:

> Suppose that the above story is true. When he uses the name "Gödel," whom do you think he is really talking about?
> (A) The person who really discovered the incompleteness of arithmetic (John does not know this)
> (B) The person who is widely believed to have discovered the incompleteness of arithmetic but in reality obtained the manuscript and claimed that it is his work

Looking at the reverse-translated Clarified Narrator's Perspective probe, we guessed that the problem had something to do with how the parenthetical remark in answer (A) was being translated. Specifically, we found that the parenthetical in the (A) answer choice reads awkwardly.

To test whether some infelicity in the reverse-translated (A) answer choice explains the drop in (B) answers that we found for the reverse-translated Clarified Narrator's Perspective probe, we conducted a further study in which we set the (A) answer choice back to its original form: "The person who (unbeknownst to John) really discovered the incompleteness of arithmetic." The probe was otherwise the same as the version used in Stage 3. We gave the revised probe to 143 American participants online, adopting the same restrictions used in Stages 1 and 3.[12] We found that with the corrected (A) answer, the percentage of (B) answers rose from 54.1% to 69.9%, which is in line with the 68.5% found for the pre-translation probe and a significantly greater percentage of (B) answers than we observed among Japanese participants.[13]

[12] Participants were 74.1% female, with an average age of 34.7 years, and ranging in age from 18–78 years old.
[13] A 95% confidence interval is given by (0.616, 0.772). A 95% credible interval based on a uniform prior is given by (0.621, 0.770). A one-sided χ^2 test of proportions showed that the proportion of (B) answers in the corrected Clarified Narrator's Perspective probe was statistically significantly smaller for Japanese participants than for American participants (χ^2=12.05, df=1, p=2.60e-4). We carried out a sub-sampling scheme identical to the one

Identifying the awkward parenthetical as the problem in the reverse-translated version of the Clarified Narrator's Perspective probe, however, does not tell us whether it was introduced in the original translation, which should reduce the confidence one has in our results, or in the reverse-translation, which should not make any difference to the confidence one has in our results.[14] Fortunately, there is good reason to think that the problem arises in the reverse-translation.

One issue with the parenthetical in answer choice (A) in the reverse-translation is that it is unclear what it is that John does not know. Specifically, the word "this" at the end of the parenthetical feels like it should be followed by another word that makes clear what it refers to, and several options are available—including "this fact," "this person," and "this theorem." Not all of the plausible referents serve to emphasize the narrator's perspective, however. In contrast, the Japanese translation of the answer choice is not ambiguous in this way. Thus, "このこと" (kono-koto) in the parenthetical literally means "this fact." Hence, while there are several possible referents for the "this" in the reverse-translated answer choice, there is only one for the Japanese answer choice. As such, we are convinced that the comparison between the baseline English version and the Japanese translation is sound.[15] Replacing the result for the Clarified

---

described in Footnote 7 in order to control for differences in gender composition. We found that in 1,000 out of 1,000 repetitions, the percentage of Japanese (B) answers was statistically significantly smaller than the percentage of American (B) answers at the 0.05 level, and we found that in 513 of those 1,000 repetitions, the percentage was smaller at the 0.001 significance level.

[14] We think the *degree* to which one should lose confidence in the overall results is small even if the Clarified Narrator's Perspective probe is dropped from consideration, since the other three probes did not appear to have any translation problems.

[15] To further test the source of the problem with the reverse-translated Clarified Narrator's Perspective probe, we simply added the word "fact" to the end of the parenthetical used in Stage 3. We gave the revised probe to 142 American participants online, again adopting the same restrictions used in Stages 1 and 3. Participants were 62.7% female, with an average age of 43.0 years, and ranging in age from 18–82 years old. As predicted, the percentage of (B) answers for the revised version of the probe—60.6%—is higher than we found for the unrevised probe in Stage 3. A 95% confidence interval is given by (0.520, 0.686). A 95% credible interval based on a uniform prior is given by (0.524, 0.683). More importantly, the percentage of (B) answers is not statistically significantly different from the percentage found for the English baseline version of the probe in Stage 1. A χ^2 test of proportions showed that the proportion of (B) answers for the version of the reverse-translated Clarified Narrator's Perspective probe with "fact"

Narrator's Perspective probe in Stage 3 with the result for the edited version, the pattern of results is especially clear, as seen in the plot in Figure 2.

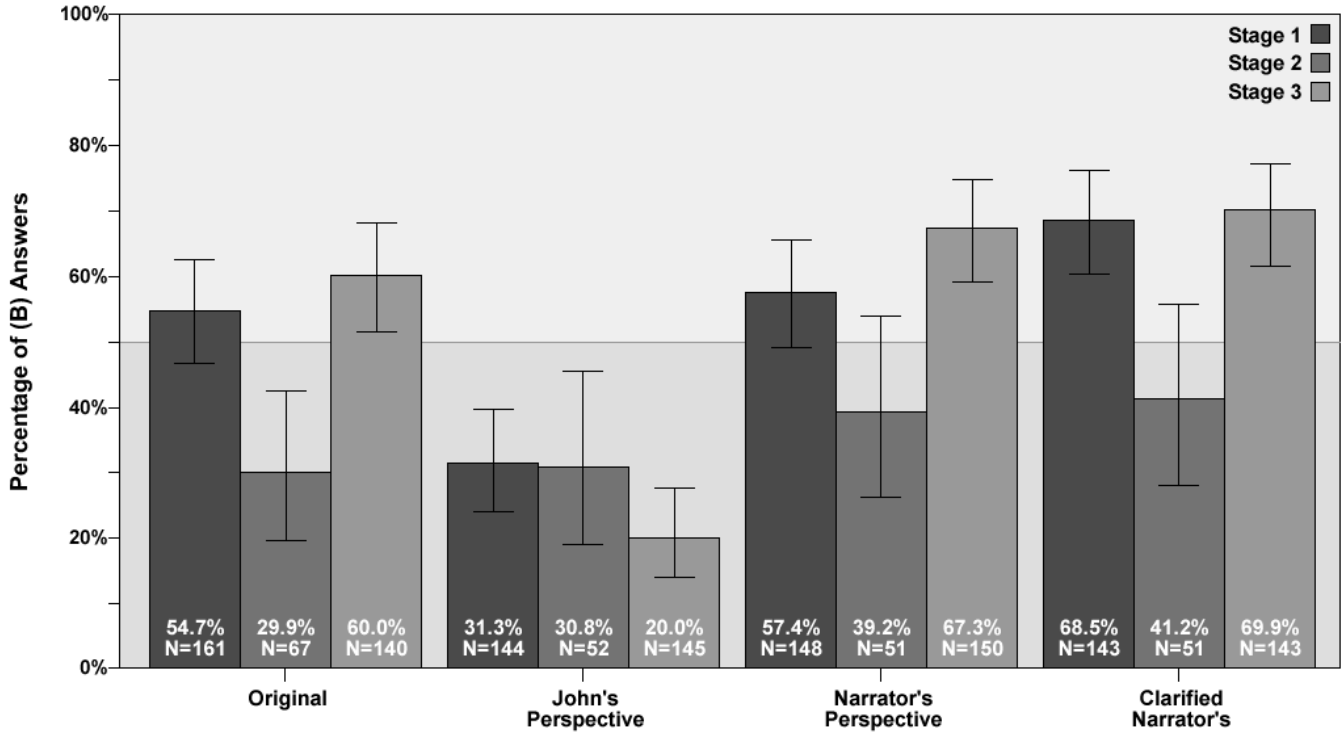

Figure 2: Plot of the results of our cross-cultural using the percentage for the edited version of the clarified narrator's probe in Stage 3, with 95% confidence intervals.

3. Conclusion

Our results support the claim that the uniformity conjecture does not hold for Kripke's Gödel case, and hence, our results confirm MMNS's conclusions, at least in rough outline. Japanese participants, a group of East Asians not previously studied by MMNS or their critics, were

---

added was not statistically significantly different from the proportion for participants in Stage 1 (χ^2=1.64, df=1, p=0.200). Further, it is statistically significantly higher than the percentage found for the translated probe in Stage 2. A one-sided χ^2 test of proportions showed that the proportion of (B) answers for the version of the reverse-translated Clarified Narrator's Perspective probe with "fact" added was statistically significantly higher than the proportion for Japanese participants in Stage 2 (χ^2=4.95, df=1, p=0.013). In 1,000 repetitions designed to control for the difference in gender composition, we found a statistically significant difference in the predicted direction 943 times at the 0.05 level but never at the 0.001 significance level.

significantly less likely to give responses consistent with the causal-historical intuition about the Gödel case than were American participants—even after controlling for the perspectival ambiguity noted by Sytsma and Livengood (2011).

Philosophers of language interested in defending the standard methodology with respect to theories of reference may have taken comfort from Sytsma and Livengood (2011), reasoning as follows: Perhaps East Asians are simply more likely than Westerners to adopt John's perspective in responding to MMNS's original probe. If so, then the cross-cultural variation found by MMNS might simply be taken to reflect cultural differences with regard to perspective-taking, not cultural differences with regard to semantic intuitions. And this is consistent with the findings of Nisbett and colleagues that MMNS appeal to.[16] Thus, it might be thought that if East Asians and Westerners could both be made to read the Gödel probe from the narrator's perspective, then they would respond to it in the same way. Our cross-cultural study substantially undermines this reply.

As such, our results support MMNS's challenge to the standard methodology. But we want to be careful on this point in two ways. First, MMNS's challenge admits various formulations that trade off boldness and security. More securely, our results challenge the use of one's own intuitions about cases like Kripke's Gödel example (i.e., cases involving ignorance and error) as evidence for a given account of reference. More boldly, the results could fuel (a) a challenge to the use of one's own intuitions as evidence in theorizing about reference more

[16] For example, Sytsma and Livengood (2011, 330-331) argue: "Before seeing any data, there was just as much reason to predict that East Asians would be more likely than Westerners to answer (A) on the basis of the epistemic ambiguity as there was for Machery et al. to predict a difference on the basis of different semantic intuitions. Interestingly, this claim follows from the very same body of empirical work that Machery et al. point to in framing their prediction that Westerners and East Asians would differ in their intuitions about the Gödel probe.... In particular, they call on the work of Nisbett and colleagues (2001) indicating a range of cultural differences, which they collect under the heading of 'holistic vs. analytic thought'.... Holistic thought is supposed to be characteristic of Easterners.... Analytic thought is supposed to be characteristic of Westerners.... The more holistic way to read Machery et al.'s Gödel probe is in terms of the beliefs that would be ascribed to John by his interlocutors. The more analytic way to read the probe is in terms of the beliefs that we the readers have as informed by an omniscient narrator."

generally, (b) a challenge to the use of one's own intuitions as evidence in theorizing about semantics, or even (c) a challenge to the use of one's own intuitions as evidence in philosophy as a whole. While we are generally skeptical of the use of one's own intuitions as evidence in philosophical theorizing—and while our results add to a growing body of data that casts doubt on this practice (see, for example, the discussion in Alexander, Mallon, and Weinberg 2010)—we will restrict ourselves to the more secure claim.

Second, it should be repeated that MMNS's empirical challenge targets *one* justification that has been offered for the practice of using one's own intuitions as evidence in theorizing about reference. Specifically, this practice has been justified by claiming that the uniformity conjecture holds for the intuitions at issue. On the basis of their empirical results, MMNS contend that the uniformity conjecture does not hold for the Gödel case, and our results support this contention. Nonetheless, the standard methodology might be justified in other ways. For example, Devitt (2011) argues that the intuitions of philosophers about the cases at issue are more reliable than those of non-philosophers because philosophers are more expert with regard to these issues. While we are skeptical of so-called "expertise defense," our results do not bear on this alternative justification.[17] As such, we again restrict ourselves to the more cautious conclusion in this paper: Responsible philosophers working on reference should not continue to assume without empirical support that the uniformity conjecture holds for cases like Kripke's Gödel example.

[17] The expertise defense is discussed briefly in MMNS's original paper; see also Machery 2012a, 2012b.

References

Alexander, J., R. Mallon, and J. Weinberg (2010). "Accentuate the Negative," *Review of Philosophy and Psychology* 1(2), 297-314.

Cullen, S. (2010). "Survey-Driven Romanticism," *Review of Philosophy and Psychology* 1(2), 275-296.

Deutsch, M. (2009). "Experimental Philosophy and the Theory of Reference," *Mind & Language* 24(4), 445-466.

Devitt, M. (2011). "Experimental Semantics," *Philosophy and Phenomenological Research* 82(2), 418-435.

Devitt, M. (2012a). "Whither Experimental Semantics?" *Theoria* 27(1), 37-54.

Devitt, M. (2012b). "Semantic Epistemology: Response to Machery," *Theoria* 27(2), 229-233.

Ichikawa, J., I. Maitra, and B. Weatherson (2012). "In Defense of a Kripkean Dogma," *Philosophy and Phenomenological Research* 85(1), 56-68.

Kripke, S. (1972). *Naming and Necessity*, Oxford: Blackwell.

Lam, B. (2010). "Are Cantonese Speakers Really Descriptivists? Revisiting Cross-Cultural Semantics," *Cognition* 115(2), 320-332.

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

Machery, E. (2012a). "Expertise and Intuitions about Reference," *Theoria* 27(1), 37-54.

Machery, E. (2012b). "Semantic Epistemology: A Brief Response to Devitt," *Theoria* 27(2), 223-227.

Machery, E., R. Mallon, S. Nichols, and S. Stich (2004). "Semantics, Cross-Cultural Style," *Cognition* 92: B1-B12.

Machery, E., C. Olivola, and M. de Blanc (2009). "Linguistic and Metalinguistic Intuitions in the Philosophy of Language," *Analysis* 69(4), 689-694.

Machery, E., M. Deutsch, R. Mallon, S. Nichols, J. Sytsma, and S. Stich (2010). "Semantic Intuitions: Reply to Lam," *Cognition* 117(3), 361-366.

Machery, E., J. Sytsma, and M. Deutsch (forthcoming). "Speaker's Reference and Cross-Cultural Semantics," in A. Bianchi (Ed.), *On Reference*. Oxford: Oxford University Press.

Machery, E., R. Mallon, S. Nichols, and S. Stich (forthcoming). "If Folk Intuitions Vary, Then What?" *Philosophy and Phenomenological Research*.

Martí, G. (2009). "Against Semantic Multiculteralism," *Analysis* 69(1), 42-49.

Motulsky, H. (2010). *Intuitive Biostatistics*, 2nd Edition. Oxford: Oxford University Press.

Nisbett, R., K. Peng, I. Choi, and A. Norenzayan (2001). "Culture and Systems of Thought: Holistic Versus Analytic Cognition," *Psychological Review* 108, 291–310.

Russell, B. (1905). "On Denoting," *Mind* 14, 479–93.

Russell, B. (1919). "Knowledge by Acquaintance and Knowledge by Description," in *Mysticism and Logic*: London: George Allen and Unwin, 1917.

Sytsma, J. and J. Livengood (2011). "A New Perspective Concerning Experiments on Semantic Intuitions," *Australasian Journal of Philosophy* 89, 315-332.

Appendix

Stage 1, Pre-Translation:

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.

Question for Original Probe:

When John uses the name "Gödel," is he 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?

Question for John's Perspective Probe:

Having read the above story and accepting that it is true, when John uses the name "Gödel," does John think he is talking about:
(A) the person who the story says really discovered the incompleteness of arithmetic?
(B) the person who the story says got hold of the manuscript and claimed credit for the work?

Question for Narrator's Perspective Probe:

Having read the above story and accepting that it is true, when John uses the name "Gödel," is he actually talking about:
(A) the person who the story says really discovered the incompleteness of arithmetic?
(B) the person who the story says got hold of the manuscript and claimed credit for the work?

Question for Clarified Narrator's Perspective Probe:

Having read the above story and accepting that it is true, when John uses the name "Gödel," would you take him to actually be talking about:
(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?

Stage 2, Japanese Translation:

ジョンという人物がいたとしよう。ジョンは大学時代に、ゲーデルとは、算術の不完全性と呼ばれる、ある重要な数学の定理を証明した人物である、と教わった。ジョンは数学が大変得意であり、不完全性定理についてその正確な内容を述べることができる。ジョンはその定理の発見者はゲーデルだと思っているが、彼がゲーデルについて聞いたことがあるのはこれだけだった.

ここで、ゲーデルはこの定理の考案者ではなかったと想定してみよう。実際には、「シュミット」と呼ばれる男――その遺体はウィーンにおいて何十年も前に不可解な状況で発見された――がその業績を成し遂げたのである。シュミットの友人であったゲーデルは何らかの手段をもちいてその手稿を手に入れ、その業績を自分のものだと主張したのである。以来、その業績はゲーデルのものとされている。こうして、彼は算術の不完全性を証明した人物として知られることになった.

「ゲーデル」という名前を聞いたことのある人のほとんどはジョンと同じである。つまり、ゲーデルについて聞いたことがあるのは、ゲーデルが不完全性定理を発見したということだけである.

Question for Original Probe:

「ゲーデル」という名前を使うとき、ジョンが語っているのは誰についてか?
(A) 算術の不完全性を本当に発見した人物
(B) 手稿を手に入れ、その業績を自分のものだと主張した人物

Question for John's Perspective Probe:

上のストーリーが実話だったとしよう。「ゲーデル」という名前を使うとき、ジョンは自分が語っているのは誰についてだと思っているか?
(A) 算術の不完全性を本当に発見した人物
(B) 手稿を手に入れ、その業績を自分のものだと主張した人物

Question for Narrator's Perspective Probe:

上のストーリーが実話だったとしよう。「ゲーデル」という名前を使うとき、ジョンは実際のところ誰について語っているのか?
(A) 算術の不完全性を本当に発見した人物
(B) 手稿を手に入れ、その業績を自分のものだと主張した人物

Question for Clarified Narrator's Perspective Probe:

上のストーリーが実話だったとしよう。ジョンが「ゲーデル」という名前を使うとき、あなたは彼が語っているのは実際のところ誰についてだと思うか?
(A) 算術の不完全性を本当に発見した人物（ジョンはこのことを知らない）
(B) 算術の不完全性を発見したと広く信じられているが、実際には、手稿を手に入れ、その業績を自分のものだと主張した人物

Stage 3, Reverse-Translation:

Suppose that there is a person named John. In college, John was taught that Gödel is the person who proved an important theorem in mathematics, called the incompleteness of arithmetic. John is very good at mathematics and is able to describe the precise content of the incompleteness theorem. John believes that the discoverer of the theorem is Gödel, but this is all he has heard about Gödel.

Now suppose that Gödel was not the originator of the theorem. In reality, the man named "Schmidt"—whose body was found many decades ago in Vienna under mysterious circumstances—accomplished the work in question. Gödel, who was Schmidt's friend, somehow obtained the manuscript and claimed that it was his work. The work is thereafter attributed to Gödel. In this way, he has come to be known as the one who proved the incompleteness of arithmetic.

Most of the people who have heard the name "Gödel" are like John. That is, what they have heard about Gödel is only that Gödel discovered the incompleteness of arithmetic.

Question for Original Probe:

When he uses the name "Gödel," whom is John talking about?
(A) The person who really discovered the incompleteness of arithmetic
(B) The person who obtained the manuscript and claimed that it is his work

Question for John's Perspective Probe:

Suppose that the above story is true. When he uses the name "Gödel," whom does John think he is talking about?
(A) The person who really discovered the incompleteness of arithmetic
(B) The person who obtained the manuscript and claimed that it is his work

Question for Narrator's Perspective Probe:

Suppose that the above story is true. When he uses the name "Gödel," whom is John really talking about?
(A) The person who really discovered the incompleteness of arithmetic
(B) The person who obtained the manuscript and claimed that it is his work

Question for Clarified Narrator's Perspective Probe:

Suppose that the above story is true. When he uses the name "Gödel," whom do you think he is really talking about?
(A) The person who really discovered the incompleteness of arithmetic (John does not know this)
(B) The person who is widely believed to have discovered the incompleteness of arithmetic but in reality obtained the manuscript and claimed that it is his work

Question for First Edited Version of Clarified Narrator's Perspective Probe:

> Suppose that the above story is true. When he uses the name "Gödel," whom do you think he is really talking about?
> (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 in reality obtained the manuscript and claimed that it is his work

Question for Second Edited Version of Clarified Narrator's Perspective Probe:

> Suppose that the above story is true. When he uses the name "Gödel," whom do you think he is really talking about?
> (A) The person who really discovered the incompleteness of arithmetic (John does not know this fact)
> (B) The person who is widely believed to have discovered the incompleteness of arithmetic but in reality obtained the manuscript and claimed that it is his work