Intuitive Expertise and Intuitions About Knowledge

Joachim Horvath & Alex Wiegmann

December 30, 2015
Forthcoming in Philosophical Studies*

Abstract Experimental restrictionists have challenged philosophers’ reliance on intuitions about thought experiment cases based on experimental findings. According to the expertise defense, only the intuitions of philosophical experts count—yet the bulk of experimental philosophy consists in studies with lay people. In this paper, we argue that direct (experimental) strategies for assessing the expertise defense are preferable to indirect (inductive) strategies. A direct argument in support of the expertise defense would have to show: first, that there is a significant difference between expert and lay intuitions; second, that expert intuitions are superior to lay intuitions; and third, that expert intuitions accord with the relevant philosophical consensus. At present, there is only little experimental evidence that bears on these issues. To advance the debate, we conducted two new experiments on intuitions about knowledge with experts and lay people. Our results suggest that the intuitions of epistemological experts are superior in some respects, but they also pose an unexpected challenge to the expertise defense. Most strikingly, we found that even epistemological experts tend to ascribe knowledge in fake-barn-style cases. This suggests that philosophy, as a discipline, might fail to adequately map the intuitions of its expert practitioners onto a disciplinary consensus.

Keywords intuitions, thought experiments, knowledge, expertise defense, intuitive expertise, experimental philosophy, experimental restrictionism

* The final publication is available at Springer via http://dx.doi.org/10.1007/s11098-016-0627-1
1 Introduction

Findings from experimental philosophy suggest that lay people’s intuitions about philosophical thought experiments vary with a number of seemingly irrelevant factors, such as cultural background (Machery, Mallon, Nichols, & Stich, 2004), order of presentation (Liao, Wiegmann, Alexander, & Vong, 2012; Swain, Alexander, & Weinberg, 2008; Wiegmann & Waldmann, 2014), affective content (Nichols & Knobe, 2007), or heritable personality traits (Cokely & Feltz, 2009; Feltz & Cokely, 2009). Based on these findings, experimental philosophers of a ‘restrictionist’ bent have challenged the trustworthiness of thought experiment intuitions (Alexander, Mallon, & Weinberg, 2009; Alexander & Weinberg, 2007; Feltz & Cokely, 2012; Machery et al., 2004; Weinberg, 2007; Weinberg, Nichols, & Stich, 2001). Other philosophers have responded in a number of ways to the experimental restrictionist challenge (Bengson, 2013; Cappelen, 2012; Deutsch, 2009, 2010; Grundmann, 2010; Horvath, 2010; Kauppinen, 2007; Ludwig, 2007, 2010; Nagel, 2012; Sosa, 2007, 2009, 2010)—some more promising, some less promising (Alexander, 2012; Alexander & Weinberg, 2007; Deutsch, 2015; Horvath, 2010; Mizrahi, 2015; Nado, 2014, 2015b; Weinberg, Alexander, Gonnerman, & Reuter, 2012; Weinberg, Gonnerman, Buckner, & Alexander, 2010).1

The so-called ‘expertise defense’ emerged as one of the most fruitful and hotly debated reactions to the experimental restrictionists’ challenge (see Nado, 2014 for a recent survey). Proponents of the expertise defense have variously argued that professional philosophers are the relevant experts with respect to the intuitive evaluation of thought experiments, who can thus be expected to be largely resistant to the influence of irrelevant factors (Devitt, 2006, 2011; Grundmann, 2010; Hales, 2006; Hofmann, 2010; Horvath, 2010; Kipper, 2010; Ludwig, 2007; Turri, 2013; Williamson, 2010).

---

1 We use the term ‘intuition’ fairly broadly in this paper, i.e., as potentially covering a wide range of spontaneous cognitive responses to thought experiment cases (see, e.g., Pust, 2012 for an overview). This broad use is open to the objection that only intuitions of a specific kind matter for philosophical methodology and thus count as genuine intuitions (see, e.g., Bengson, 2013; Ludwig, 2007). We do not have the space to discuss this objection in more detail here. However, we would like to note that we are skeptical about the methodological significance of the ‘genuine intuitions objection’, since it is not clear whether even professional philosophers typically respond to thought experiment cases with “genuine intuitions” in their everyday practice (for discussion, see, e.g., Alexander, 2012, Chapter 5; Alexander & Weinberg, 2007; Horvath, 2010; Weinberg & Alexander, 2014).
2005, 2007, 2011; Wright, 2010). Since most findings from experimental philosophy concern only lay intuitions, these findings can accordingly be ignored as irrelevant to philosophical practice, which mainly relies on the intuitions of well-trained professional philosophers.  

There are direct and indirect strategies for assessing the expertise defense (Schulz, Cokely, & Feltz, 2011, p. 1724). The seminal restrictionist response by Weinberg et al. (2010) pursues an indirect strategy. Based on a survey of the psychological literature on expertise, Weinberg et al. argue that the intuitive abilities of professional philosophers lack some of the features that constitute genuine expertise in other domains. In contrast, proponents of direct strategies investigate the intuitions of the alleged philosophical experts more directly, that is, by performing suitable experiments with those expert subjects.

In this paper, we first argue in favor of direct strategies for assessing the expertise defense, both for dialectical and substantive reasons (direct and indirect strategies need not be in conflict, however, and can sometimes even be mutually reinforcing).

To advance the debate about direct strategies for assessing the expertise defense, we then present the results of two new experiments on expert intuitions about knowledge. Despite the fact that “Normativity and Epistemic Intuitions” (Weinberg et al., 2001)—arguably the “founding document” of experimental philosophy—reports a number of findings about knowledge-related intuitions, there has been no attempt so far to investigate the knowledge intuitions of philosophical experts more directly. This is a failure that we want to rectify in this paper. Our experiments were guided by three ideas: (1) to compare expert intuitions about knowledge with lay intuitions, (2) to

---

2 For a qualified empirical defense of philosophers’ intuitive expertise, see De Cruz (2015) on the basis of dual process psychology, and Buckwalter (2014) on the basis of the psychology of expert biases and limitations. See also Nado (2015a) for a qualified endorsement of philosophical expertise that does not, however, aim to rebut the experimental restrictionist challenge.

3 Some philosophers have argued that the findings of experimental philosophy are methodologically irrelevant because intuitions about thought experiment cases do not, on closer inspection, play any significant methodological role in philosophical practice (Cappelen, 2012, 2014; Deutsch, 2009, 2010, 2015). In this case, the expertise defense might be obsolete simply because there would be no challenge from experimental philosophy in the first place. The controversy over this important issue is still ongoing, however, and a number of philosophers have offered forceful replies to various arguments for the methodological irrelevance of intuitions (see, e.g., Bengson, 2014; Boghossian, 2014; Brogaard, 2014; Chalmers, 2014; Ichikawa, 2014; Nado, 2015b; Weinberg, 2014). For the purposes of this paper, we will therefore simply assume that intuitions about thought experiment cases, broadly understood, do play a significant methodological role in philosophical practice.

4 One reason might be that the results of Weinberg, Nichols, and Stich failed to replicate (Machery et al., 2015; Nagel, 2012; Nagel, Mar, & San Juan, 2013; Nagel, San Juan, & Mar, 2013; Seyedsayamdost, 2015).
measure their quality against the relevant epistemological consensus, and (3) to check whether the experts’ intuitions are in line with the consensus of their own discipline.

In contrast to most other studies on philosophical expert intuitions, our results offer some degree of support for the expertise defense. For example, we found that the knowledge-related intuitions of epistemological experts differ systematically from those of lay people in a number of critical cases, and that they differ in a way that one would expect on the basis of the relevant literature. On the other hand, our results have a puzzling aspect as well, for our expert subjects tended to ascribe knowledge even in cases where epistemological orthodoxy would deny knowledge, such as fake-barn-style cases. Thus, our findings also raise the troubling new question whether epistemologists have been collectively unaware of what their own intuitive expertise really tells them about certain key thought experiments about knowledge.

2 Ways of Assessing the Expertise Defense

2.1 Indirect and Direct Strategies for Assessing the Expertise Defense

Let us first consider the merits of indirect and direct strategies for supporting or challenging the expertise defense. As mentioned above, the early restrictionist response by Weinberg et al. (2010) pursues an indirect strategy. Weinberg et al. extract certain findings about genuine expertise from the relevant psychological literature and then suggest an ampliative inference from these findings to the alleged expert intuiters in philosophy. One of their key points is that, in a wide range of domains, the development of genuine expertise requires clear, timely, and reliable feedback during the process of training in a large number of training situations (see also Clarke, 2013). Weinberg et al. then argue that the training of professional philosophers in the intuitive evaluation of thought experiments falls considerably short of satisfying these requirements on the development of genuine expertise, and they conclude that professional philosophers’ intuitions about thought experiments are unlikely to result from genuine intuitive expertise. The crux in their inductive argument is the assumption that intuitive expertise with respect to thought experiments is relevantly similar to intuitive expertise in other domains, because otherwise findings about the latter would not be projectible to the
former. But there are reasons to be skeptical about this assumption of projectibility (see also Andow, 2015).

For example, paradigmatic cases of intuitive expertise in the psychological literature, such as fire ground commanders (G. Klein, Calderwood, & Clinton-Cirocco, 1986) or nurses in neonatal intensive care units (Crandall & Getchell-Reiter, 1993), are prima facie quite disanalogous to philosophical thought experimenters. While the intuitive judgments of the former are in the business of making causal predictions in highly complex, uncertain situations, intuitive judgments about thought experiments typically do not concern causal relations or uncertain outcomes. Rather, in a typical philosophical thought experiment, all relevant features of the intended scenario are either explicitly stipulated or otherwise implicitly assumed by the thought experimenter. If important facts about the scenario still remain open or unclear, then the thought experiment in question is simply not a good one. Thus, the thought experimenter's main job is to evaluate a scenario that already contains enough information to foreclose any uncertainty with respect to the target question, e.g., whether Smith knows that someone in his office owns a Ford (Lehrer, 1965). This is quite disanalogous to the informational situation of a fire fighter in a burning building, or a neonatal nurse who cares for premature infants.

Whether one shares our reservations about indirect strategies or not, they can hardly provide the most compelling way of assessing the expertise defense. One reason is that the relevant inductive arguments are almost always epistemically riskier than a more direct investigation of professional philosophers' intuitive expertise. Direct strategies are also dialectically more effective, for in the face of indirect arguments only, many professional philosophers will surely continue to think that they simply must be better and less biased in their intuitive evaluations of thought experiments than lay people. This expectation is actually supported by psychological research on the bias blind spot, i.e., by the finding that almost everyone takes her- or himself to be less biased than other people (Armor, 1998; Pronin, 2007; Pronin, Gilovich, & Ross, 2004)—a tendency that many experts exhibit in their own area of expertise as well (Chi, 1978; Glenberg & Epstein, 1987). Direct strategies for assessing the expertise defense should thus take priority over indirect strategies at the present stage of the debate.

---

5 It should be noted, however, that—due to the mediating role of operationalization—even the most direct psychological investigation of philosophical judgments and intuitions will still be indirect in comparison to, e.g., the observation of overt human behavior or bodily movements. For this reason, one should think of the distinction between direct and indirect strategies more as a continuum than as a strict dichotomy.
2.2 The Task for Direct Strategies

What needs to be shown by *direct strategies* for supporting the expertise defense?

First, there needs to be a **significant difference** between the relevant intuitions of professional philosophers and lay people. In the absence of such a difference, the question whether professional philosophers have **better** intuitions than lay people will typically be moot. Of course, wide intuitive agreement between alleged experts and lay people is not sufficient to rule out intuitive expertise, nor does substantial intuitive disagreement between experts and lay people already establish intuitive expertise (both experts and lay people might arrive at their intuitive verdicts in epistemically dubious ways). But on the assumption that philosophical experts possess genuine intuitive expertise, some significant difference between expert and lay intuitions is surely the most natural expectation, and its absence would seem to constitute strong prima facie evidence against the expertise defense.⁶

One direct strategy for challenging the expertise defense would therefore be to show that lay people and professional philosophers have more or less the same intuitions in a given domain. Extant studies that reveal a significant difference between lay people and professional philosophers are Sytsma and Machery (2010) on intuitions about phenomenal consciousness, Schulz, Cokely, and Feltz (2011) on incompatibilist intuitions about free will, and Machery (2012) on expert intuitions about reference.⁷ Other pertinent studies (Hitchcock & Knobe, 2009; Schwitzgebel & Cushman, 2012, 2015; Tobia, Buckwalter, & Stich, 2013; Tobia, Chapman, & Stich, 2013; Vaesen, Peterson, & Van Bezooijen, 2013) did not report any significant difference between philosophers and lay people. Our own experimental results indicate that there is indeed a significant difference between professional epistemologists and lay people concerning intuitions about knowledge (see below).

Second, a significant difference between experts and lay people by itself is not enough to support the expertise defense, of course. The relevant difference must also involve a **significant improvement** of expert intuitions over lay intuitions, for example, by

---

⁶ Thanks to an anonymous reviewer for prompting these clarifications.

⁷ After this paper was accepted for publication, we learned about a recent study by J. Adam Carter, Martin Peterson, and Bart van Bezooijen (2015) that inter alia suggests that philosophical experts are less willing than lay people to self-ascribe knowledge of simple analytic truths versus knowledge of widely accepted empirical truths (thanks to Martin Peterson for the pointer).
revealing that expert intuitions are less susceptible to the influence of some of the irrelevent factors that experimental philosophers have identified. So far, there is only little experimental evidence that this might be the case. While the results of Sytsma and Machery (2010) are at least consistent with expert superiority, the findings of Schulz, Cokely, and Feltz (2011) indicate that philosophers’ compatibilist intuitions about free will are equally affected by the innate personality trait of extraversion as those of lay people. And Machery (2012) presents evidence that the referential intuitions in various subgroups of linguistic experts point in rather different directions. Turri (2013) offers at least some indirect support for an improvement of intuitions about knowledge through expertise, since he found that lay people’s judgments about Gettier cases can be brought in line with the epistemological consensus by a more perspicuous presentation of their structure. Our own results also suggest that various knowledge-related intuitions of professional epistemologists might be better than those of lay people insofar as they come significantly closer to the textbook consensus on cases of the relevant type (see below). As we will see, however, they might still not come close enough.

Third, a direct strategy in support of the expertise defense would ideally show that the expert intuitions are in line with the textbook consensus on the intuitive evaluation of thought experiments of the relevant type. If the intuitions of professional philosophers turned out to be significantly different from the relevant consensus in the literature, this would seem to be a considerable embarrassment, because it would indicate that philosophy, as a discipline, might be deluded about what the intuitions of its expert practitioners actually say. For example, even if the intuitions of professional epistemologists turned out to be free of all known biases, but also came out in favor of ascribing knowledge in Gettier situations, this would still be a troubling result for epistemology as a discipline. For such a result would suggest that the discipline of epistemology is dysfunctional insofar as it fails to map the intuitions of its expert practitioners onto a disciplinary consensus in the right kind of way. Our own results indicate that something like this might in fact be the case (see below).

In conclusion, it must be conceded that the presently available experimental evidence does not substantiate direct strategies for supporting the expertise defense. And while our own experiments—to be reported below—offer some support for the superiority of expert intuitions about knowledge, they also suggest that some of these intuitions deviate from the textbook consensus in unexpected and troublesome ways.
3 Experiments

3.1 Experiment 1

3.1.1 Rationale and Material

A first idea behind our experiment is simply to compare the intuitions of expert epistemologists and lay people concerning a number of systematically important types of thought experiments about knowledge.

A second idea is to measure the quality of lay people’s and expert epistemologists’ intuitions against how close they come to the relevant textbook consensus on thought experiments of the same kind. We therefore approach the issue of intuitive expertise from a slightly different angle than usual. The standard procedure would be to check whether professional philosophers are less susceptible to some philosophically irrelevant factor that was found to influence lay people’s intuitions (see, e.g., Schwitzgebel & Cushman, 2012, 2015). In contrast, we aim for a positive evaluation of the quality of expert intuitions about knowledge from the viewpoint of the present epistemological consensus. Experiments that “only” test for the susceptibility of expert intuitions to philosophically irrelevant factors can “at best” establish a purely negative result, i.e., that professional philosophers are not susceptible to the influence of such factors. Our experiments thus bear on the expertise defense in ways that differ from experiments that follow the standard procedure and, as a consequence, lead to interestingly different results, as we will argue below.8

It might be objected that regarding the textbook consensus as an adequate standard for evaluating intuitions about knowledge begs the question in favor of intuitions that support the textbook consensus. In this case, however, we face the unusually difficult task of assessing the quality of intuitions about knowledge in the absence of some independent standard. Moreover, the insistence that we have to bracket even those cases where we do find a robust epistemological consensus is arguably an unduly skeptical maneuver (Sosa, 2007; Williamson, 2004, 2011). For these reasons, the assumption that the textbook consensus provides at least a defeasible or

---

8 Thanks to an anonymous reviewer for encouraging us to make our procedure more explicit.
prima facie standard for assessing the quality of intuitions about knowledge does not seem especially problematic or question begging.

Even if one should reject the idea that the textbook consensus can be used as an adequate standard of correctness for intuitions about knowledge, one can still acknowledge a *third idea* behind our experiment, namely, to test whether the intuitions of expert epistemologists are in line with the relevant textbook consensus of their own discipline.

In order to truly engage the expertise of our epistemological experts, we confronted them with unfamiliar variations of thought experiments that already figure in the epistemological literature, although less prominently than, e.g., Gettier’s original cases (1963) or the fake-barn case (Goldman, 1976). Using such well-known cases might have the unwanted effect that epistemologists merely recall their earlier intuitive verdicts from memory, or that they simply reproduce the textbook consensus (see also Rini, 2015, sec. 5.1). Of course, someone who knows the epistemological literature really well will probably be able to identify the cases that inspired our vignettes. But using new variations of familiar cases should at least raise the chance that even professional epistemologists truly engage their intuitive abilities.

Let us now introduce the vignettes that we used in our first experiment. In each case, we will first reproduce the vignettes in exactly the form that we used in the experiment, alongside with the target question that the participants were asked to answer. After that, we describe the point of the vignette in a more abstract way, and then cite the cases on which it was modeled, as well as the consensus on the evaluation of cases of this kind according to the epistemological literature.

A security guard monitors a number of video screens that show what is going on in an office building across the street. At midnight, the guard checks her video screens and sees, on five different screens that show five different offices, that people are still working in their offices. She thinks to herself: “There is still someone in the building.” Four of the five screens are in fact working properly. But due to some unusual malfunction, one of the five screens actually shows a videotape from the night before, and today this particular office is already empty.

**How much do you disagree or agree with the following claim:**

*The guard knows that there is still someone in the building.*
Monitor is a case where the subject has multiple pieces of evidence for the target proposition, each of which would be sufficient for knowledge. In addition, the subject also has one faulty piece of evidence for the target proposition, in this case: the one malfunctioning video screen. In the epistemological literature, cases of this kind are regarded as clear cases of knowledge that refute the early no-false-assumption solution (Clark, 1963) to Gettier’s (1963) original counterexamples to the standard analysis of knowledge as justified true belief (cf. Lehrer, 1965). Such cases then gave rise to the no-essential-false-assumption solution (see, e.g., Harman, 1973; Lehrer, 1974; Lycan, 2006). In Monitor, for example, the security guard’s justification does not essentially depend on the malfunctioning video screen, since she equally relies on the four screens that function properly.

A CEO has owned a famous painting for a long time. Just this morning, she showed it to a colleague in her office where she can watch the painting every day. In the evening, the CEO returns home from work and is about to check her email. What she will find there is an urgent note by her secretary that says that her painting was stolen shortly after she had left her office. In fact, the CEO’s colleague merely wanted to play a trick on her, and so it was actually him who used the secretary’s email account to send her this note. The painting itself is still on the CEO’s office wall, just as it always was.

How much do you disagree or agree with the following claim:

At the time when the CEO returns home, but before she actually checks her email, she knows that the painting is still on her office wall.

Painting is a case where evidence that the subject does not (yet) possess—here, an unread email—threatens to undermine the subject’s justification for believing the target proposition that the painting is still on her office wall. Gilbert Harman, who introduced cases of this kind into the epistemological literature, regarded them as intuitive cases of non-knowledge (Harman, 1968, p. 172, 1973, pp. 144–145). Painting is modeled after the case of Donald and the unopened letter (Harman, 1973, p. 143), which is less discussed today than the more familiar assassination case (Harman, 1968, p. 172, 1973, pp. 143–144). It is fair to say, however, that many epistemologists did not follow Harman’s intuitive assessment of such cases, at least not without qualifications (see, e.g., P. Klein, 1981; Lycan, 1977; Pritchard, 2005). Accordingly, the present consensus is that cases of this kind are intuitively unclear (Shope, 2002, p. 32).
The director of a sculpture museum is so impressed with recent improvements of hologram images that she decides to perform a secret test on the visitors of her museum. To this end, she orders hologram images that even art experts cannot visually distinguish from the real sculptures in her museum, and she replaces all but one of the sculptures by their hologram image. As the director had expected, no one realizes any difference between the hologram images and the real sculptures. One day, the world's greatest Rodin expert is visiting her museum. The expert is standing in front of a famous marble sculpture by Rodin, which is the only real sculpture that is presently on display in the museum, and she thinks to herself: "I'm facing one of Rodin's famous marble sculptures now."

**How much do you disagree or agree with the following claim:**

*The Rodin expert knows that the sculpture in front of her is one of Rodin's famous marble sculptures.*

In *Sculpture*, the subject has a true belief about a marble sculpture on the basis of visual perception. However, the marble sculpture is surrounded by visually indistinguishable hologram images of sculptures in the same museum. In a sense, then, it is a matter of luck that the subject has acquired a true belief in this situation, for she could have easily looked at another sculpture and thereby acquired a false belief. This case is partly inspired by Lehrer and Paxson's (1969, pp. 234–235) case of Mr. Promoter, which they present as a counterexample to Unger's (1968) early anti-luck analysis of knowledge, and mainly by Carl Ginet's famous fake-barn case (Goldman, 1976, pp. 772–773), which Goldman uses as a counterexample to his earlier causal analysis of (perceptual) knowledge (Goldman, 1967). The fake-barn case was originally presented as a clear non-instance of knowledge, and this intuitive verdict is widely accepted in the epistemological literature (see, e.g., Engel Jr., 2015; Hetherington, 2015; Ichikawa & Steup, 2014; Shope, 2004; Steup, 2014), with only very few exceptions (see, e.g., Gendler & Hawthorne, 2005; Heathcote, 2006; Hetherington, 1999; Lycan, 1977, 2006).

A little girl likes to play a game with flipping a coin. She sometimes gets a “special feeling” that the next flip will come out heads. When she gets this “special feeling”, she is

---

9 Steup, for example, explicitly notes that there is “… broad agreement among epistemologists that Henry’s belief [in the fake-barn case] does not qualify as knowledge” (Steup, 2014, sec. 1.2).
right about half the time, and wrong about half the time. Just before the next flip, the little girl gets that "special feeling", and the feeling leads her to believe that the coin will land heads. She flips the coin, and it does land heads.

How much do you disagree or agree with the following claim:

*The little girl knew that the coin was going to land heads.*

This clear case of a non-instance of knowledge, *Clear Non-Knowledge 1*, was already used several times in experimental studies (see, e.g., Nichols, Stich, & Weinberg, 2003; Swain et al., 2008; Weinberg et al., 2001). In each case, the vast majority of the tested lay subjects classified it as a case of non-knowledge. A significant difference between expert intuitions and lay intuitions concerning this case would thus be strong evidence for the superiority (or inferiority) of expert intuitions.

Before leaving the house, a woman wants to check whether the light in her bathroom is off. She opens the bathroom door, and it is completely dark inside. She thinks to herself: “Alright, the light in my bathroom is off.”

How much do you disagree or agree with the following claim:

*The woman knows that the light in her bathroom is off.*

This newly devised case, *Clear Knowledge 1*, is intended as a clear, everyday case of knowledge on the basis of visual perception. As in the case of *Clear Non-Knowledge 1*, a significant difference between expert and lay intuitions about this case would be strong evidence for the superiority (or inferiority) of expert intuitions.

3.1.2 Subjects

Overall, the data of 224 subjects were included in our main analysis. 82 were identified as expert subjects (mean age was 38 years; 82% male) and 142 as lay subjects (mean age was 39 years; 37% male).

The expert subjects were recruited via a call for participation on an electronic mailing list for philosophers (PHILOS-L) and on the Experimental Philosophy blog (http://philosophycommons.typepad.com/xphi/) that contained a link that directed subjects to the experiment. In order to be included as expert subjects, participants had to indicate that, firstly, they have a PhD in philosophy and, secondly, epistemology is one
of their areas of specialization or competence. Furthermore, we excluded the data of subjects who did not complete the survey or completed it in less than one minute. Out of initially 421 people, 82 met these criteria, which were not announced in the call for participation or during the survey. To motivate people to take part in our study we announced that a book about experimental philosophy would be raffled among all participants. To enter this optional raffle, subjects were asked to provide their email address at the end of the survey.

Lay subjects were recruited via a database located in the UK. They were invited via an email that contained a link that directed subjects to the experiment. Each subject received £0.50. We included the data of subjects who indicated that they have no prior experience with philosophy and that they are native speakers of English. Moreover, we excluded the data of subjects who did not complete the survey or completed it in less than one minute. Out of initially 284 people, 142 met these criteria, which were again not announced in the call for participation or during the survey.

3.1.3 Design and Procedure
The experiment was conducted on the Internet. Upon clicking on a link which subjects received via email or on the blog post, they were redirected to a website containing the experiment. Subjects first read general instructions. These familiarized them with the 6-point Likert scale ranging from 1 (‘strongly disagree’) to 6 (‘strongly agree’), asked them to read the subsequent descriptions of scenarios carefully, and appealed to them to take the task seriously. After that, the three cases Monitor, Painting, and Sculpture were presented in a randomized order. The Clear Non-Knowledge 1 case and the Clear Knowledge 1 case were always presented on fourth and fifth position, respectively. We presented the relatively clear cases after the less clear cases in order to avoid contrast effects as found in other studies (cf. Swain et al., 2008). On the final page, we asked a number of demographic questions and assessed the participants’ level of philosophical education and expertise.

3.1.4 Results
The results of our first experiment are summarized in Figure 1. A mixed ANOVA (between subjects: lay subjects vs. expert subjects; within subject: the five cases) showed a very significant interaction of subjects and cases, indicating that the response pattern of lay subjects and expert subjects differed strongly over the five cases, F(4,
The average time spent to complete the survey did not differ significantly between lay people (322 seconds) and experts (353 seconds), $F(1, 220) = .77287, p = .38.\textsuperscript{10}

![Bar chart](image)

**Figure 1.** Mean ratings for the five cases divided in lay and expert judgments. The scale ranged from 1 ('strongly disagree' with the claim that the agent knew the relevant proposition) to 6 ('strongly agree'). Error bars represent 95% confidence intervals.

A more fine-grained analysis revealed that for three of the five cases the judgments of lay subjects and expert subjects differed significantly, with the experts' judgments being closer to the evaluation of corresponding cases in the epistemological literature. For *Monitor*, lay subjects ($M = 4.72, SD = 1.51$) and expert subjects ($M = 5.10, SD = 1.11$) both tended to agree with the claim that the guard knows that there is still someone in the building. The level of agreement for this case, which corresponds to cases that are regarded as a cases of knowledge in the epistemological literature, was significantly higher for expert subjects, $F(1, 222) = 3.93, p < .05, \eta^2 = .02.\textsuperscript{11}$ For *Sculpture*, lay subjects' ($M = 4.79, SD = 1.20$) agreed significantly stronger with the claim that the Rodin expert knows that the sculpture in front of her is one of Rodin's famous marble

\textsuperscript{10}Two subjects (one expert subject, one lay subject) were excluded from this analysis because they spent over 10,000 seconds longer than all other subjects.

\textsuperscript{11}We calculated all comparisons assuming equal and unequal variances. Since the results differed only marginally (it was never the case that a certain difference was significant using one assumption but not significant when the other assumption was used) we chose to report our findings in the most common form, i.e., without adjusting the degrees of freedom.
sculptures than expert subjects \( (M = 3.77, SD = 1.67), F(1, 222) = 28.06, p < .001, \eta^2 = .11 \). Although experts were more skeptical than lay subjects about the claim that this case constitutes a case of knowledge, they were also more inclined to agree with this claim than one would expect from the consensus in the epistemological literature (see Figure 2 for the distribution of expert judgments for Sculpture). For Clear Non-Knowledge 1, expert subjects \( (M = 1.41, SD = 0.74) \) disagreed significantly stronger than lay subjects \( (M = 2.64, SD = 1.44) \) with the claim that the little girl knew that the coin was going to land heads, \( F(1, 222) = 51.55, p < .001, \eta^2 = .19 \). A further interesting difference between lay subjects and expert subjects is that lay subjects’ evaluations for the first three cases (Monitor, Painting, and Sculpture) did not differ significantly, \( F(2, 282) = 2.00, p = .14 \), while the evaluations of expert subjects differed strongly, \( F(2, 162) = 27.32, p < .001, \eta^2 = .25 \) and in line with the relevant consensus in the epistemological literature. For the Painting case and Clear Knowledge 1, no significant difference between lay subjects and expert subjects was found \( (p = .52 \) and \( p = .09 \), respectively).

---

12 Since Sculpture is modeled after fake-barn cases (see above), our finding that lay people tend to ascribe knowledge in this case basically confirms a key result of the pioneering study on fake-barn cases by Colaço et al. (2014).

13 After this paper was accepted for publication, we learned about an unpublished study by J. Adam Carter, Duncan Pritchard, and Joshua Sheperd (ms) that inter alia compares expert and lay intuitions concerning fake-barn-style cases, and that mostly confirms our own findings. For example, Carter, Pritchard, and Sheperd also found that, on average, lay people and experts do ascribe knowledge in such cases, even though experts are less inclined to ascribe knowledge than lay people (thanks to Joshua Shepherd for the pointer).

14 Since the ratings of 3 and 4 were labeled with 'mildly disagree' and 'mildly agree', respectively, one might interpret ratings of 3 or below as cases of disagreement and ratings of 4 or above as cases of agreement. Applying this classification, only 3.8% of the expert subjects agreed with the claim that the little girl knew the outcome of the coin flip, while 33.8% of the lay subjects agreed with this claim.
Moreover, the result pattern of experts who indicated that epistemology is one of their areas of specialization versus experts who merely indicated that epistemology is one of their areas of competence did not differ significantly, $F(4, 320) = 1.67, p = .16$ (for the interaction of level of competence * cases). In order to see whether there is a significant difference between our expert subjects and those subjects who “merely” hold a PhD in philosophy, yet without having epistemology as one of their areas of specialization or competence, we compared the ratings of the latter ($N = 74$, not included in previous analyses) with the ratings of our experts subjects. The response pattern of these two groups differed significantly, $F(4, 468) = 4.51, p = .001, \eta^2 = .04$ (for the interaction of epistemological expertise * cases).\footnote{Thanks to Shen-yi Liao and Hannes Rusch for pressing this point on independent occasions.} At the level of individual cases, however, the difference between expert subjects ($M = 3.77, SD = 1.67$) and “mere” philosophy PhDs ($M = 4.54, SD = 1.32$) was only clearly significant for Sculpture, $F(1, 117) = 15.20, p = .01, \eta^2 = .05$, and it went in the direction that one would expect on the basis of the epistemological literature. This difference between expert epistemologists and “mere” philosophy PhDs lends some support to the domain-specificity of intuitive expertise in epistemology, which is a well-established feature of expertise more generally (cf. Ericsson & Lehmann, 1996; Gobet & Simon, 1996; Vicente & Wang, 1998).
3.2 Experiment 2

3.2.1 Rationale and Material

The main idea behind our second experiment was to replicate and corroborate two striking results of our first experiment.

On the one hand, this was the finding that even the epistemological experts classified *Sculpture* as a case of knowledge (in contrast to how cases of this kind are commonly assessed in the epistemological literature). To this end, we repeated *Sculpture* in the second experiment and added two new cases, *Exam* and *Dollar*, that should also be regarded as non-instances of knowledge according to the epistemological literature (see below).

On the other hand, we tried to explain the significant difference between epistemological experts and lay people in their intuitive evaluation of *Clear Non-Knowledge 1*. We speculated that this surprising difference might result from the fact that the agent sees the result, i.e., that the agent comes to know that she was right about the outcome of her coin flip (labeled as “success bias” in the following—see below). Therefore, we replaced *Clear Non-Knowledge 1* by an analogous case, *Clear Non-Knowledge 2*, which only differs insofar as the agent does not come to know that her prediction about the outcome of the coin flip was correct.

Let us now introduce and explain the additional vignettes that we used in our second experiment, following the same schema as above (cf. section 3.1.1).

A professor of medicine learns that one in a billion students die while taking their final exam. In about five minutes, the professor is set to substitute for a sick colleague in conducting a final exam of a student that she had never met before. The professor thinks to herself: “The student will survive her final exam.” And in fact, the student does survive her final exam.

How much do you disagree or agree with the following claim:

*Befor the exam began, the professor knew that the student would survive her final exam.*
In *Exam*, a professor acquires a belief about the survival of a student on the sole basis of overwhelming statistical evidence. This case is modeled after a lottery case, where one acquires the belief that one’s lottery ticket will lose merely on the basis of knowing the extremely high odds for winning the lottery (Harman, 1968, p. 166, 1973, p. 118). Such cases are typically treated as clear non-instances of knowledge in the epistemological literature (see, e.g., DeRose, 1996; Hawthorne, 2004; Pritchard, 2005; Williamson, 2000).

A waiter was recently hired by a restaurant in a remote part of the country where, unbeknownst to everyone, most of the circulating dollar bills are fake. The restaurant manager owes the waiter ten dollars for an extra hour that he worked. She takes a ten dollar bill from the cash register and hands it over to the waiter, and this leads the waiter to believe that the restaurant manager has just given him ten dollars. In fact, this particular ten dollar bill happens to be one of the few genuine dollar bills in the area.

*How much do you disagree or agree with the following claim:*

*The waiter knows that the restaurant manager has given him ten dollars.*

*Dollar* is a variation on the fake-barn case that is used less frequently than the original fake-barn case (see above). Just like the fake-barn case, it is treated as a clear non-instance of knowledge in the epistemological literature (see, e.g., Littlejohn, 2012, p. 193; Sutton, 2005, p. 360).

---

16 There is a disanalogy between *Exam* and standard lottery cases that might be relevant to their knowledge-related assessment, however. For on the natural assumption that the student is in normal physical condition, there might be no close possible world where this particular student dies during her final exam. And in that case, the professor’s belief that the student will survive her final exam could not easily have been false, i.e., the professor’s belief is *safe*. For this reason, one might count cases like *Exam* as bona fide cases of knowledge—unlike standard lottery cases, where we assume that every lottery ticket wins in some close possible world (see, e.g., Pritchard, 2005). A more systematic investigation of (expert) intuitions about lottery-style cases would thus have to compare cases like *Exam* with cases that are more closely analogous to standard lottery cases (thanks to Jens Kipper for pressing this point).

17 After this paper was accepted for publication, Aaron Meskin pointed out to us that the phrase ‘dollar bills’—without the qualifier ‘ten’—is naturally taken to refer to *one* dollar bills only. So taken literally, our vignette would trigger a reading according to which only the one dollar bills—but not the ten dollar bills—are fake in the scenario described, and this would weaken the analogy with the original fake-barn case. However, none of the people who took the study, or who saw the vignette in conference presentations—and not even the native speakers among our proof-readers—seem to have noticed the mistake (or, more cautiously, none of them felt the need to point it out to us). Moreover, our results for *Dollar* are almost identical to other results for fake-barn style cases (see above and below). Therefore, it seems reasonable to assume that the participants in our experiment simply accommodated the mistake in line with the intended reading of the case description, which is clearly suggested by the final sentence of our vignette—a sentence that makes little sense on the assumption that only the one dollar bills in the scenario are fake. In all fairness, however, it would be useful to have a follow-up study that confirms the insignificance of our mistaken phrasing (thanks to Joshua Alexander and Aaron Meskin for discussion).
A little girl likes to play a game with flipping a coin. She sometimes gets a "special feeling" that the next flip will come out heads. When she gets this "special feeling", she is right about half the time, and wrong about half the time. Just before the next flip, the little girl gets this "special feeling", and the feeling leads her to believe that the coin will land heads. She flips the coin, and it does land heads. But she has inadvertently flipped the coin out of the window, and therefore she cannot—and will not—see the result.

**How much do you disagree or agree with the following claim:**

*The little girl knew that the coin was going to land heads.*

The main point of *Clear Non-Knowledge 2* was to test a striking result concerning lay people's intuitions about *Clear Non-Knowledge 1* from our first experiment, namely, that a significant number of lay subjects (33.8%) did in fact ascribe knowledge in this case. Our main hypothesis for explaining this finding was that lay people might be susceptible to something like a “success bias”, i.e., that they might be influenced by the fact that, after flipping the coin, the little girl learns that her "special feeling" about the outcome of the coin flip turned out to be correct (see, e.g., Baron & Hershey, 1988). A related phenomenon that might explain why a third of all lay subjects ascribed knowledge to the little girl is called “protagonist projection” and describes people’s tendency to project themselves into the point of view of the protagonist (see, e.g., Holton, 1997; Nagel, San Juan, & Mar, 2013). To test these hypotheses, we replaced the clear case of non-knowledge from *experiment 1* with a case where the agent does not and, in fact, cannot see the result.

### 3.2.2 Subjects

Overall, the data of 227 subjects were included in our analysis. 50 were identified as expert subjects (mean age was 33 years; 76% were male) and 177 as lay subjects (mean age was 46 years; 53% were male).

The recruiting and classification procedure for expert subjects was the same as in *experiment 1*, with the following exceptions. In order to gain enough expert subjects, we included not only participants who indicated that they have a PhD in philosophy, but also those participants who indicated that they have an MA in philosophy (in addition to having epistemology as one of their areas of specialization or competence). Subjects
who indicated that they had participated in our first experiment were excluded. Out of initially 186 people, 50 met these criteria.

Apart from the fact that we only invited subjects who did not participate in our first experiment, the recruiting and classification procedure for lay subjects was the same as in experiment 1. Out of initially 293 people, 177 met the relevant criteria.

3.2.3 Design and Procedure

The design and procedure was the same as in experiment 1, i.e., participants saw the first three cases (Exam, Sculpture, and Dollar) in a randomized order while the cases Clear Non-Knowledge 2 (see above) and Clear Knowledge 1 (same as in experiment 1) were always presented in fourth and fifth position, respectively.

3.2.4 Results

The results of our second experiment are summarized in Figure 3. A mixed ANOVA (between subjects: lay subjects vs. expert subjects; within subject: the five cases) showed a very significant interaction of subjects and cases, indicating again that the response pattern of lay subjects and expert subjects differed strongly over the five cases, F(4, 900) = 13.80, p < .001, η² = .06. There was a significant difference in the average time spent to complete the survey between lay people (267 seconds) and experts (370 seconds), F(1, 224) = 7.54, p > .01, η² = .03.\footnote{One expert subject was excluded from this analysis because she or he spent over 1,000 seconds longer (2,554 seconds) than all other subjects.}
Figure 3. Mean ratings for the five cases divided in lay and expert judgments. The scale ranged from 1 (‘strongly disagree’ with the claim that the agent knew the relevant proposition) to 6 (‘strongly agree’). Error bars represent 95% confidence intervals.

A more fine-grained analysis revealed that, as in experiment 1, the judgments of lay subjects and expert subjects differed significantly for three of the five cases, with the experts’ judgments being closer to the evaluation of corresponding cases in the epistemological literature. In case of Dollar, lay subjects (M = 5.10, SD = 1.31) and expert subjects (M = 3.84, SD = 1.73) both tended to agree with the claim that the waiter knows that the restaurant manager has given him ten dollars. The level of agreement for this case, which corresponds to cases that are regarded as cases of non-knowledge in the epistemological literature, was significantly lower for expert subjects, F(1, 225) = 28.97, p < .001, η² = .11. Interestingly, and in contrast to the textbook consensus, the majority of expert subjects (60%, cf. footnote 9) evaluated Dollar as a case of knowledge. For Sculpture, we obtained the same result pattern as in our first experiment. Lay subjects’ (M = 4.94, SD = 1.35) agreed significantly stronger with the claim that the Rodin expert knows that the sculpture in front of her is one of Rodin’s famous marble sculptures than expert subjects (M = 3.72, SD = 1.86), F(1, 225) = 26.83, p < .001, η² = .11. Again, expert subjects were more inclined to agree (56%) with this claim than one would expect from the consensus in the epistemological literature, which is an almost exact replication of the surprising result from our first experiment. For Clear Non-Knowledge 2, the result pattern was similar to the one obtained for the corresponding case in our first experiment. Expert subjects (M = 1.40, SD = 0.67) disagreed significantly stronger than lay subjects (M = 2.94, SD = 1.63) with the claim that the little girl knew that the coin was going to land heads, F(1, 225) = 42.52, p < .001, η² = .16. Again, a surprisingly high number of lay subjects (40%) agreed with this claim. The result speaks against our hypothesis that the high number of agreeing lay subjects in our first experiment was due to the fact that the protagonist of the case comes to know the outcome of her coin flip. We discuss this surprising finding below (see Discussion).

For the cases Exam and Clear Knowledge 1, there was no significant difference between lay subjects and expert subjects (p = .73 and p = .19, respectively). Interestingly and again in tension with the textbook consensus, the majority of expert subjects (68%) considered Exam as a case of knowledge.
Moreover, the result pattern of those experts who indicated that they have a PhD versus those experts who indicated that their highest degree in philosophy was an MA did not differ significantly, F(4, 192) = 1.00, p = .40 (for the interaction of degree * cases). In addition, we compared our expert subjects with those subjects who hold at least an MA in philosophy but indicated that they have no competence in epistemology (N=32, not included in previous analyses). The response pattern of these two groups differed significantly, F(4, 320) = 3.18, p = .01, η² = .04 (for the interaction of epistemological expertise * cases). At the level of individual cases, however, the difference between these two groups did not turn out to be significant.

4 Discussion

What is the upshot of our experiments with respect to the three motivating ideas explained above, i.e., (1) to compare expert with lay intuitions, (2) to measure their quality against the relevant textbook consensus, and (3) to check whether expert intuitions are in line with the textbook consensus?

First, there is indeed a significant expert-lay difference in the majority of the tested cases. This concerns the cases Monitor, Sculpture, and Clear Non-Knowledge 1 in experiment 1, and the cases Dollar, Sculpture, and Clear Non-Knowledge 2 in experiment 2. In contrast, we found no significant expert-lay difference for Painting (experiment 1), Exam (experiment 2), and both instances of Clear Knowledge 1. In the latter case, expert-lay agreement may not come as a big surprise, but expert-lay agreement on the more probing cases Painting and Exam suggests that there is no straightforward general pattern of expert-lay agreement or disagreement. Overall, however, the expert-lay differences that we found are prima facie encouraging for proponents of the expertise defense.

Second, when the quality of expert and lay intuitions is measured against the relevant textbook consensus, the results are again prima facie favorable for the expertise defense. In all cases where we found a significant expert-lay difference the mean expert ratings were closer to the relevant consensus in the epistemological literature. In case of Sculpture, for example, one would expect that the mean expert rating is significantly closer to the 'disagree'-end of the scale than that of lay people, given that Sculpture is inspired by fake-barn cases, which are widely considered as cases
of non-knowledge in the epistemological literature (see above). And indeed, that is what we found in both of our experiments.

On closer inspection, however, the distribution pattern of expert responses over the rating scale raises certain doubts about their intuitive expertise, in particular with respect to the fake-barn-style cases Sculpture and Dollar. In both cases, we found the same striking pattern of expert responses. First, the slight majority of expert ratings fall into the ‘agree’-half of the scale. Second, the two middle options (‘mildly disagree’ and ‘mildly disagree’) were not the favorite options, indicating that most of our expert participants were pretty confident in their judgments (see Figure 2 for a typical distribution pattern of expert judgments about these cases). The fact that most of our expert participants tended to classify these cases as instances of knowledge raises questions of its own, which will be discussed below. But even apart from this surprising finding, the fact that the experts’ intuitions are divided in the way described does not really count in favor of their intuitive expertise, at least when we consider them as a group. Thus, when one takes distribution patterns into account, the initial impression that our experiments favor the expertise defense already begins to fade.

A further striking result was the highly significant expert-lay difference in case of Clear Non-Knowledge 1, which was used as a control case in earlier studies, with the explicit purpose of weeding out participants who did not understand the probes correctly (see, e.g., Swain et al., 2008). The degree of expert-lay differences for this case, both in mean ratings and absolute numbers, is about the strongest evidence for the superiority of expert intuitions that one can get from our experiments. Given that the case of someone who correctly guesses the outcome of a random coin flip is such a clear non-instance of knowledge, the high number of lay people who tend to ascribe knowledge in this case calls for an explanation.

The first hypothesis that we tested in experiment 2 was the presence of a “success bias” in lay people, i.e., the tendency to ascribe knowledge to someone who learns, on independent grounds, that her target belief is true—irrespective of how problematically her belief was initially acquired. In order to suppress this tendency, we added the following sentence to our follow-up vignette Clear Non-Knowledge 2: ‘But she has inadvertently flipped the coin out of the window, and therefore she cannot—and will not—see the result.’ Since we found the same expert lay-difference as in experiment 1,

---

19 Thanks to Shen-yi Liao and Hannes Rusch for pressing the importance of distribution patterns.
our hypothesis of a "success bias" in lay people was clearly disconfirmed. The same holds for the hypothesis of "protagonist projection" in lay people, i.e., the tendency to project oneself into the point of view of the protagonist, because in Clear Non-Knowledge 2 the outcome of the coin flip is not accessible from the point of view of the little girl.

The second hypothesis, which we tested in an additional follow-up experiment with lay participants, turns on a subtle difference between our version of the coin flip case and the standard formulation that was used in earlier studies. In our vignettes Clear Non-Knowledge 1 and 2, we referred to the protagonist of the case as 'a little girl', while in earlier studies the protagonist was simply called 'Dave', and one might speculate that lay people are more willing to ascribe knowledge to little children than to normal human adults. To test this hypothesis, we confronted lay people with six versions of the coin flip case, where the only difference was that—in the five additional cases—the phrase ‘a little girl’ was replaced with ‘a little boy’, ‘a woman’, ‘a man’, ‘an old woman’, and ‘an old man’, respectively (pronouns were adjusted accordingly). We found no significant difference in lay responses to those cases, but we could again replicate the high level of lay agreement for this clear non-instance of knowledge.

A third hypothesis trades on Alvin Goldman’s claim that the English word 'know' is ambiguous between a weak and a strong reading (Goldman, 1999, 2001). According to Goldman, ‘know’ in the weak sense is more or less synonymous with ‘true belief’, while ‘know’ in the strong sense also entails justification or other epistemic properties (see also Hawthorne, 2002). In light of Goldman’s claim about the meaning of ‘know’, one might hypothesize that a significant number of lay people interpret occurrences of ‘know’ in the coin flip case in the weak sense of ‘true belief’, given that they were not in any way prompted to focus on the strong sense of ‘know’. Epistemological experts, on the other hand, are trained to interpret ‘know’ in the strong sense in the context of philosophical thought experiments about knowledge, and so they should not be inclined to the weak reading of ‘know’ in such contexts. Our findings are highly consistent with this hypothesis. Moreover, if one considers the distribution of lay responses in all of our

---

20 Thanks to Jonathan Weinberg for the suggestion.
21 The lay participants for this additional experiment were selected in the same way as in our two main experiments (see above). 178 lay subjects were included in our analysis. The main difference in the design of the experiment was that each participant only responded to one randomly selected version of the six versions of the coin flip case, which helped to prevent order effects (thanks to Shen-yi Liao for suggesting that our original findings might involve order effects). The mean ratings for the six cases did not differ significantly, F(5, 172) = .41, p = .84, with means varying between 2.45 and 2.92 (overall = 2.74) and percentages of agreement between 24% and 46% (overall = 36%). So the additional experiment clearly disconfirms our second hypothesis.
experiments with coin flip cases, one finds that there is a peak at the ‘strongly disagree’-end of the scale, and another peak around ‘mildly agree’. It seems natural to assume that the participants in the former peak read ‘know’ in the strong sense, while participants in the second peak read ‘know’ in the weak sense, although with some hesitation (maybe because the coin flip case is such a clear non-instance of ‘know’ in the strong sense). Nevertheless, confirming the ambiguity hypothesis would require additional experiments that target the idea of an ambiguity between a weak and a strong sense of ‘know’ more directly. In the meantime, it seems advisable to suspend the use of coin flip cases as control cases in experiments with lay people.

Irrespective of what the best explanation for the striking rates of lay agreement for Clear Non-Knowledge 1 and 2 and other versions of the coin flip case might be, it should be pointed out that all considered hypotheses favor expert over lay judgments with respect to such cases.

Third, are expert intuitions in line with the consensus in the epistemological literature? In experiment 1, mean expert ratings were in line with the relevant consensus for Monitor, and for Clear Non-Knowledge 1 and Clear Knowledge 1 (the results for those clear cases were also replicated in experiment 2). With respect to the remaining cases, Painting, Sculpture, Exam, and Dollar, our two experiments suggest, however, that the intuitions of expert epistemologists are not in line with the consensus in the literature. Even if we put Painting aside, which the literature would classify as an unclear case, we still get the striking result that three cases that the epistemological literature would regard as non-instances of knowledge are nevertheless classified as knowledge by our experts. What might explain this striking mismatch between the intuitions of individual expert epistemologists and the corresponding disciplinary consensus?

A first and somewhat friendly suggestion would be that only the intuitions of epistemological “super-experts” really count, who might roughly be understood as those philosophers who publish papers on knowledge in philosophical top-journals—and it is arguably those “super-experts” who shape the relevant disciplinary consensus.22 This hypothesis is tentatively disconfirmed by the fact that we found no significant difference between various groups of “first-rate” vs. “second-rate” experts (e.g., epistemology as area of specialization vs. area of competence, or PhD vs. MA in philosophy—see above). So, unless the development of intuitive expertise is highly discontinuous, the hypothesis

---

22 Thanks to Eddy Nahmias for this suggestion (in a comment on our post “Expert Intuitions About Knowledge” on the Experimental Philosophy blog; cf. Horvath & Wiegmann (2013)).
of intuitive “super-expertise” in epistemology is rendered prima facie implausible by our results. Yet even if true, this relatively friendly hypothesis would still require far-reaching modifications of our actual philosophical practice, because it suggests that everyone should defer to a really small group of “elite intuiters” when it comes to intuitions about thought experiments on knowledge.

A second and less friendly suggestion might be that contemporary epistemology falls squarely within the “puzzle-solving” paradigm of philosophy, like most analytic philosophy in general, according to some authors (cf. Rorty, 2007). And of course, the more difficult the relevant puzzles are, the more interesting the activity of puzzle-solving becomes. Therefore, one might hypothesize that epistemologists are overly liberal in accepting fake-barn-style or lottery cases as non-instances of knowledge—despite their intuitive tendency to the contrary—because this makes the puzzle of analyzing knowledge a lot harder and thus a lot more interesting to solve. It goes without saying that this hypothesis, if correct, would significantly undermine the current practice of professional epistemology.

A third and even less friendly suggestion might be that the disciplinary consensus in epistemology with respect to, e.g., fake-barn or lottery cases is simply the result of some kind of social conformity. Maybe certain epistemological “opinion leaders”, such as Alvin Goldman in case of fake-barn cases, initially classify the relevant cases as non-instances of knowledge. The bulk of epistemologists then simply follow the intuitive evaluations of such authorities, for example, in order to be taken seriously in professional discussions, or to get their papers published in leading journals (cf. Turri, forthcoming; Weinberg, 2007, p. 337).

Our results do not actually tell us which of these hypotheses, if any, provides the best explanation for the mismatch between expert intuitions and the disciplinary consensus that we found in experiments 1 and 2. However, none of these explanations would provide a vindication for our actual disciplinary practice, for both explanations appeal to irrelevant factors in the experimental restrictionist sense, that is, to factors that are irrelevant to the truth of the relevant first-order claims about knowledge. Our experiments therefore reveal a surprising problem with expert intuitions about knowledge that is difficult to explain without unsettling the expertise defense. Despite the fact that our experiments provide some evidence that favors the expertise defense, they also give rise to a challenging new problem—and ultimately raise more questions than they answer.
5 Conclusion

In this paper, we first distinguished direct and indirect strategies for assessing the claim that professional philosophers are intuitive experts for evaluating philosophical thought experiment cases. A problem for indirect strategies is that the paradigms of empirical research on intuitive expertise are very unlike thought-experimenting philosophers, because the former, unlike the latter, primarily engage in causal predictions. In contrast, direct strategies that are based on experimental data about philosophers’ intuitions are epistemically less risky and dialectically more effective. In order to advance the debate about direct arguments for assessing the expertise defense, we conducted two new experiments with expert epistemologists and lay people.

In the first experiment, we tested a number of unfamiliar thought experiment cases about knowledge with expert epistemologists and lay people. In the majority of the tested cases, the experts’ intuitions were indeed significantly different from those of lay people, and the difference was always in the direction that one would expect on the basis of the relevant textbook consensus. However, concerning the fake-barn-style vignette Sculpture we found two things that challenge the intuitive expertise of our epistemological experts: first, on average, they tended to ascribe knowledge in this case, and second, their overall response pattern exhibits two opposing peaks. We also found a striking difference between expert and lay intuitions concerning the coin flip case Clear Non-Knowledge 1. A significant number of lay people—unlike the experts—tended to ascribe knowledge in this case.

The point of our second experiment was to test and corroborate the more surprising results of our first experiment. To this end, we tested two further non-instances of knowledge: Dollar, another fake-barn-style case, and Exam, a case that is inspired by lottery cases. On average, the epistemological experts again ascribed knowledge in these cases, and again with the same polarized distribution pattern.

To explain the high rate of lay knowledge ascriptions in the coin flip case, we also tested the hypothesis of a “success bias” in lay people, which was not confirmed by our second experiment. In an additional experiment with lay people, we tested the further hypothesis that lay people might be more willing to ascribe knowledge to little children than to adults, which was disconfirmed as well. The high rate of lay approval in coin flip
cases therefore remains puzzling. The most promising explanation might be in terms of an ambiguity between a weak and a strong sense of 'know'.

Our most troubling result from the viewpoint of the expertise defense is the striking mismatch between the intuitions of our expert subjects and the relevant consensus in the epistemological literature. This finding suggests that there might be something wrong with philosophy as a discipline, given that it apparently fails to map the intuitions of its expert practitioners onto a disciplinary consensus in the right kind of way. Friendly explanations of this mismatch are hard to come by, and so our experiments indicate a problem that has not been acknowledged so far—and that challenges the expertise defense in unexpected and troublesome ways.

Acknowledgments We would like to thank Joshua Alexander, James Andow, Steve Clark, Jens Kipper, Shen-yi Liao, Aaron Meskin, Jonas Nagel, Eddy Nahmias, Jennifer Nado, Martin Peterson, Hannes Rusch, Joshua Sheperd, Jonathan Weinberg, and three anonymous reviewers for very helpful comments and discussions. Thanks also to our audiences at the lecture series Gedankenexperimente—Kann man aus dem Lehnsuhl die Welt erforschen? at Universität Zürich in May 2014, the X-Phi Workshop Vienna at Universität Wien in June 2014, the conference Möglichkeiten und Grenzen der Experimentellen Philosophie at Philips Universitat Marburg in June 2014, and the conference Investigating the Nature and our Understanding of Causality, Morality, Language, Mind, and Aesthetics—the inaugural meeting of the Experimental Philosophy Group Germany—at Ruhr-Universität Bochum in November 2015.

References


http://doi.org/10.1111/nous.12110


Vaesen, K., Peterson, M., & Van Bezooijen, B. (2013). The Reliability of Armchair


