Registered Replication Report: A Large Multilab Cross-Cultural Conceptual Replication of Turri et al. (2015)

Braeden Hall[1], Kathleen Schmidt[2], Jordan Wagge[3],
Savannah C. Lewis[2,4], Sophia C. Weissgerber[5], Felix Kiunke[5],
Gerit Pfuhl[6], Stefan Stieger[7], Ulrich S. Tran[8],
Krystian Barzykowski[9], Natalia Bogatyreva[10], Marta Kowal[11],
KarlIJn Massar[12], Felizitas Pernerstorfer[7], Piotr Sorokowski[13],
Martin Voracek[8], Christopher R. Chartier[2], Mark J. Brandt[14],
Jon E. Grahe[15], Asil A. Özdoğru[16,17], Michael R. Andreychik[18],
Sau-Chin Chen[19], Thomas R. Evans[20], Caro Hautekiet[21],
Hans IJzerman[22,23], Pavol Kačmár[24], Anthony J. Krafnick[25],
Erica D. Musser[26], Evie Vergauwe[21], Kaitlyn M. Werner[27],
Balazs Aczel[28], Patrícia Arriaga[29], Carlota Batres[30],
Jennifer L. Beaudry[31,32], Florian Cova[33], Simona Ďurbisová[24],
Leslie D. Cramblet Alvarez[34], Gilad Feldman[35], Hendrik Godbersen[36],
Jaroslav Gottfried[37], Gerald J. Haeffel[38], Andree Hartanto[39],
Chris Isloi[40], Joseph P. McFall[41], Marina Milyavskaya[42],
David Moreau[43], Ester Nosáľová[24], Kostas Papaioannou[44],
Susana Ruiz-Fernandez[45], Jana Schrötter[46], Daniel Storage[47],
Kevin Vezirian[22], Leonhard Volz[48], Yanna J. Weisberg[49],
Qinyu Xiao[35], Dana Awlia[2], Hannah W. Branit[47], Megan R. Dunn[50],
Agata Groyecka-Bernard[11], Ricky Haneda[15], Julita Kielinska[9],
Caroline Kolle[5], Paweł Lubomski[9], Alexys M. Miller[51],
Martin J. Mækelæ[6], Mytro Pantazi[52], Rafael R. Ribeiro[29],
Robert M. Ross[53], Agnieszka Sorokowska[11],
Christopher L. Aberson[54], Xanthippi Alexi Vassiliou[55],
Bradley J. Baker[56], Miklos Bognar[28], Chin Wen Cong[57],
Alex F. Danvers[58], William E. Davis[59], Vilius Dranseika[60],
Andrei Dumbravă[61,62], Harry Farmer[20], Andy P. Field[63],
Patrick S. Forscher[64], Aurélien Graton[65], Nandor Hajdu[28],
Peter A. Howlett[66], Radosław Kabut[9], Emmett M. Larsen[67],
Sean T. H. Lee[68], Nicole Legate[50], Carmel A. Levitan[69],

Corresponding Authors:
Erin Buchanan, Department of Analytics, Harrisburg University of Science and Technology, Harrisburg, Pennsylvania
Email: ebuchanan@harrisburg.edu

Jordan Wagge, Department of Psychology, Avila University, Kansas City, Missouri
Email: jordan.wagge@avila.edu

Neil Levy[70] [iD]**, Jackson G. Lu[71]** [iD]**, Michał Misiak[11],**
Roxana E. Morariu[72], Jennifer Novak[73], Ekaterina Pronizius[8] [iD]**,**
Irina Prusova[74] [iD]**, Athulya S. Rathnayake[75]** [iD]**, Marina O. Romanova[76],**
Jan P. Röer[77], Waldir M. Sampaio[78] [iD]**, Christoph Schild[79]** [iD]**,**
Michael Schulte-Mecklenbeck[80] [iD]**, Ian D. Stephen[81], Peter Szecsi[28],**
Elizabeth Takacs[2], Julia N. Teeter[38], Elian H. Thiele-Evans[32] [iD]**,**
Julia Valeiro-Paterlini[2] [iD]**, Iris Vilares[82]** [iD]**, Louise Villafana[5],**
Ke Wang[83] [iD]**, Raymond Wu[84], Sara Álvarez-Solas[85]** [iD]**,**
Hannah Moshontz[86] [iD]**, and Erin M. Buchanan[87]** [iD]

[1]Department of Human and Behavioral Sciences, Southern Illinois University Carbondale, Carbondale, IL; [2]Department of Psychology, Ashland University, Ashland, OH; [3]Department of Psychology, Avila University, Kansas City, Missouri; [4]Department of Psychology, The University of Alabama, Tuscaloosa, Alabama; [5]Institute of Psychology, University of Kassel, Kassel, Germany; [6]Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway; [7]Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences; [8]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria; [9]Institute of Psychology, Jagiellonian University, Krakow, Poland; [10]Department of Psychology, University of Oxford, Oxford, England; [11]IDN Being Human Lab - Institute of Psychology, University of Wrocław, Wrocław, Poland; [12]Department of Work & Social Psychology, Maastricht University, Maastricht, the Netherlands; [13]Institute of Psychology, University of Wrocław, Wrocław, Poland; [14]Department of Psychology, Michigan State University, East Lansing, Michigan; [15]Department of Psychology, Pacific Lutheran University, Tacoma, Washington; [16]Department of Psychology, Marmara University, Istanbul, Türkiye; [17]Department of Psychology, Üsküdar University, Instanbul, Türkiye; [18]Department of Psychology, Fairfield University, Fairfield, Connecticut; [19]Department of Human Development and Psychology, Tzu-Chi University, Taiwan; [20]Institute for Lifecourse Development and School of Human Sciences, University of Greenwich, London, England; [21]Department of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland; [22]LIP/PC2S, Université Grenoble Alpes, Grenoble, France; [23]Institut Universitaire de France, Paris, France; [24]Department of Psychology, Pavol Jozef Šafárik University in Košice, Košice, Slovakia; [25]Department of Psychology, Dominican University, River Forest, Illinois; [26]Department of Psychology, Florida International University, Miami, Florida; [27]Department of Psychology, University of Toronto; [28]Department of Psychology, Eötvös Loránd University, Budapest, Hungary; [29]Department of Psychology, ISCTE - University Institute of Lisbon; [30]Department of Psychology, Franklin and Marshall College, Lancaster, Pennsylvania; [31]Research Development and Support, Flinders University, Adelaide, Australia; [32]Department of Psychological Sciences, Swinburne University of Technology; [33]Centre Interfacultaire en Sciences Affectives, University of Geneva, Geneva, Switzerland; [34]Office of Teaching and Learning, University of Denver, Denver, Colorado; [35]Department of Psychology, The University of Hong Kong, Hong Kong; [36]Study Centre Stuttgart, FOM University of Economics and Management, Essen, Germany; [37]Department of Psychology, Masaryk University; [38]Department of Psychology, University of Notre Dame, Notre Dame, Indiana; [39]School of Social Sciences, Singapore Management University, Singapore; [40]Unaffiliated researcher, London, England; [41]Department of Psychology, University of Rochester, Rochester, New York; [42]Department of Psychology, Carleton University, Ottawa, Canada; [43]Centre for Brain Research Auckland, University of Auckland, Auckland, New Zealand; [44]Department of Political Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece; [45]Department of Psychology, Brandenburg University of Technology, Cottbus-Senftenberg, Germany; [46]1st Department of Psychiatry, Pavol Jozef Šafárik University in Košice, Košice, Solvakia; [47]Department of Psychology, University of Denver, Denver, Colorado; [48]Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; [49]Department of Psychology, Linfield University, McMinnville,Oregon; [50]Department of Psychology, Illinois Institute of Technology, Chicago, Illinois; [51]Department of Psychology, Adams State University, Rosario, Argentina; [52]Social Psychology Programme Group, University of Amsterdam, Amsterdam, The Netherlands; [53]Department of Psychology, Macquarie University, Sydney, Australia; [54]Department of Psychology, Cal Poly Humboldt, Arcata, California; [55]Psychology Research Institute, University of Amsterdam, Amsterdam, the Netherlands; [56]Department of Sport and Recreation Management, Temple University, Philadelphia, Pennsylvania; [57]Unaffiliated researcher, Penang, Maylasia; [58]Department of Psychology, University of Arizona, Tucson, Arizona; [59]Department of Psychology, Wittenberg University, Springfield, Ohio; [60]Interdisciplinary Center for Ethics & Institute of Philosophy, Jagiellonian University; [61]George I. M. Georgescu Institute of Cardiovascular Diseases, Iasi, Romania; [62]Alexandru Ioan Cuza University, Iasi, Romania; [63]School of Psychology, University of Sussex, Brighton, England; [64]Research and Innovation Division, Busara Center for Behavioral Economics, Nairobi, Kenya; [65]LIP/PC2S, Université Savoie Mont-Blanc, Chambery, France; [66]Department of Psychology, State University of New York at Fredonia, Fredonia, New York; [67]Department of Psychology, Stony Brook University, Stony Brook, New York; [68]School of Social and Health Sciences, James Cook University, Singapore;

[69]Department of Cognitive Science, Occidental College, Los Angeles, California; [70]Department of Philosophy, Macquarie University, Sydney, Australia; [71]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts; [72]Babeş-Bolyai University, Cluj-Napoca, Romania; [73]Department of Psychology, Kassel University Faculty 05 Social Sciences, Kassel, Germany; [74]School of Psychology, HSE University, Moskva, Russian Federation; [75]Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand; [76]School of Psychology, Higher School of Economics School of Psychology, Moskva, Russian Federation; [77]Department of Psychology and Psychotherapy, University of Witten/Herdecke, Witten, Germany; [78]Department of Psychology, Universidade Federal de São Carlos, São Carlos, Brazil; [79]Department of Psychology, University of Münster, Münster, Germany; [80]Consumer Behavior, University of Bern, Bern, Switzerland; [81]Department of Psychology, Nottingham Trent University, Nottingham, England; [82]Department of Psychology, University of Minnesota, Minneapolis, MN; [83]Darden School of Business, University of Virginia, Charlottesville, Virginia; [84]Department of Psychology, The University of British Columbia, Vancouver, Canada; [85]Grupo de Investigación de Biogeografia y Ecología Espacial (BioGeoE2), Universidad Regional Amazónica IKIAM, Tena, Ecuador; [86]Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin; and [87]Analytics, Harrisburg University of Science and Technology, Harrisburg, Pennsylvania

Abstract

According to the justified true belief (JTB) account of knowledge, people can truly know something only if they have a belief that is both justified and true (i.e., knowledge is JTB). This account was challenged by Gettier, who argued that JTB does not explain knowledge attributions in certain situations, later called "Gettier-type cases," wherein protagonists are justified in believing something to be true, but their belief was correct only because of luck. Laypeople may not attribute knowledge to protagonists with justified but only luckily true beliefs. Although some research has found evidence for these so-called Gettier intuitions, Turri et al. found no evidence that participants attributed knowledge in a counterfeit-object Gettier-type case differently than in a matched case of JTB. In a large-scale, cross-cultural conceptual replication of Turri and colleagues' Experiment 1 ($N = 4{,}724$) using a within-participants design and three vignettes across 19 geopolitical regions, we did find evidence for Gettier intuitions; participants were 1.86 times more likely to attribute knowledge to protagonists in standard cases of JTB than to protagonists in Gettier-type cases. These results suggest that Gettier intuitions may be detectable across different scenarios and cultural contexts. However, the size of the Gettier intuition effect did vary by vignette, and the Turri et al. vignette produced the smallest effect, which was similar in size to that observed in the original study. Differences across vignettes suggest that epistemic intuitions may also depend on contextual factors unrelated to the criteria of knowledge, such as the characteristics of the protagonist being evaluated.

The justified true belief (JTB) account of knowledge (or alternative versions of it) has been an important explanation of propositional knowledge in philosophical discourse for the past 2 millennia (e.g., Jacquette, 1996; Moser, 2002); however, some have challenged how widely accepted it has truly been (Dutant, 2015; Turri, 2016). The JTB analysis states that a claim, or proposition, is considered knowledge if it meets three conditions (Gettier, 1963). Specifically, a person (*S*) knows a proposition (*p*) if and only if (a) *S* believes that *p* is true, (b) *p* is in fact true, and (c) *S* is justified in believing *p* is true.

In other words, to know something, people not only must believe a claim that is indeed true; they also must have sufficient reason for believing the claim to be true.

Specifically, to know something, a person must believe a true claim that was reasonably inferred from an observation or "entailed proposition" (i.e., a truth claim that is used to infer the truth of a subsequent claim). Thus, a lucky guess that happens to reflect the truth should not be considered knowledge. However, many philosophers have argued that people's "epistemic intuitions" (i.e., intuitions about knowledge) rely on more than just the presence of JTBs. Accordingly, they have investigated the extent to which other factors, such as luck, may play a crucial role in lay epistemology.

Gettier (1963) challenged the sufficiency of the JTB account to explain propositional knowledge by presenting two strong counterexamples that are inconsistent

with its predictions. These counterexamples, later referred to as "Gettier-type cases," are situations in which a person has a belief that is both true and well supported by evidence (i.e., meets all three conditions of JTB), yet that person is not judged as possessing knowledge. In many Gettier-type cases, protagonists reasonably infer a true belief (*p*) from an entailed proposition (*e*); however, in a lucky turn of events, the validity of using *e* to infer *p* is called into question despite *p* still turning out to be true.

In one of his original counterexamples, Gettier (1963) described a scenario in which two men, Smith and Jones, have applied to the same job at a company. Much to Smith's disappointment, the president of the company has told Smith that Jones will ultimately get the job (entailed proposition, *e1*). Smith then notices that Jones has 10 coins in his pocket (entailed proposition, *e2*). Smith then infers from *e1* and *e2* the belief (*p*) that the man who gets the job, whom he assumes will be Jones, will have 10 coins in his pocket. This belief is well founded by evidence (i.e., he counted the coins in Jones's pocket himself) and, therefore, is justified. However, unexpectedly, Smith gets the job himself. Coincidentally, Smith discovers that he also has 10 coins in his own pocket. Although the specifics of this outcome were not expected, his inferred belief (*p*) that the man who has 10 coins in his pocket will get the job was still true. Smith reasonably inferred a true belief (*p*) from *e1* and *e2*, but neither *e1* nor *e2* actually produce the truth of *p*. Even though Smith's belief was both true and justified, Gettier argued that Smith does not have knowledge in this case—he just got lucky. Many similar scenarios (i.e., Gettier-type cases) have since been employed to demonstrate the insufficiency of JTBs to fully explain knowledge attributions.[1]

Epistemic intuitions that prevent people from attributing knowledge to Gettier-type case protagonists, such as Smith, have since been referred to as Gettier intuitions (DePaul & Ramsey, 1998; Machery, Stich, Rose, Chatterjee, et al., 2017; Sosa, 2007). Past research has revealed some evidence that people have a universal tendency to demonstrate Gettier intuitions for some Gettier-type scenarios (e.g., Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013). However, the extent to which people demonstrate Gettier intuitions may be influenced by other factors that have not been widely investigated. Turri et al. (2015) presented evidence that people demonstrate different epistemic intuitions for Gettier-type cases depending on how the entailed proposition (*e*) used to infer a justified true belief (*p*) is challenged, which they argued may explain the apparent inconsistencies in past work.

In the present research, we aimed to (a) provide a robust test of Gettier intuitions for counterfeit-object

Gettier-type cases, (b) explore explanations for why Gettier intuitions vary across different scenarios, and (c) explore possible cultural and demographic differences in Gettier intuitions. A secondary goal of this project was to allow psychology students to actively contribute to replication research; students engaged in data collection and other activities as part of dozens of student-lead teams across 19 geopolitical regions.

The Role of Luck in Epistemic Intuitions

Prior work suggests that people generally exhibit Gettier intuitions for at least some Gettier-type cases. Such findings indicate that people's conception of knowledge requires more than justification, truth, and belief (e.g., Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013). However, past results have been mixed (e.g., Powell et al., 2015). In a study by Machery, Stich, Rose, Alai, et al. (2017), participants attributed knowledge to protagonists in cases of luckily true justified belief (i.e., Gettier-type cases) significantly less than in clear cases of true justified belief. Colaço et al. (2014) also found that participants were significantly less likely to attribute knowledge in a Gettier-type case than in a similarly matched knowledge control case (i.e., a clear case of JTB).

However, people may not demonstrate Gettier intuitions for some Gettier-type cases (i.e., intentionally replaced evidence cases; e.g., Powell et al., 2015). Starmans and Friedman (2012) found that participants were similarly likely to attribute knowledge in a "replacement-by-backup" Gettier-type case, in which the subject of the belief was replaced by a replica, as in a clear case of knowledge (Gettier intuition not demonstrated); yet Turri et al. (2015) found that participants were less likely to attribute knowledge in a replacement-by-backup Gettier-type case than in a clear case of knowledge (Gettier intuition demonstrated). Turri et al. also found that participants attributed knowledge in a "counterfeit-object" Gettier-type case, in which the subject of the belief could have been an indistinguishable but not identical object, no differently than in a clear case of knowledge (Gettier intuition not demonstrated); however, Powell et al. (2015) found that participants attributed knowledge less in a counterfeit-object Gettier-type case than in a clear case of knowledge (Gettier intuition demonstrated).[2]

In the experiment replicated in the present research, Turri et al. (2015; Experiment 1) tested whether laypeople demonstrate Gettier intuitions when a salient threat to the truth of a judgment fails. Turri et al. asked participants whether a protagonist in one of three conditions (i.e., a "threat" Gettier condition, a "no-threat" knowledge condition, and a "no-detection" ignorance

condition) knew or only believed a claim. In the experimental Gettier condition, participants read a story in which a protagonist named "Darrel" correctly identifies the species of an animal (i.e., target species) despite it being the only animal of that species among many animals of a different, almost identical species (i.e., counterfeit species). Participants in the other two conditions read the same story with slight changes: In the knowledge control version, the story never mentions the other identical species (i.e., no counterfeit), and in the ignorance control version, the protagonist incorrectly identifies the counterfeit species as the target species. Turri et al. then compared the rate of knowledge attributions between participants in the Gettier condition and participants in the two control conditions. They found no evidence of Gettier intuitions; participants in the Gettier condition attributed knowledge at rates no different from participants in the knowledge control condition, $\chi^2(1, N = 98) = 2.63$, Fisher's exact $p = .164$, Cramér's $V = .164$ (Gettier intuition not demonstrated). These findings suggest that luckily true justified beliefs may be consistent with laypeople's conception of knowledge under certain conditions and highlight the need for further research on epistemic intuitions in Gettier-type cases.

The average size of Gettier-intuition effects and the conditions under which they emerge are currently unknown. According to Turri (2016), knowledge-attribution rates for different Gettier-type cases vary from lower than 20% (Gettier intuition demonstrated) to higher than 80% (Gettier intuition not demonstrated); although, the sources of these estimates are unclear. Such inconsistencies in knowledge-attribution rates are perhaps due to two major reasons: (a) people's epistemic intuitions, which lead them to make different judgments about various types of Gettier-type cases based on the characterization of the luckily true justified belief, and (b) variation in experimental designs, including differences in matched controls and some possibly underpowered samples (see Colaço et al., 2014; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013; Nagel, Mar, & Juan, 2013; Powell et al., 2015; Starmans & Friedman, 2012; Turri et al., 2015; Weinberg et al., 2001).

Although the literature on epistemic intuitions has demonstrated varying attribution rates across different types of Gettier-type cases, Powell et al. (2015) and Nagel, Juan, and Mar (2013) provided evidence for Gettier intuitions using counterfeit-object Gettier-type cases. Unlike Turri et al. (2015), Nagel, Juan, and Mar found that participants were more likely to attribute knowledge to a protagonist in a standard JTB condition than a protagonist in a Gettier condition. In reply, Starmans and Friedman (2013) argued that Nagel, Juan, and Mar employed a questioning method that biased participants to deny knowledge, did not properly evaluate the responses of participants who may have attributed knowledge to protagonists in Gettier-type cases, misconstrued the distinction between "apparent" and "authentic" evidence, and used scenarios that did not feature the structure that characterizes most Gettier-type cases. Starmans and Friedman concluded that Nagel, Juan, and Mar's findings are fully compatible with the claim that laypeople attribute knowledge in Gettier-type cases (Gettier intuition not demonstrated; cf. Nagel, Mar, & Juan, 2013).

The Current Study

Some previous research suggests that laypeople may be more likely to attribute knowledge to protagonists who have nonlucky JTBs than to protagonists who have JTBs because of luck alone, thus demonstrating Gettier intuitions (e.g., Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013). However, other investigations have found no differences in knowledge attributions between these conditions (e.g., Starmans & Friedman, 2012; Turri et al., 2015). Because of such inconsistencies in the literature, we sought to estimate the prevalence of Gettier intuitions in a large, highly powered, and international conceptual replication of Turri et al.'s (2015) Experiment 1. In this study, we examined one subset of Gettier-type cases, counterfeit-object cases, using a variety of vignettes, carefully matched controls, and a large cross-cultural sample. Like Turri et al.'s original experiment, in the current study, we explored the frequency of knowledge attribution in response to a protagonist making a correct inference from a false belief.

First, we tested whether participants attributed knowledge to a protagonist differently across three conditions: when the protagonist's belief is justified and true (i.e., in the no-threat or knowledge condition), when the protagonist's belief is justified but true only because of luck (i.e., in the threat or Gettier condition), and when the protagonist's justified belief is false (i.e., in the no-detection or ignorance condition). Following the results of Turri et al. (2015), we predicted that the Gettier condition would produce knowledge attributions at rates no different from the knowledge condition but more frequent than the ignorance condition. Second, we compared participant ratings of the belief's reasonableness by condition to see if, like Turri et al., we would find no condition differences in participant perceptions of what was reasonable for the protagonist to believe. For the original knowledge-attribution and reasonableness results, see Figure 1. We also attempted to replicate Turri and colleagues' findings that participants were more likely than chance to attribute knowledge to protagonists in the no-threat (i.e., knowledge) condition ($p < .001$)
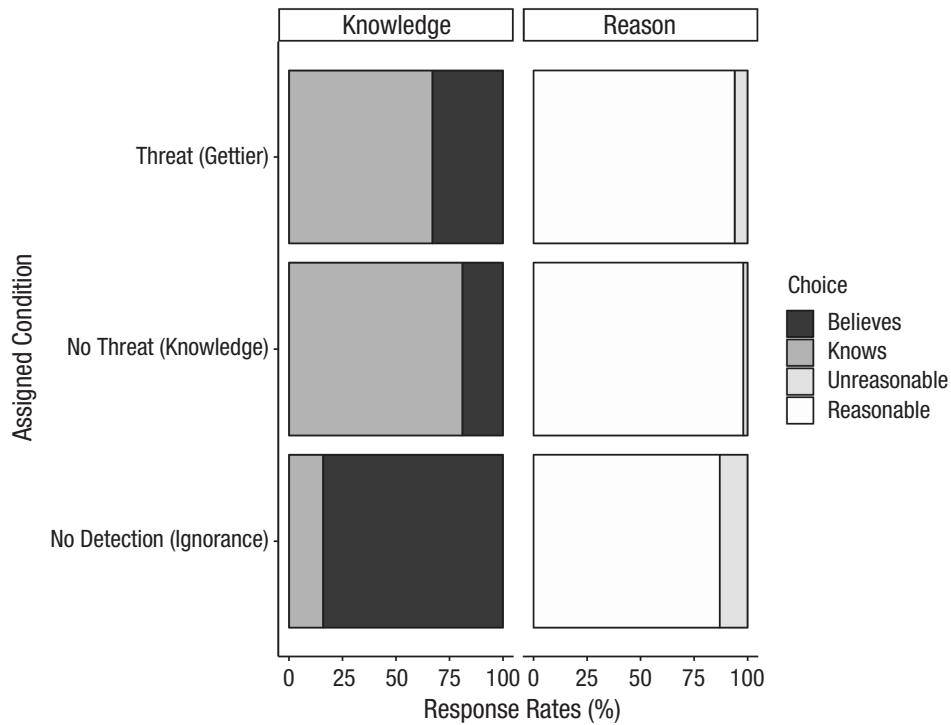
Fig. 1. Results of Turri et al. (2015), Experiment 1.

and in the threat (i.e., Gettier) condition ($p < .001$) but less likely than chance to attribute knowledge in the no-detection (i.e., ignorance) condition ($p = .021$). Finally, to increase the contribution of our replication, we tested the extent to which Turri et al.'s findings generalize across different data-collection sites and vignettes.

Differences from Turri et al. (2015)

Past experimental philosophy research has provided several methodological explanations for inconsistencies in Gettier-intuition research, such as design, measurement, and culture. We modified the original Turri et al. (2015) experiment to address these concerns.

Design considerations. The consensus for explaining inconsistencies in Gettier-intuition research is that the epistemological structure of Gettier-type cases varies depending on the tested vignette or case type (Turri, 2016). The two original counterexamples Gettier used in his 1963 article each described a protagonist who forms an initially justified but false belief from which a true claim is then inferred (Gettier, 1963). Some philosophers now use the term "Gettier case" (or Gettier-type case) to refer to any instance that is intended to illustrate the nonequivalence of JTB and knowledge, wherein a given JTB is supposed to be viewed as not being consistent with knowledge (Nagel, Juan, & Mar, 2013). Alternatively, others have used

the term more specifically to denote cases of the particular inference-from-false-belief type structure featured in Gettier's original article regardless of whether the case itself is viewed as consistent with knowledge (e.g., Weatherson, 2013). We do not define Gettier-type cases as instances that are intended to show a disparity between JTB and knowledge, as Nagel, Juan, and Mar (2013) suggested. Instead, we adopted the latter interpretation by defining Gettier-type cases as scenarios with the structure featured in Gettier's original article, which we used to guide our selection of additional related Gettier-type cases to test.

Ignoring the stimulus variation present in the experimental-philosophy literature would limit the generalizability of our results (Nagel, Juan, & Mar, 2013; Starmans & Friedman, 2012; see also Judd et al., 2012; Yarkoni, 2022). Thus, we attempted to conceptually replicate the original Turri et al. (2015) experiment using additional counterfeit-object Gettier vignettes from the literature (i.e., "Fake Barn" vignette from Colaço et al., 2014; "Diamond" vignette from Nagel, Juan, & Mar, 2013). In these vignettes, a protagonist makes a true inference from a false belief by unknowingly and luckily choosing a true, genuine object among many convincing counterfeits. Doing so allowed us to test the generalizability of Turri and colleagues' Experiment 1 Darrel manipulation to other similar counterfeit-object cases while reducing stimulus sampling error. We decided to test these different vignettes using a mixed design rather than a

between-participants design. Participants were randomly assigned without replacement to each condition and each vignette, resulting in each participant being presented with three vignette/condition combinations. This approach allowed us to parse out the within-participants variation, thereby increasing the statistical power of our analyses to detect and estimate the Gettier-intuition effect.

Measurement considerations. Turri et al. (2015) used a binary measure to assess knowledge attribution. However, in personal correspondence (J. Turri, personal communication, March 10, 2018), Turri stated that participants in knowledge control condition and the Gettier condition may not have differed in their knowledge attributions in the to-be-replicated study because of the study's underpowered sample size and the binary format of the knowledge probe. If laypeople evaluate the knowledge of others along a spectrum, then employing a more scaled measure may reveal differences that could be missed by a dichotomous measure. Subsequent research by one of the original authors measured knowledge with a 7-point Likert-type scale on which participants rated their agreement with a statement claiming a protagonist knew a given proposition (Turri, 2016, Study 2). Although this study used a slightly different vignette than Turri et al.'s Experiment 1, Turri (2016) found a sizable difference ($d = 0.73$) in participant knowledge attributions between a threat (i.e., Gettier) condition and an appropriately matched knowledge control condition.[3] Potentially, the use of a scaled measure allowed for the detection of the Gettier-intuition effect. In the present research, we employed a visual analogue scale (VAS) ranging from 0 to 100 in lieu of the original binary-response (i.e., knows/only believes) variable. The VAS may be as efficacious as a Likert-type response scale and provides more fine-grained data for analysis via parametric statistics than alternatives by allowing for more variability in responding (Bishop & Herron, 2015). Although using a VAS departs from the original study and from how these kinds of judgments are typically made in everyday life, our pretest using a VAS found that participants responded to the control conditions in the expected way with this measure (i.e., knowledge controls and ignorance controls demonstrated paradigmatic rates; see https://osf .io/3ygsk/).

Another addition to our replication was the inclusion of an exploratory knowledge probe. Differences in knowledge attribution may depend on how participants are asked whether a target has knowledge (e.g., Nagel, Juan, & Mar, 2013). To check for these differences in knowledge attribution based on the form of the knowledge question, we asked an exploratory binary knowledge-attribution question after the primary knowledge-attribution question. We also added an exploratory item to assess perceptions of luck and ability that may moderate knowledge attributions in response to

Gettier-type cases (e.g., Turri, 2016). See the Materials and Measures section below for details.

Cultural considerations. Researchers have examined potential cultural sources of variation in knowledge attribution (e.g., Buckwalter & Stich, 2010; Kim & Yuan, 2015; Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013; Nichols et al., 2003; Seyedsayamdost, 2015; Turri, 2013; Turri et al., 2015; Weinberg et al., 2001). For example, Weinberg et al. (2001) reported evidence that participants with Western cultural backgrounds demonstrate Gettier intuitions more often than participants with Eastern cultural backgrounds. However, this preliminary study was underpowered and lacked control conditions; subsequent cross-cultural studies (that also lacked matched controls) found no such cultural differences (e.g., Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Seyedsayamdost, 2015). In one of the largest of these cross-cultural studies, Machery, Stich, Rose, Alai, et al. (2017) provided evidence that people exhibit Gettier intuitions across quite different cultures and languages (i.e., United States, Brazil, India, and Japan); they argued that humans have a "species-typical core folk epistemology" wherein justification, truth, and belief are insufficient for knowledge attribution (p. 12).

Comparisons among these past findings are difficult because of the use of different control conditions that varied in how closely matched they were to the experimental Gettier condition. Although more recent studies have used both knowledge and ignorance control conditions in which participants are exposed to paradigmatic cases of knowledge and ignorance, respectively, most cross-cultural studies have not used closely matched control stimuli (e.g., Kim & Yuan, 2015; Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017; Seyedsayamdost, 2015). For example, Machery, Stich, Rose, Chatterjee, et al. (2017) used a between-participants design with entirely different vignettes and different protagonists for each condition. By contrast, Turri et al. (2015) used slight variations of the same vignette for each condition. Because the versions of the Darrel vignette used in Turri et al. differed only in the words necessary to alter the condition of the protagonist's belief, we also ensured that the two added vignettes (i.e., the "Fake Barn/Gerald" vignette and the "Diamond/Emma" vignette) were implemented with closely matched control conditions. For full details, see Appendix B in the Supplemental Material available online.

Pedagogical goals

A second aim of this project was to provide psychology students across the globe with the opportunity to contribute to a rigorous large-scale research study. We

implemented the model of the Collaborative Replications and Education Project (CREP; Grahe et al., 2014; Wagge et al., 2019) and initiated a collaboration between the CREP and the Psychological Science Accelerator (PSA; Moshontz et al., 2018). The purpose of the CREP is to provide experiential learning opportunities for psychology students while addressing the need for direct replication work in the field of psychology by using the collective power of student research projects. The PSA is an international network of collaborators with a mission to expedite the accumulation of reliable and generalizable evidence in psychological science (Moshontz et al., 2018). The CREP and PSA partnership involved the CREP selecting a study, developing materials, overseeing the quality of the replications using standard CREP procedures, and using the existing PSA network to increase participation among labs. In addition, the PSA's extensive network of experts has supported lab recruitment, translations, data management, and navigating international collaborative research.

Although both the CREP and the PSA have been successful models of multisite collaboration, this project was neither solely a CREP study nor solely a PSA study. The study differed from the typical CREP project in the following ways: (a) It was not a direct replication, (b) it involved a Registered Report, (c) almost all of the data collection was centralized, and (d) students were encouraged but not required to conduct site-level data analysis to earn a CREP completion certificate. The study also differed from the typical PSA project in the following ways: (a) It had significant pedagogical goals, (b) some data were collected independently by labs rather than with a centralized survey, and (c) teams were more autonomous in how they implemented the project. At times, methodological decisions pitted scientific priorities against pedagogical priorities, and pedagogy was prioritized. For example, we allowed students to collect data via Qualtrics surveys that they had created themselves, which allowed for more autonomy and opportunities for students to develop skills but also resulted in some data loss and processing difficulties (see Method section and Appendix A in the Supplemental Material).

Summary

Previous research has produced mixed evidence regarding the presence and size of Gettier-intuition effects. Some of this variation may be explained by differences in the design, measurement, and cultural contexts found across previous investigations. Using counterfeit-object Gettier-type cases, we sought to estimate the effect size of Gettier intuitions across a variety of geopolitical contexts while attempting to address methodological concerns (i.e., measurement sensitivity, lack of matched controls, and stimulus variation). Our results provided evidence regarding the prevalence of Gettier intuitions among lay participants, the extent to which Gettier intuitions are shared across cultures, and the stability of Gettier intuitions across similar scenarios with different protagonists in different contexts.

Disclosures

Preregistration

This study was provisionally accepted as a Registered Replication Report and subsequently preregistered on OSF (see https://osf.io/4bfs7).

Data, materials, and online resources

Study materials, de-identified raw data, de-identified data with exclusions, and analysis code and output are available on our master OSF page (https://osf.io/n5b3w/). Many project teams also posted data on their team's OSF page linked to our master page.

Reporting

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (see Simmons et al., 2011).

Ethical approval

All contributing project teams were required to submit their local institutional ethics approval (if applicable) before data collection as part of their preregistration and CREP review process.

Method

Deviations from provisionally accepted protocol

The protocol for this study was accepted as a Stage 1 Registered Replication Report (https://osf.io/37p8t/; see also Appendix A in the Supplemental Material). In this section, we describe the method as implemented and deviations from the protocol, including minor adjustments to language, corrections of factual inaccuracies, and methodological alterations. The primary deviations from the approved protocol, albeit minor, consisted of changes to study procedure and the analysis plan because of error and adaptations required for valid statistical inference. As detailed below, we changed the methodology according to how surveys were programmed and implemented, how we measured luck attribution, how we measured race/ethnicity, and how we determined the inclusion of data from the student-led

teams. We additionally chose to drop two of the planned covariates, whether the study was conducted individually versus in a group setting and in person versus online, because they were unusable.[4] A number of aspects were not sufficiently described in the original protocol; we therefore clarified the analysis plan in terms of exclusion criteria and data assumption-checking procedures.

Project teams

Each student-led project team prepared a study protocol for approval by a CREP reviewer to ensure quality control. Teams could not contribute to data collection until their protocol was approved. For more information about this process and detailed descriptions of logistical considerations, see Appendix A in the Supplemental Material.[5] In total, 65 student-led teams (i.e., unique teams with OSF pages) signed up to collect data for this project, and 51 student-led project teams were approved to begin data collection using CREP procedure guidelines. Only 47 of these teams contributed to the full data set, which represented 38 data-collection sites. For a summary of the sites and their data-collection features, see Table 1. Teams were not included in the full data set either because they did not collect any data (e.g., because of campus closures during the COVID-19 pandemic) or because the data they collected were unusable for analyses (e.g., vignettes were not properly randomized). After applying the participant-level exclusions described below, the final data set included 45 student-led project teams across 37 data-collection sites. Of those 45 teams, 22 received CREP completion certificates.[6] Although we initially planned to include only the data from teams that received completion certificates, we decided to include all usable data from teams that were approved to start data collection (see Analytic Approach).

Participants

In the analysis sample (i.e., after the exclusions described below), participants were 4,826 adults recruited to participate by student researchers at 37 data-collection sites in various geopolitical contexts across geographical regions (i.e., Northern America, Eastern Europe, Western Europe, Northern Europe, Southern Europe, Australia and New Zealand, Western Asia, Southeastern Asia, Eastern Asia). For sample sizes by geopolitical region, see Table 2. Data collection took place between January 1, 2019, and June 1, 2021.[7] Data-collection sites contributed a median of 81 participants to the analysis sample (minimum = 28, maximum = 588); six sites collected fewer participants than the target of 50. On average, participants were young (age: $M = 24.84$ years, $SD = 9.91$; $n = 4,826$) and had completed some college, as measured by years of education (education: $M = 13.84$ years, $SD = 2.59$; $n = 4,771$).[8] Most participants (70.37%; $n = 3,396$) identified as White.[9] Over half of participants identified as female (70.56%; $n = 3,405$), and most other participants identified as male (29.01%; male: $n = 1,400$; neither: $n = 21$). The plurality of participants completed the survey in English (47.53%; $n = 2,294$). Participation details, such as compensation and the sampled population, varied by data-collection site. For a summary, see Table 3.

Exclusions. Of the 9,440 participants who completed the survey, data from 48.88% ($n = 4,614$) were excluded from the analytic sample. Of this total, 2,187 participants (23.17%) were flagged for exclusion based on multiple criteria. All listed exclusions were preregistered with one exception (i.e., maximum age).[10] Participants were excluded for the following reasons.

Age. The participant did not provide an age, listed an age greater than or equal to 100, or was not the age of majority of their geopolitical region, operationalized as at least 18 in all regions except Taiwan, where the age of majority is 20 (total excluded: $n = 2,118$; missing: $n = 2,040$; 22.44% of participants met this exclusion criterion).

Prior participation. The participant had taken part in a previous version of this study or in another contributor's replication of the same study ($n = 238$; 2.52% of participants met this exclusion criterion).

Comprehension. The participant failed to answer all three of the vignette comprehension questions correctly (e.g., did not correctly identify whether Darrel was looking at a squirrel or a prairie dog; total excluded: $n = 4,376$; missing: $n = 1,490$; 46.36% of participants met this exclusion criterion).[11] For rates of correct responses by vignette and condition combination, see Table 4.

Knowledge of hypothesis. Participants correctly and explicitly articulated knowledge of the specific hypotheses or specific conditions of this study when asked what they thought the study hypothesis was ($n = 203$; 2.15% of participants met this exclusion criterion).

Language proficiency. Participants reported their understanding of the language the survey was presented in as "not well" or "not well at all" (total excluded: $n = 2,093$; missing: $n = 2,003$; 22.17% of participants met this exclusion criterion; for criteria, see Vickstrom et al., 2015).

For item details, see the Materials and Measures section below. The rate at which participants were excluded because of failed comprehension in the present study (46%) was consistent with prior cross-cultural Gettier-intuition research (e.g., rates between 21% [Machery,

Table 1. Characteristics of Data-Collection Sites

Language	Geopolitical region	N teams	Full N	Final N	Sample	In person?	In a group?	Compensation	Site-level analysis
Chinese (traditional)	Taiwan	1	452	89	Undergraduates	No	No	Money	ANOVA (VAS) + chi-square (binary)
English	Australia	1	215	122	Undergraduates	No	No	Credit	None
		1[a]	165	119	Undergraduates	No	No	Credit	None
	Canada	1	551	258	Undergraduates	No	No	Credit	None
	United Kingdom	2	340	132	Undergraduates	No	No	Credit	Friedman
	Greece	1	98	52	Both	No	No	None	None
	New Zealand	1	58	42	Undergraduates	Yes	No	Lottery	ANOVA
	Singapore	1	78	52	Undergraduates	Yes	No	Credit	None
	United States	1	124	57	Undergraduates	No	No	Credit	ANOVA
		1	387	221	Undergraduates	Yes	Unclear	Credit	Linear mixed model
		2	402	201	Undergraduates	No	No	Credit	ANOVA
		1	91	48	Undergraduates	Both	No	Credit	None
		1	164	93	Undergraduates	No	No	Credit	ANOVA
		1	64	43	Undergraduates	Yes	Unclear	Credit	ANOVA
		1	187	78	Undergraduates	No	No	Credit	Descriptives only
		1	129	56	Undergraduates	No	No	Credit	ANOVA
		1[a]	510	356	Community	No	No	Money	Linear mixed model
		1	213	103	Undergraduates	No	No	Credit	ANOVA
		1	135	63	Both	No	No	Students: credit; community: none	None
		1	162	93	Undergraduates	No	No	Credit	None
		1	103	36	Undergraduates	No	No	Credit	None
		1	112	56	Undergraduates	No	No	Credit	Descriptives only
		1	95	54	Undergraduates	No	No	Credit	None
		1	8	NA	Undergraduates	Unclear	Unclear	Unclear	None
French	Switzerland	1	58	28	Undergraduates and community	Both	Both	Undergraduates: credit; community: money	Bayesian ANOVA

Table 1. *(continued)*

Language	Geopolitical region	N teams	Full N	Final N	Sample	In person?	In a group?	Compensation	Site-level analysis
German	Austria	1	159	76	Both	No	No	None	None
		1	143	75	Both	No	No	None	None
	Germany	7	1102	588	Undergraduates	Both	Both	Varied across teams (sweets, money, credit)	One team used McNemar for binary and Quade test for VAS; otherwise none
Hungarian	Hungary	1	332	184	Undergraduates	No	No	Credit	None
		1	658	449	Undergraduates	Both	Both	Credit	None
Norwegian	Norway	1	147	76	Both	No	No	Students: sweets; community: none	None
Polish	Poland	1	179	72	Both	Both	Both	Students: credit; community: none	MANOVA
		1	243	121	Both	No	No	None	None
Portuguese	Portugal	1[a]	149	81	Both	No	No	Students: credit; community: none	Chi-square and binomial tests
Romanian	Romania	1	661	371	Undergraduates	No	No	Credit	None
Russian	Russia	1	233	99	Both	No	No	None	None
Slovak	Slovakia	1	229	105	Both	No	No	None	ANOVA
Turkish	Türkiye	1	304	77	Undergraduates	No	No	Credit	None
Total		46	9,440	4,826					

Note: Full and Final N indicate sample size before and after exclusions. Data-collection context variables were gleaned from OSF page documentation and confirmed by the team when possible. "Unclear" indicates lack of documentation. Site-level analyses were conducted independently and not included in the analyses presented here. See the team OSF pages for details and results. ANOVA = analysis of variance; VAS = visual analogue scale; MANOVA = multivariate analysis of variance.
[a]Team collected data using Qualtrics instead of SoSciSurvey.

Table 2. Number and Percentage of Participants in the Analysis Data Set (After Exclusions) by Geopolitical Region

Geopolitical region	*n*	% of total
United States	1,558	32.28
Germany	772	16.00
Hungary	449	9.30
Romania	371	7.69
Canada	258	5.35
Australia	241	4.99
Poland	193	4.00
Austria	151	3.13
United Kingdom	132	2.74
Slovakia	105	2.18
Russia	99	2.05
Taiwan	89	1.84
Portugal	81	1.68
Türkiye	77	1.60
Norway	76	1.57
Greece	52	1.08
Singapore	52	1.08
New Zealand	42	0.87
Switzerland	28	0.58

Note: "Geopolitical region" refers to the location of the data-collection site except for one team that collected data through Amazon Mechanical Turk in another geopolitical region (i.e., the United States). For all other data-collection sites, participants were recruited from the geopolitical region of the site.

Stich, Rose, Chatterjee, et al., 2017] and 47% [Machery, Stich, Rose, Alai, et al., 2017]). Across Gettier-intuition studies more broadly, such exclusions have rarely had an impact on results (for review, see Popiel, 2016).

Power analysis. We conducted an a priori power analysis using the *powerCurve* function in the *simr* package (Green & MacLeod, 2016) in R to estimate the sample size required to detect an effect of knowledge condition on participants' knowledge attributions with 90% power at $\alpha = .05$.[12] To estimate the effect size, we considered (a) the effects observed in our pilot-test data (difference between Gettier and knowledge, $\beta = 0.32$; difference between Gettier and ignorance, $\beta = -0.44$), (b) both the difference between the Gettier condition and knowledge condition (Cramér's $V = .509$) and the small nonsignificant difference between the Gettier condition and ignorance condition (Cramér's $V = .16$) from Experiment 1 of Turri et al. (2015), and (c) the small effects sometimes found in the literature (e.g., Machery, Stich, Rose, Alai, et al., 2017). To be conservative, we selected a standardized fixed effect in the multilevel model analysis described below of .1 for our power analyses.

The model tested included random intercepts for data-collection site, vignette, and participants such that vignettes were nested within participants, who were

nested within sites. We simulated data using a standardized fixed-effect regression parameter of .1. In these simulations, the number of participants per site was allowed to vary, but the number of vignettes (three) and the number of collection sites (nine) were held constant. Results suggested that at least 32 participants per data-collection site (i.e., 288 total participants; 864 total observations) would be necessary to detect the identified fixed-effect regression parameter (.1) 90% of the time with an alpha of .05. Considering the potential for attrition (e.g., because of lack of comprehension) and effect-size heterogeneity between data-collection sites (Kenny & Judd, 2019), we set a target sample size of 50 participants per data-collection site. Of the 46 data-collection sites included in analyses, 45 met this target before exclusions, and 40 met the target after exclusions.

Materials and measures

As described in the approved protocol, we planned to collect all data using a single SocSciSurvey survey programmed to accommodate lab-specific variations. However, eight student-lead teams used Qualtrics surveys programmed by student researchers; some Qualtrics teams used versions created by other Qualtrics teams. The majority of the data collected via Qualtrics was not included in the full data set because of logistical challenges (e.g., no access to raw survey data); only three of the teams included in the analysis data set used Qualtrics surveys ($n = 556$ after exclusions).[13] All materials used in this replication are available in Appendix B in the Supplemental Material and at https://osf.io/n5b3w.

Vignettes. In addition to the "Squirrel/Darrel" vignette from Turri et al. (2015), two vignettes were selected on the basis of their similarity to the original vignette, their quality, and their prevalence in the literature: the "Fake Barn/Gerald" vignette (Colaço et al., 2014; altered to more closely match

Table 3. Number and Percentage of Participants by Data-Collection Context Variables

Variable	*n*	% of total
Compensated for participation	3,533	73.21
Recruited through Amazon Mechanical Turk	356	7.38
Completed the centralized survey	4,270	88.48

Note: Variables are not exclusive. Information about the compensation method was obtained by examining each student-led team's institutional review board approval, confirming with the students or principal investigators at each site, and making inferences based on the data-collection site's specific surveys when neither source was available. Three data-collection teams included in analyses used Qualtrics to distribute their surveys instead of the centralized survey programmed in SoSciSurvey.

Table 4. Comprehension Question Correct Answer Rates by Condition and Vignette Combination

	Gettier		Ignorance		Knowledge	
	Total	Correct	Total	Correct	Total	Correct
Darrel	2,821	1,986 (70.40%)	3,153	2,119 (67.20%)	2,972	2,174 (73.15%)
Emma	2,982	2,009 (67.37%)	3,034	2,104 (69.35%)	2,930	2,085 (71.16%)
Gerald	3,143	1,942 (61.79%)	2,759	2,001 (72.53%)	3,044	2,035 (66.85%)
Missing across vignettes and condition	494		494		494	

Note: Participants were excluded from analyses if they incorrectly answered any of the comprehension questions.

the "Squirrel/Darrel" vignette) and the "Diamond/Emma" vignette (Nagel, Juan, & Mar, 2013). The vignettes as administered in this study are reported in full in Appendix B in the Supplemental Material. The vignettes were pretested to ensure they effectively manipulated the target construct and produced sufficient participant comprehension (see https://osf.io/3ygsk/). Four student-lead teams participated in an optional extension that included a fourth vignette after the main study protocol to test the effects of perceived expertise on Gettier intuitions (see Larkin & Andreychik, 2019). However, we did not use the data from this extension in any of the analyses reported in this article.

For each vignette, participants were randomly assigned without replacement to one of three conditions: a Gettier-type condition in which the vignette subject correctly identified the target but not because of the reason the subject thought it to be true (i.e., the threat condition in Turri et al., 2015), a knowledge control condition in which the subject correctly identified the target because of the subject's knowledge (i.e., the no-threat condition in Turri et al.), and an ignorance control condition in which the protagonist incorrectly identified the target (i.e., the no-detection condition in Turri et al.).

Dependent measures. After each vignette, two primary and two exploratory dependent variables were measured. In line with the approved protocol, all student-led teams included the default VAS ranging from 0 to 100 for three of these variables (i.e., knowledge attributions, reasonableness judgments, and attributions to luck vs. ability). However, six teams also participated in an optional extension that randomly assigned participants to take the study with either entirely continuous-scale measures or entirely binary-choice measures for these variables.[14] Overall, for each of the three measures, 86.52% of responses used in analyses were originally measured on the continuous scale. For the exact question text, see Appendix B in the Supplemental Material.

Knowledge attributions. Participants were asked whether the protagonist believes or knows the stated proposition.

Reasonableness judgments. Participants were asked to rate the extent to which the protagonist's belief was unreasonable or reasonable.

Luck/ability attributions. For this exploratory measure, participants were asked two questions relevant for evaluating their attributions of outcomes to luck or ability. First, participants were asked whether the protagonist got the "right" or "wrong" answer. Then, participants were asked whether the protagonist's "right" or "wrong" answer was due to the protagonist's ability/inability or good luck/bad luck on one of the two scales.[15] If participants selected the incorrect answer to the first part of the question, they were subsequently excluded from the luck-attribution analyses because their response indicated that they did not comprehend whether the protagonist held the given true belief.

Alternative knowledge attribution. In addition, participants were asked a binary alternative-knowledge probe in which participants chose whether the protagonist either knew what the target of identification was or felt like the protagonist knew what the target was but did not actually know. For example, after the Darrel vignette, participants were asked, "In your view, which of the following sentences better describes Darrel's situation?" Participants could then select one of two response options: "Darrel knows that the animal he saw is a red speckled ground squirrel" or "Darrel feels like he knows that the animal he saw is a red speckled ground squirrel, but he doesn't actually know that it is."

Demographics and participation characteristics. Participants were asked to report their age, gender, geopolitical region (i.e., "What country do you currently live in?" and "What is your country of birth?"), the number of years they had attended school, and their race or ethnicity. Because of differences in how student-led teams measured these items, we matched item answers across different implementations of the survey. Participants also completed a 12-question study-experience questionnaire that was not used in analyses (see Appendix C in the Supplemental Material).

Education level. All participants were asked a question about their education. Participants who completed the study in SocSciSurvey were asked about the number of years they had been in school (truncated at 18). Participants who completed the survey in Qualtrics were asked about their educational attainment. Education (in years) was imputed for participants who reported their educational attainment from these three sites ($n = 553$).[16] The years of education for these sites was also truncated to match how this item was measured in SocSciSurvey such that any value above 17 was recoded as 18.

Compensation. Participants were asked whether they were compensated for their participation (i.e., "Will you receive any kind of compensation or reward for taking part in this study?") and indicated the type of compensation (e.g., the number of course credits, the amount of money). Some student-led teams opted not to include this question in their survey because all participants were compensated the same way. The method of compensation described in the site's approved institutional review board (IRB) protocol was imputed for those missing responses. Among participants who were asked about their compensation, responses were sometimes missing or discrepant with the documented method of compensation. For student-led teams in which fewer than 50% of participants in the final data set agreed on a method of compensation, the method of compensation described in the data-collection site's approved IRB protocol was imputed for all participants if a single method of compensation was described.

Comprehension and language proficiency. Participants were asked to indicate the true correct answer for each vignette as a comprehension check that was used for listwise exclusions. Participants were also asked to rate their proficiency for the survey language. The original article asked participants whether they were native English speakers but did not seem to exclude participants on this basis. Given that the tasks in the present study were highly dependent on language comprehension and proficiency and that participants had a 12.5% chance (i.e., 1 in 8) of passing all three comprehension questions based on guesses, we decided an additional check of self-reported language proficiency would be helpful in excluding participants who did not understand or may not have understood the task completely.

Prior participation and knowledge of study. We also asked participants to describe what they thought the hypothesis of the study was (used for exclusions), provide their impression of study materials (not used in any analyses), and indicate whether they had participated in a similar study (used for exclusions). The original study did not contain these three questions, but the researchers excluded Amazon Mechanical Turk (MTurk) workers if they had already participated. Evaluating the hypothesis and prior-participation exclusion criterion required subjective judgments about open-ended responses. Each nonmissing observation was evaluated by three raters who spoke the language of the provided response. These three raters did not translate responses but instead directly evaluated responses with respect to the exclusion criteria. Responses marked "yes" (i.e., meets criteria) were assigned 2 points, responses marked "maybe" (i.e., may meet criteria) were assigned 1 point, and responses marked "no" (i.e., does not meet criteria) were assigned 0 points. After summing points for each response across the three raters, we excluded cases with 4 or more points on either response. See Appendix D in the Supplemental Material for the instructions given to raters and http://osf.io/gs29c for the ratings data. Responses identified by raters as test cases (e.g., "TEST") were excluded (study purpose: $n = 222$; previously participated: $n = 170$).[17] Responses that were not coherent were labeled but not excluded (study purpose: $n = 5$; previously participated: $n = 3$).

Procedure

After providing informed consent, participants read and answered questions about three vignettes that described counterfeit-object cases. Each participant responded to three condition and vignette combinations randomly assigned on each factor without replacement such that all participants saw each vignette (Darrel, Emma, Gerald) and each condition (ignorance, knowledge, Gettier) exactly once. After reading each vignette, participants responded to a series of items in a fixed order on separate screens. Items were presented as follows: knowledge attribution, comprehension check, reasonableness judgment, luck attribution (two items), and alternative-knowledge probe. Next, participants answered questions related to their experience completing the study, data-exclusion criteria, and demographics, respectively. Finally, participants were debriefed and compensated if applicable.

Analytic approach

Analyses were conducted on combined raw data collected in SocSciSurvey and Qualtrics. In the original protocol, we planned to evaluate the quality of each student-led team's data, including the raw data, analysis scripts, codebooks, cleaned data sets, and narrative summaries of results. We also planned that data would be included in analyses only if teams received a CREP completion certificate after these products passed a quality check. However, the original protocol did not describe clear criteria that would be used to detect and correct errors, and many teams did not submit their
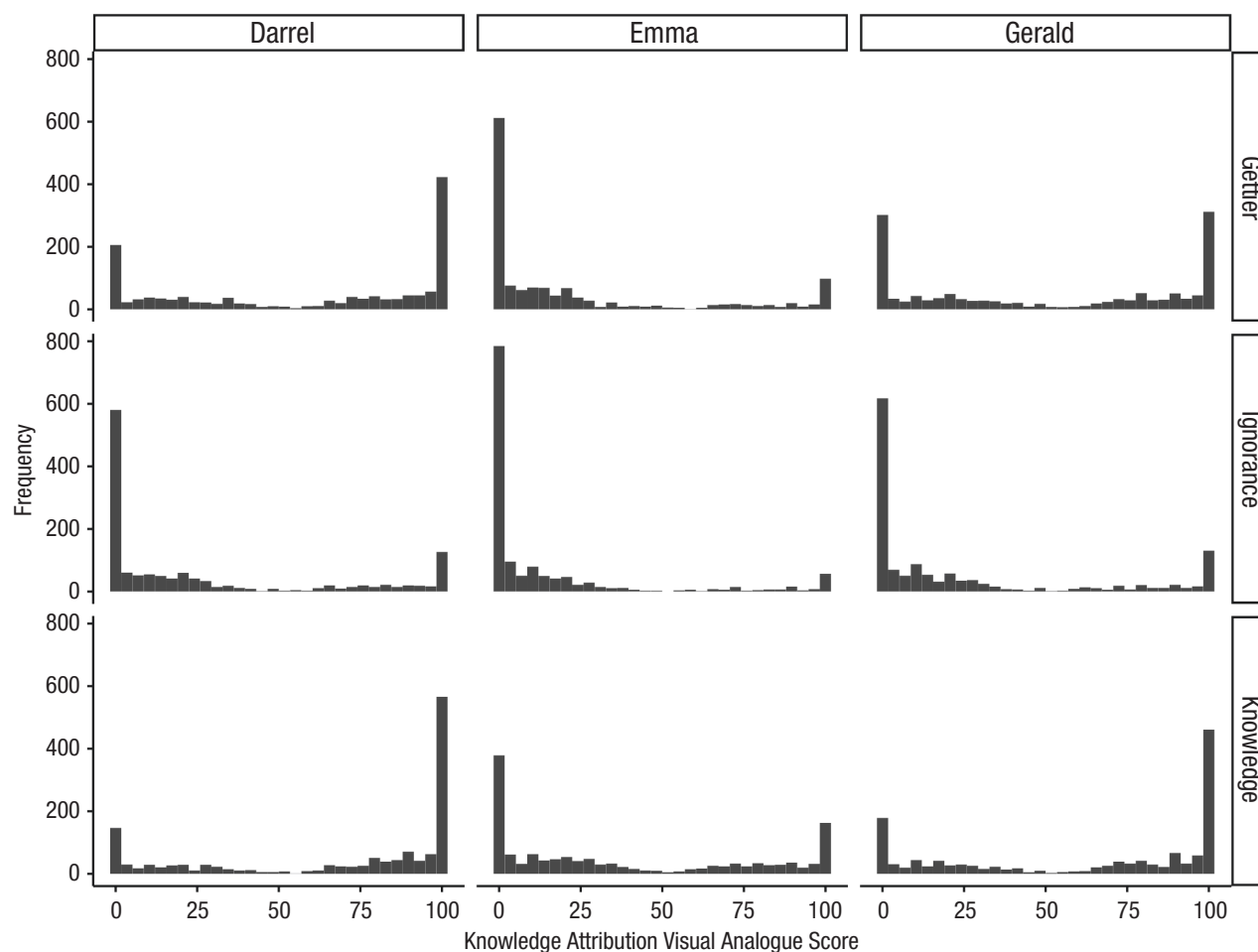
Fig. 2. Knowledge-attribution visual analogue scores by vignette and condition.

projects for final CREP review. To conduct reproducible, transparent analyses, we chose not to exclude data from teams who failed to meet the target sample size or did not receive completion certificates. All teams were required to receive CREP approval before commencing data collection; this process included preparing an OSF page with all materials and videos of their procedure, submitting the page for review by CREP reviewers, and making any revisions as necessary. If data-collection teams received approval and collected their data using the centralized survey, their data were also included in analysis. Because of this oversight and the strict data-quality exclusions implemented at the level of participants, we were not concerned about team-level variation in data quality. Still, we repeated our primary analyses excluding data from the teams that did not receive completion certificates. Generally, we observed the same patterns of results (see https://osf.io/nvfbm).[18] A summary of how the teams independently analyzed their data (i.e., the test used for the effect of condition on knowledge attribution) is reported in the last column

of Table 1, and those results can be found on their OSF pages.

Multilevel models were used to evaluate our hypotheses. The unit of analysis was the question response, and cross-classified random intercepts for the vignette, participant, and data-collection site were included to account for the nesting of responses within these groups.[19] Exact model specification can be found at https://osf.io/8ut6e/.

Assumptions and transformations. Although the approved protocol described testing assumptions before conducting analyses, it did not detail criteria that would be used for testing assumptions or approaches to handling model-convergence issues. No convergence issues emerged during analyses. Here, we describe the approach taken to test assumptions. Assumptions of and related to linearity are primarily relevant for the analysis of the continuously measured dependent variables. The continuous knowledge-attribution variable was bimodal overall and within vignette and condition combinations (see Fig. 2).

To examine normality, homogeneity, and linearity, we used linear mixed models that predicted continuously measured knowledge, reasonableness, and luck attribution as a function of condition with covariates of compensation, age, gender, and education. The residual distributions were also bimodal or heavily skewed, indicating violations of the residual normality assumption. Furthermore, plots of residuals by fitted values suggested that residuals varied as a function of predicted values, indicating violations of the homoscedasticity assumption. Last and most important, the linearity assumption was not met for any dependent variable that each showed a sigmoid function similar to binary outcome data.

Transforming continuous variables into discrete variables for analysis is not generally recommended (MacCallum et al., 2002; Maxwell & Delaney, 1993). For the present analyses, however, this approach was necessary because of the already bimodal distribution of the dependent variables and the suggested sigmoid function from the residual data-screening results. Thus, we split the continuously measured versions of the three dependent variables such that scores at and below 40 and scores at and above 60 were classified into discrete categories. Higher scores were coded 1 to indicate knowledge, reasonableness, or ability, and lower scores were coded 0 to indicate belief, unreasonableness, or luck. We chose these points so that participants clearly had indicated a side (i.e., 41–59 were considered neutral), and very few data points were lost in this middle range. Of the nonmissing responses on each continuous measure, 359 (2.87%) responses were dropped for the knowledge-attribution variable, 279 (2.23%) responses were dropped for the reasonableness-attribution variable, and 683 (5.85%) responses were dropped for the luck-attribution variable.

This approach allowed us to validly interpret model results and also test whether the method of measurement (continuous or binary) affected results. Data screening was examined for logistic models with the same parameters as above; the assumptions of logistic regression were met: no empty or small categories, linearity of the logit for continuous predictors, and additivity of the predictors. We repeated our primary analyses with the continuous dependent measures using linear regressions to see whether this deviation affected our findings. Overall, we found the same pattern of results.[20] See https://osf.io/nvfbm for details.

Model steps. A series of multilevel logistic regression models were fit predicting knowledge attributions and reasonableness judgments. Transformed and originally binary responses were analyzed together. Each model was fit including all participants with no missing data on that model's variables. After estimating a baseline intercept-only model (Model 1), we fit models with random intercepts for vignette (Model 2), person (Model 3), and data-collection site (Model 4) added sequentially. In Model 5, participant age, compensation, gender, and education (in years) were added as fixed effects. These variables served as covariates and were included in our original analysis plan given previous research that demonstrated their impact on knowledge attribution. Finally, the knowledge-condition variable was added in Model 6. To see if the effect of condition varied by vignette, the interaction between vignette and condition was added as a fixed effect in Model 6A. Additional models were fit to test the moderating effects of participant source (Model 6B; MTurk vs. lab), luck attributions (Model 6C; luck vs. ability), and the original measurement scale (Model 6D; binary vs. continuous). The conceptual models presented in Models 1 through 6B were preregistered, maintaining independent and random-effects variables in the updated analysis plan. Model 6D was added when the data screening indicated the VAS results were not continuous as expected, and the dependent variables were dichotomized. The exact implementation of the multilevel models (i.e., model order and interpretation) were updated from our preregistered plan to ensure appropriate statistical inference (for full details, see Appendix A in the Supplemental Material).

Results

To better test our research questions, we implemented analyses that differed from those we originally planned.[21] All deviations are summarized in Appendix A in the Supplemental Material. The Results section as it appeared in the approved protocol is also included in Appendix A in the Supplemental Material with updated statistics where possible. Although the results below indicate that components of the random structure (i.e., intercepts of participant and site) do not add to or improve the models, we included these facets to match the preregistered plan and to maintain independence of observations (i.e., participant intercepts are arguably necessary for a repeated measures design). The lack of participant variance suggests that individuals did not systematically vary in their responding across vignette-condition combinations; the lack of site variance suggests that results were consistent across data-collection sites.

For each focal model, we report the model-fit statistics and parameter estimates. Parameter estimates for logistic models can be interpreted in a similar fashion to linear regression models: Negative values indicate that increasing the predictor decreases the likelihood of the dependent variable (e.g., the choice coded 1, therefore, increasing the likelihood of the choice coded 0), and positive values indicate that increases in the predictor correspond to increases in the likelihood of the dependent variable (e.g., the choice coded 1). When predictors are also categorical, increasing the predictor indicates a

Table 5. Knowledge-Attribution Model Summaries

Parameter estimate or statistic	Model						
	1	2	3	4	5	6	6A
Fixed effects							
Intercept	−0.44*** (0.02)	−0.49 (0.35)	−0.49 (0.35)	−0.49 (0.35)	−0.37 (0.36)	−0.18 (0.40)	0.56*** (0.13)
Age					0.003 (0.00)	0.004 (0.00)	0.004 (0.00)
Gender					−0.07 (0.04)	−0.08* (0.04)	−0.09* (0.04)
Education					−0.02* (0.01)	−0.02* (0.01)	−0.02* (0.01)
Compensation					0.02 (0.04)	0.02 (0.04)	0.02 (0.04)
Condition: ignorance						−1.31*** (0.05)	−1.60*** (0.08)
Condition: knowledge						0.61*** (0.04)	0.50*** (0.08)
Vignette: Emma							−1.93*** (0.08)
Vignette: Gerald							−0.40*** (0.07)
Ignorance × Emma							0.98*** (0.13)
Ignorance × Gerald							0.21* (0.11)
Knowledge × Emma							0.40*** (0.11)
Knowledge × Gerald							0.02 (0.11)
Random effects							
Site				< 0.001	< 0.001	< 0.001	< 0.001
Participant			< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Vignette		0.600	0.600	0.600	0.601	0.669	< 0.001
Akaike information criterion	18,881.09	17,834.75	17,836.75	17,838.75	17,554.31	15,871.99	15,807.69

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increasing likelihood of knowledge attribution. For condition, the comparison group was Gettier, and for vignette, the comparison group was Darrel. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*p < .05. ***p < .001.

comparison between the predictor group coded as 0 and the predictor group coded as 1. All pseudo-R^2 values were calculated with the *MuMIn* package (Bartoń, 2020) using formulas for fixed and random effects from Nakagawa et al. (2017).

Knowledge attribution

The goal of the present research was to provide a well-powered estimate of the magnitude and prevalence of Gettier intuitions (i.e., the difference in knowledge attribution between Gettier and knowledge conditions) across different vignettes and testing sites in a replication and extension of Turri et al. (2015). Models were fit in steps to determine whether participants attributed knowledge to the protagonist at different rates as a function of condition. For a summary of model results, see Table 5. Compared with the baseline Model 1 (Akaike information criterion [AIC] = 18,881.09), the model including random intercepts for vignette (AIC = 17,834.75) explained more variance (pseudo R^2s = .08–.10). Participants attributed knowledge most frequently in response to the Darrel vignette (52.16%) and least frequently in response to the Emma vignette (20.94%). For differences by vignette extracted from Model 2, see Table 6.

Table 6. Knowledge Attributions From Model 2 Overall and by Vignette

	Overall	Darrel	Emma	Gerald
Believes	8,595 (60.92%)	2,268 (48.12%)	3,716 (78.95%)	2,611 (55.70%)
Knows	5,513 (39.08%)	2,445 (51.88%)	991 (21.05%)	2,077 (44.30%)

Table 7. Knowledge Attributions From Model 6 Overall and by Condition

	Overall	Knowledge	Ignorance	Gettier
Believes	8,476 (61.00%)	2,005 (43.41%)	3,833 (82.06%)	2,638 (57.29%)
Knows	5,419 (39.00%)	2,614 (56.59%)	838 (17.94%)	1,967 (42.71%)
$\chi^2(1)$		80.29***	1920.37***	97.77***
Darrel				
Believes	2,239 (48.24%)	454 (28.73%)	1,170 (76.82%)	615 (39.99%)
Knows	2,402 (51.76%)	1,126 (71.27%)	353 (23.18%)	18923 (60.01%)
Gerald				
Believes	2,570 (55.68%)	558 (36.83%)	1,255 (79.63%)	757 (49.64%)
Knows	2,046 (44.32%)	957 (63.17%)	321 (20.37%)	768 (50.36%)
Emma				
Believes	3,667 (79.06%)	993 (65.16%)	1,408 (89.57%)	1,266 (82.10%)
Knows	971 (20.94%)	531 (34.84%)	164 (10.43%)	276 (17.90%)

Note: The χ^2 tests comparing participant knowledge attributions in each condition with chance were conducted using data from Model 6.
***$p < .001$.

The model nesting vignette within participants (Model 3; AIC = 17,836.75) explained similar amounts of variance (pseudo R^2s = .08–.10) as Model 2. The addition of the random effect of data-collection site in Model 4 (AIC = 17,838.75) likewise did not improve model fit (pseudo R^2s = .08–.10). The model including the covariates predicting knowledge attributions as fixed effects (Model 5; AIC = 17,554.31) was more useful in explaining variance in knowledge attribution than previous models. Age predicted knowledge attribution such that as age increased, participants were more likely to attribute knowledge to the protagonists. Education was a negative predictor; rates of knowledge attribution decreased as reported education increased. However, these fixed effects accounted for a very small proportion of the variance, pseudo $R^2 < .001$.

Model 6 served as the key replication test of Turri et al. (2015). The knowledge condition was added as a fixed effect (AIC = 15,539.57). This model performed better than the previous model and revealed an effect of condition on knowledge attribution (pseudo R^2s = .12–.15). See Table 5 for model statistics and Table 7 for knowledge-attribution rates by condition. Participants were more likely to attribute knowledge to the protagonist in the knowledge-condition vignette than to the protagonists in the ignorance- and Gettier-condition vignettes; furthermore, the ignorance condition differed from the Gettier condition. Thus, we did not fully replicate the results of Turri et al., who found no difference in knowledge attribution between the knowledge and Gettier conditions. Using the data from this model, each condition was examined for difference from chance using χ^2 tests. In the knowledge condition, participants were more likely than chance to attribute knowledge to the protagonist. Participants were less likely than chance

to attribute knowledge to the protagonists in the ignorance- and Gettier-condition vignettes, all ps < .001 (see Table 7).

To better understand whether the effect of condition varied as a function of the vignette's content, Model 6A was estimated including an interaction between vignette and condition (AIC = 15,807.69). This model fit the data better (pseudo R^2s = .20–.24) than Model 6. As shown in Figure 3, the pattern of results was the same for every vignette; however, values suggest that the interaction between condition and vignette accounted for some of the variance in knowledge attributions. The size of the differences between conditions (and between vignettes) depended on the vignette-condition combinations.

In responding to the Darrel vignette, participants attributed knowledge at different rates according to the vignette's condition, $\chi^2(2) = 781.00$, $p < .001$. Participants were more likely to attribute knowledge when responding to the Gettier-condition version ($\hat{p} = .60$) than in the ignorance condition version ($\hat{p} = .23$; Cramér's $V = .37$, 95% confidence interval [CI] = [.34, .41], $\chi^2[1] = 425.61$, $p < .001$). They were also more likely to attribute knowledge to Darrel when responding to the knowledge condition version ($\hat{p} = .71$) than in the Gettier = condition version ($\hat{p} = .60$; Cramér's $V = .12$, 95% CI = [.08, .15], $\chi^2[1] = 43.30$, $p < .001$).

The pattern of responding was similar for the Emma vignette; the likelihood that participants attributed knowledge to Emma differed according to the vignette's condition, $\chi^2(2) = 291.42$, $p < .001$. Participants were more likely to attribute knowledge when responding to the Gettier condition of the Emma vignette ($\hat{p} = .18$) than in the ignorance condition of the Emma vignette ($\hat{p} = .10$; Cramér's $V = .11$, 95% CI = [.07, .14], $\chi^2[1] = 35.15$, $p < .001$). The likelihood of knowledge attribution was higher
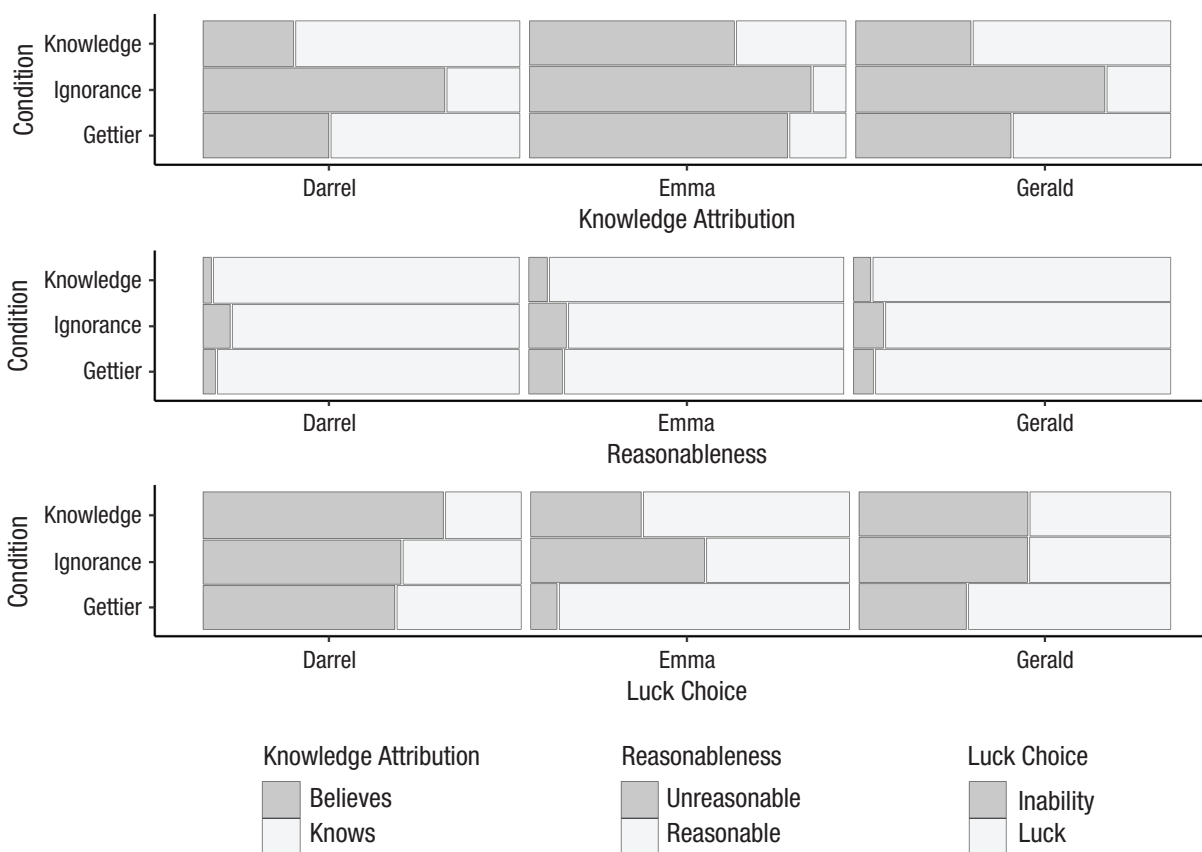
Fig. 3. Knowledge attribution, reasonableness, and luck/(in)ability rates by vignette and condition.

for the knowledge version of the vignette ($\hat{p}$ = .35) than for the Gettier version ($\hat{p}$ = .18; Cramér's V = .19, 95% CI = [.16, .23], $\chi^2[1]$ = 112.59, $p < .001$).

In response to the Gerald vignette, participant knowledge attributions similarly differed according to vignette condition, $\chi^2(2)$ = 607.03, $p < .001$. Participants were more likely to attribute knowledge in response to the Gettier-condition version of the Gerald vignette ($\hat{p}$ = .50) than to the ignorance-condition version of the Gerald vignette ($\hat{p}$ = .20; Cramér's V = .31, 95% CI = [.28, .35], $\chi^2[1]$ = 304.67, $p < .001$). In addition, they were more likely to attribute knowledge to Gerald in the knowledge-condition version ($\hat{p}$ = .63) than in the Gettier-condition version ($\hat{p}$ = .50; Cramér's V = .13, 95% CI = [.09, .17], $\chi^2[1]$ = 50.27, $p < .001$).

To interpret the condition by vignette interaction, we examined Cramér's V for the analyses of each vignette. This approach revealed that the likelihood of knowledge attributions in the Gettier and ignorance conditions differed less for the Emma vignette than for the Darrel and Gerald vignettes. In addition, the Gettier and knowledge conditions of the Darrel vignette produced a smaller difference in likelihood than that for those conditions

of the other two vignettes. Thus, participants demonstrated Gettier intuitions in all three vignettes (i.e., participants were more likely to deny knowledge in the Gettier condition than in the knowledge condition, a case of JTB), but these Gettier intuitions were weakest in response to the Darrel vignette and strongest in response to the Emma vignette.

Reasonableness judgments

As a secondary dependent measure, judgments of reasonableness were predicted in a series of logistic regression models paralleling those for knowledge attributions. For a summary of model results, see Table 8. Compared with a baseline intercept-only model (Model 1: AIC = 7,343.35), a model with a random intercept for vignette (Model 2: AIC = 7,286.55) explained more variance. The likelihood of the protagonist being judged as reasonable varied by vignette (pseudo R^2s = .00–.02); although, overall, participants were far more likely to respond that the protagonist was reasonable than unreasonable in all three vignettes. Collapsing across conditions, participants were more likely to judge Emma as unreasonable

Table 8. Reasonableness-Judgment Model Summaries

Parameter estimate or statistic	Model						
	1	2	3	4	5	6	6A
Fixed effects							
Intercept	2.56*** (0.03)	2.59*** (0.16)	2.59*** (0.16)	2.59*** (0.16)	1.81*** (0.24)	1.84*** (0.25)	2.43*** (0.22)
Age					−0.005 (0.00)	−0.005 (0.00)	−0.005 (0.00)
Gender					−0.18* (0.07)	−0.18* (0.07)	−0.1*8 (0.07)
Education					0.06*** (0.01)	0.06*** (0.01)	0.06*** (0.01)
Compensation					0.25** (0.07)	0.25*** (0.07)	0.25*** (0.07)
Condition: ignorance						−0.40*** (0.08)	−0.86*** (0.16)
Condition: knowledge						0.43*** (0.09)	0.39 (0.20)
Vignette: Emma							−1.10*** (0.16)
Vignette: Gerald							−0.52** (0.17)
Ignorance × Emma							0.74*** (0.20)
Ignorance × Gerald							0.42* (0.21)
Knowledge × Emma							0.26 (0.25)
Knowledge × Gerald							−0.24 (0.26)
Random effects							
Site				< 0.001	< 0.001	0.023	< 0.001
Participant			< 0.001	< 0.001	< 0.001	0.091	0.046
Vignette		0.274	0.274	0.274	0.278	0.277	< 0.001
Akaike information criterion	7,343.35	7,286.55	7,288.56	7,290.55	7,144.10	7,047.13	7,025.80

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increasing likelihood of reasonableness judgments. For condition, the comparison group was Gettier, and for vignette, the comparison group was Darrel. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*p < .05. **p < .01. ***p < .001.

than Gerald. Participants were more likely to judge Gerald as unreasonable than Darrel (see Table 9).

A model with a random intercept for vignette nested within participant (Model 3: AIC = 7,288.56) explained similar amounts of variance (pseudo R^2s = .00–.02) as Model 2. The model with a random intercept for vignette nested in participant nested in data-collection site (Model 4: AIC = 7,290.55) did not explain more variance (pseudo R^2s = .00–.02) than previous models. In Model 5, covariates were added as fixed effects (AIC = 7,144.10). Relative to Model 4, this model was more useful in explaining variance in judgments of reasonableness

(pseudo R^2s = .01–.04). Participant compensation, gender, and education were associated with reasonableness judgments. Participants who were compensated and female participants were more likely to judge the protagonist as reasonable than uncompensated and male participants. As participants' years of education increased, the likelihood that they would judge the protagonist as reasonable increased.

Finally, we estimated a model including knowledge condition as a fixed effect (Model 6: AIC = 7,047.13). This model performed better than Model 5 and revealed an effect of condition on reasonableness judgment

Table 9. Reasonableness Judgments From Model 2 Overall and by Vignette

	Overall	Darrel	Emma	Gerald
Unreasonable	1,021 (7.19%)	237 (5.01%)	447 (9.48%)	337 (7.10%)
Reasonable	13,173 (92.81%)	4,493 (94.99%)	4,269 (90.52%)	4,411 (92.90%)

Table 10. Reasonableness Judgments From Model 6 Overall and by Condition

	Overall	Knowledge	Ignorance	Gettier
Unreasonable	1,007 (7.21%)	217 (4.65%)	467 (10.02%)	323 (6.94%)
Reasonable	12,967 (92.79%)	4,447 (95.35%)	4,192 (89.98%)	4,328 (93.06%)

(pseudo R^2s = .01–.05). Participants were more likely to judge the protagonist in the knowledge-condition vignette as reasonable than the protagonists in the other two conditions (see Table 10). Protagonists in the ignorance-condition vignette were less likely to be judged as reasonable than protagonists in the knowledge- and Gettier-condition vignettes.

To test whether the effect of condition on reasonableness judgments varied by vignette, a model was estimated that included an interaction between vignette and condition (Model 6A: AIC = 7,025.80). This model explained more variance than the model without the interaction term. As shown in Figure 3, although the general pattern was the same for all vignettes, the magnitudes of the differences varied by vignette (pseudo R^2s = .02–.08).

The likelihood that participants judged the protagonist as reasonable varied by condition in response to the Darrel vignette, $\chi^2(2) = 781.00$, $p < .001$; Emma vignette, $\chi^2(2) = 36.36$, $p < .001$; and Gerald vignette, $\chi^2(2) = 21.10$, $p < .001$. Participants were more likely to judge Darrel to be reasonable in the Gettier-condition vignette ($\hat{p} = .96$) than in the ignorance-condition vignette ($\hat{p} = .91$; Cramér's $V = .10$, 95% CI = [.06, .13], $\chi^2[1] = 28.84$, $p < .001$), but we found no evidence that reasonableness judgments differed between participants responding to the Gettier and knowledge conditions of that vignette (Cramér's $V = .03$, 95% CI = [.02, .07], $\chi^2[1] = 3.44$, $p = .064$). The same pattern of results appeared in response to the Gerald vignette; participants were more likely to judge Gerald as reasonable when responding to the Gettier-condition vignette ($\hat{p} = .94$) as opposed to the ignorance-condition vignette ($\hat{p} = .91$; Cramér's $V = .06$, 95% CI = [.03, .09], $\chi^2[1] = 10.49$, $p = .001$), but the Gettier and knowledge vignettes produced similar rates of reasonableness judgments, ($\hat{p} = .94$; Cramér's $V = .02$, 95% CI = [.02, .05], $\chi^2[1] = 0.77$, $p = .381$).

The condition by vignette interaction in predicting judgments of reasonableness appears to have emerged because of the condition differences produced by the Emma vignette. Although participants were equally likely to judge Emma as reasonable in the Gettier and ignorance conditions (Cramér's $V = .02$, 95% CI = [.02, .06], $\chi^2[1] = 1.12$, $p = .291$), participants were more likely to judge Emma as reasonable in response to the knowledge-condition vignette ($\hat{p} = .94$) than in response to the Gettier-condition vignette ($\hat{p} = .89$; Cramér's $V = .09$,

95% CI = [.05, .12], $\chi^2[1] = 22.44$, $p < .001$). Thus, condition differences were found between the Gettier and ignorance versions of the Darrel and Gerald vignettes but not the Emma vignette and between the Gettier and knowledge versions of the Emma vignette but not the Darrel and Gerald vignettes.

Participant recruitment

Data were collected from MTurk workers and participants recruited from individual labs. Because the MTurk sample more likely represented the sample originally collected by Turri et al. (2015), we examined whether participant recruitment moderated the effect of condition on knowledge attributions and reasonableness judgments. Although Model 6B (AIC = 15,850.16) was superior to Model 6, the interaction term was not a significant predictor of knowledge attributions (Δ pseudo R^2s = .00–.01). Next, we estimated the same model (Model 6B) in predicting judgments of reasonableness (AIC = 7,017.37). Although this model performed better than Model 6, the interaction between condition and recruitment type was not significant (Δ pseudo R^2s = .00–.01). For summary of results, see Table 11.

Exploratory analyses

In addition to the hypotheses and research questions outlined in the approved protocol, we conducted additional exploratory analyses to examine three additional research questions and assess the influence of original measurement characteristics (binary vs. continuous).

"Direct" replication analysis. As previously explained, the design of our study substantially differed from that of Turri et al. (2015, Experiment 1). Rather than encountering one of three conditions of the "Darrel/Squirrel" vignette, our participants viewed three conditions matched with three vignettes in a within-participants design. Perhaps our observation of a Gettier-intuition effect, which was not found in the original experiment, can be explained by these methodological changes. To explore this possibility, we compared the knowledge-attribution rates of participants who viewed the Darrel vignette first ($n = 2,538$) in an analysis devised to closely approximate Turri et al.'s original test.[22] Overall, participants attributed knowledge at different rates according to condition, $\chi^2(2) = 252.57$,

Table 11. Participant Recruitment Moderation Model (6B) Summaries

Parameter or statistic	Measure	
	Knowledge	Reasonableness
Fixed effects		
Intercept	–0.07 (0.41)	2.07*** (0.25)
Age	< 0.001 (0.00)	–0.01*** (0.003)
Gender	–0.11** (0.04)	–0.23** (0.07)
Education	–0.02* (0.01)	0.06*** (0.01)
Compensation	–0.03 (0.05)	0.13 (0.08)
Condition: ignorance	–1.29*** (0.05)	–0.38*** (0.08)
Condition: knowledge	0.59*** (0.05)	0.44*** (0.09)
Participant source	0.32* (0.13)	1.39*** (0.37)
Source × Ignorance	–0.33 (0.19)	–0.66 (0.43)
Source × Knowledge	0.30 (0.17)	–0.44 (0.51)
Random effects		
Site	< 0.001	0.049
Participant	< 0.001	< 0.001
Vignette	0.670	0.278
Akaike information criterion	15,850.16	7,017.37

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increased likelihood of knowledge attributions or reasonableness judgments. Source was coded with lab participants as the comparison group. For condition, the comparison group was Gettier. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*$p < .05$. **$p < .01$. ***$p < .001$.

$p < .001$, Cramér's $V = .34$, 95% CI = [.30, .38], and the pattern of effects mirrored those of our primary analysis. Participants responding to the Gettier condition were more likely to attribute knowledge to Darrel ($\hat{p} = .59$) than those responding to the ignorance condition ($\hat{p} = .32$), $\chi^2(1) = 103.61$, $p < .001$, Cramér's $V = .26$, 95% CI = [.22, .32]. However, participants were less likely to attribute knowledge in response to the Gettier-condition vignette than to the knowledge-condition vignette ($\hat{p} = 0.72$), $\chi^2(1) = 30.48$, $p < .001$, Cramér's $V = .14$, 95% CI = [.10, .20]. Thus, this analysis provided further evidence for Gettier intuitions despite more closely approximating Turri et al.'s original test than our planned analysis. These effects were similar for the Gerald vignette when presented as the first vignette (i.e., same effect size and pattern) and the Emma vignette (i.e., same pattern and half the effect size).

Luck attributions. Attributions of luck were predicted in a series of multilevel logistic regressions models. These models were fit in the same fashion as the models focused on the two dependent variables with one notable difference: Observations in which the participant did not correctly answer the first part of our two-part luck-attribution measure were excluded. That is, the luck versus ability attributions that followed incorrect identification responses

were excluded from analyses ($n = 952$; 6.58%). For summary of Models 1 through 6A, see Table 12. Compared with the baseline intercept-only model (Model 1: AIC = 11,269.61), a model with a random intercept for vignette (Model 2: AIC = 10,613.78) explained more variance. The likelihood that outcomes were attributed to luck varied according to vignette (pseudo R^2s = .08–09). Although the Darrel vignette produced more attributions to ability than luck, the Emma vignette produced more attributions to luck than ability (see Table 13).

A model with a random intercept for vignette nested within participants (Model 3: AIC = 16,773.30) explained similar amounts of variance as the previous model (pseudo R^2s = .08–.09). Nesting within the data-collection site (Model 4: AIC = 16.775.30) did not improve the model fit (pseudo R^2s = .08–.09). Next, covariates were added to the model as fixed effects (Model 5: AIC = 16,489.60). Relative to Model 4, Model 5 explained more variance in luck attributions (pseudo R^2s = .08–.10). Years of education, age, and compensation independently predicted luck attributions (see Table 12).

Finally, we estimated a model including condition as a fixed effect (Model 6: AIC = 15,896.17). This model performed better than the previous models; the likelihood of luck attributions differed according to condition (pseudo R^2s = .05–.06). Participants were more likely to attribute the outcome to luck in the Gettier condition than in the other two conditions (see Table 14). In response to both the knowledge condition and the ignorance condition, participants were more likely to attribute outcomes to the protagonist's ability than to luck, but they were more likely to make luck attributions than ability attributions in response to the Gettier-condition vignette.

Vignette interactions. To better understand whether the effect of condition on luck attributions varied as a function of vignette, we estimated a model including an interaction between vignette and condition (Model 6A: AIC = 15,458.37). This model explained more variance (pseudo R^2s = .20–.23) than Model 6. As shown in Figure 3, each vignette demonstrated a different pattern of effects. Post hoc analyses suggested that the vignette by condition interaction was driven by responses to the Gettier condition. The difference in likelihoods of luck attributions between the Gettier and ignorance conditions was absent for the Darrel vignette (Cramér's $V = .02$, $p = .315$), moderate for the Gerald vignette (Cramér's $V = .20$, $p < .001$), and large for the Emma vignette (Cramér's $V = .50$, $p < .001$). The difference in luck attributions between the Gettier and knowledge conditions was largest in responses to the Emma vignette (Cramér's $V = .32$, $p < .001$) but of similar size in response to the Darrel vignette (Cramér's $V = .16$, $p < .001$) and Gerald vignette (Cramér's $V = .20$, $p < .001$).

Table 12. Luck/(In)Ability-Attribution Model Summaries

Parameter estimate or statistic	Model						
	1	2	3	4	5	6	6A
Fixed effects							
Intercept	–0.04* (0.02)	–0.04 (0.34)	–0.04 (0.34)	–0.04 (0.34)	0.40 (0.36)	–0.25 (0.37)	0.91*** (0.13)
Age					0.004 (0.00)	0.004* (0.00)	0.004* (0.00)
Gender					0.01 (0.04)	0.01 (0.04)	–0.004 (0.04)
Education					–0.03*** (0.01)	–0.03*** (0.01)	–0.03*** (0.01)
Compensation					–0.17*** (0.04)	–0.18* (0.04)	–0.20*** (0.04)
Condition: ignorance						1.03*** (0.05)	0.08 (0.08)
Condition: knowledge						0.94*** (0.05)	0.72*** (0.08)
Vignette: Emma							–2.85*** (0.11)
Vignette: Gerald							–1.07*** (0.08)
Ignorance × Emma							2.53*** (0.14)
Ignorance × Gerald							0.72*** (0.11)
Knowledge × Emma							1.07*** (0.14)
Knowledge × Gerald							0.10 (0.11)
Random effects							
Site				< 0.001	< 0.001	0.099	0.022
Participant			< 0.001	< 0.001	< 0.001	0.066	0.040
Vignette		0.581	0.581	0.581	0.585	0.611	< 0.001
Akaike information criterion	17,776.67	16,771.30	16,773.30	16,775.30	16,489.60	15,896.17	15,458.37

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increasing likelihood of ability attributions. For condition, the comparison group was Gettier, and for vignette, the comparison group was Darrel. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*p < .05. ***p < .001.

Luck/(in)ability as a moderator. Next, we explored whether attributions of outcomes to luck versus ability influence knowledge attributions, as suggested by prior research (Turri, 2016, 2017). Turri 2016 (Experiment 7) found a strong positive correlation between knowledge attributions and attributions to ability rather than luck ($r = .622$) and a moderating effect of luck attributions on Gettier intuitions; participants attributed knowledge less frequently when protagonists were perceived as having arrived at a truth because of a lucky guess rather than because of ability ($\eta_p^2 = .353$; Turri, 2016, Experiment 7).

We tested whether luck attributions moderated the effect of condition on knowledge attribution among participants who accurately identified that the protagonist was correct (in the Gettier and knowledge conditions) or incorrect (in the ignorance conditions) in their identification of the object as real or counterfeit. The main effect of luck attributions and the interaction between

Table 13. Luck (In)ability Attributions From Model 2 Overall and by Vignette

	Overall	Darrel	Emma	Gerald
Luck	6,551 (51.08%)	1,434 (33.36%)	2,910 (67.50%)	2,207 (52.34%)
(In)ability	6,275 (48.92%)	2,864 (66.64%)	1,401 (32.50%)	2,010 (47.66%)

Table 14. Luck/(In)Ability Attributions From Model 6 Overall and by Condition

	Overall	Knowledge	Ignorance	Gettier
Luck	6,451 (51.04%)	1,888 (44.53%)	1,784 (42.71%)	2,779 (65.81%)
(In)ability	6,189 (48.96%)	2,352 (55.47%)	2,393 (57.29%)	1,444 (34.19%)

condition and luck attributions were added to Model 6 of the knowledge-attributions analysis (Model 6C: AIC = 13,363.98). This model (pseudo R^2s =.24–.28) explained more variance in knowledge attributions than Model 6. For model summary, see Table 15.

Condition affected knowledge attributions when participants attributed the protagonists' (in)correct identification to bad or good luck, $\chi^2(2) = 211.03$, $p < .001$. Participants were more likely to attribute knowledge to the protagonist in the Gettier-condition vignette ($\hat{p} = .31$) than in the ignorance-condition vignette ($\hat{p} = .17$; Cramér's $V = .16$, 95% CI = [.13, .19], $\chi^2[1] = 116.03$, $p < .001$). They were also more likely to attribute knowledge in the knowledge-condition vignette ($\hat{p} = .38$) than in the Gettier-condition vignette ($\hat{p} = .31$; Cramér's $V = .07$, 95% CI = [.05, .10], $\chi^2[1] = 24.54$, $p < .001$).

Likewise, condition affected knowledge attributions when participants attributed the protagonists' (in)correct identification to (in)ability, $\chi^2(2) = 1,737.19$, $p < .001$. Participants in this group were more likely to attribute knowledge to the protagonist in the Gettier-condition vignette ($\hat{p} = .66$) than in the ignorance-condition vignette ($\hat{p} = .16$; Cramér's $V = .51$, 95% CI = [.48, .54], $\chi^2[1] = 972.07$, $p < .001$). These participants were also more likely to attribute knowledge in the knowledge-condition vignette ($\hat{p} = .73$) than in the Gettier-condition vignette ($\hat{p} = .66$; Cramér's $V = .08$, 95% CI = [.05, .11], $\chi^2[1] = 21.36$, $p < .001$). Although the knowledge-attribution difference between the Gettier and ignorance conditions was larger when participants made ability attributions (Cramér's $V = .51$) than when they made luck attributions (Cramér's $V = .16$), effect sizes were similar for the differences between the Gettier and knowledge conditions (Cramér's Vs = .08 vs. .07). Thus, unlike in previous research (Turri, 2016, 2017), luck attributions did not decrease the likelihood of participants demonstrating Gettier intuitions.

Alternative-knowledge probe. We also assessed whether question wording affected participants' knowledge attributions, as has been suggested by previous research (e.g., Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013). Participants may be more likely to deny knowledge to a protagonist when they are asked a more nuanced question (whether protagonists knew or only felt like they knew but did not actually know; Nagel, Juan, & Mar, 2013) than when they are asked a simpler question (whether protagonists knew or did not know).

In our exploratory analyses of the alternative-knowledge probe (i.e., following Model Steps 1 through 6), we found a pattern of results similar to those for the analyses of our primary knowledge measure (Model 6: AIC = 16,332.68; pseudo R^2s = .16–.21). For model summary, see Table 16. Participants were more likely to choose the knowledge option in response to the Gettier condition than in response to the ignorance condition. The likelihood of choosing knowledge was also higher in response to the knowledge condition than in response to the Gettier and ignorance conditions. Thus, participants demonstrated Gettier intuitions as measured by the alternative-knowledge probe as well.

Measurement characteristics. We examined whether condition effects were influenced by measurement characteristics, specifically, if the outcome was originally measured on a binary scale or VAS. For model summaries, see Table 17. Adding measurement and its interaction with condition to the model predicting knowledge attribution

Table 15. Luck/(In)Ability Attribution Moderation Model (6C) Summary

Parameter or statistic	Knowledge
Fixed effects	
Intercept	–0.48 (0.36)
Age	0.003 (0.00)
Gender	–0.13** (0.05)
Education	–0.02* (0.01)
Compensation	0.02 (0.05)
Condition: ignorance	–1.00*** (0.08)
Condition: knowledge	0.35*** (0.66)
Luck/ability	1.03*** (0.07)
Ignorance × Luck/Ability	–1.13*** (0.11)
Knowledge × Luck/Ability	0.20 (0.10)
Random Effects	
Site	< 0.001
Participant	< 0.001
Vignette	0.574
Akaike information criterion	13,363.98

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increased likelihood of knowledge attributions. Luck/(in)ability was coded so that 0 indicated luck and 1 indicated (in)ability. For condition, the comparison group was Gettier. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 16. Alternative-Knowledge Probe Model 6 Summary

	Measure
Parameter or statistic	Knowledge probe
Fixed effects	
Intercept	–0.67 (0.38)
Age	0.01** (0.00)
Gender	–0.09* (0.04)
Education	–0.01 (0.01)
Compensation	0.22*** (0.04)
Condition: ignorance	–1.18*** (0.05)
Condition: knowledge	0.41*** (0.04)
Random effects	
Site	< 0.001
Participant	< 0.001
Vignette	0.628
Akaike information criterion	16,332.68

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest an increased likelihood of choosing knowledge. For the condition variable, Gettier was the comparison group. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*$p < .05$. **$p < .01$. ***$p < .001$.

did not produce moderation effects or improve model fit (Model 6D: AIC = 15,876.57; pseudo R^2s = .21–.25). Next, we estimated the same model (Model 6D) in predicting judgments of reasonableness (AIC = 7,041.29). Although

this model (pseudo R^2s = .02–.07) performed better than Model 6, the interactions between condition and measurement type were not significant. Finally, we estimated a model that included an interaction between condition and measurement type predicting luck attributions (Model 6D: AIC = 15,862.09). This model (pseudo R^2s = .14–.16) performed better than Model 6 and revealed an interaction effect for the ignorance condition compared with the Gettier condition.

Condition affected the likelihood of luck attributions on responses to the binary measure, $\chi^2(2) = 120.98$, $p < .001$. Participants were more likely to attribute outcomes to luck in the Gettier condition ($\hat{p} = .68$) than in the ignorance condition ($\hat{p} = .37$; Cramér's $V = .31$, 95% CI = [.26, .37], $\chi^2[1] = 118.14$, $p < .001$). Participants were also more likely to attribute outcomes to luck in the Gettier condition ($\hat{p} = .55$) than in the knowledge condition ($\hat{p} = .37$; Cramér's $V = .18$, 95% CI = [.13, .24], $\chi^2[1] = 41.00$, $p < .001$).

Condition similarly affected luck attributions as measured by the VAS, $\chi^2(2) = 454.78$, $p < .001$. Participants were more likely to attribute outcomes to luck in the Gettier condition ($\hat{p} = .66$) than in the ignorance condition ($\hat{p} = .45$; Cramér's $V = .22$, 95% CI = [.20, .24], $\chi^2[1] = 341.27$, $p < .001$). Participants were also more likely to attribute outcomes to luck in the Gettier condition ($\hat{p} = .66$) than in the knowledge condition ($\hat{p} = .45$; Cramér's $V = .22$, 95% CI = [.20, .24], $\chi^2[1] = 345.90$, $p < .001$). The

Table 17. Measurement Moderation Model (6D) Summaries

	Measure		
Parameter or statistic	Knowledge	Reasonableness	Luck/(In)Ability
Fixed Effects			
Intercept	–0.23 (0.41)	1.57*** (0.28)	–0.15 (0.38)
Age	0.00 (0.00)	–0.01 (0.00)	0.00* (0.00)
Gender	–0.08* (0.04)	–0.18* (0.07)	0.01 (0.04)
Education	–0.02* (0.01)	0.06*** (0.01)	–0.03*** (0.01)
Compensation	0.02 (0.04)	0.25*** (0.07)	–0.18*** (0.04)
Condition: ignorance	–1.27*** (0.14)	–0.34 (0.18)	1.42*** (0.13)
Condition: knowledge	0.59*** (0.12)	0.48* (0.22)	0.82*** (0.12)
Measurement: visual analogue scale	0.06 (0.09)	0.35* (0.15)	–0.13 (0.10)
Measurement × Ignorance	–0.06 (0.15)	–0.07 (0.20)	–0.45*** (0.14)
Measurement × Knowledge	0.02 (0.13)	–0.06 (0.24)	0.14 (0.13)
Random effects			
Site	< 0.001	0.086	0.072
Participant	< 0.001	0.070	0.010
Vignette	0.669	0.277	0.613
Akaike information criterion	15,876.57	7,041.30	15,862.10

Note: Estimates and their standard errors, in parentheses where applicable, are provided for each variable in the model. Positive values suggest increased likelihood of knowledge attributions, reasonableness judgments, or attributions to (in)ability. For the condition variable, Gettier was the comparison group. For the measurement variable, binary was the comparison group. For full model statistics, see the analysis folder at https://osf.io/8ut6e/.
*$p < .05$. ***$p < .001$.

effect size of the difference between the Gettier and knowledge conditions was smaller when attributions to luck were measured continuously, but the CIs of the continuous-measure effect sizes overlapped with those produced by the binary measure.

Gettier scores. Finally, at the request of a reviewer, we compared the rates of knowledge attribution across the Gettier and knowledge conditions by examining so-called Gettier scores. Starmans and Friedman (2020) devised this approach to account for baseline skepticism in comparing differences in knowledge attribution according to condition across subsamples. Gettier scores are calculated by dividing the percentage of participants who attribute knowledge in the Gettier condition by the percentage of participants who attribute knowledge in the knowledge condition. Using the values from Model 6 (see Table 7), we computed a Gettier score of 75.47, which suggests that participants, on average, attributed knowledge in response to the Gettier condition 75.47% as often as they did in response to the knowledge condition. Considering just the Darryl vignette data for participants who responded to it first (i.e., the "direct" replication) yielded a Gettier score of 80.98. These scores highlight the somewhat similar rates of knowledge attribution across the two conditions.

Discussion

Past cross-cultural research has suggested that nonphilosophers may rely on a shared epistemic intuition (i.e., a core folk epistemology) that leads them to deny knowledge to protagonists in Gettier-type cases more often than to protagonists in cases of JTB, thereby demonstrating Gettier intuitions (e.g., Machery, Stich, Rose, Chatterjee, et al., 2017). In the present research, we examined the prevalence of Gettier intuitions in counterfeit-object Gettier-type cases by replicating and extending Experiment 1 of Turri et al. (2015). Our international multisite study employed three counterfeit-object Gettier vignettes to compare how participants attribute knowledge to protagonists in Gettier, knowledge, and ignorance vignette conditions. Overall, we observed a small Gettier-intuition effect. Participants were more likely to attribute knowledge to protagonists in standard cases of JTB (i.e., the knowledge conditions) than in special cases of JTB in which protagonists formed a true belief based on a true observation of an authentic object despite the presence of a salient but failed threat to their ability to detect its authenticity (i.e., the Gettier conditions). This result did not correspond to that found by Turri et al., who failed to detect a significant difference in knowledge attribution between these two conditions. Note that the size of the Gettier-intuition effect varied by vignette in our research; the Darrel vignette from the original study produced the smallest effect size

and was similar to that we calculated using the nonsignificant result from Turri et al. Therefore, we did find effect sizes in the same range as Turri et al. when directly comparing like conditions; however, the null result did not replicate. The Emma vignette produced the largest effect size of the conditions; yet few participants attributed knowledge to Emma regardless of epistemological condition. Our results align with research that suggests that participant perceptions of the protagonist contribute to differences in knowledge-attribution rates in Gettier-intuition research (e.g., Disher et al., 2021).

Knowledge attribution

Our results did not correspond to those found by Turri et al. (2015) in a potentially substantive way. In the original study, participants who read the Gettier version of the "Darrel/Squirrel" vignette attributed knowledge to the protagonist at higher rates than participants who read the ignorance version. However, the rates of knowledge attribution did not differ between participants in the Gettier and knowledge control conditions. Although we similarly found a large difference between the Gettier and the ignorance conditions in our replication, our analyses also revealed a difference in rates of knowledge attribution between the Gettier and knowledge conditions (i.e., the Gettier-intuition effect). This discrepancy could be explained by the low power of the original study (i.e., $N = 135$ in a between-participants design with three levels). Indeed, the original authors suspected that their experiment may have failed to demonstrate a difference between these two conditions because of insufficient power (personal communication, J. Turri, March 10, 2018). To further examine this possibility, we estimated an effect size for their original analysis for comparison purposes. Although nonsignificant, the original effect (odds ratio [*OR*] = 2.00, 95% CI = [0.77, 5.21]) was similar in magnitude to the one we found in our analyses (*OR* = 1.86, 95% CI = [1.78, 1.94]). Thus, although we did not replicate Turri et al.'s null result, they potentially could have also found a significant effect with a sufficiently powered experiment.

Despite this similarity in effect sizes, we argue that our findings do contradict Turri et al.'s (2015) conclusion that "a salient but failed threat to the truth of a judgment does not significantly affect whether it is viewed as knowledge" (p. 381). Given our evidence that participants demonstrated Gettier intuitions for two other similar counterfeit-object Gettier-type cases, which also featured failed threats to the truth of a judgment, we challenge their claim that knowledge attributions are insensitive to such threats. In this way, our results best align with those of other researchers who have found similar effects and concluded that protagonists with luckily true beliefs are less likely to elicit knowledge

attributions than protagonists in clear cases of knowledge (Colaço et al., 2014; Machery, Stich, Rose, Chatterjee, et al., 2017; Nagel, Juan, & Mar, 2013). Still, the small size of the effect suggests that Gettier intuitions were not prevalent in our research.

Changes in the methods, design, and analytic approach may also account for differences between our results and those of Turri et al. (2015). One major difference between our replication and the original study was the inclusion of two additional vignettes as part of a within-participants design. The inclusion of these unique stimuli and design features changed the context of the task and may explain some results discrepancies. Unlike in the original study, which had a between-participants design, participants in our study responded to all three conditions randomly matched to each vignette in a single experimental session; therefore, participants' responses to a vignette condition may have anchored or led to contrast effects on responses to subsequent vignette conditions. However, participants in the present research were more likely to attribute knowledge to protagonists in the knowledge control condition than in the Gettier condition across all three vignettes, including the one used by Turri et al. In fact, our exploratory analysis of the Darrel vignette that closely approximated Turri et al.'s original analysis found evidence for Gettier intuitions among participants who responded to that vignette first. Furthermore, and likely because participants were presented with the vignette-condition combinations in a random order, contextual order effects were minimal, and order did not interact with condition in predicting outcomes (see https://osf.io/uz8te).

Prior research on epistemic intuitions has demonstrated the presence of Gettier intuitions among nonphilosophers (e.g., Colaço et al., 2014; Nagel, Juan, & Mar, 2013) and across cultures and geographic regions (e.g., Machery, Stich, Rose, Alai, et al., 2017; Machery, Stich, Rose, Chatterjee, et al., 2017). Specifically, the limited research using counterfeit-object Gettier-type cases has found that people are generally less likely to attribute knowledge to a protagonist when the protagonist's true and justified belief is formed on the basis of misleading evidence than in a parallel case when the true and justified belief is formed on the basis of clear evidence (e.g., Nagel, Juan, & Mar, 2013; Weinberg et al., 2001).

In the present research, participants likewise demonstrated Gettier intuitions in these cases across different geographic regions. Small Gettier-intuition effects were detected on a variety of measures, and knowledge attributions were only minimally (but not meaningfully) affected by participant characteristics such as gender, age, and years of education. Although prior research has suggested that differences in knowledge attribution may depend on how participants are asked whether a target has knowledge (e.g., Machery, Stich, Rose, Chatterjee,

et al., 2017; Nagel, Juan, & Mar, 2013), we found the same pattern of results on the continuous measure, the original binary measure (knows vs. only believes), and the alternative-knowledge attribution measure (knows vs. feels like they know but does not know). Thus, the present research supports the view that at least some nonphilosophers generally demonstrate Gettier intuitions and may to some extent rely on a shared core folk epistemology (i.e., intuitions about knowledge) when assessing the knowledge of others. However, our findings using counterfeit-object Gettier-type cases may not generalize broadly to other categories of Gettier-type cases (e.g., reliabilist, apparent evidence), which may elicit different epistemic intuitions. Furthermore, a notable number of participants (43.41%; see Table 7) denied knowledge to protagonists even in clear cases of JTB; thus, this supposed "core folk epistemology" is not universally shared. After accounting for such baseline skepticism, participants, on average, attributed knowledge in response to the Gettier condition 75.47% as often as they did in response to the knowledge condition. Although Gettier protagonists were deemed ignorant more often than not, Gettier intuitions were by no means common. Finally, given the small size of the observed effect, the theoretical significance of this result is debatable.

Ancillary findings

Reasonableness judgments. According to the JTB account of knowledge, protagonists must be perceived as having met all three criteria (i.e., justification, truth, and belief) to be attributed knowledge (Jacquette, 1996). To test whether Gettier-type challenges to standard JTBs produce different rates of knowledge attribution in counterfeit-object cases, we evaluated whether conditions were perceived as having met the appropriate criteria for the JTB analysis of knowledge. In the present research, the vignette comprehension questions served as the belief criteria by ensuring that participants could report that protagonists held the relevant belief. The truth of the protagonists' belief varied by condition (i.e., only the protagonist in the ignorance condition held a false belief). The reasonableness-judgment measure assessed whether participants judged the protagonists' beliefs to be justified (i.e., reasonable). In the original study, Turri et al. (2015) found no difference between the three epistemological conditions in participants' reasonableness judgments (i.e., how reasonable the participant rated the protagonist for holding a given belief). The authors interpreted this null result as evidence that differences in knowledge attribution could not be explained by differences in judgments of the protagonists' reasonableness by condition.

In the present research, condition did minimally affect whether participants judged protagonists as reasonable. Participants were more likely to judge protagonists in

the Gettier conditions as reasonable than protagonists in the ignorance conditions. They were also more likely to judge protagonists as reasonable in the knowledge condition than in the Gettier condition. Although we did detect small differences in judgments of reasonableness between conditions, the vast majority of participants responded that protagonists were reasonable in all conditions. Thus, participants generally perceived the protagonists as being justified in their belief regardless of vignette or condition. Furthermore, the high statistical power of our study allowed us to detect very small effects of condition on reasonableness judgments. Such small differences were unlikely to have had much impact on knowledge attributions; however, we did not directly examine this causal pathway.

Luck attributions. Prior research suggests that attributions of true beliefs to luck may moderate the extent to which Gettier intuitions are demonstrated; when Gettier protagonists are perceived as lucky (as opposed to able), the likelihood they are denied knowledge appears to increase (Turri, 2016, 2017). In the present research, participants attributed outcomes to luck more frequently in the Gettier condition than in the other two conditions. As expected, we found a negative relationship between the likelihood of luck attributions and the likelihood of knowledge attributions. However, we failed to find evidence that the magnitude of the Gettier-intuition effect was moderated by luck attributions. Although results suggested a moderating effect of attributions to luck or (in)ability on the difference between the ignorance and Gettier conditions, the difference in knowledge attributions between the Gettier and knowledge conditions did not differ according to whether participants attributed truth outcomes to luck or to ability. Seemingly, the likelihood of Gettier intuitions did not depend on participants attributing the protagonist's true belief to luck. However, the stark differences in luck attributions between vignettes may have dampened moderation effects that could have been found if we had examined a single scenario.

Differences between vignettes

In prior research, Gettier intuitions have been investigated using a variety of different Gettier-type cases and methodologies. Across the types of Gettier-type cases (e.g., "replacement by backup," "counterfeit object," "authentic evidence," "apparent evidence"), research results often contradict one another. Previous research suggests that heterogeneous findings can sometimes be explained by methodological features of the research, such as the stimuli used (e.g., Kenny & Judd, 2019; Landy et al., 2020). In line with this view, in the present research, we found that vignette moderated the effect of condition on all considered dependent measures to varying degrees.

Despite possessing the same epistemological structure, the three tested vignettes produced different rates of knowledge attribution both overall and according to condition (see Fig. 3). Participants attributed considerably less knowledge to Emma in the fake-diamond vignette than to the protagonists in the other two vignettes. These findings align with prior research that provided evidence for the prevalence of Gettier intuitions using the "Emma/Diamond" vignette (Disher et al., 2021; Nagel, Juan, & Mar, 2013; Powell et al., 2015). For example, although Powell et al. (2015) found different rates of knowledge attribution among participants in the ignorance, Gettier, and knowledge conditions, few participants attributed knowledge to Emma overall (e.g., just 25% of participants in the knowledge condition). However, in Experiment 4 reported by Turri et al. (2015), participants in the knowledge condition of a similar "Emma/Diamond" vignette attributed knowledge at a similar rate (90%) to participants in the Gettier condition involving a failed threat (83%). The epistemological structure of the Gettier condition in Turri et al. (Experiment 4) differed from that employed in the present research. Thus, the strength of Gettier intuitions we observed for the Emma vignette appears to cohere with prior research.

Knowledge attributions for the Gerald vignette were overall more split compared with the other two vignettes. However, making comparisons with past empirical research that used the "Gerald/House" vignette is difficult given that prior studies that have used it relied on very different methodology and study materials (Colaço et al., 2014; Disher et al., 2021; Swain et al., 2008; Ziółkowski, 2016). Some researchers have found differences in knowledge attributions between Gettier conditions and knowledge conditions for this vignette, albeit using different methodologies (Colaço et al., 2014; Disher et al., 2021; Ziółkowski, 2016). Thus, in line with our findings, most research using the Gerald vignette has found evidence for Gettier intuitions. Besides the original study (Turri et al., 2015) and the present replication, only one other study (Disher et al., 2021) has employed the "Darrel/Squirrel" vignette to our knowledge; Disher et al. (2021) did not find evidence for Gettier intuitions in response to this vignette, although they used a different name.

One reason why our vignettes elicited different rates of knowledge attribution may relate to perceptions of luck; vignette moderated the effect of condition on both knowledge and luck attributions. For luck attributions, differences between the Gettier and ignorance conditions were considerably smaller for the Darrel and Gerald vignettes (Cramér's $V = .02$ and Cramér's $V = .20$, respectively) than for the Emma vignette (Cramér's $V = .50$), and the luck attribution differences between the Gettier and knowledge conditions were also smaller for the

Darrel and Gerald vignettes (Cramér's *V* = .17 and Cramér's *V* = .20, respectively) than for the Emma vignette (Cramér's *V* = .32). Overall, Emma's outcomes were attributed most often to luck, and Darrel's outcomes were attributed most often to ability. Thus, the reason why vignettes differed in their overall level of both knowledge and luck attributions may relate to the perceived characteristics of the target protagonist or the protagonist's situation. In further support of this view, a separate extension of the present research manipulated the gender of the target protagonist and found that a female protagonist's knowledge outcome was more likely to be attributed to luck (as opposed to ability) than that of a male protagonist's across all conditions and vignettes (Disher et al., 2021). Thus, the gender of the protagonist may have potentially served as a cue that participants used to assess the ability of a protagonist when deciding whether the protagonist possessed knowledge.

However, in the present research, differences in the results produced by the Emma vignette compared with the other vignettes cannot easily be attributed to protagonist gender alone. Other factors unique to the Emma vignette may also partially explain the differences in response rates across vignettes. For instance, the Emma vignette introduced skeptical pressure in ways the other two vignettes did not. Specifically, participants in all conditions read that Emma "could not tell the difference between a real diamond and a cubic zirconium fake," suggesting a lack of expertise and subsequent knowledge. In an extension carried out by collaborators (Larkin & Andreychik, 2019; see also Appendix E in the Supplemental Material), an additional vignette that manipulated the perceived expertise level of protagonists (i.e., expert or novice) and the condition (i.e., knowledge, Gettier, or ignorance) was tested as part of our data collection in a fully between-participants design. Their results demonstrated that the perceived expertise of protagonists affected knowledge-attribution rates; protagonists with high expertise were more likely to be attributed knowledge than protagnoists with low expertise. Given that the Darrel vignette features a protagonist that is described as being an ecologist (i.e., an expert) and the Emma vignette features a protagonist that is described as not able to evaluate whether a diamond is authentic (i.e., not an expert), differences in attribution rates between these vignettes may be due to their perceived level of expertise.

Finally, the Emma vignette also featured a scenario with which most participants were more likely to have personal experience (i.e., shopping). In contrast, the Darrel vignette featured a scenario with which most participants were less likely to have personal experience (i.e., ecological research). Nagel, Juan, and Mar (2013) argued that epistemic egocentrism, or the tendency of people to evaluate others as though the others know what the people know (Birch, 2005; Birch & Bloom, 2007; Camerer et al., 1989; Nickerson, 1999), may play a substantial role in how participants evaluate the knowledge of others. If participants have differing levels of preexisting knowledge about vignette scenarios, they may be differently equipped to evaluate protagonists in each scenario based on their assumed shared knowledge. Perhaps, participant familiarity with the context of shopping in the Emma vignette allowed participants to consider ways in which she could have better evaluated her belief. Participant familiarity with the context of ecological research was likely comparatively low; they may not have generated alternative approaches for Darrel to evaluate his belief. Because we did not manipulate these features of the tested vignettes, such interpretations remain speculative. Parsing out the effects of these different sources of stimulus variation would be a valuable aim for future research.

Implications

Previous research on epistemic intuitions has primarily focused on whether laypeople deny knowledge to targets in philosophical problems based on the epistemological structure of the problem. Secondarily, research has investigated whether lay denials of knowledge in these sorts of problems differ based on the identity of the rater/participant (e.g., participants' gender, class, or culture). Our results demonstrate that epistemological structure and participant identity alone cannot fully account for the rate at which people deny or attribute knowledge. Even standard cases of JTB were attributed knowledge at different rates between these vignettes.

In the present research, all scenarios represented the same type of Gettier-type case (i.e., counterfeit-object cases) and thus featured the same epistemological structure. If people's epistemic intuitions rely only on all of the same epistemological criteria (e.g., justification, truth, and belief), then they should have denied knowledge similarly across these scenarios as a function of whether those criteria were met. Instead, our results suggest that people attribute knowledge in ways that deviate from these theoretical expectations. Specifically, characteristics of the protagonists and situations presented in the vignettes seem to moderate attributions of knowledge.

Although participants' knowledge attributions may have been sensitive to the nuances of the tested vignettes, the way in which participants attributed knowledge was fairly straightforward. Most participants attributed knowledge on a continuous VAS that allowed for but did not reveal considerable variability in the degree of knowledge attributed to the protagonist. Instead, participants responded in a clearly binary manner, as revealed by

the bimodal distribution of the knowledge variable: Protagonists were generally perceived as either having knowledge or not. These findings in and of themselves demonstrate that people make judgments about knowledge in a very dichotomous manner.

Pedagogical considerations

As a partnership between the PSA and CREP, this project had a central goal of serving a pedagogical function with support through the PSA's network and resources. Experiment 1 from Turri et al. (2015) was selected by the CREP team as a study that was feasible for students to directly replicate; the original study had relatively simple materials (i.e., three variations of one "Darrel" vignette), measurements (i.e., dichotomous "knows/believes" judgments), and analyses (i.e., chi-square goodness-of-fit tests). In the process of submitting and revising a Registered Report for the study, the materials, measurements, and analyses all became more complex and, importantly, more useful for the underlying empirical questions than the original. However, we observed some trade-offs between rigor and pedagogy because of this increase in complexity.

In a typical CREP project, students prepare their materials and OSF pages, submit their pages for initial review, collect data, clean and analyze data, interpret their results, and submit their pages for final review. The increase in design complexity resulted in the need for centralized data collection to guarantee adherence to the randomization and counterbalancing procedures. Instead, students worked with the project administration team to incorporate their own information (e.g., informed consent, compensation) into the centralized survey where needed. The increased analytic complexity meant that students (and instructors) faced challenges in completing their site-level analyses. The majority of undergraduate- and master's-level students have likely not been trained in mixed analysis of variance or multilevel modeling.

For this project, students generally did not prepare their own materials or analyze their own site's data (see the Site Level Analysis column in Table 1). However, most or all students completed many traditional CREP steps: creating OSF accounts and following instructors to create study pages for their site; recording videos of the study procedures; posting all materials, including ethics approval; requesting reviews; and revising projects as necessary. All teams with data included in the present study completed at least these minimum requirements; some teams did more than the minimum required, including the evaluation of extension hypotheses. In large part, however, teams just completed the minimum requirements.

In general, we believe that student contributors may have received less training by participating in this project than they would have during a typical CREP project. We have planned a follow-up survey to assess self-reported learning among student collaborators. Although we can compare the results of that survey to similar surveys following other CREP projects, we cannot determine whether participation in the project would have produced different learning outcomes for students had it been implemented as originally planned.

The trade-offs between the scientific and pedagogical aims of this study had other consequences. Our attempt to provide flexibility for teams resulted in data loss and energy-, time-, and resource-draining data-processing procedures. For instance, some contributors requested the ability to prepare their own project materials via Qualtrics, and in consultation with the Registered Replication Report editor, we decided to support the pedagogical goals of those researchers. This effort to allow for experiential learning while adhering to the approved methods and analysis plan led to complications. Data from some of the teams who administered a Qualtrics survey proved unusable because of lack of adequate documentation.

If this had been a purely PSA study, then students would presumably have had fewer opportunities to participate in educational activities, such as using the OSF or communicating with reviewers before data collection. Students also would have had less flexibility in data-collection methods and extension variables. On the other hand, data processing and documentation would have been much easier. If we were interested only in addressing the empirical questions of this research or if we were interested only in training students how to do replications or research, our approach would not have been appropriate. We exchanged time, resources, and energy for the opportunity to satisfy both empirical and pedagogical goals. Creative strategies, such as requiring students to prepare materials on their own before being given access to the centralized data-collection link, may satisfy the needs of both pedagogy and rigor in future large-scale collaborations.

Despite these trade-offs, we would recommend doing big-team science with student researchers in the future. Likely, some of our challenges may have been less pronounced without a Registered Report process that placed a priority on the empirical question and resulted in a complicated design. At the very least, the students who collaborated as researchers on this project learned about preregistration, Registered Reports, and the OSF. General research literacy can be improved by learning about these practices, and for those students who will continue to do research in graduate school or as part of their profession, incorporating these practices into

their tool kit at an early stage may improve the rigor and transparency of their future contributions (Pownall et al., 2023).

Limitations

Although the present research represents the largest multisite empirical study of Gettier intuitions to date and was conducted across multiple geographic regions using multiple minimally matched stimuli, our conclusions are limited by (a) inconsistencies in data documentation and collection, (b) methodological decisions, (c) strict a priori exclusion criteria, and (d) generalizability.

Given the pedagogical goals of this project, trade-offs between research quality and accessibility to students were made at various stages of the project that led to inconsistencies in data documentation and collection. Exceptions to the accepted protocol were granted for several student teams (e.g., some teams implemented the study independently in Qualtrics rather than using the vetted SoSciSurvey survey). Thus, some of the samples collected as part of this project were excluded because of data-quality concerns. However, despite losses in data because of these exclusions, permitting flexibility in data collection allowed for more students to experience being part of a large multisite research project that enriched their research education.

Methodological complications further limit our results. The original experiment used binary-response options for the dependent measures; as planned, we implemented VASs instead. This difference may have affected the results that we found before and after converting those continuous responses to a binary format. Exploratory analyses suggested that a binary knowledge measure, a randomly assigned alternative implemented by some teams, did not produce meaningfully different results from those we obtained using the dichotomized continuous knowledge measure. Furthermore, using the untransformed continuous measures in analyses produced a similar pattern of results as those we reported (see https://osf.io/nvfbm/). Still, our findings may have been different if all participants were asked to respond to the knowledge question in a response format that better reflects the binary way in which people appear to make these kinds of determinations. In addition, the exploratory luck versus ability measure was originally planned to be a single question that required two responses. We changed how the question was displayed to alleviate participant confusion, but this deviation may have affected responding. Finally, we were unable to include two of the planned test-setting covariates (i.e., online vs. in person and individually vs. in a group) in our analyses because of unforeseen challenges in data collection (e.g., changing modalities because of the COVID-19 pandemic). The omission of these variables may have affected our results.

The large number of participant exclusions is another potential limitation of this research. According to our strict a priori exclusion criteria, many participants were excluded because they responded incorrectly to at least one of the vignette comprehension questions (46.36% of participants met this criteria), had missing or invalid data for age (22.44% of participants met this criteria), and/or did not respond to the language-proficiency question or reported low proficiency (22.17% of participants met this criteria). These three exclusion criteria resulted in nearly half of participants being excluded from analyses. Failed vignette-comprehension checks accounted for most of the exclusions, likely because of inattention or the intellectually challenging content. However, the direct replication analysis using data from only the Turri et al. (2015) squirrel vignette that excluded only participants who got the corresponding comprehension question wrong closely mirrored our primary findings. Additional exploratory analyses excluding participants who failed a specific comprehension question, rather than employing listwise exclusions, demonstrated a similar pattern of results (see https://osf.io/nvfbm/). Furthermore, although nearly half the participants were excluded, potentially limiting the generalizability of our results, our strict criteria arguably increased the validity of our findings by including only participants who understood the scenarios.

Comprehension exclusion rates have varied widely in previous Gettier-intuition investigations (e.g., 2%–47%; Machery, Stich, Rose, Alai, et al., 2017; Starmans & Friedman, 2012), but those studies used between-participants designs in which participants responded to a single scenario. Our relatively high rates of comprehension exclusions (i.e., 46%) may have resulted from our listwise exclusion of participants if they responded incorrectly to any one of the three vignettes' comprehension questions. However, other cross-cultural studies in this domain have produced similar comprehension exclusion rates with between-participants designs (e.g., 47%, Machery, Stich, Rose, Alai, et al., 2017). Perhaps cultural variation in conceptual familiarity or linguistic forms reduced comprehension or memory of the tested vignettes (see Machery, Stich, Rose, Alai, et al., 2017). Regardless, according to a review of Gettier-intuition studies (Popiel, 2016), participant exclusions typically have no effect on study results. Still, we cannot easily draw conclusions about laypeople's epistemic intuitions given their difficulty engaging with our scenarios. This potential limitation may broadly apply to the field of experimental philosophy. Often, experimental-philosophy research introduces participants to highly abstract and intricate scenarios with underlying assumptions that

laypeople struggle to understand or do not accept (e.g., Bergenholtz et al., 2021; Murray et al., 2022).

We chose to not execute an additional planned exclusion, which would have removed participants from sites for which teams did not receive a CREP completion certificate. As discussed in the Method section, we decided not to exclude data from teams that were approved for data collection and used the centralized survey even if they did not receive certificates. Requiring completion of the remaining pedagogical tasks would have further reduced our sample size without meaningfully increasing quality assurance. Furthermore, as previously explained, implementing this additional exclusion criteria did not substantively affect results (see https://osf.io/nvfbm).

Finally, because most of our participants were drawn from university samples, our findings may not generalize beyond the small subset of educated, socioeconomically advantaged young adults—at least participants able to pass comprehension checks (for evidence regarding socioeconomic differences, see Nichols et al., 2003; for educational differences, see Starmans & Friedman, 2012; for age differences, see Colaço et al., 2014). However, our results indicated that age and years of education had only very small associations with knowledge-attribution rates that were not robust to changing model specifications. In addition, given that our sample of Gettier-type cases from the epistemology literature was limited to specific types of counterfeit-object scenarios, inferences made from our findings should be applied only to intuitions in that subset of Gettier-type cases. Other forms of Gettier-type cases (e.g., evidence-replacement cases) may produce different epistemic intuitions. For example, prior literature has demonstrated that participants are less likely to attribute knowledge to protagonists in Gettier-type cases that present "apparent" evidence (e.g., Turri, 2013) and more likely to attribute knowledge in cases that present "authentic" evidence (e.g., Starmans & Friedman, 2013). We have no reason to believe that the results presented in this article were dependent on other characteristics of the participants, materials, or context (Simons et al., 2017).

Conclusion

Turri et al. (2015) interpreted their original findings as supporting the view that a salient but failed threat to the truth of a judgment does not affect whether it is viewed as knowledge. The results from this Registered Replication Report suggest that this view should be amended. Contrary to Turri et al.'s claim, our participants attributed knowledge significantly more often to protagonists in standard JTB cases than in counterfeit-object Gettier-type cases. However, the overall effect was small, and we did observe a smaller Gettier-intuition effect in the

vignette used in the original study than in the other vignettes we employed. Overall, our results suggest that attributions of knowledge may be affected by contextual characteristics unrelated to the knowledge criteria met by protagonists, such as perceptions about protagonists' ability and expertise. Future research on epistemic intuitions should focus on identifying the moderating role of contextual characteristics to better understand the conditions necessary for people to attribute knowledge to others.

Mark J. Brandt: Conceptualization; Formal analysis; Methodology; Supervision; Writing – review & editing.

Jon E. Grahe: Conceptualization; Investigation; Methodology; Supervision; Writing – review & editing.

Asil A. Özdoğru: Investigation; Methodology; Resources; Supervision; Writing – review & editing.

Michael R. Andreychik: Investigation; Methodology; Supervision; Writing – review & editing.

Sau-Chin Chen: Investigation; Methodology; Supervision; Writing – review & editing.

Thomas R. Evans: Investigation; Resources; Supervision; Writing – review & editing.

Caro Hautekiet: Data curation; Investigation; Resources; Writing – review & editing.

Hans IJzerman: Conceptualization; Methodology; Supervision; Writing – review & editing.

Pavol Kačmár: Investigation; Resources; Supervision; Writing – review & editing.

Anthony J. Krafnick: Data curation; Investigation; Supervision; Writing – review & editing.

Erica D. Musser: Investigation; Methodology; Supervision; Writing – review & editing.

Evie Vergauwe: Investigation; Methodology; Supervision; Writing – review & editing.

Kaitlyn M. Werner: Investigation; Methodology; Supervision; Writing – review & editing.

Balazs Aczel: Investigation; Supervision; Writing – review & editing.

Patrícia Arriaga: Investigation; Resources; Supervision; Writing – review & editing.

Carlota Batres: Investigation; Supervision; Writing – review & editing.

Jennifer L. Beaudry: Investigation; Supervision.

Florian Cova: Methodology; Resources; Writing – review & editing.

Simona Ďurbisová: Investigation; Resources; Writing – review & editing.

Leslie D. Cramblet Alvarez: Investigation; Supervision; Writing – review & editing.

Gilad Feldman: Investigation; Supervision; Writing – review & editing.

Hendrik Godbersen: Investigation; Resources; Writing – review & editing.

Jaroslav Gottfried: Investigation; Supervision; Writing – review & editing.

Gerald J. Haeffel: Investigation; Supervision; Writing – review & editing.

Andree Hartanto: Investigation; Supervision; Writing – review & editing.

Chris Isloi: Investigation; Resources; Supervision; Writing – review & editing.

Joseph P. McFall: Investigation; Supervision; Writing – review & editing.

Marina Milyavskaya: Investigation; Supervision; Writing – review & editing.

David Moreau: Investigation; Supervision; Writing – review & editing.

Ester Nosáľová: Investigation; Resources; Writing – review & editing.

Kostas Papaioannou: Investigation; Resources; Supervision; Writing – review & editing.

Susana Ruiz-Fernandez: Investigation; Resources; Supervision; Writing – review & editing.

Jana Schrötter: Investigation; Resources; Writing – review & editing.

Daniel Storage: Investigation; Resources; Supervision; Writing – review & editing.

Kevin Vezirian: Resources; Supervision; Writing – review & editing.

Leonhard Volz: Investigation; Supervision; Writing – review & editing.

Yanna J. Weisberg: Resources; Supervision; Writing – review & editing.

Qinyu Xiao: Resources; Visualization; Writing – review & editing.

Dana Awlia: Investigation; Writing – review & editing.

Hannah W. Branit: Investigation; Writing – review & editing.

Megan R. Dunn: Investigation; Writing – review & editing.

Agata Groyecka-Bernard: Investigation; Writing – review & editing.

Ricky Haneda: Investigation; Writing – review & editing.

Julita Kielinska: Resources; Writing – review & editing.

Caroline Kolle: Investigation; Writing – review & editing.

Paweł Lubomski: Resources; Writing – review & editing.

Alexys M. Miller: Investigation; Writing – review & editing.

Martin J. Mækelæ: Resources; Writing – review & editing.

Mytro Pantazi: Resources; Writing – review & editing.

Rafael R. Ribeiro: Investigation; Writing – review & editing.

Robert M. Ross: Investigation; Writing – review & editing.

Agnieszka Sorokowska: Investigation; Writing – review & editing.

Christopher L. Aberson: Supervision; Writing – review & editing.

Xanthippi Alexi Vassiliou: Resources; Writing – review & editing.

Bradley J. Baker: Supervision; Writing – review & editing.

Miklos Bognar: Resources; Writing – review & editing.

Chin Wen Cong: Resources; Writing – review & editing.

Alex F. Danvers: Software; Writing – review & editing.

William E. Davis: Investigation; Writing – review & editing.

Vilius Dranseika: Methodology; Writing – review & editing.

Andrei Dumbravă: Investigation; Resources; Writing – review & editing.

Harry Farmer: Supervision; Writing – review & editing.

Andy P. Field: Methodology; Writing – review & editing.

Patrick S. Forscher: Formal analysis; Writing – review & editing.

Aurélien Graton: Resources; Writing – review & editing.

Nandor Hajdu: Investigation; Writing – review & editing.

Peter A. Howlett: Investigation; Writing – review & editing.

Radosław Kabut: Resources; Writing – review & editing.

Emmett M. Larsen: Resources; Writing – review & editing.

Sean T. H. Lee: Investigation; Writing – review & editing.

Nicole Legate: Investigation; Writing – review & editing.

Carmel A. Levitan: Supervision; Writing – review & editing.

Neil Levy: Investigation; Writing – review & editing.

Jackson G. Lu: Resources; Writing – review & editing.

Michał Misiak: Investigation; Writing – review & editing.

Roxana E. Morariu: Resources; Writing – review & editing.

Jennifer Novak: Investigation; Writing – review & editing.

Ekaterina Pronizius: Resources; Writing – review & editing.

Irina Prusova: Investigation; Writing – review & editing.

Athulya S. Rathnayake: Investigation; Writing – review & editing.

Marina O. Romanova: Investigation; Writing – review & editing.

Jan P. Röer: Resources; Writing – review & editing.

Waldir M. Sampaio: Resources; Writing – review & editing.

Christoph Schild: Resources; Writing – review & editing.

Michael Schulte-Mecklenbeck: Resources; Writing – review & editing.

Ian D. Stephen: Investigation; Writing – review & editing.

Peter Szecsi: Investigation; Writing – review & editing.

Elizabeth Takacs: Investigation; Writing – review & editing.

Julia N. Teeter: Investigation; Writing – review & editing.

Elian H. Thiele-Evans: Investigation; Writing – review & editing.

Julia Valeiro-Paterlini: Investigation; Writing – review & editing.

Iris Vilares: Investigation; Writing – review & editing.

Louise Villafana: Investigation; Writing – review & editing.

Ke Wang: Resources; Writing – review & editing.

Raymond Wu: Investigation; Writing – review & editing.

Sara Álvarez-Solas: Investigation; Writing – review & editing.

Hannah Moshontz: Data curation; Formal analysis; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.

Erin M. Buchanan: Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing – review & editing.

Declaration of Conflicting Interests

Funding

Open Practices

This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges

ORCID iDs

Braeden Hall ⬦ https://orcid.org/0000-0002-4157-3054
Kathleen Schmidt ⬦ https://orcid.org/0000-0002-9946-5953
Jordan Wagge ⬦ https://orcid.org/0000-0002-5105-2084
Savannah C. Lewis ⬦ https://orcid.org/0000-0002-9948-1195
Felix Kiunke ⬦ https://orcid.org/0000-0003-2344-853X
Gerit Pfuhl ⬦ https://orcid.org/0000-0002-3271-6447
Stefan Stieger ⬦ https://orcid.org/0000-0002-7784-6624
Ulrich S. Tran ⬦ https://orcid.org/0000-0002-6589-3167
Marta Kowal ⬦ https://orcid.org/0000-0001-9050-1471
Martin Voracek ⬦ https://orcid.org/0000-0001-6109-6155
Christopher R. Chartier ⬦ https://orcid.org/0000-0002-4568-4827
Mark J. Brandt ⬦ https://orcid.org/0000-0002-7185-7031
Jon E. Grahe ⬦ https://orcid.org/0000-0001-6970-0863
Asil A. Özdoğru ⬦ https://orcid.org/0000-0002-4273-9394
Sau-Chin Chen ⬦ https://orcid.org/0000-0001-6092-6049
Thomas R. Evans ⬦ https://orcid.org/0000-0002-6670-0718
Hans IJzerman ⬦ https://orcid.org/0000-0002-0990-2276
Pavol Kačmár ⬦ https://orcid.org/0000-0003-0076-1945
Anthony J. Krafnick ⬦ https://orcid.org/0000-0002-1692-0413
Evie Vergauwe ⬦ https://orcid.org/0000-0002-7339-2370
Kaitlyn M. Werner ⬦ https://orcid.org/0000-0003-4475-5256
Balazs Aczel ⬦ https://orcid.org/0000-0001-9364-4988
Patrícia Arriaga ⬦ https://orcid.org/0000-0001-5766-0489
Carlota Batres ⬦ https://orcid.org/0000-0002-3833-7667
Jennifer L. Beaudry ⬦ https://orcid.org/0000-0003-1596-6708
Florian Cova ⬦ https://orcid.org/0000-0001-9360-8290

Leslie D. Cramblet Alvarez (iD) https://orcid.org/0000-0002-0830-9606

Gerald J. Haeffel (iD) https://orcid.org/0000-0002-4029-1493

Chris Isloi (iD) https://orcid.org/0000-0001-5295-287X

Marina Milyavskaya (iD) https://orcid.org/0000-0002-0510-4891

David Moreau (iD) https://orcid.org/0000-0002-1957-1941

Kostas Papaioannou (iD) https://orcid.org/0000-0002-4661-5743

Susana Ruiz-Fernandez (iD) https://orcid.org/0000-0002-1709-1506

Jana Schrötter (iD) https://orcid.org/0000-0002-9830-6184

Daniel Storage (iD) https://orcid.org/0000-0001-7147-2206

Kevin Vezirian (iD) https://orcid.org/0000-0003-4013-7725

Qinyu Xiao (iD) https://orcid.org/0000-0002-9824-9247

Agata Groyecka-Bernard (iD) https://orcid.org/0000-0002-1932-4828

Ricky Haneda (iD) https://orcid.org/0000-0002-8300-8408

Caroline Kolle (iD) https://orcid.org/0000-0002-4647-627X

Alexys M. Miller (iD) https://orcid.org/0000-0002-3063-9361

Martin J. Mækelæ (iD) https://orcid.org/0000-0002-6791-1218

Mytro Pantazi (iD) https://orcid.org/0000-0002-8347-3115

Rafael R. Ribeiro (iD) https://orcid.org/0000-0001-6290-3182

Robert M. Ross (iD) https://orcid.org/0000-0001-8711-1675

Agnieszka Sorokowska (iD) https://orcid.org/0000-0003-3999-8851

Xanthippi Alexi Vassiliou (iD) https://orcid.org/0000-0002-6058-3389

Bradley J. Baker (iD) https://orcid.org/0000-0002-1697-4198

Miklos Bognar (iD) https://orcid.org/0000-0003-2717-1769

Chin Wen Cong (iD) https://orcid.org/0000-0003-4592-0727

William E. Davis (iD) https://orcid.org/0000-0002-8119-821X

Vilius Dranseika (iD) https://orcid.org/0000-0002-1144-1624

Andrei Dumbravă (iD) https://orcid.org/0000-0002-7981-8707

Andy P. Field (iD) https://orcid.org/0000-0003-3306-4695

Patrick S. Forscher (iD) https://orcid.org/0000-0002-7763-3565

Aurélien Graton (iD) https://orcid.org/0000-0001-7533-576X

Nandor Hajdu (iD) https://orcid.org/0000-0003-4012-4932

Emmett M. Larsen (iD) https://orcid.org/0000-0002-8516-5538

Sean T. H. Lee (iD) https://orcid.org/0000-0002-4059-9738

Nicole Legate (iD) https://orcid.org/0000-0001-8086-9643

Carmel A. Levitan (iD) https://orcid.org/0000-0001-5403-444X

Neil Levy (iD) https://orcid.org/0000-0002-5679-1986

Jackson G. Lu (iD) https://orcid.org/0000-0002-0144-9171

Ekaterina Pronizius (iD) https://orcid.org/0000-0003-1446-196X

Irina Prusova (iD) https://orcid.org/0000-0001-9298-2408

Athulya S. Rathnayake (iD) https://orcid.org/0000-0002-8870-4138

Waldir M. Sampaio (iD) https://orcid.org/0000-0002-6066-4314

Christoph Schild (iD) https://orcid.org/0000-0002-6668-5773

Michael Schulte-Mecklenbeck (iD) https://orcid.org/0000-0002-0406-8809

Elian H. Thiele-Evans (iD) https://orcid.org/0000-0001-8008-3165

Julia Valeiro-Paterlini (iD) https://orcid.org/0000-0003-1669-2534

Iris Vilares (iD) https://orcid.org/0000-0003-1500-3864

Ke Wang (iD) https://orcid.org/0000-0002-5776-0815

Sara Álvarez-Solas (iD) https://orcid.org/0000-0002-8267-9816

Hannah Moshontz (iD) https://orcid.org/0000-0003-4345-3715

Erin M. Buchanan (iD) https://orcid.org/0000-0002-9689-4189

Acknowledgments

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/25152459241267902

Notes

1. What makes a scenario a true Gettier-type case has been widely debated in the literature; however, for the purpose of this predominantly empirical article, we loosely refer to scenarios from this class of philosophical thought experiments as Gettier-type cases, which we operationalize for our research below.

2. We use the term "counterfeit-object" in line with Powell et al. (2015) to describe "fake barn" Gettier cases in which the protagonist could have identified a counterfeit object but happened to identify the real object by luck alone.

3. The no-threat (i.e., knowledge control) and threat (i.e., Gettier) conditions were structurally similar to the conditions used in the replicated study. Both studies featured protagonists in the woods trying to identify an animal. In the threat conditions, the protagonist identifies the animal correctly but only because of some kind of situational luck.

4. The COVID-19 pandemic changed and significantly limited how students could carry out their replication studies. After it began, our data collection was shifted to almost entirely online (and individual) participation. As shown in Table 1, most sites had online and individual sessions, some of the sites had both session types for one or both of the two variables, and some sites were missing documentation. Thus, using the covariates as intended would have been impossible.

5. The Stage 1 registered report manuscript included sections that described the recruitment and approval of collaborators who would collect data. We have restructured the Method section to more closely resemble that of a typical empirical article. The original text, updated to reflect the study's completion, is available in Appendix A in the Supplemental Material available online.

6. As described in our Stage 1 manuscript and as per CREP policy, completion certificates are typically granted to project teams that achieve a target sample size and pass a post-data-collection

review of the team's OSF page to ensure it includes anonymized data, a brief writeup of their site-level analysis, and a completion pledge. Because this project required more sophisticated statistical analysis than is typical of CREP projects, we decided not to require the written results, although 18 teams did complete this step. We also introduced some flexibility regarding the sample size because of COVID limitations and other concerns related to recruitment on small campuses.

7. In the approved protocol, we described a plan for data collection whereby each lab preregistered a target sample size of 50 to 100 and stopped collecting data on April 1, 2020, or once all contributors reached their preregistered target sample size. Because of the COVID-19 pandemic, this plan was not followed. The deadline for data collection was extended to June 1, 2021. Many data-collection sites stopped collecting data earlier.

8. There may be measurement error in participants' reported years of education. Less than the equivalent of a high school diploma was reported by 620 participants, 52 of whom reported 1 year of education.

9. Although we planned to measure participants' racial and ethnic identities using an open-ended response, racial and ethnic identity was measured using nonexclusive categories with an open-ended fill-in option for reasons that were not documented. Student research teams designed different response options tailored to their geographic region (see all variations in Appendix C in the Supplemental Material). All data-collection sites allowed people to select multiple racial and ethnic identities, and all asked whether participants identified as White (either "White/ European," "White/European descent," or "European descent").

10. We did not preregister the exclusion of people who reported their age as over 100; only seven people were excluded on the basis of this criteria alone (i.e., they did not meet any other exclusion criteria). These responses may have been errors in data entry or unlabeled test responses.

11. Turri et al. (2015) used the same type of question for the same purpose and excluded 15 of 135 participants on this basis.

12. The approved protocol described a power analysis conducted before data collection. The text from the original protocol is reproduced in full in Appendix A in the Supplemental Material and is summarized here.

13. A set of multilevel models examined if the data source (Qualtrics vs. SoSciSurvey) interacted with experimental condition in predicting knowledge, reasonableness, and luck judgments. No interaction was found in these analyses, which can be viewed at https://osf.io/nvfbm. Therefore, all data were combined into one large data set after matching variables.

14. Teams that participated in this extension were required to collect twice as many participants ($n > 100$; half in the continuous condition and half in the binary condition) so that they could meet the sample-size requirement ($n = 50$) for participants using only the preapproved continuous measure. However, because we converted all continuous responses to binary responses (see Analytic Approach section below for more details), the binary responses collected using this extension were also included with the converted binary responses in analyses.

15. The two-part luck/ability attribution was planned as a single item with two responses presented on a single screen. The presentation of the measure was modified to reduce participant confusion by splitting the two parts across two items on separate screens.

16. For participants from the United States, less than a high school education was coded as 10 years, a high school diploma was coded as 12 years, some college or a 2-year college degree was coded as 14 years, a 4-year college degree was coded as 16 years, a master's degree was coded as 18 years, and a doctorate or professional degree was coded as 20 years. For participants from Portugal, the labels and coding were the same except that a 3-year college degree was coded as 15 years, and a doctorate degree was coded to 21 years.

17. Data-collection sites were not given instructions about avoiding or clearly identifying test responses. At many data-collection sites, the students and other researchers executing the study tested their survey link multiple times (e.g., as inferred by responses to open-ended questions marked "test").

18. Analyses were repeated using the original exclusion criteria, which included five additional participants who reported ages 100 or above and excluded participants from sites without CREP completion certifications. One minor difference in results was found. For the reasonableness dependent variable, the vignette by condition interaction was not observed in one of the tested models, likely because of the smaller sample size after exclusions.

19. In the approved protocol, data collection was described as taking place in labs. Labs were described as uniquely identifying data-collection sites. However, at some data-collection sites, multiple student-led teams joined this project (e.g., under the mentorship of the same principal investigator, multiple students joined the project as "labs"). Observations were labeled as belonging to both a "lab" (which we describe as a "student-led team") and a data-collection site. For analyses, the data-collection site was used in place of the "lab" variable described in the approved protocol.

20. The only difference we found in comparing results of the linear versus logistic models was in the sample-source analyses for the reasonableness and knowledge dependent measures. The linear models found interaction effects between condition and sample source (MTurk vs. not MTurk), whereas the logistic models did not. Examination of the patterns of results indicated the same condition differences for both data sources with slightly weaker effects for the MTurk data than the non-MTurk data.

21. In the approved protocol, the results section focused heavily on the project's logistics and structured results reporting in ways that would not allow for a transparent and thorough description of model fit and other important aspects of results, such as assumption checks. Furthermore, some model-specification details in the approved protocol conflicted with stated research questions (e.g., we specified that the null model would include the focal predictor, which would have rendered the null model invalid because null models are not supposed to include any predictors).

22. Only participants who missed the Darrel comprehension-check question ($n = 1,138$) were excluded from this analysis to replicate the exclusion criteria implemented in the original experiment.

References

Bartoń, K. (2020). *MuMIn: Multi-model inference* (1.43.17) [Computer software]. https://CRAN.R-project.org/package=MuMIn

Bergenholtz, C., Busch, J., & Praëm, S. K. (2021). Exclusion criteria in experimental philosophy. *Erkenntnis*, *86*(6), 1531–1545. https://doi.org/10.1007/s10670-019-00168-5

Birch, S. A. J. (2005). When knowledge is a curse: Children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, *14*(1), 25–29. https://doi.org/10.1111/j.0963-7214.2005.00328.x

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science*, *8*(3), 297–302.

Buckwalter, W., & Stich, S. (2010). Gender and philosophical intuition. *SSRN*. https://doi.org/10.2139/ssrn.1683066

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254. https://doi.org/10.1086/261651

Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme*, *11*(2), 199–212. https://doi.org/10.1017/epi.2014.7

DePaul, M. R., & Ramsey, W. M. (Eds.). (1998). *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry*. Rowman & Littlefield.

Disher, N. G., Guerra, A. L., & Haeffel, G. J. (2021). Men have ability, women are lucky: A pre-registered experiment examining gender bias in knowledge attribution. *British Journal of Social Psychology*, *60*(3), 808–825. https://doi.org/10.1111/bjso.12443

Dutant, J. (2015). The legend of the justified true belief analysis. *Philosophical Perspectives*, *29*(1), 95–145. https://doi.org/10.1111/phpe.12061

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123. https://doi.org/10.1093/analys/23.6.121

Grahe, J., Brandt, M., IJzerman, H., & Cohoon, J. (2014). Replication education. *APS Observer*, *27*(3). https://www.psychologicalscience.org/observer/replication-education

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Jacquette, D. (1996). Is nondefectively justified true belief knowledge? *Ratio*, *9*(2), 115–127. https://doi.org/10.1111/j.1467-9329.1996.tb00100.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69. https://doi.org/10.1037/a0028347

Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes: Implications for power, precision, planning of research, and replication. *Psychological Methods*, *24*(5), 578–589. https://doi.org/10.1037/met0000209

Kim, M., & Yuan, Y. (2015). No cross-cultural differences in the Gettier car case intuition: A replication study of Weinberg et al. 2001. *Episteme*, *12*, 355–361. https://doi.org/10.1017/epi.2015.17

Landy, J. F., Jia, M. (Liam), Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M.,

Derks, K., Wagenmakers, E.-J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., . . . Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451–479. https://doi.org/10.1037/bul0000220

Larkin, K., & Andreychik, M. (2019). *Materials*. OSF. https://osf.io/fh52n/

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989x.7.1.19

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., Grinberg, M., Hannikainen, I., Hashimoto, T., Horowitz, A., . . . Zhu, J. (2017). The Gettier intuition from South America to Asia. *Journal of Indian Council of Philosophical Research*, *34*(3), 517–541. https://doi.org/10.1007/s40961-017-0113-y

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (2017). Gettier across cultures. *Noûs*, *51*(3), 645–664. https://doi.org/10.1111/nous.12110

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181–190. https://doi.org/10.1037/0033-2909.113.1.181

Moser, P. K. (2002). *The Oxford handbook of epistemology*. Oxford University Press.

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., . . . Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Murray, S., Dykhuis, E. D., & Nadelhoffer, T. (2022). *Do people understand determinism? The tracking problem for measuring free will beliefs*. PsyArXiv. https://doi.org/10.31234/osf.io/kyza7

Nagel, J., Juan, V. S., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, *129*(3), 652–661. https://doi.org/10.1016/j.cognition.2013.02.008

Nagel, J., Mar, R. A., & Juan, V. S. (2013). Authentic Gettier cases: A reply to Starmans and Friedman. *Cognition*, *129*(3), 666–669. https://doi.org/10.1016/j.cognition.2013.08.016

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), Article 20170213. https://doi.org/10.1098/rsif.2017.0213

Nichols, S., Stich, S., & Weinberg, J. (2003). Metaskepticism: Meditations in ethno-epistemology. In *The skeptics* (pp. 227–247). https://doi.org/10.1093/acprof:oso/9780199733477.003.0010

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Popiel, M. (2016). *A systematic review of studies using Gettier-type thought experiments* (Publication No. 1216) [Doctoral dissertation, University of Denver]. Electronic Theses and Dissertations. https://digitalcommons.du.edu/etd/1216

Powell, D., Horne, Z., Pinillos, N. Á., & Holyoak, K. J. (2015). A Bayesian framework for knowledge attribution: Evidence from semantic integration. *Cognition*, *139*, 92–104. https://doi.org/10.1016/j.cognition.2015.03.002

Pownall, M., Azevedo, F., König, L. M., Slack, H. R., Evans, T. R., Flack, Z., Grinschgl, S., Elsherif, M. M., Gilligan-Lee, K. A., Oliveira, C. M., Gjoneska, B., Kanadadze, T., Button, K. S., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Dechterenko, F., Alzahawi, S., & Baker, B. J., . . . FORRT. (2023). Teaching open and reproducible scholarship: A critical review of the evidence base for current pedagogical methods and their outcomes. *Royal Society Open Science*, *10*(5), Article 221255. https://doi.org/10.1098/rsos.221255

Seyedsayamdost, H. (2015). On normativity and epistemic intuitions: Failure of replication. *Episteme*, *12*(1), 95–116. https://doi.org/10.1017/epi.2014.27

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, *132*(1), 99–107. https://doi.org/10.1007/s11098-006-9050-3

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124*(3), 272–283. https://doi.org/10.1016/j.cognition.2012.05.017

Starmans, C., & Friedman, O. (2013). Taking 'know' for an answer: A reply to Nagel, San Juan, and Mar. *Cognition*, *129*(3), 662–665. https://doi.org/10.1016/j.cognition.2013.05.009

Starmans, C., & Friedman, O. (2020). Expert or esoteric? Philosophers attribute knowledge differently than all other academics. *Cognitive Science*, *44*(7), Article e12850. https://doi.org/10.1111/cogs.12850

Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on truetemp. *Philosophy and Phenomenological Research*, *76*(1), 138–155. https://doi.org/10.1111/j.1933-1592.2007.00118.x

Turri, J. (2013). A conspicuous art: Putting Gettier to the test. *SSRN*. https://doi.org/10.2139/ssrn.3643930

Turri, J. (2016). Vision, knowledge, and assertion. *Consciousness and Cognition*, *41*, 41–49. https://doi.org/10.1016/j.concog.2016.01.004

Turri, J. (2017). Knowledge attributions in iterated fake barn cases. *Analysis*, *77*(1), 104–115. https://doi.org/10.1093/analys/anx036

Turri, J., Buckwalter, W., & Blouw, P. (2015). Knowledge and luck. *Psychonomic Bulletin & Review*, *22*(2), 378–390. https://doi.org/10.3758/s13423-014-0683-5

Vickstrom, E. R., Shin, H. B., Collazo, S. G., & Bauman, K. J. (2015). *How well—still good? Assessing the validity of the American community survey English-ability question* (Working Paper SEHSD-WP2015-18). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2015/demo/SEHSD-WP2015-18.pdf

Wagge, J. R., Baciu, C., Banas, K., Nadler, J. T., Schwarz, S., Weisberg, Y., IJzerman, H., Legate, N., & Grahe, J. (2019). A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect. *Collabra: Psychology*, *5*(1), Article 5. https://doi.org/10.1525/collabra.177

Weatherson, B. (2013). Margins and errors. *Inquiry*, *56*(1), 63–76. https://doi.org/10.1080/0020174X.2013.775015

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, *29*(1), 429–460. https://doi.org/10.5840/philtopics2001291/217

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Ziółkowski, A. (2016). Folk intuitions and the no-luck-thesis. *Episteme*, *13*, 343–358. https://doi.org/10.1017/epi.2015.49