

AUTHENTIC AND APPARENT EVIDENCE GETTIER CASES ACROSS AMERICAN AND INDIAN NATIONALITIES

Chad Gonnerman
Banjit Singh
Grant Toomey

Abstract

We present three experiments that explore the robustness of the *authentic-apparent effect*—the finding that participants are less likely to attribute knowledge to the protagonist in apparent- than in authentic-evidence Gettier cases. The results go some way towards suggesting that the effect is robust to assessments of the justificatory status of the protagonist's belief. However, not all of the results are consistent with an effect invariant across two cultural contexts: the US and India.

Keywords

Experimental epistemology, Gettier intuitions, cross cultural, protagonist projection, knowledge, epistemic justification, restrictionist challenge

Gettier cases are a set of vignettes of special interest to epistemologists. The intuitions that they are assumed to elicit are often taken to show that having a justified true belief does not suffice for knowledge, a conclusion which runs counter, or so some say, to a view about the conditions of knowledge held by most all philosophers for nearly two millennia. More recent empirical work has shown that the psychology of intuitive reactions to Gettier cases are complicated. For one, the current record indicates that these reactions are not universal. By declaring as much, at least at this point, we aren't so much referring to work on potential cross-cultural differences in *Gettier intuitions or judgments*—ones to the effect that the protagonist in a Gettier case lacks knowledge despite having a justified true belief. Rather, we have in mind a kind of variability in Gettier reactions that Jennifer Nagel appears to deny when she writes, "Ordinary people do tend to agree with professional philosophers in judging that the agents in Gettier-type cases lack knowledge" (2014, p. 111). There is reason to think that this sentence is too strong. For there appears to be some degree of cross-*case* variability in Gettier intuitions. Some Gettier cases seem worse than others at generating an intuition to the effect that the protagonist has a justified true belief that doesn't amount to knowledge. So, the extent to which regular folk agree with professional philosophers may well depend on which Gettier case we are talking about.

One of the clearest illustrations of the uneven potency of Gettier cases comes from Starman and Friedman (2012). After reporting the results of four experiments in which most of their participants ascribed knowledge to multiple Gettier protagonists (and thus already violated the expectations of many philosophers), Starman and Friedman describe one final experiment. Its purpose was to compare reactions to two kinds of Gettier vignettes: *authentic-* versus *apparent-evidence cases*. Exactly what distinguishes these two isn't perfectly clear. But one salient difference pertains to the evidence on which the protagonist relies. In an apparent-evidence case, this evidence results in a belief that is counterfactually fragile in the following sense: if the

protagonist were aware of certain facts about the nature of her evidence, it is unlikely that she would have reached the conclusion she did. This form of fragility seems to be absent in an authentic-evidence case. Below, we will have more to say about this distinction. For now, the main point is that participants seem to track the distinction. Starmans and Friedman found that participants were far more likely to ascribe knowledge in authentic-evidence cases than in apparent-evidence cases.

The *authentic-apparent effect* is an interesting one, psychologically and philosophically. That non-philosophers are more inclined to treat a large family of Gettier cases as instances of knowledge than other Gettier cases calls for an explanation. And, for philosophers who want to put a lot of weight on folk intuitions in building a case against the justified true belief (JTB) account of knowledge, the effect also calls for some degree of explaining away. After all, we might wonder, why argue against the JTB account by appeal to one family of intuitions rather than defend it by appeal to another?

But, before the relevant psychological and philosophical communities get into such thorny issues, perhaps it is advisable to explore the authentic-apparent effect in greater depth. How robust is the effect? More specifically, we might ask at least two questions. First, does the effect persist when efforts are made to control for participants' assessments of the justificatory status of the belief in question? After all, the effect may be reducible to an effect on attributions of epistemic justification. Second, are there any cross-cultural or cross-nationality differences in the effect? This question is particularly acute considering arguments to the effect that we cannot assume "that findings from one population apply across the board" (Henrich, Heine, & Norenzayan, 2010a, p. 29). This is especially so when the findings stem from people who are WEIRD—Western, educated, industrialized, rich, and democratic (Henrich, 2020; Henrich, Heine, & Norenzayan, 2010b), as they appear to be in Starmans and Friedman's (2012) studies (see p. 274).

In this paper, we present three studies that explore the robustness of the authentic-apparent effect along the two dimensions just identified. The results go some way towards suggesting that the effect is robust to participants' justificatory assessments; that is, in those conditions where the effect emerges, it appears to be not (entirely) reducible to an effect on assessments of the justificatory status of the protagonist's belief. In addition, the results provide some reason to think that the effect is not fully robust to participants' nationalities. In one experiment, the effect clearly emerged across two cultural contexts—the US and India; in another experiment, however, the effect clearly manifested only among American participants, and, in another, the effect was stronger among Indian participants. As we might summarize the overall trends in the data we report, they provide some reason to think that the authentic-apparent effect is robust to participants' attributions of epistemic justification but not so robust to their nationality.

1. A Methodological Innovation

It is fair to say that Weinberg, Nichols, and Stich (2001) initiated a new mode of reflection on *Gettier cases*—thought experiments, vignettes, or scenarios in which it seems, at least to many philosophers, that the protagonist has a justified true belief that *p* but lacks knowledge that *p*. Yes, prior to Weinberg et al., there was a voluminous literature, with long historical roots, on

Gettier cases and the possibility that knowledge may outstrip admixtures of justification, truth, and belief. For example, according to Google scholar, Edmund Gettier's short paper had over 500 citations before 2001. And, as for its historical roots, it is often noted that Gettier cases were discussed in ancient India. For instance, Nagel (2014) presents two thought experiments given by the Indian philosopher Dharmottara around 1200 years ago. But what sets Weinberg et al. (2001) apart from the preceding is its use of scientific methods to investigate lay cognition as it applies to Gettier cases.

It is well known that Weinberg et al. (2001) reported a cross-cultural difference in Gettier reactions. Their results suggested that university students of Western cultural backgrounds were more likely than those of East Asian and South Asian backgrounds to say that the Gettier protagonist does not know. But efforts to replicate their result have been unsuccessful (Kim & Yuan, 2015; Nagel et al., 2013; Seyedsayamdost, 2015). For instance, Kim and Yuan (2015) found no difference between East Asians' and Caucasians' responses to a Gettier case involving a subject who has recently changed the make of their American car. 85% of their East Asian participants and 86% of their Caucasian participants denied knowledge in the case.

Also, attempts to expand on the result reported in Weinberg et al. (2001) have not been especially favorable to the idea of cross-cultural differences in Gettier intuitions. In a remarkable series of papers that goes a long way towards answering the experimental epistemological equivalent of Henrich et al.'s (2010a, 2010b) calls for research that explores various sides of the WEIRD/non-WEIRD divides, Edouard Machery and colleagues (2017a, 2017b, 2018) have examined Gettier responses across a number of populations. These include (a) industrialized and small-scale populations, (b) Western and non-Western populations, and (c) student and non-student populations. Their results indicate little cross-cultural variability in Gettier intuitions when participants are asked to determine whether the protagonist knows or only thinks (or feels like) she knows. True, there are specifics in their data that may complicate interpretations of their results. To begin with, most Bedouins did not share the Gettier intuition, though the sample size for this population was small ($n = 21$) (Machery et al., 2017b). Moreover, it may be important that, when simply asked whether the protagonist knows or doesn't know, Machery and colleagues did find a fair number of cultures in which most of their participants did attribute knowledge to the protagonist: 1 of 4 in Machery et al. (2017a) and 10 of 24 in Machery et al. (2017b). Nevertheless, based on their results, Machery et al. (2017b, p. 519) propose that "the Gettier intuition may well be part of a core epistemology, a universal way of thinking about epistemic matters." And Machery et al. (2017a, p. 655) write, "We have provided evidence that the Gettier intuition is universal."

Thus, an examination of the empirical record developed in the wake of Weinberg et al. (2001) suggests that their main finding has not withstood the test of time. Still, the central innovation of their paper has fared quite well: there is much to learn from experimental research on lay and non-lay reactions to Gettier cases. This sentiment is supported by even the briefest of surveys of the extant record, which contains an impressive array of work on the potential effects of a wide range of variables on Gettier reactions. These include variables pertaining to the demographics of experimental participants such as their gender (e.g., Adleberg, Thompson, & Nahmias, 2015; Machery et al., 2017b), age (Colaço, Buckwalter, Stich, & Machery, 2014; Machery et al., 2017b), academic background (Starmans & Friedman, 2020), need for cognition (Kneer, Colaço,

Alexander, & Machery, forthcoming; Machery et al., 2017b; Weinberg, Alexander, Gonnerman, & Reuter, 2012), and personality (Holtzman, 2013; Machery et al., 2017b). In addition, there is a large body of work examining the impact of non-demographic variables on Gettier responses. Among these variables are the valence of the potentially known proposition (Beebe & Shea, 2013; Buckwalter, 2014b; Turri, 2014), the holistic versus piecemeal presentation of Gettier vignettes (Turri, 2013), and the narrative details of the case (Machery et al., 2018).

2. Two Types of Gettier Cases

As impressive as the empirical record on Gettier intuitions is, the record is somewhat erratic. On the one hand, there are many studies in which participants seem perfectly willing to attribute knowledge to the Gettier protagonist. Starmans and Friedman (2012), for example, observed rates of knowledge attributions ranging between 69% and 83% in four of their five experiments. And they report similar rates of around 65% and 69% among their lay participants in two more recent experiments (2020). But there are also many studies in which participants seem rather reluctant to ascribe knowledge in Gettier cases. For instance, when given the option between saying that the protagonist knows or only feels like s/he knows, Machery et al. (2017a) reported that only 10.2% of their cross-cultural sample attributed knowledge to the protagonist in response to the first of their two Gettier cases. Nagel et al. (2013) also observed relatively low rates of attributions. Only 35% of their participants ascribed knowledge in the first two of their three Gettier cases.

The variability present in the empirical record on Gettier intuitions likely stems, in part, from differences in data analytic decisions. But differences in experimental materials and procedures probably matter as well. One possibility that Starmans and Friedman (2012) helps to motivate is that some of the variability may hinge on the type of Gettier case that participants are asked to evaluate. To illustrate the difference that they have in mind, we might turn to the likes of a Brown-in-Barcelona case (Gettier, 1963), contrasting it with a Fake Barn scenario (Goldman, 1976). In the former, the agent comes to believe the disjunction that either Jones owns a Ford or Brown is in Barcelona, based on strong evidence in favor of the left disjunct and despite a state of complete ignorance regarding the truth-value of the right disjunct. It turns out that Jones doesn't own a Ford, but Brown is in Barcelona. Thus, the agent's disjunctive belief is true and, evidently, justified. But many philosophers are disinclined to treat the belief as knowledge. Contrast this case with a Fake Barn scenario. Here, the agent is in a sea of barn facades but happens to look upon the only real barn in the scene. He believes that it is a barn. Apparently, he is justified in believing as much. But many philosophers deny that the agent knows it is a barn.

Exactly what epistemologically differentiates the likes of a Brown-in-Barcelona from a Fake Barn is unclear. But at least one difference appears to pertain to the evidence had by the protagonists. Somehow, the evidence had by the protagonist in Brown-in-Barcelona seems more problematic than that had in Fake Barn. More fully, following Starmans and Friedman (2012, p. 280), we might note that, in the former case, the protagonist is unaware of the fact that his disjunctive belief is based on a false disjunct, but, if he were aware of this fact about his evidence, he probably wouldn't have endorsed the disjunction in the first place. It seems that his evidence fails to track the fact that eventually renders his disjunctive belief true. Parallel claims don't apply to Fake Barn, at least not nearly so cleanly. Yes, if the protagonist knew that he was

in fake barn country, he might not come to believe that a barn sits before him, at least not without further examination. But that's not so much a fact about his evidence, as it is about the evidential context. Importantly, unlike Brown-in-Barcelona, his evidence is tracking the fact that makes his belief true. So, as Starmans and Friedman are inclined to put, in Brown-in-Barcelona, the protagonist's evidence is *apparent*; in Fake Barn, however, his evidence is *authentic*.

There is some reason to think that the difference between apparent and authentic evidence cases can matter to lay attributions of knowledge. Some of the clearest of these reasons come from Starmans and Friedman (2012, Experiment 3). They reported that participants who received an apparent-evidence case were far less likely to attribute knowledge to the protagonist (30%) than those who saw an authentic-evidence case (67%) (see also Nagel et al., 2013). Turri, Buckwalter, and Blouw (2015) report converging results. They argue that Gettier cases come in five distinct flavors: Failed Threat, Detect Similar, Detect Dissimilar, No Detect Similar, and No Detect Dissimilar. And their results provide some reason to think that these differences can matter to regular folks. Most important for current purposes is that they found that their participants were more likely to ascribe knowledge in Similar Replacement and Detect Similar Replacement cases than in No Detection and No Detect Similar cases. This is a distinction that they assimilate to the authentic-/apparent-evidence difference.

It seems, then, that the apparent-/authentic-evidence distinction can matter for the experimental epistemology of Gettier intuitions. But how robust is the effect? In this paper, we endeavor to deepen our understanding of the effect by exploring its robustness along two dimensions. We describe and motivate these two in the following section.

3. Two Dimensions and an Alternative Probe

3.1. Justification

Experimental epistemologists have recently begun to explore the relationship that knowledge bears to epistemic justification. The orthodox position among (Western?) philosophers is that knowledge entails justification. Do regular people concur? There is some data suggesting that the answer is, no. Across nine vignettes involving true beliefs that many philosophers would treat as unjustified, Sackris and Beebe (2014) found an oversized tendency among their participants to ascribe knowledge. But, more recently, Nolte, Rose, and Turri (forthcoming) identify a shortcoming in these studies as well as in the research builds on Sackris and Beebe's work (e.g., Gonnerman et al., 2020). And that is, these studies fail to ask participants whether, in their estimation, the protagonist's belief is justified. To the chagrin of many epistemologists, it could be that a fair number of non-philosophers are disposed to treat the beliefs at play in the vignettes as justified, which would likely make these individuals more disposed to classify the beliefs as cases of knowledge (on this possibility, see also Gonnerman et al., 2020, p. 209).

This recent exchange in the experimental epistemology of knowledge and justification helps to highlight a possibility relevant to the authentic-apparent effect. It may be that the effect is not so much an effect on folk attributions of knowledge as it is an effect on lay assessments of epistemic justification. Perhaps the reason why rates of knowledge attributions were so low in Starmans and Friedman's apparent-evidence cases was that, contrary to the expectations of many

of philosophers, participants were strongly inclined to treat the beliefs at issue as unjustified. And maybe there was also a strong tendency to treat the beliefs at play in the authentic-evidence cases as justified. If so, then the possibility begins to emerge that the authentic-apparent effect on knowledge attributions is entirely reducible to an effect on justification attributions. Thus, the following question arises: to what extent does the effect persist when we control for attributions of justification? Is the effect robust to justificatory assessments?

3.2. Nationality

A second dimension that we consider in connection with the potential robustness or fragility of the authentic-apparent effect is cross cultural. Henrich et al. (2010) note that psychologists often formulate their results in rather general terms. For instance, when reporting the results of an experiment seeming to show that a variable of interest influences an outcome of interest, psychologists often frame their conclusions about people in general. Thus, they might declare that their findings show that people (note: no qualification) are inclined to judge X under conditions C. Such declarations about people in general can feel under-substantiated when one recognizes that they are based on data derived from a very narrow sample of the world's populations. In many cases, it is hard to see why we should think that what is true of, say, American undergraduate psychology students should carry over to the species.

This tendency to frame empirical results derived from narrow samples in terms of very broad populations is also found in experimental epistemology. It is not uncommon to see experimental epistemologists draw (tentative) conclusions about lay cognition, folk concepts, or ordinary usage based on English-speaking samples alone (e.g., as regards [redacted for blind review]). Indeed, in this very paper, we defaulted to a broad characterization of empirical results, without due qualifications or cautionary notes. For example, in the introduction, we implicitly characterized the authentic-apparent effect as the finding that non-philosophers are more likely to ascribe knowledge in authentic-evidence cases than in apparent-evidence cases. Nowhere did we note that this characterization summarizes the results of an experiment done in English and probably on Western participants. (Unfortunately, Starmans and Friedman don't specify the cultural or national makeup of their sample in their experiment, though it is reasonable to bet that their sample was WEIRDish; see p. 274.)

Broad generalizations based on narrow samples are of questionable merit, of course. This is especially so when (a) there is good reason to suspect the possibility of cross-demographic differences in the domain of cognition or behavior being explored and (b) we have only the beginnings of a theory for predicting when these differences are apt to make for a difference in the task being explored. As regards epistemic cognition, there are reasons to suspect some cross-cultural differences. Waterman, Gonnerman, Yan, and Alexander (2018), for example, report results suggesting that the effect that salient alternatives have on knowledge attributions varies across American, Chinese, and Indian populations. Moreover, given that even very basic forms of cognitive processing such as susceptibility to the Müller-Lyer illusion can vary across cultures (Henrich et al., 2010b, p. 4), it seems reasonable to suspect some degree of cross-cultural variability with respect to a good chunk of non-basic cognition as well, including epistemic cognition. This line of argumentation leads to the conclusion that characterizations of the authentic-apparent effect in terms of people in general, or even laypeople somewhat more

narrowly, are of questionable merit. At least as much as so, given that we lack even the beginnings of a good theory outlining when cross-cultural differences are likely to make a difference in tasks like the one deployed by Starmans and Friedman.

What the preceding helps to motivate is a cross-cultural—or, here, a cross-nationality—investigation of the authentic-apparent effect. The main question that we pursue on this front is, does the effect manifest, in comparable size and direction, among American and Indian populations? In other words, is the effect robust across American and Indian nationalities? Of course, our work here is only one step towards understanding the American and Indian dynamics of the authentic-apparent effect. All the more is true of the effect's dynamics in other populations.

3.3. *Contrastive probe*

One worry that some may have with the authentic-apparent effect, as reported in Starmans and Friedman (2012), is that it might be artificially inflated. In recent years, experimental epistemologists have become concerned with the possibility that many of the knowledge attributions observed in their studies are non-literal attributions. Typically, this concern centers on *protagonist projection*. This is a phenomenon in language use and interpretation in which the speaker uses words that a salient protagonist might deploy to describe her situation and the audience interprets the speaker accordingly (Holton, 1997). Applied to the experimental epistemology of Gettier cases, the worry is that, when participants select a response option such as “actually knows” instead of “merely believes”, it may be that she had interpreted the “actually knows” option through the lens of protagonist projection. If so, then her response is not an indication that she agrees with the knowledge statement literally construed. A better interpretation is that she agrees that the protagonist *thinks* he knows (Buckwalter, 2014a). Could protagonist projection be responsible for the high rates of knowledge attributions witnessed by Starmans and Friedman in response to their authentic-evidence cases? If so, then the authentic-apparent effect might be, at least in part, an artifact stemming from a greater tendency to engage in protagonist projection when considering authentic-evidence cases.

The usual strategy for controlling for protagonist projection is to give participants a *contrastive probe* (e.g., Buckwalter, 2014a; Machery et al., 2017a, 2017b; Nagel et al., 2013; Nolte, Rose, & Turri, forthcoming). That is, instead of asking whether the protagonist knows or not, or whether she actually knows the proposition or merely believes it, what the participant must do is to decide between two options like the following: (i) the protagonist actually knows the proposition or (ii) she merely thinks that (or feels like) she knows the proposition. The thought behind giving participants these two is that the second option gives those inclined towards a projective interpretation a place to park their non-literal ascription, whereas the first option remains open to those who are apt to ascribe knowledge to the protagonist under a literal interpretation of ‘knows’. In the three experiments that we report below, we adopt the usual strategy for controlling for projective interpretations of the knowledge probes.

4. Experiment 1

4.1. *Participants, materials, and procedures*

Two hundred sixteen individuals were recruited through Prolific. Participants were paid \$0.27 USD to complete the study. Prescreening options were set so that only individuals who had indicated to Prolific that their nationality is the United States or India were allowed to participate.¹ Prior to analysis, exclusion criteria were determined. Individuals who failed the attention check or who reported that they were not fluent in English were excluded from the main analyses of the paper. (As far as we know, there are no exclusion criteria internal to Prolific itself.) This decision gave us a sample size of $N = 215$ (Age $M = 30.69$, Female = 49%, American $n = 131$, Indian $n = 84$).

Participants were randomly assigned to one of two vignettes: (i) Yogurt—Authentic Evidence or (ii) Yogurt—Apparent Evidence, taken from Starmans and Friedman (2012, p. 282). Both cases describe a protagonist who puts what she believes is a container of yogurt in her fridge. It's just that, whereas her belief is true at this point in the authentic-evidence case, it is false in the apparent-evidence case. However, in both, a neighbor enters the picture, surreptitiously replacing the original container with one that contains yogurt. By the end of the story, then, the protagonist's belief that there is a container of yogurt in her fridge is true in both cases.

Following the story, participants saw five questions. They were presented in a fixed order.

- (1) A comprehension question: "At the end of the story, is there a container of yogurt in Julie's fridge?" The response options were "Yes" and "No".
- (2) A question about epistemic justification: "In your view, does the evidence justify Julie in thinking that there is a container of yogurt in her fridge?" Again, the response options were "Yes" and "No".
- (3) A confidence measure: "How confident or unconfident are you in your answer?" Responses were recorded on a 7-point scale (1 = very unconfident; 7 = very confident)²

¹ Participants complete several demographic questions when they sign up for Prolific. One is, what is your nationality? Prolific states that participants can choose only one nationality. And they leave this question to the participants to interpret. It is worth noting that recruiting participants who report that their nationality is the US or India does raise a possibility that may complicate the results presented in this paper. An anonymous reviewer observes that reporting an Indian nationality is consistent with residing in the West. This includes Indians who have long lived in the US or the UK, even during their formative years. The complication that this observation presents is reasonably clear. To the extent that our Indian sample is comprised of individuals whose psychologies have been substantially shaped by Western environments, and thus qualify as WEIRD, it is reasonable to worry that our experiments are unlikely to find any cross-cultural differences. Indeed, any evidence from our experiments seeming to suggest that the authentic-apparent effect is robust across American and Indian nationalities could be because we were simply comparing, more or less, comparing the responses of one WEIRD sample with those of another WEIRD or WEIRDish sample. We acknowledge that it is possible that the Indian-specific results reported in this paper may stem, in large part, from Indians living in the West. As such, any results that appear to indicate a cross-culturally robust authentic-apparent effect needs to be interpreted with this complication in mind (along with the fact that all the materials were presented in English, which may be masking important differences as well). With that said, the flip of the line of reasoning just presented is that any evidence of cross-cultural differences reported in the paper may be an underrepresentation of the magnitude of the difference among WEIRD Americans and less WEIRD Indians. And thus, the experiments of this paper could be seen as operating under somewhat stringent conditions insofar as building a positive case for cross-cultural differences is concerned. As we will see, such differences did emerge in Experiments 2 and 3.

² An anonymous reviewer noted an ambiguity in the first confidence measure. It isn't perfectly clear to which question this one is referring. Is it the comprehension question or the one about epistemic justification? The most natural interpretation is the latter, since it is the one that immediately precedes the confidence measure. But

(4) A contrastive probe about knowledge: “In your view, which of the following better describes Julie's situation?” Response options were: “Julie actually does know that there is yogurt in her fridge” and “Julie only thinks she knows that there is yogurt in her fridge”.

(5) A second confidence measure, with responses also recorded on a seven-point scale.

The story remained at the top of the screen while participants answered the five questions. Afterwards, they received a short demographic questionnaire.

4.2. Results

Figure 1 captures the rates of knowledge and justification attributions observed among the American and Indian participants.

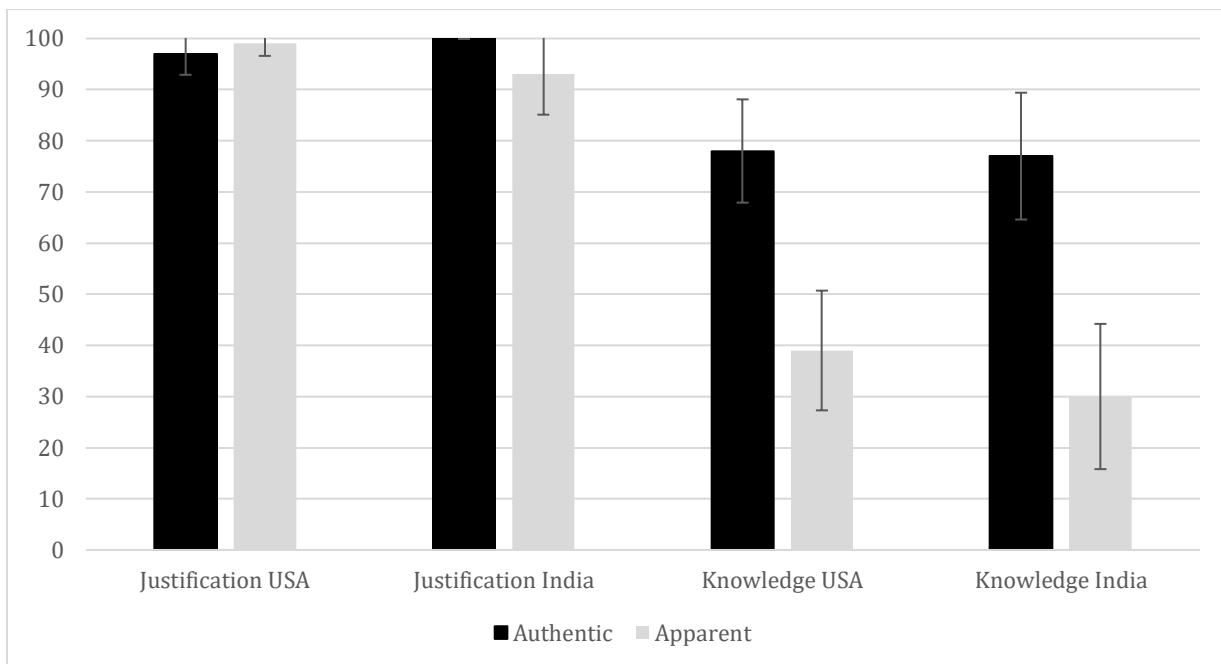


Figure 1. Rates of Justification and Knowledge Attributions in Experiment 1

Prior to analysis, we decided to focus our analyses on participants' *composite scores*. Inspired by Starmans and Friedman (2012), scores were calculated as follows. If a participant answered “yes” in response to the epistemic justification question, then we coded their response as +1; if they answered “no”, we coded the response as -1. Indications that “Julie actually does know that there is yogurt in her fridge” were also coded +1. And selections of “Julie only thinks she knows that there is yogurt in her fridge” were coded -1. These values were then multiplied by the corresponding confidence rating. The result was a *justification score* and a *knowledge score*, each ranging from -7 to +7.

admittedly, some participants may have reached a different interpretation, which would introduce some degree of measurement noise in the experiment (and the next). In Experiment 3, we clarify the question.

We carried out a 2×2 ANOVA on participants’ justification scores with case type and nationality as between-subject factors. The analysis revealed a main effect of case type on justification scores, $F(1, 211) = 4.28, p = .040, \eta_p^2 = .020$, as well as a main effect of nationality on justification scores, $F(1, 211) = 4.31, p = .039, \eta_p^2 = .020$. An interaction effect between the two factors did not emerge, $F(1, 211) = 1.69, p = .196, \eta_p^2 = .008$.

We also performed a 2×2 ANOVA on knowledge scores with case type and nationality as between-subject factors. A main effect of case type emerged, $F(1, 211) = 45.76, p < .001, \eta_p^2 = 0.178$. But neither a main effect of nationality nor an interaction of nationality and case type were observed, $F(1, 211) = 0.66, p = .419, \eta_p^2 = .003$ and $F(1, 211) = 0.21, p = .651, \eta_p^2 = .001$. Indeed, the simple effects of case type on knowledge scores were quite similar for the two nationalities (Americans: $F(1, 211) = 25.53, p < .001, d = 0.88$; Indians: $F(1, 211) = 21.36, p < .001, d = 1.01$). Table 1 summarizes the mean knowledge and justification scores.

Table 1.
Mean Justification and Knowledge Scores and Standard Deviations for Experiment 1

| | Justification | | Knowledge | |
|----------|---------------|-------------|-------------|--------------|
| | Authentic | Apparent | Authentic | Apparent |
| American | 6.11 (2.06) | 5.87 (2.02) | 3.55 (5.02) | -1.28 (5.88) |
| Indian | 5.86 (1.83) | 4.80 (3.19) | 3.27 (5.39) | -2.25 (5.54) |

The final analyses that we conducted were a series of simple mediation analyses using the PROCESS macro for SPSS (Hayes, 2013). Our main goal in conducting these analyses was to test whether case type directly influenced knowledge scores while controlling for justification scores. In doing so, we also tested whether case type indirectly influenced knowledge scores through justification scores. In addition to running the mediation analysis on the entire data set, we ran a separate analysis for each of the two nationalities, per the recommendations of an anonymous reviewer.

Similar patterns emerged in the three analyses. A statistically significant relationship between case type and justification scores was not found in any of the analyses. Justification scores were not found to be related to knowledge scores. The bias-corrected bootstrap confidence intervals for the indirect effects, based on 5,000 bootstrap samples, included zero in all three cases. But the direct effect of case type on knowledge scores was statistically significant for all three. Table 2 summarizes the mediation results.

Table 2.
Mediation Results for Experiment 1

| | Consequent |
|--|------------|
|--|------------|

| | Antecedent | | M (JUST.) | | | Y (KNOW.) | | | |
|-----------|------------|----------|-----------|------|------|--------------------------------------|-------|------|-------|
| | | | Coeff. | SE | p | Coeff. | SE | p | |
| ALL Ps | X (CASE) | <i>a</i> | -0.54 | 0.31 | .083 | <i>c'</i> | -4.98 | 0.75 | <.001 |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | 0.18 | 0.16 | .262 |
| | | | | | | <i>ab</i> = 0.10, CI = [-0.42, 0.04] | | | |
| AMERICANS | X (CASE) | <i>a</i> | -0.24 | 0.35 | .495 | <i>c'</i> | -4.83 | 0.96 | <.001 |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | -0.01 | 0.24 | .960 |
| | | | | | | <i>ab</i> < 0.01, CI = [-0.25, 0.24] | | | |
| INDIANS | X (CASE) | <i>a</i> | -1.06 | 0.56 | .061 | <i>c'</i> | -5.16 | 1.21 | <.001 |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | 0.34 | 0.23 | .151 |
| | | | | | | <i>ab</i> = 0.35, CI = [-1.11, 0.01] | | | |

4.3. Discussion

When given a contrastive formulation of the knowledge prompt, participants were found to be less likely to ascribe knowledge to the Gettier protagonist in the apparent-evidence case than in its authentic cousin. We witnessed this pattern in both nationalities. No clear indication emerged that the pattern was especially stronger in one group than in the other. Thus, the results of Experiment 1 go some way towards suggesting that the authentic-apparent effect is robust across American and Indian nationalities, at least in response to Yogurt-like vignettes.

As for the potential robustness of the effect to attributions of epistemic justification, the results that most matter are the mediations results. The question is, is the direct effect of case type on knowledge scores significant when controlling for justification scores?³ All three mediation analyses found some such effect. This goes some way towards suggesting that the authentic-apparent effect on knowledge attributions is not entirely explicable in terms of an effect on attributions of justification. Otherwise put, there is some reason to think that, at least when it comes to the yogurt cases, there is something beyond, or perhaps in addition to, participants' justificatory assessments that drives the authentic-apparent effect.

Thus, the results of Experiment 1 are consistent with a rather robust authentic-apparent effect. At least this is so, when it comes to an American versus Indian nationality and to participants'

³ Indirect effects are also interesting, of course. That the indirect effect of case type on knowledge scores through justification scores was not significant is somewhat surprising. In the authentic-evidence case of this experiment, the belief at issue is inferred from a falsehood, whereas, in the apparent-evidence case, it comes from a truth. This difference seems to matter to participants. A main effect of case type on justification scores did emerge. And yet the indirect pathway was not significant in any of the mediation analyses. Why is that? One possibility is that the authentic-apparent effect on knowledge is not mediated through justificatory assessments, at least not in response to the vignettes used in this experiment. But there are other possibilities as well. Simulations by Fritz and MacKinnon (2007) suggest that bias-corrected bootstrapping tests of mediation, of which PROCESS is an example, can call for rather large sample sizes to achieve a power of at least 0.80. Indeed, in some conditions, the required sample size is nearly double the size of our overall sample in this experiment. So, another possibility is that the non-significant indirect effects that emerged in this experiment are due to a lack of statistical power. We thus caution the reader against drawing any strong conclusions from the null results that we witnessed.

assessments of the justificatory status of the targeted belief. That said, Experiment 1 is but one experiment. A different pair of vignettes might generate rather different results. In the next experiment, we take the design of Experiment 1 and apply it to another pair of cases. As we will see, the patterns of robustness that we witnessed in Experiment 1 did not emerge so clearly in the next round.

5. Experiment 2

5.1. Participants, materials, and procedures

Two hundred twenty-one individuals were recruited through Prolific. Payment for completing the study was as before, as were the prescreening options and exclusion criteria. The result was a sample size of $N = 197$ (Age $M = 30.27$, Female = 55%, American $n = 123$, Indian $n = 74$).

Participants were randomly assigned to one of Starmans and Friedman's (2012, pp. 282-283) coin vignettes: (i) Coin—Authentic Evidence or (ii) Coin—Apparent Evidence. In these, the protagonist believes that a certain quarter in his piggybank is from 1936. He thus comes to believe that there is a coin from 1936 in his piggybank. The first of these beliefs is true in the authentic-evidence case; it is false in the other. However, in both, there is, unbeknownst to the protagonist, a quarter buried deep in his piggybank that is from 1936. So, in both cases, the protagonist's belief that there is a coin from 1936 is true from the outset.

After reading the story, participants received, in a fixed order, five questions like those used in Experiment 1: (1) a binary comprehension question; (2) a binary question about epistemic justification ("In your view, does the evidence justify Corey in thinking that there is a coin from 1936 in his piggybank?"); (3) a seven-point confidence measure for justification responses; (4) a binary contrastive probe about knowledge ("In your view, which of the following better describes Corey's situation? Corey actually does know that there is a coin from 1936 in his piggybank/Corey only thinks he knows that there is a coin from 1936 in his piggybank"); and (5) a seven-point confidence measure for knowledge responses. The story remained on the screen while participants answered the five questions.

5.2. Results

Figure 2 summarizes the rates of justification and knowledge attributions that we observed in the American and Indian samples.

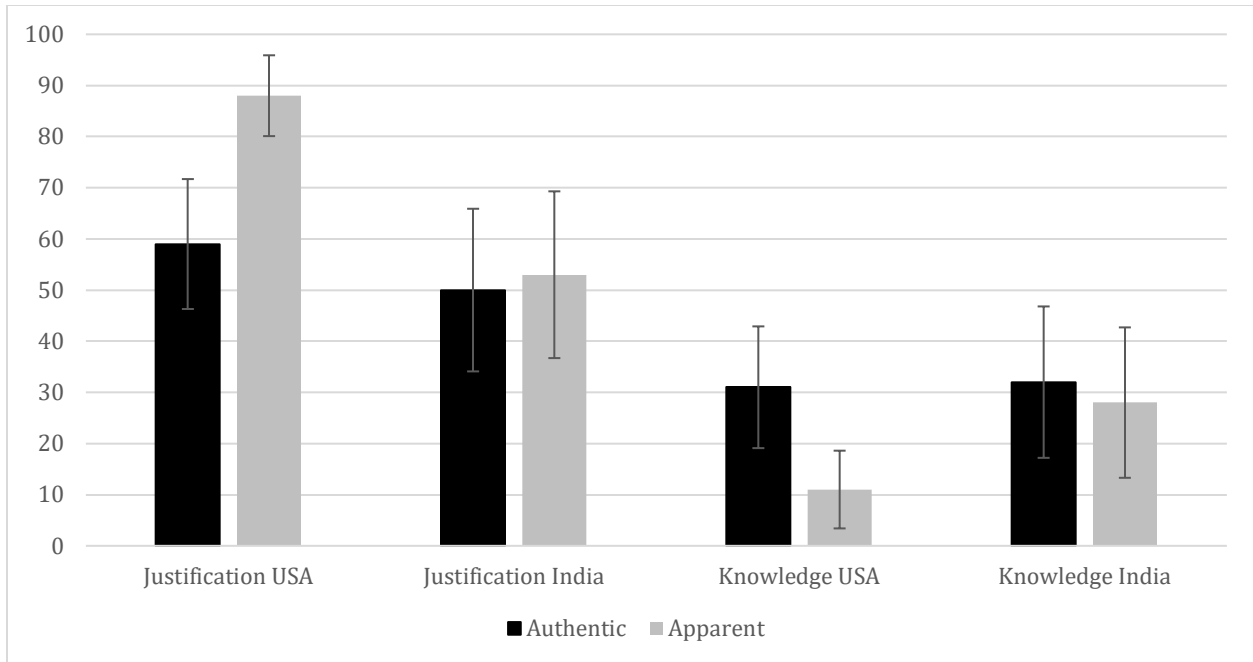


Figure 2. Rates of Justification and Knowledge Attributions in Experiment 2

As in Experiment 1, we focused our analyses on participants' composite scores, which were computed as before. A 2×2 ANOVA on participants' justification scores with case type and nationality as between-subject factors revealed a main effect of nationality on justification scores, $F(1, 193) = 10.24, p = .002, \eta_p^2 = .050$, as well as an interaction effect between nationality and case type on justification scores, $F(1, 193) = 4.96, p = .027, \eta_p^2 = .025$. A main effect of case type did not emerge, $F(1, 193) = 3.19, p = .076, \eta_p^2 = .016$.

A second 2×2 ANOVA with case type and nationality as between-subject factors found a main effect of case type on knowledge scores, $F(1, 193) = 6.89, p = .009, \eta_p^2 = 0.034$. A main effect of nationality did not emerge, $F(1, 193) = 2.34, p = .128, \eta_p^2 = 0.012$. And we did not observe an interaction effect, $F(1, 193) = 2.49, p = .116, \eta_p^2 = 0.013$. The simple effects of case type on knowledge scores for the two nationalities were as follows: Americans— $F(1, 193) = 11.74, p = .001, d = 0.66$; Indians— $F(1, 193) = 0.44, p = .508, d = 0.14$. See Table 3 for the mean knowledge and justification scores.

Table 3.

Mean Justification and Knowledge Scores and Standard Deviations for Experiment 2

| | Justification | | Knowledge | |
|----------|---------------|-------------|--------------|--------------|
| | Authentic | Apparent | Authentic | Apparent |
| American | 1.26 (5.71) | 4.28 (3.74) | -1.47 (5.37) | -4.48 (3.62) |

Indian 0.53 (5.44) 0.19 (5.84) -1.50 (5.25) -2.25 (5.52)

The final set of analyses that we ran were three simple mediation analyses: one on the sample as a whole and the other two, per a referee’s request, on the American and Indian samples. The main goal was to test for a direct effect of case type on knowledge scores while controlling for justification scores. The analyses revealed that case type was related to justification scores in the whole sample and in the American sample, but this relationship did not emerge in the Indian sample. The analyses did not find a significant relationship between justification and knowledge scores in any of the three samples. Moreover, the bootstrap confidence intervals for the indirect effects 1, using 5000 bootstrap samples, included zero in the whole sample and in the Indian sample but not in the American sample. Finally, the direct effect of case type on knowledge scores was significant in the whole sample and the American sample but not for the Indian sample. Table 4 details the results.

Table 4.
Mediation Results for Experiment 2

| | | | Consequent | | | | | | | |
|-----------|------------|----------|------------|------|------|---------------------------------------|-------|------|------|--|
| | | | M (JUST.) | | | Y (KNOW.) | | | | |
| | Antecedent | | Coeff. | SE | p | Coeff. | SE | p | | |
| ALL Ps | X (CASE) | <i>a</i> | 1.85 | 0.75 | .015 | <i>c'</i> | -2.36 | 0.71 | .001 | |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | 0.09 | 0.07 | .200 | |
| | | | | | | <i>ab</i> = 0.17, CI = [-0.07, 0.61] | | | | |
| AMERICANS | X (CASE) | <i>a</i> | 3.01 | 0.86 | .001 | <i>c'</i> | -3.52 | 0.85 | .001 | |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | 0.17 | 0.09 | .052 | |
| | | | | | | <i>ab</i> = 0.51, CI = [0.02, 1.33] | | | | |
| INDIANS | X (CASE) | <i>a</i> | -0.33 | 1.31 | .801 | <i>c'</i> | -0.72 | 1.25 | .569 | |
| | M (JUST.) | | --- | --- | --- | <i>b</i> | 0.09 | 0.11 | .410 | |
| | | | | | | <i>ab</i> = -0.03, CI = [-0.67, 0.25] | | | | |

5.2. Discussion

Unlike Experiment 1, Experiment 2 fails to provide positive support for an authentic-apparent effect that is robust across American and Indian nationalities. While the effect did emerge in the American sample, the same is not true of the Indian sample. That is, unlike the American participants, our Indian participants were not less likely to attribute knowledge in the apparent-evidence case than in the authentic-evidence case. True, the means that we observed in Experiment 2 were in the expected direction. But still a simple effects analysis did not reveal a difference between the two means. Indeed, the effect size that arose out of the Indian sample, $d = 0.14$, falls below the threshold that many psychologists consider, at least at first blush, to be reportable ($d = 0.20$). But it is worth noting that the 95% confidence interval around this effect size is large, [-0.59, 0.32]. So, the data of Experiment 2 are consistent with an authentic-apparent

contrast that results in a small positive effect among Indian participants, at least when it comes to responses to contrastive knowledge probes about Coin-like vignettes. (The data are also consistent with a medium-sized *reverse* effect.) Could it be that the authentic-apparent effect is weaker (or even reversed) in the Indian population when considering Coin-like cases? Some high-powered studies would help to answer this question.

As regards the potential robustness of the authentic-apparent effect to justificatory assessments, the results of Experiment 2 are more in line with Experiment 1. In both, a mediation analysis applied to the American samples revealed a direct effect of case type on knowledge scores when controlling for justification scores. These findings go some way towards suggesting that the authentic-apparent contrast tends to impact the knowledge attributions of Americans in ways that go beyond their attributions of epistemic justification. Of course, unlike Experiment 1, in this one, the results of a mediation analysis did not support a direct effect in the Indian sample. Then again, unlike Experiment 1, the Indian data of Experiment 2 didn't even reveal a clear-cut authentic-apparent effect in the first place.⁴ So, it is perhaps unlikely that the data would support a significant direct effect anyway. In any event, the mediation analyses of Experiment 2 did go some way towards supporting the following: in those cases where an authentic-apparent effect is present, it seems to be robust to the justificatory attributions of those in the relevant group(s).

Overall, then, like Experiment 1, the results of Experiment 2 help to support an authentic-apparent effect that is robust to people's justificatory assessments, at least conditions where an effect arises. Unlike Experiment 1, however, the results of Experiment 2 fail to provide positive evidence in favor of an authentic-apparent effect that is robust to American versus Indian nationalities. Instead, the results of Experiment 2 are merely consistent with an authentic-apparent effect along these lines. Perhaps with more data a clearer picture of the robustness or fatality of the effect will emerge. The goal of Experiment 3 is to go some way towards adding to the emerging picture.

6. Experiment 3

6.1. Participants, materials, and procedures

⁴ Another difference emerged from the nationality-specific mediation results of Experiments 1 and 2. In Experiment 1, among neither nationality did clear evidence emerge that the authentic-apparent contrast exerts an indirect influence on knowledge attributions through attributions of justification. In Experiment 2, matters are different. A simple mediation analysis revealed a significant indirect effect in the American sample but not in the Indian sample. Might the internal dynamics of the authentic-apparent contrast play out differently for Americans than for Indians, at least in response to Coin-like vignettes? The results of Experiment 2 provide some hints that the answer is, yes. This answer stems from two considerations working in concert. The first is the pattern of results just discussed: whereas a simple mediation analysis revealed, for the American sample, a significant indirect pathway from the authentic-apparent contrast to knowledge attributions via justification assessments, the same is not true of the Indian sample. The second consideration relates to the first set of ANOVA results reported in the previous section. Recall that an interaction effect between case type and nationality on justification scores emerged. The increased tendency to ascribe justification to the protagonist in the apparent- than in the authentic-evidence case was stronger among American than Indian participants. This combination of considerations suggests that, as we go from authentic- to apparent-evidence Coin-like cases, there is a stronger willingness among Americans to treat the protagonist's belief as justified, which appears to have downstream effects on their knowledge attributions. The first leg of this indirect pathway is either not present at all in the Indian sample or it is weaker.

Two hundred sixty individuals were recruited through Prolific. The call was to participate in a study that paid \$0.27 USD to complete. Prescreening options restricted the subject pool to individuals who reported that both their nationality and country of birth is either the United States or India. We used the same exclusion criteria as the previous experiments. This resulted in a sample size of $N = 170$ (Age $M = 25.85$, Female = 72%, American $n = 97$, Indian $n = 73$).

Participants received one of two vignettes. Both were derived from a case in the epistemological literature (Ichikawa & Steup, 2018). The following captures the content of the vignettes, with the difference between the two displayed in curly brackets. The part before the forward slash was included in the authentic-evidence case; the part after was in the apparent-evidence case.

While James is relaxing on a bench in a park, he sees what looks like a dog in a nearby field. [¶] What James actually saw was a new {breed of dog/robot dog}. James is unaware that such {hybrid/robot} dogs exist. A Japanese {dog breeder/toy manufacturer} has only recently developed them. [¶] While James was looking at his phone, the {hybrid/robot} dog ran down the backside of the hill and into the woods next to the hill. The {hybrid/robot} dog is now no longer on the hill. However, on the backside of the hill, concealed from James's view, is another dog. [¶] James' friend messages him, "Anything interesting going on in the park today?" James responds, without looking up at the hill again, "There's a dog on the hill."

Following the story, participants received six questions in a fixed order: (1) a comprehension question; (2) a question about epistemic justification; (3) a justification confidence measure; (4) a contrastive probe about knowledge; (5) a knowledge confidence measure; and (6) a question about how easy or difficult the story was to understand. For the most part, the wording of the first five questions was like the wording used in the previous experiments. For example, the knowledge probe went as follows: "In your view, which of the following better describes James's situation? James actually does know that there is a dog on the hill/James only think that he knows that there is a dog on the hill." The biggest difference in question wording was in the confidence measures. Each specified that we are interested in how confident or unconfident participants were in their responses "to the immediately preceding question".

The final question asked was, "How easy or difficult was it for you to understand the story?" Responses were recorded on a seven-point scale (1 = very difficult; 7 = very easy). We included this question because it is reasonable to worry that the contents of the vignettes used in the experiments of this paper include details that are more local to Western contexts. For example, in Experiment 2, the scenario refers to quarters, which are not used in India. It is possible that culturally specific details along these lines may have gotten in the way of Indians' engagement with the vignettes and thereby affected their responses to survey questions. To go some way towards assessing this possibility as regards the vignettes of this experiment, we asked participants to report on their experiences of processing the vignette they received.

6.2. Results

Analyses began with participants' responses to the questions about how easy or difficult the story was to understand. An independent samples t -test found no differences in American and Indian

responses to this question ($M = 5.26, SD = 1.37$ vs. $M = 5.21, SD = 1.51$), $t(168) = 0.24, p = .81, d = 0.04$.

Figure 3 displays the rates of justification and knowledge attributions that emerged from the American and Indian samples.

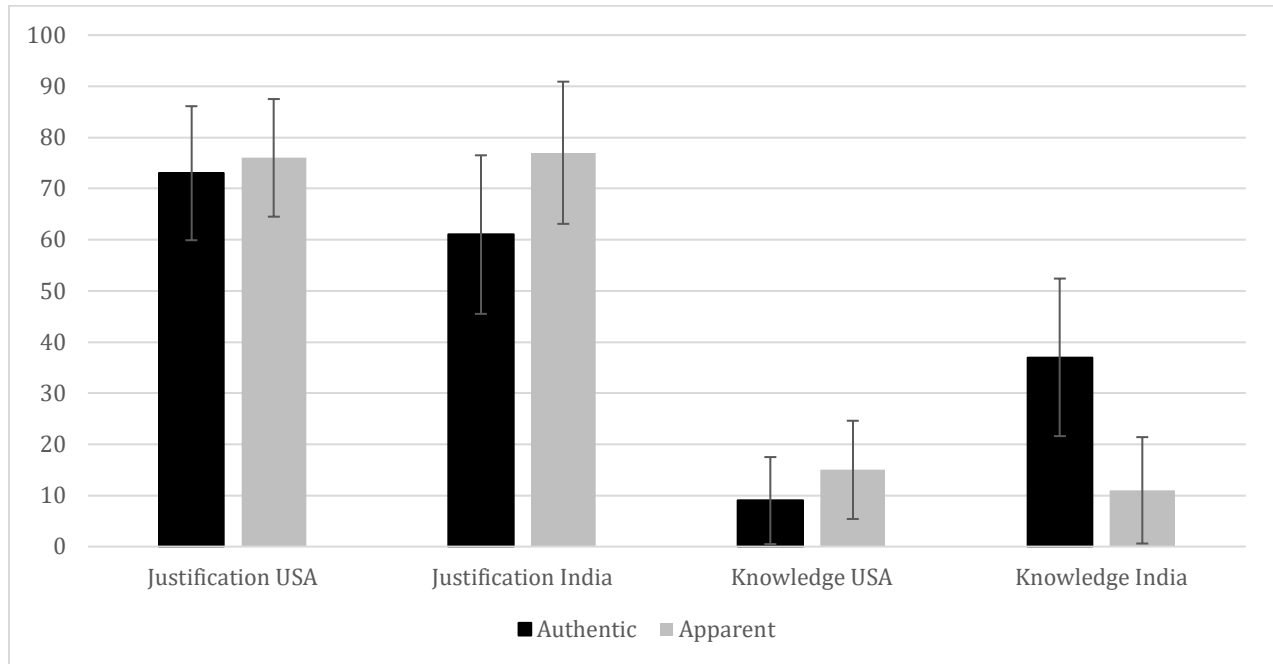


Figure 3. Rates of Justification and Knowledge Attributions in Experiment 3

A 2×2 ANOVA on participants' justification scores with case type and nationality as between-subject factors did not find a main effect of nationality on justification scores, $F(1, 166) = 0.48, p = .486, \eta_p^2 = .0003$ nor did it find a main effect of case type, $F(1, 166) = 1.34, p = .245, \eta_p^2 = .008$. Finally, an interaction effect between the two factors did not emerge, $F(1, 166) = 1.08, p = .301, \eta_p^2 = .006$.

As for knowledge scores, a 2×2 ANOVA with case type and nationality as between-subject factors did not reveal a main effect of case type on knowledge scores, $F(1, 166) = 1.82, p = .179, \eta_p^2 = 0.011$; however, a main effect of nationality did emerge, $F(1, 166) = 4.05, p = .046, \eta_p^2 = 0.024$, as did an interaction effect between the two factors, $F(1, 166) = 6.11, p = .014, \eta_p^2 = 0.036$. The simple effect analyses showed that an authentic-apparent effect arose among the Indian participants but not for the American participants: Americans— $F(1, 166) = 0.73, p = .394, d = -0.19$; Indians— $F(1, 166) = 6.42, p = .012, d = 0.55$. Table 5 displays the relevant means.

Table 5.

Mean Justification and Knowledge Scores and Standard Deviations for Experiment 3

| | Justification | | Knowledge | |
|----------|---------------|-------------|--------------|--------------|
| | Authentic | Apparent | Authentic | Apparent |
| American | 2.64 (4.79) | 2.74 (4.97) | -4.59 (3.46) | -3.83 (4.46) |
| Indian | 1.29 (5.36) | 3.00 (4.87) | -1.55 (5.43) | -4.14 (3.91) |

We ended our analyses with a series of mediation analyses of the same sort carried out in Experiments 1 and 2. One was conducted on the entire sample, and the other two were applied to the American and Indian samples separately. All three analyses revealed a significant relationship between justification scores and knowledge scores. None of them found a significant relationship between case type and justification scores or an indirect effect of case type on knowledge scores. Finally, a direct effect of case type on knowledge scores emerged only out of the Indian sample. See Table 6.

Table 6.
Mediation Results for Experiment 3

| | Antecedent | <i>a</i> | Consequent | | | | | | |
|---------------|------------------|----------|------------------|-----------|----------|--------------------------------------|--------|-----------|----------|
| | | | <i>M</i> (JUST.) | | | <i>Y</i> (KNOW.) | | | |
| | | | Coeff. | <i>SE</i> | <i>p</i> | <i>c'</i> | Coeff. | <i>SE</i> | <i>p</i> |
| ALL <i>Ps</i> | X (CASE) | <i>a</i> | 0.83 | 0.77 | .281 | <i>c'</i> | -0.96 | 0.67 | .152 |
| | <i>M</i> (JUST.) | | --- | --- | --- | <i>b</i> | 0.23 | 0.07 | <.001 |
| | | | | | | <i>ab</i> = 0.19, CI = [-0.13, 0.61] | | | |
| AMERICANS | X (CASE) | <i>a</i> | 0.10 | 1.00 | .921 | <i>c'</i> | 0.74 | 0.81 | .360 |
| | <i>M</i> (JUST.) | | --- | --- | --- | <i>b</i> | 0.19 | 0.08 | .026 |
| | | | | | | <i>ab</i> = 0.02, CI = [-0.38, 0.42] | | | |
| INDIANS | X (CASE) | <i>a</i> | 1.71 | 1.21 | .159 | <i>c'</i> | -3.15 | 1.07 | .004 |
| | <i>M</i> (JUST.) | | --- | --- | --- | <i>b</i> | 0.33 | 0.10 | .002 |
| | | | | | | <i>ab</i> = 0.56, CI = [-0.11, 1.80] | | | |

6.2 Discussion

The results of Experiment 3 suggest that the authentic-apparent effect is not entirely robust to respondents' nationalities. More fully, we found that, when given a contrastive formulation of the knowledge prompt, Indian participants were less likely to ascribe knowledge in the apparent-evidence case than in the corresponding authentic-evidence case; however, we did not observe this pattern in the American sample. True, the 95% confidence interval around the effect size emerging from the American sample, $d = -0.19$, is quite large, [-0.59, 0.21]. So, our results are consistent with a small effect being present among Americans (as well as a medium-sized

reverse effect). With that said, the significant interaction that arose suggests that the magnitude, if not the direction, of the effect is sensitive, and thus not entirely robust to, the nationality of the participant, at least not for Americans versus Indian responses to Dog-like vignettes. Moreover, the fact that we found no difference in American and Indian reports on the ease or difficulty of understanding the vignettes helps to suggest that any nationality-related differences found in this experiment are not due to processing difficulties, such as those that may be brought about by any culturally specific details in the contents of the vignettes.

As for justificatory assessments, Experiment 3 appears to tell a story like the one told by Experiments 1 and 2. A mediation analysis found, among Indian participants, a direct effect of case type on knowledge scores when controlling for justification scores. This suggests that the authentic-apparent contrast tended to exert an influence on the knowledge attributions of our Indian participants that goes beyond their justificatory assessments. Yes, it is worth noting that a direct effect did not emerge out of the American sample. Then again, neither did an authentic-apparent effect. So, it is not clear that we should expect a significant direct effect out of them in the first place. What we might say, then, is that the results of Experiment 3 help to support a more guarded claim. Where an authentic-apparent effect emerges, it appears to be robust to the justificatory assessments of folks in the relevant population(s).

7. Discussion

7.1. Replication failure or success? Neither

There is some reason to think that experimental philosophy enjoys a higher replication rate than some other areas of empirical inquiry. For instance, in an impressive multi-team effort, Cova et al. (2021) attempted to replicate 40 studies taken to be representative of experimental philosophy between 2003 and 2015. They found that over 70% of the studies successfully replicated. This estimate compares favorably to the replication rates sometimes reported for psychology, which the Open Science Collaboration (2015) places between 23% and 29%.

At first blush, it might seem that the results of the first two experiments reported in this paper makes for an uneven contribution to experimental philosophy's replication rate. On the one hand, collapsing across nationalities, we observed rates of knowledge attributions in Experiment 1 quite comparable to those reported by Starmans and Friedman (2012, Experiment 3) (see Figure 4). On the other hand, in Experiment 2, again collapsing across nationalities, we witnessed much lower rates than those found by Starmans and Friedman. The pattern of results a question. Is Experiment 1 a replication success while Experiment 2 is a replication failure?

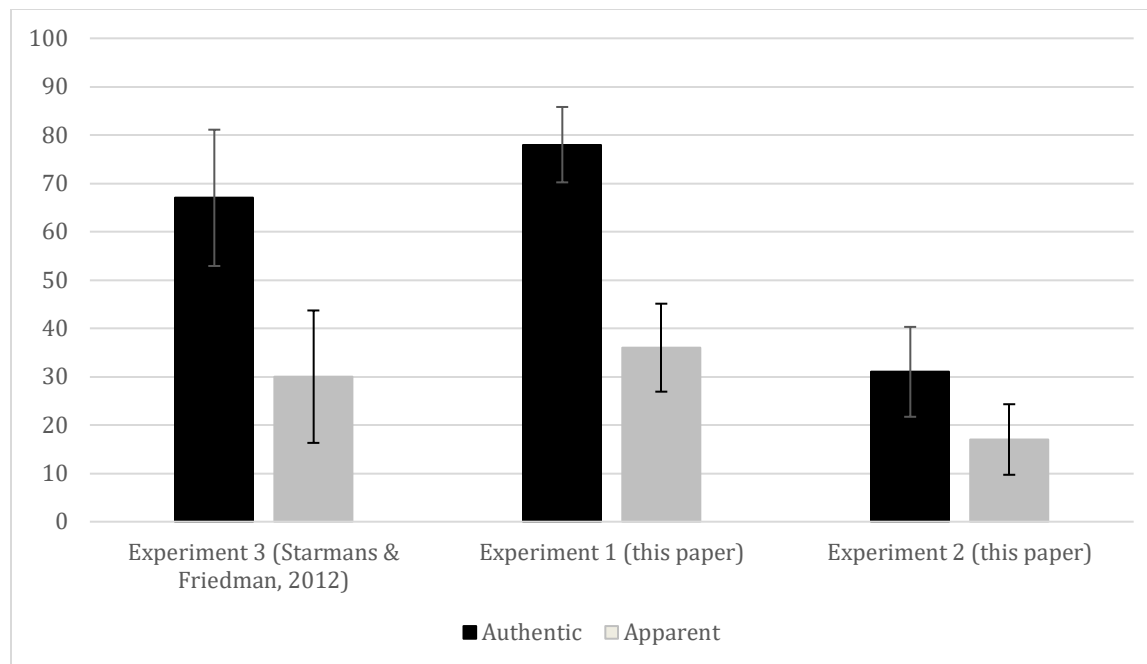


Figure 4. Rates of Knowledge Attributions in Yogurt and Coin Cases Across Three Experiments

There are many reasons why a replication might fail. Machery, Grau, and Pury (2020) discuss some of these reasons in connection to replication efforts in experimental philosophy, with an emphasis on statistical power. In the case of Experiment 2, it is likely that the reason why we observed much lower rates of knowledge attributions than those reported by Starmans and Friedman has to do with the probes used to elicit participants' reactions. In our experiment, we used a contrastive probe. We asked participants to determine whether the Gettier protagonist actually knows or merely thinks she knows. Starmans and Friedman used a standard probe. Their participants had to determine whether the protagonist "really knows" or "only believes". Prior research indicates that contrastive probes tend to lead to substantially lower rates of knowledge ascriptions than more standard probes (e.g., Gonnerman et al., 2020; Machery et al., 2017a, 2017b; Turri, 2015). For this reason, we think it is better not to treat our results from Experiment 2 as a failure to replicate Starmans and Friedman's work. In terms of the machinery developed by Machery (2020), the experimental components of Experiment 2 (here, the measurements) are not the same as those of Starmans and Friedman (2012, Experiment 3), and so we aren't looking at a mere resampling of the original experimental components.

That said, the line of reasoning articulated in the previous paragraph does leave the results of Experiment 1 looking rather odd. For, in that experiment, we also used a contrastive probe. And yet we observed similar rates of attributions as those reported by Starmans and Friedman. Why is that? Shouldn't we have seen much lower rates? Future research into the conditions where contrastive probes do and don't lead to reduced rates of knowledge ascriptions are likely to help in answering this question.⁵

⁵ Future research into the mechanisms whereby contrastive probes depress rates of knowledge ascriptions is also likely to help. The explanation that will occur to many is protagonist projection. Evidently, or so the thought goes, the reason that contrastive probes tend to lead to lower rates of knowledge ascriptions is that they control for non-

7.2. *Two dimensions, armchair practice, and universality*

In all three experiments reported in this paper, mediation analyses determined that the direct effect of case type on knowledge scores, controlling for justification scores, was significant. Or at least this is so in those cases where there was a significant authentic-apparent effect. These results go some way towards supporting the robustness of the effect to participants' justificatory assessments. In other words, in conditions where the effect emerges, it appears that it outstrips assessments of the justificatory status of the protagonist's belief. Thus, the authentic-apparent effect on knowledge attributions doesn't seem to be merely a downstream effect on attributions of justification.

The results of the experiments reported here are less friendly to the robustness of the effect to participants' nationalities. In Experiment 1, the effect did emerge in the American and Indian samples. But, in Experiment 2, the same is true only of the American sample. Plus, in Experiment 3, we witnessed an interaction such that Indian participants displayed a much stronger authentic-apparent effect than American participants. Thus, insofar as Henrich et al. (2010a, 2010b)'s call for research that explicitly spans WEIRD/non-WEIRD divides is concerned, our results are a step towards the conclusion that the size, if not the presence, of the authentic-apparent effect is not a universal effect in knowledge attributions.

Still, we have only taken a very small step towards supporting the non-universality of the effect. Our Indian participants don't represent the world's many non-WEIRD populations. This is especially so when it comes to small-scale societies and populations in which English isn't as prevalent as it is in India. Would the effect emerge among, say, the Bedouins? If Machery et al. (2017b)'s results are representative of Bedouins in general (recall that they found a tendency to ascribe knowledge to the protagonist in an apparent-evidence Gettier case, even in response to a contrastive prompt), it might be that there's not enough room for a stable authentic-apparent effect to arise. Specifically, it could be that, in apparent-evidence cases, rates of knowledge attributions are high enough that it is hard to get reliably higher rates in authentic-evidence cases. (Reverse effects are also a possibility.) Moreover, it is worth emphasizing that it is far from clear that we should treat data derived from American participants as representative of Western populations. Heinrich et al. (2010b, pp. 14-15) emphasize this point by highlighting some ways in which Americans seem rather unlike other Western populations (e.g., in their propensity towards individualism and analytic reasoning).

In view of the role that Weinberg et al. (2001) has played in animating various parts of the experimental epistemology of Gettier intuitions, some may wonder what, if anything, the data reported here mean for experimental critiques of the method of cases. How do our cross-nationality results interact with ongoing discussions of the extant practice of deploying case-

literal knowledge ascriptions where their content is not that the protagonist actually does know but rather only that she thinks she knows. We acknowledge that this explanation is a reasonable one. But we also hasten to add that there are other possibilities as well. One is that, rather than controlling for a response option that spontaneously occurs to many participants, contrastive probes do something closer to the opposite: they present a response option that doesn't spontaneously occur to many when they simply consider standard probes. That could matter. For, as work on formal models of categorization makes clear, the response probabilities that these models generate will depend on the alternative categories under consideration (e.g., Nosofsky, 2011).

based intuitions as evidence in analytic epistemology? A cynic's take on the history of these discussions is one of extensive dismissal followed by widespread acceptance. It seems that when the cross-cultural results looked unfavorable to the extant practice, many armchair philosophers were quick to come up with reasons to ignore the input of regular people. But matters appear to have changed dramatically in light of more recent results that seem to indicate that Gettier intuitions are cross-cultural universals. It is as if the views of the unreflective and undereducated rabble have become the wisdom of crowds. To be sure, this cynical history is too simple. Still, there is little reason to think that, as more data have come in, confirmation bias has not shaped the responses of defenders (and critics!) of the method of cases.

A case can be made that the cross-cultural results that we report here should provide little comfort to the defender of philosophy's intuition-deploying practices as they currently stand. Sure, if we're counting noses, in those cases where most of our American participants had a certain Gettier reaction, most of our Indian participants had the same reaction. But, at least on some formulations of the experimental challenge to philosophy's intuitive practices, where the majorities lie in various demographic categories doesn't matter so much (e.g., see Alexander & Weinberg, 2014; Weinberg, 2007; Weinberg, Gonnerman, Buckner, & Alexander, 2010; for a different but very compelling approach, see Machery, 2017). In other words, the challenge on this formulation doesn't hinge on the majority in one demographic group agreeing to a philosophical prompt with the majority in another demographic group disagreeing. Although such results are consistent with the challenge, and sometimes rather compelling fodder what matters more are demographic differences (as well as presentational differences). For, then, we have evidence that, among the sampled populations, people's philosophical intuitions, and, perhaps by extension, philosophers' intuitions, are sensitive to irrelevant factors—here, characteristics of the intuiter that don't figure in the truth conditions of the intuition's propositional content. Noisy evidence is not necessarily a serious obstacle to truth-oriented inquiry. When it is embedded in a rich set of error detection and correction practices, inquiry can progress towards the truth (Weinberg, 2007). The worry is that philosophy's intuition-deploying practices are not sufficiently rich to predict where and how noisy intuitions are apt to arise, much less do our practices have the mechanisms in place for reliably weeding out or correcting noisy intuitions.

We didn't find evidence of cross-cultural differences in Gettier intuitions that “crossed the midpoint”. Again, if most Americans had a certain response to a Gettier case, we found that most Indians had the same response. Thus, those like Joshua Knobe (2019), rightfully struck by results consistent with cross-demographic universality in philosophical intuitions, might regard our results as another entry in this emerging pattern. But, as regards Gettier intuitions, it is too early to declare that they are cross-cultural universals. There is barely more than a half dozen studies exploring the matter. And only one, Machery et al. (2017b), has examined reactions in small-scale societies—a total of two of such societies. And, when it comes to this study, we get an indication of a cross-cultural universality in Gettier intuitions only if we ignore participants' responses to one of the two knowledge questions that was asked and only if we explain away the responses given by the participants in one of the two small-scale societies. Given an evidential record of this sort, any attempt to work out the metaphilosophical implications of the universality of Gettier intuitions across cultures has got to be highly provisional at best. And somewhat odd. It feels a bit like trying to place a college football team among history's very best after they have

only played a few games (while also possibly ignoring or explaining away those many games where they lost).

8. Conclusion

There has been a lot of excellent empirical research on Gettier intuitions over the last couple decades. Among the very best of this work is a series of studies carried out by Machery and colleagues. Their findings are, in some ways, rather complicated. But one reasonable summary of the findings is well captured by Machery et al. (2017a) when they write, “We have provided evidence that the Gettier intuition is universal” (p. 655). When conjoined with their explicit notes of caution, Machery et al.’s decision to summarize their findings as supporting the universality of Gettier intuitions is perfectly reasonable. However, the results of the three experiments reported here reinforce the importance of interpreting their summary with their cautionary notes in mind. Building off research first carried out by Starmans and Friedman (2012) and developed by the likes of Nagel et al. (2013) and Turri et al (2015), our results help to show that dispositions to Gettier intuitions are not universal across Gettier cases. Their unevenness is manifest in the authentic-apparent effect, an effect that, in view of the results that we report here, appears to be fairly robust to participants’ attributions of epistemic justification but also less robust across nationalities, at least when it comes to Americans and Indians.

Works Cited

- Adleberg, T., Thompson, M., & Nahmias, E. (2015). Do men and women have different philosophical intuitions? Further data. *Philosophical Psychology*, 28, 615-641.
- Alexander, J., & Weinberg, J. M. (2014). The “unreliability” of epistemic intuitions. In E. Machery & E. O’Neill (Eds.), *Current controversies in experimental epistemology* (pp. 128-145). New York: Routledge.
- Beebe, J. R., & Shea, J. (2013). Gettierized Knobe effects. *Episteme*, 10, 219-240.
- Buckwalter, W. (2014a). Factive verbs and protagonist projection. *Episteme*, 11, 391-409.
- Buckwalter, W. (2014b). Gettier made ESEE. *Philosophical Psychology*, 27, 368-383.
- Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme*, 11, 199-212.
- Cova, F., Strickland, B., Abatista, A., Andow, J., Attie, M., Beebe, J., Berniūnas, R., Boudesseul, J., Colombo, M., Cushman, F., Diaz, R., van Dongen, N. N. N., Dranseika, V., Earp, B. D., Gaitán Torres, A., Hannikainen, I., Hernández-Conde, J. V., Hu, W., et al. (2021). Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology*, 12, 9-44.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18, 233-239.
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23, 121-123.
- Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, 73, 771-791.
- Gonnerman, C., Poag, L., Redden, L., Robbins, J., & Crowley, S. (2020). In our shoes or the protagonist’s? Knowledge, justification, and projection. In T. Lombrozo, S. Nichols, & J. Knobe (Eds.), *Oxford studies in experimental philosophy* (Vol. 3, pp. 189-212). Oxford: Oxford University Press.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. New York: Farrar, Straus and Giroux.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, 466, 29.

- Henrich J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-83.
- Holton, R. (1997). Some telling examples: A reply to Tsohatzidis. *Journal of Pragmatics*, *28*, 625-628.
- Holtzman, G. (2013). Do personality effects mean philosophy is intrinsically subjective? *Journal of Consciousness Studies*, *20*, 27-42.
- Ichikawa, J. J., & Steup, M. (2018). The analysis of knowledge. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018 Edition).
<<https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>>.
- Kim, M., & Yuan, Y. (2015). No cross-cultural differences in Gettier car case intuitions: Replication study of Weinberg *et al.* 2001. *Episteme*, *12*, 355-361.
- Kneer, M., Colaço, D., Alexander, J., & Machery, E. (forthcoming). On second thought: Reflections on the reflection defense. In T. Lombrozo, S. Nichols, & J. Knobe (Eds.), *Oxford studies in experimental philosophy* (Vol. 4). Oxford: Oxford University Press.
- Knobe, J. (2019). Philosophical intuitions are surprisingly robust across demographic differences. *Epistemology & Philosophy of Science*, *56*, 29-36.
- Machery, E. (2009). *Doing without concepts*. Oxford: Oxford University Press.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: Oxford University Press.
- Machery, E. (2020). What is a replication? *Philosophy of Science*, *87*, 545-567.
- Machery, E., Grau, C., & Pury, C. L. (2020). Love and power: Grau and Pury (2014) as a case study in the challenges of x-phi replication. *Review of Philosophy and Psychology*, *11*, 995-1011.
- Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (2017a). Gettier across cultures. *Noûs*, *51*, 645-664.
- Machery, E., Stich, S., Rose, D., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E. E., Chatterjee, A., Cheon, H., Cho, I.-R., Cohnitz, D., Cova, F., Dranseika, V., Lagos, Á. E., Ghadakpour, L., Gringerg, M., Hannikainen, I., Hashimoto, Y., Horowitz, A., Hristova, E., et al. (2017b). The Gettier intuition from South America to Asia. *Journal of Indian Council of Philosophical Research*, *34*, 517-541.
- Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (2018). Gettier was framed! In S. Stich, M. Mizumoto, & E. McCready (Eds.), *Epistemology for the rest of the world* (pp. 123-148). Oxford: Oxford University Press.
- Nagel, J. (2014). *Knowledge: A very short introduction*. Oxford: Oxford University Press.

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, *129*, 652-661.

Nolte, A., Rose, D., & Turri, J. (forthcoming). Experimental evidence that knowledge entails justification. In T. Lombrozo, S. Nichols, & J. Knobe (Eds.), *Oxford studies in experimental philosophy* (Vol. 4). Oxford: Oxford University Press.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M Pothos & A. J. Willis (Eds.), *Formal models in categorization* (pp. 18-39). Cambridge: Cambridge University Press.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *6251*, aac4716.

Sackris, D., & Beebe, J. (2014). Is justification necessary for knowledge? In J.R. Beebe (Ed.). *Advances in Experimental Epistemology: Series: Advances in Experimental Philosophy* (pp. 175-192). London: Bloomsbury Academic.

Seyedsayamdost, H. (2015). On normativity and epistemic intuitions: Failure of replication. *Episteme*, *12*, 95-116.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124*, 272-283.

Starmans, C., & Friedman, O. (2020). Expert or esoteric? Philosophers attribute knowledge differently than all other academics. *Cognitive Science*, *44*, e12850.

Turri, J. (2013). A conspicuous art: Putting Gettier to the test. *Philosophers' Imprint*, *13*, 1-16.

Turri, J. (2014). The problem of ESEE knowledge. *Ergo*, *1*, 101-127.

Turri, J., Buckwalter, W., & Blouw, P. (2015). Knowledge and luck. *Psychonomic Bulletin Review*, *22*, 378-390.

Waterman, J., Gonnerman, C., Yan, K., & Alexander, J. (2018). Knowledge, subjective certainty, and philosophical skepticism: A cross-cultural study. In M. Mizumoto, S. Stich, & E. McCready (Eds.), *Epistemology for the rest of the world* (pp. 187-214). Oxford: Oxford University Press.

Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy*, *31*, 318-343.

Weinberg, J. M., Alexander, J., Gonnerman, C., & Reuter, S. (2012). Restrictionism and reflection. Challenge deflected, or simply redirected? *The Monist*, *95*, 200-222.

Weinberg, J. M., Gonnerman, C., Buckner, C., & Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, *23*, 331-355.

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