THOUGHT
A JOURNAL OF PHILOSOPHY

Thought ISSN 2161-2234

ORIGINAL ARTICLE

# Knowledge from Falsehood: An Experimental Study

John Turri

University of Waterloo

Philosophers have debated whether it is possible to knowledgeably infer a conclusion from a false premise. For example, if a fan believes that the actress's dress is blue, but the dress is actually green, can the fan knowledgeably infer "the dress is not red" from "the dress is blue?" One aspect of this debate concerns what the intuitively correct verdict is about specific cases such as this. Here, I report a simple behavioral experiment that helps answer this question. The main finding is that people attribute knowledge in cases where a true conclusion is inferred from a false premise. People did this despite judging that the premise was false and unknown. People also viewed the agent as inferring the conclusion from the premise. In closely matched conditions where the conclusion was false, people did not attribute knowledge of the conclusion. These results support the view that the ordinary knowledge concept includes in its extension cases of knowledge inferred from false premises.

**Keywords** knowledge; truth; folk epistemology; inference

DOI:10.1002/tht3.417

## 1 Introduction

Some philosophers contend that one cannot gain knowledge by reasoning from false or unknown premises (e.g., Clark 1963; Harman 1973; Audi 1998, p. 159). One motivation for this view comes from its potential to explain intuitions about knowledge in certain types of cases. For example, suppose that Robert hears a loud sound which, based on background experience, he identifies as a vehicle backfiring, from which he infers that there is a vehicle operating nearby. But Robert is unaware that local people often ignite firecrackers that make the exact same sound. In that case, some philosophers claim, even if it is true that a vehicle backfired nearby, Robert does not know this because the basis for his inference "just as well might [have been] a firecracker" (Audi 1998, p. 159). In other words, Robert does not know that the sound was a vehicle backfiring, so from the premise, "That sound was a vehicle backfiring," he cannot knowledgeably infer that a vehicle is operating nearby. Generalizing from this case, we might conclude that even if inference sometimes extends knowledge, it cannot produce knowledge of a conclusion unless the premises are known. If a more general theoretical defense of this principle is desired, it might be argued that it prevents objectionable forms of luck from generating knowledge.

*Correspondence to:* E-mail: john.turri@gmail.com

In contrast, other philosophers contend that one can gain knowledge by reasoning from false or unknown premises (e.g., Fitelson 2010; Klein 2008; Luzzi 2014; Saunders and Champawat 1964; Turri 2011, 2012; Warfield 2005). Again, one motivation for this view is that it helps explain intuitions about knowledge in certain types of cases, such as these:

> The news channel breaks in with a live report: "The President is speaking now to supporters in Utah." Ted reasons, "The President is in Utah; therefore he is not attending today's NATO talks in Brussels." But the President is speaking at a "border rally" at the Utah-Nevada border and the speaking platform he is standing on is in Nevada. The crowd listening to the speech is in Utah. (Adapted from Warfield 2005, p. 408)

> Bill awaits Monica's arrival. He wonders whether she'll wear a scarlet dress. He hears a step on the staircase and swings around to see Monica enter the room. "What a dazzling indigo dress!" he thinks and concludes, "Monica's dress isn't scarlet." And he's right: her dress isn't scarlet. But it isn't indigo either. It's ultramarine. (Turri 2012, p. 217)

Regarding each case, it is claimed, the agent intuitively knows the inferred conclusion despite the fact that the premise is false and, as a result, unknown. For instance, Ted knows that the President is not in Brussels, despite inferring it from the false premise that the President is in Utah. And Bill knows that Monica is not wearing a scarlet dress, despite inferring it from the false premise that she is wearing an indigo dress. If a more general theoretical defense of this interpretation of the cases is desired, it might be argued that, in each case, the agent does enough to discover that the conclusion is true and, on the ordinary understanding, such discovery suffices for knowledge, even if it goes through false premises.

One question relevant to the debate regarding alleged knowledge from falsehood pertains to intuitively correct verdicts: what, if anything, is the intuitive verdict about such cases? Although some contributors to the literature propose complicated meta-philosophical theories about what a relevant "intuition" is (e.g., Schnee 2015, pp. 57–8), it is widely accepted that philosophical theories should respect broadly shared intuitive judgments about particular cases. The preference for respecting intuitive judgments is evident in philosophy generally (for some citations and review, see Buckwalter 2017; Turri 2018a) and among epistemologists contributing specifically to the debate over knowledge from falsehood. For instance, in a review article covering the debate over alleged knowledge from falsehood, one epistemologist remarks that accounting for the "intuitions these cases engender" is important for developing a theory of knowledge (Engel 2015). But recent research has shown that philosophers are not always good at identifying which verdicts are actually intuitive (e.g., Colaco et al. 2014; for reviews, see Turri 2016a; Turri in press), even among the community of professional philosophers (e.g., Horvath and Wiegmann 2016). Accordingly, it is worthwhile to investigate what the intuitive verdicts actually are regarding knowledge in such cases.

The present research was designed to accomplish this. I conducted a simple controlled experiment to test people's judgments about cases structurally similar to ones mentioned earlier in this introduction. Hundreds of adult participants first read one of several brief stories about an agent who infers a conclusion; then they judged whether the agent knows the conclusion. The critical test cases involved an agent who reasons from a false premise to a true conclusion.

To provide a fuller context for interpreting results from the cases of principal interest — that is, cases involving an inference from a false premise to a true conclusion — I also did three other things. First, I tested cases involving a true conclusion and a true premise. Such cases provide a useful control condition, on the assumption that they are naturally judged to be examples where the agent knows both the premise and the conclusion. Second, I tested cases involving a false conclusion and a false premise. Such cases provide another useful control condition, on the assumption that they are naturally judged to be examples where the agent knows neither the premise nor the conclusion. Third, I collected judgments about the truth-value of the premise and the conclusion. This allows us to assess whether participants correctly identified the truth-values of the underlying propositions. Failing to assess this would be unwise, for the following reason. Suppose we find that participants attribute knowledge of the conclusion in cases where the premise is false. Does it follow that they naturally judge that knowledge from falsehood is possible? No, it does not. Perhaps they judged the premise to be true because it was, overall, close (enough) to the truth. Responsibly addressing this possibility requires collecting judgments about truth-value in addition to judgments about knowledge. Additionally, I also piloted the test cases to check that people interpret the agent as inferring the conclusion from the relevant premise.

The results from the present study could support one of several conclusions. On one hand, the clear central tendency might be to attribute knowledge of a true conclusion when the premise is judged to be false and unknown. This would support the view that such cases intuitively involve knowledge. On another hand, the clear central tendency might be to deny knowledge of a true conclusion when the premise is judged to be false and unknown. This would support the view that such cases intuitively do not involve knowledge. On yet another hand, there might be no central tendency in such cases, which would be the hardest outcome to interpret. One possible interpretation would be that there is no intuitive verdict about knowledge in such cases, and philosophers who claim otherwise are being influenced by their theoretical commitments. Another possible interpretation would be that there are slightly different knowledge concepts shared by different groups of people — one allowing knowledge from falsehood, another disallowing it — and philosophers who intuitively disagree about such cases belong to these slightly different linguistic communities.

## 2 Experiment

The data for this experiment are available through an Open Science Foundation project (osf.io/v496q).

### 2.1 Method

Two hundred forty participants were tested (aged 18–79, mean age = 34 years; 96 female; 92% reporting English as a native language). Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk (AMT) and Qualtrics, and compensated $0.35 for approximately 2 minutes of their time. Repeat participation was prevented (by AMT worker ID).

Participants were randomly assigned to one of six conditions in a 3 (truth-value: true–true, false–true, false–false) × 2 (story: gift, dress) between-subjects design. All participants read a brief story, responded to four test items, then completed a brief demographic questionnaire. The truth-value factor manipulated whether the agent in the story reasoned from a true premise to a true conclusion, from a false premise to a true conclusion, or from a false premise to a false conclusion. I had no expectations for the story factor and included it as a robustness check (i.e., to ensure that any observed effects were not due principally to peculiarities of any particular scenario). Here is the complete text for each story (truth-value manipulation in brackets: [true–true/false–true/false–false]):

> (Gift) For her birthday gift, Michelle expects that her parents will give her a baby gerbil. When she opens the gift, it looks like a baby gerbil. [And it is a baby gerbil, which is a member of the rodent family. / But it is a baby mouse, which is a member of the rodent family. / But it is a pygmy possum, which is not a member of the rodent family.] Michelle says, "Gerbils are rodents, so it is a rodent."

> (Dress) Michael bet one of his friends that their favorite actress would not wear a red dress to tonight's award ceremony. A leading fashion site just posted images of the actress arriving at the ceremony. Given the lighting, the dress looks blue to Michael, [and it is blue / but it is green / but it is red]. Michael says, "Her dress is blue, so it is not red."

A pilot study confirmed that people interpret the agents in these stories as inferring the conclusion from the relevant premise. (See Appendix B for details on the pilot study.)

Beneath the story participants rated their agreement with two knowledge attributions in a question matrix (order randomized). For the gift story, the attributions were,

> Michelle knows that it is a gerbil. ("premise knowledge")

> Michelle knows that it is a rodent. ("conclusion knowledge")

For the dress story, they were,

> Michael knows that the dress is blue. ("premise knowledge")

> Michael knows that the dress is not red. ("conclusion knowledge")

Responses were collected on a standard 7-point Likert scale, 1 "strongly disagree" to 7 "strongly agree," left-to-right on the participant's screen.

Participants then advanced to a new screen and rated their agreement with two statements in a question matrix (order randomized). These were the propositions embedded in the two knowledge attributions above (i.e., the prejacents). The purpose of this was to ensure that the truth-value manipulations were effective. For the gift story, the two statements were,

It is a gerbil. ("truth of premise")

It is a rodent. ("truth of conclusion")

For the dress story, they were,

The dress is blue. ("truth of premise")

The dress is not red. ("truth of conclusion")

Responses were collected using the same scale described above.

### 2.2 Results

Preliminary regression analyses revealed no effect of participant age or sex on any of the dependent variables, so I omitted these demographic variables from all other analyses and will not discuss them any further. Participant response to the prejacents indicated that the truth-value manipulations were extremely effective: people recognized true propositions as true, and false propositions as false. (see Table A1 in Appendix A.) Analysis of variance revealed that the story factor had only one very small, unpredicted effect on the attribution of premise knowledge: the story factor explained less than 2% of the variance in the attribution of premise knowledge ($\eta_p^2 = .017$). (see Table A2 in Appendix A.) But a follow-up independent samples $t$-test exploring this effect was statistically insignificant, $t(238) = -1.73$, $p = .086$. Moreover, the story factor was not of independent theoretical interest and was included merely as a robustness check. Accordingly, the following analyses collapse across story.

Paired samples $t$-tests revealed that mean attribution of premise knowledge and conclusion knowledge did not differ in the true–true condition, $t(80) = 0.00$, $p = 1$, or in the false–false condition, $t(78) = 0.82$, $p = .418$. (see Figure 1.) Both attributions were high in the true–true condition but low in the false–false condition. By contrast, in the true–false condition, mean attribution was much lower for premise knowledge than for conclusion knowledge, $t(79) = 10.03$, $p < .001$, $d = 1.13$. Mean attribution of conclusion knowledge did not differ between the true–true and false–true conditions, $t(159) = 1.38$, $p = .170$. The results from the false–true condition are unexpected if, on the ordinary view, knowledge could not be based on an unknown, false premise. In particular, people

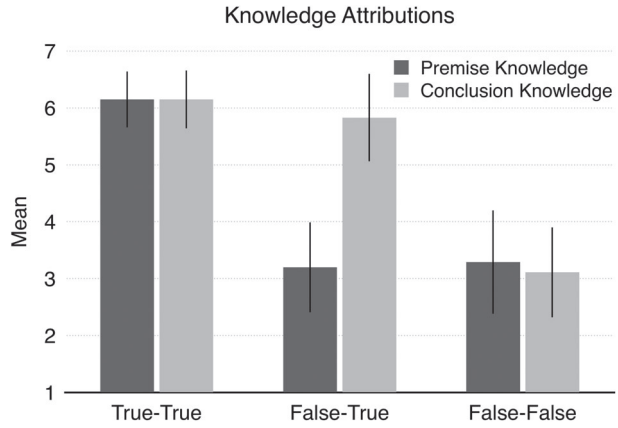


**Figure 1: Mean knowledge attributions. Scales ran 1 (*strongly disagree*) to 7 (*strongly agree*). Error bars represent bootstrapped 95% confidence intervals.**

judged that the premise was false and unknown, but they judged that the conclusion was true and known.

## 3 Conclusion

Philosophers have disagreed over whether knowledge from falsehood is possible. One part of this disagreement concerns the intuitively correct view of cases involving a true conclusion inferred from a false premise. In order to advance our understanding of this issue, I conducted a simple experiment to assess whether there was a clear central tendency in how people naturally judge such cases. In the critical condition where an agent reasons from a false premise to a true conclusion, people judged that the premise was false, and they denied that the agent knows the premise, but they nevertheless judged that the agent knows the conclusion. Moreover, people viewed the agent as reasoning from the relevant premise to the conclusion. Overall, this is not the pattern we would expect if, on the ordinary view, knowledge could not be based on an unknown, false premise.

An alternative interpretation of the results in the critical false–true conditions is that people were either biased against attributing knowledge of the specific premises featured in the cases, or biased in favor of attributing knowledge of the specific conclusions featured. If all we had were the results from the false–true conditions, then nothing in the data would speak against this interpretation. However, the results from closely matched control conditions allow us to rule it out. When the exact same premise and conclusion were both true (i.e., in the true–true conditions), people attributed knowledge of both. And when the exact same premise and conclusion were both false (i.e., in the false–false conditions), people attributed knowledge of neither. Accordingly, the results in the critical false–true conditions cannot be due to favorable or unfavorable biases toward the specific premises and conclusions tested here.

It is standard philosophical practice to prefer theories that agree with widely shared intuitive judgments (e.g., Aristotle 350 BCE/1941; Gautama 200/1913; Reid 1764/1997). Of course, sometimes widely held intuitive views are wrong, so this preference is not absolute. For instance, all things considered, we should reject the intuitive view of physical motion, which fails to properly account for, among other things, the principle of superposition (McCloskey 1983). So the preference for theories that match intuition is not absolute and I do not claim that the present results settle the debate over knowledge from falsehood. Nevertheless, the preference is not trivial either. Commonsense does not get the last word in philosophical debates, but it should at least get *a* word (Austin 1956). And the present results suggest that this word comes down clearly in favor of the possibility of knowledge from falsehood.

The present results are an important first step toward understanding the intuitive view of cases that have featured prominently in recent debates in epistemology. In particular, the results support the conclusion that the ordinary knowledge concept includes in its extension cases of knowledge inferred from a false premise. Nevertheless, further work remains to be done. For instance, some theorists speculate that the ordinary interpretation of such cases is, at once, subtler and cruder than many philosophers suppose (Montminy 2014, pp. 464–6). For instance, on the subtler side, people might not only attribute to Michael the belief, "Her dress is blue," but also the belief, "Her dress looks blue." As a result, they might judge that Michael has available to him multiple sensible inferential paths to the conclusion, "Her dress is not red." And they might even view Michael as making each of those inferences. On the cruder side, people's ordinary knowledge judgments might not encode information about the knowledge's source. For instance, people might view Michael as reasoning from "Her dress is blue" to "Her dress is not red," and also as knowing the conclusion, without representing Michael as having inferential knowledge of the conclusion. If so, then commonsense might not take a stand either way on the possibility of inferential knowledge from falsehood, because it does not exhibit any pattern of intuitions about knowledge *as* inferential.

These speculative hypotheses suggest further avenues of research on the present topic. On one hand, further studies could investigate the psychological processes involved in making knowledge attributions in cases like the ones tested here. Do people represent agents as implicitly conducting multiple inferences to the same conclusion, some involving false premises but others involving only true premises? If so, which of these representations inform people's judgment that the agent knows the conclusion? These are very challenging questions that will likely require a series of sophisticated and creative experimental studies to begin answering with confidence. On the other hand, further studies could investigate whether people represent knowledge as specifically inferential or, alternatively, whether the representation of inference and the representation of knowing remain distinct. As researchers consider this question, existing findings from cognitive science can be helpful in generating hypotheses and designing studies. For instance, existing findings from developmental psychology suggest that from a very early age, children

trust informants more or less strongly depending on the source of an informant's information (e.g., Pillow 1989). And research in linguistics shows that in some languages, markers indicating the evidential source of information are a syntactically obligatory aspect of assertions (e.g., Chafe and Nichols 1986). Such findings, I think, make it likely that people represent knowledge as deriving from specific sources, such as inference.

Prior findings demonstrate that people sometimes deny knowledge in cases that are naturally understood to involve a true conclusion inferred from a false premise. For example, researchers have studied the effects of different forms of luck on knowledge attribution (e.g., Machery et al. 2017; Starmans and Friedman 2012; Turri, Buckwalter, and Blouw 2015). A consistent finding from this literature is that people deny knowledge in cases where an agent luckily forms a true belief based on a false premise. For instance, suppose that an agent purchases a stone from the diamond exhibit of a reputable jewelry store. She puts it in her jacket pocket, leaves the store, and thinks, "I have a diamond in my pocket." Naturally, her conclusion is understood to be based on the premise that she just put a diamond in her pocket. However, she is unaware that the stone she purchased is a fake: it is cubic zirconia, not diamond. She is also unaware that, long ago, her grandmother sewed a diamond into the lining of her jacket pocket. So it is true that she has a diamond in her pocket, but people deny that she knows. Such findings on luck are fully consistent with my findings here, because the cases tested are very different. To take an obvious difference, the cases tested here do not involve improbable strokes of luck unrelated to what makes the agent's conclusion true. The findings on luck are also fully consistent with the hypothesis that it is conceptually possible to infer knowledge from a false premise. They are consistent because the possibility of knowledge from falsehood does not imply that all true beliefs inferred from false premises constitute knowledge.

Similarly, the present findings are also consistent with recent work arguing that even though it is conceptually possible for knowledge to be based on false premises, it is nevertheless impossible for knowledge to be based on irresponsibly held premises, which is thought to be a "plausible replacement" for theories claiming that premises must be known or at least true (Luzzi 2014). The argument for this alternative view also depends partly on verdicts about what is evident about specific thought experiments, which means that it could be evaluated by future research of the sort reported here.

Before closing, I would like to mention two theories that not only take the present findings at face value, but could also shed light on them. Each of these is a theory of the ordinary knowledge concept, that is, of the content of the knowledge concept used in everyday thought and talk. One theory is that knowledge is an accurate representation produced by cognitive ability (for relevant findings and discussion, see Turri 2016b, 2018b). Inferential abilities are for discovering truths based on background information. Focusing on the example involving Michael and the dress, one possibility is that people view Michael as having done enough to discover that the dress is not red. His false belief that the dress is blue is viewed as a good enough basis to discover that it is not red. A second theory is that knowledge is an accurate, actionable

proposition (for relevant findings, discussion, and citations, see Turri and Buckwalter 2017; Turri, Buckwalter, and Rose 2016; Turri in press). On this view, an agent knows a proposition when it is true and he should base his decisions on it. One possibility is that people attribute knowledge to Michael because they think he should act based on the proposition that the dress is red. For instance, Michael should settle the bet with his friend based on the dress's not being red. Of course, my purpose here is not to defend either of these theories as a general account of knowledge, nor to strongly claim that either theory definitely explains the present results—further work is clearly required to accomplish either of those things. Instead, my purpose is to connect the present findings to existing theories in a way that could identify fruitful paths to explore.

In summary, results from a behavioral experiment support the conclusion that, on the ordinary view, it is possible to infer knowledge from a false premise. Further research could undermine this conclusion by showing, for instance, that the psychological processes producing knowledge judgments in relevant cases are always much more complex than it appears. However, unless and until further evidence of that sort emerges, it is reasonable to conclude that, according to our ordinary knowledge concept, knowledge from falsehood is possible. To the extent that commonsense counts in favor of philosophical theories, this finding counts in favor of theories that allow for knowledge from falsehood. Regardless of whether commonsense counts in favor of philosophical theories, the present findings advance our understanding of the ordinary knowledge concept and commonsense epistemology.

## Acknowledgments

For helpful feedback I thank Wesley Buckwalter, Martin Montminy, and Angelo Turri. Thanks also to *Thought*'s editors and anonymous referees. This research was supported by the Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program.

## References

Aristotle. Nichomachean ethics. (R. McKeon, Ed., W. D. Ross, Trans.). 1941.

Audi, R. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. New York: Routledge, 1998.

Austin, J. L. "A plea for excuses." *Proceedings of the Aristotelian Society* **57** (1956): 1–30.

Buckwalter, W. "Epistemic Contextualism and Linguistic Behavior," in *Handbook of Epistemic Contextualism*, edited by J. J. Ichikawa. New York: Routledge, 2017; 44–56.

Chafe, W. and J. Nichols, (eds). *Evidentiality: The Linguistic Coding of Epistemology. Norwood, NJ*: Ablex, 1986.

Clark, M. "Knowledge and Grounds: A Comment on Mr. Gettier's Paper." *Analysis* **24**.2 (1963): 46–48.

Colaco, D., W. Buckwalter, S. Stich, and E. Machery. "Epistemic Intuitions in Fake-Barn Thought Experiments." *Episteme* **11**.02 (2014): 199–212. https://doi.org/10.1017/epi.2014.7

Engel, M. Epistemic luck. 2015. URL: http://www.iep.utm.edu/epi-luck/ (accessed 21 February, 2016).

Fitelson, B. "Strengthening the Case for Knowledge from Falsehood." *Analysis* **70**.4 (2010): 666–669. https://doi.org/10.1093/analys/anq078

Gautama. (200AD). Nyāya Sūtras. (B. D. Basu, Ed., V. S. scholars, Trans.). Allahabad: Sudhindranatha Vasu, 1913.

Harman, G. *Thought*. Princeton, NJ: Princeton University Press, 1973.

Horvath, J. and A. Wiegmann. "Intuitive Expertise and Intuitions about Knowledge." *Philosophical Studies* 173 (2016): 1–26. https://doi.org/10.1007/s11098-016-0627-1

Klein, P. "Useful false beliefs," in *Epistemology: New Essays*, edited by Q. Smith. Oxford: Oxford University Press, 2008: 25–63.

Luzzi, F. "What Does Knowledge-Yielding Deduction Require of its Premises." *Episteme* **11**.3 (2014): 261–275.

Machery, E., S. Stich, D. Rose, A. Chatterjee, K. Karasawa, N. Struchiner, Smita Sirker, Naoki Usui, and Takaaki Hashimoto. "Gettier across Cultures." *Nous* **51**.3 (2017): 645–664. https://doi.org/10.2307/2220147

McCloskey, M. "Intuitive Physics." *Scientific American* **284**.4 (1983): 122–130.

Montminy, M. "Knowledge despite Falsehood." *Canadian Journal of Philosophy* **44**.3–4 (2014): 463–475. https://doi.org/10.1080/00455091.2014.982354

Pillow, B. H. "Early Understanding of Perception as a Source of Knowledge." *Journal of Experimental Child Psychology* **47**.1 (1989): 116–129.

Reid, T. An inquiry into the human mind on the principles of common sense. (D. R. Brookes, Ed.). University Park, Penn.: Pennsylvania State University Press, 1997.

Saunders, J. T. and N. Champawat. "Mr. Clark's Definition of 'Knowledge'." *Analysis* **25**.1 (1964): 8–9.

Schnee, I. "There Is no Knowledge from Falsehood." *Episteme* **12**.1 (2015): 53–74.

Starmans, C. and O. Friedman. "The Folk Conception of Knowledge." *Cognition* **124**.3 (2012): 272–283. https://doi.org/10.1016/j.cognition.2012.05.017

Turri, J. "Manifest Failure: The Gettier Problem Solved." *Philosophers' Imprint* **11**.8 (2011): 1–11.

Turri, J. "In Gettier's wake," in *Epistemology: The Key Thinkers*, edited by S. Hetherington. London: Continuum, 2012, 214–229.

Turri, J. "How to Do Better: Toward Normalizing Experimentation in Epistemology," in *Advances in Experimental Philosophy and Philosophical Methodology*, edited by J. Nado. London: Bloomsbury Academic, 2016a, 35–51.

Turri, J. "A New Paradigm for Epistemology: From Reliabilism to Abilism." *Ergo* **3**.8 (2016b): 189–231.

Turri, J. "The Non-factive Turn in Epistemology: Some Hypotheses," in *The Factive Turn in Epistemology*, edited by V. Mitova. Cambridge: Cambridge University Press, 2018a, 219–228.

Turri, J. "Virtue Epistemology and Abilism on Knowledge," in *Routledge Handbook of Virtue Epistemology*, edited by H. Battaly. New York: Routledge, 2018b, 309–316.

Turri, J. "Knowledge Attributions and Lottery Cases: A Review and New Evidence," in *The Lottery Problem*, edited by I. Douven. Cambridge: Cambridge University Press, in press.

Turri, J. and W. Buckwalter. "Descartes's Schism, Locke's Reunion: Completing the Pragmatic Turn in Epistemology." *American Philosophical Quarterly* **54**.1 (2017): 25–46.

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

Turri, J., W. Buckwalter, and D. Rose. "Actionability Judgments Cause Knowledge Judgments." *Thought: a Journal of Philosophy* **5**.3 (2016): 212–222. https://doi.org/10.1002/tht3.213
Warfield, T. A. "Knowledge from Falsehood." *Philosophical Perspectives* **19** (2005): 405–417.

## Appendix A: Tables

**Table A1** Descriptive Statistics for All Dependent Variables in the Six Conditions

| | | Kp | | Kq | | P | | Q | |
|---|---|---|---|---|---|---|---|---|---|
| Condition | *N* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Gift | | | | | | | | | |
| True–true | 39 | 6.21 | 0.95 | 5.97 | 1.29 | 6.59 | 0.55 | 6.08 | 1.29 |
| False–true | 40 | 2.93 | 1.85 | 5.57 | 1.80 | 1.90 | 1.53 | 6.43 | 0.71 |
| False–false | 40 | 2.87 | 1.86 | 3.28 | 1.91 | 2.03 | 1.70 | 2.58 | 2.17 |
| Dress | | | | | | | | | |
| **T**rue–true | 42 | 6.10 | 1.32 | 6.31 | 1.09 | 6.40 | 0.77 | 6.48 | 0.74 |
| **F**alse–true | 40 | 3.48 | 1.75 | 6.08 | 1.66 | 1.80 | 1.36 | 6.80 | 0.61 |
| **F**alse–false | 39 | 3.72 | 2.00 | 2.95 | 1.75 | 2.31 | 1.78 | 2.46 | 1.78 |

Kp: the agent knows the premise. Kq: the agent knows the conclusion. P: the premise. Q: the conclusion.

## Appendix B: Pilot Study

I conducted a pilot study to check whether people interpreted the agent in the story as inferring the conclusion from the relevant premise. In light of my research question, this was especially important to check for the critical false–true condition. More specifically, in order for the results to help answer the research question, it is important for people to think that the agent's conclusion was based on the relevant false proposition. Given the way the scenarios are described, it is reasonable to assume that people did recognize this. For instance, in the last line of the scenario for the dress condition, Michael says, "Her dress is blue, so it is not red." So it would be very surprising if people did not interpret Michael as reasoning from "her dress is blue" to "it is not red." However, it is preferable to directly test this point rather than relying on my view of what is plausible or reasonable. Accordingly, I conducted a pilot study to accomplish this.

Eighty participants (aged 20–70, mean age = 38 years; 32 female; 92% reporting English as a native language) were randomly assigned to one of two conditions, gift and dress. Participants read one of the two false–true stories (gift, dress) used in the experiment reported in the main text. After reading the story, participants rated their agreement with two statements. For the gift story, the two statements were,

> Michelle concludes that the animal is a rodent because she thinks that the animal is a gerbil. ("correct description")

> Michelle concludes that the animal is a rodent because she thinks that the animal is a mouse. ("incorrect description")

**Table A2** Analyses of variance for the four dependent variables

| | Factor | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Truth-value | | | | Story | | | | Truth-value*Story | | | |
| Measure | *df* | *F* | *p* | $\eta_p^2$ | *df* | *F* | *p* | $\eta_p^2$ | *df* | *F* | *p* | $\eta_p^2$ |
| Kp | 2, 234 | 81.98 | <.001 | .412 | 1, 234 | 3.98 | .047 | .017 | 2, 234 | 1.73 | .179 | .015 |
| Kq | 2, 234 | 85.69 | <.001 | .423 | 1, 234 | 0.67 | .414 | .003 | 2, 234 | 1.48 | .231 | .012 |
| P | 2, 234 | 293.2 | <.001 | .715 | 1, 234 | <.001 | .997 | .001 | 2, 234 | 0.67 | .514 | .006 |
| Q | 2, 234 | 227.1 | <.001 | .660 | 1, 234 | 1.61 | .206 | .007 | 2, 234 | 0.92 | .400 | .008 |

Kp: the agent knows the premise. Kq: the agent knows the conclusion. P: the premise. Q: the conclusion.

For the dress story, they were,

> Michael concludes that the dress is not red because he thinks that the dress is blue. ("correct description")

> Michael concludes that the dress is not red because he thinks that the dress is green. ("incorrect description")

Responses were collected using the same 7-point scale described above.

Participants interpreted each story accurately. One sample *t*-tests showed that participants agreed with the correct description of the agent's reasoning (gift: $M = 5.23$, $SD = 1.61$, $t(39) = 4.82$, $p < .001$; dress: $M = 6.20$, $SD = 1.14$ $t(39) = 12.49$, $p < .001$), and they disagreed with the incorrect description (gift: $M = 3.03$, $SD = 1.83$, $t(39) = -3.37$, $p = .002$; dress: $M = 1.98$, $SD = 1.37$, $t(39) = -9.36$, $p < .001$) (all test values = 4). Although participants interpreted each story accurately, they were more confident in their interpretation of the dress story. Independent samples *t*-tests revealed that mean agreement with the correct description was stronger for the dress story, $t(78) = -3.15$, $p = .002$, and mean disagreement with the incorrect description was also stronger, $t(78) = 2.90$, $p = .005$. The modal response to the correct description was "*agree*" for the gift story and "*strongly agree*" for the dress story; the modal response to the incorrect description was "*disagree*" for the gift story and "*strongly disagree*" for the dress story. On reflection, this overall pattern makes sense, because the dress story is shorter and the agent's reasoning is more fully explicitly stated in the dress story. Michael is quoted as saying, "Her dress is blue," but Michelle is not quoted as saying, "The animal is a gerbil." Instead, that aspect of Michelle's reasoning, although clearly suggested, is left implicit. Overall, regardless of these differences, the results confirm the reasonable assumption that participants interpret the agents as reasoning from a false premise in the critical false–true conditions.

 