CrossMark

"Nobody would really talk that way!": the critical project in contemporary ordinary language philosophy

Nat Hansen[1]

Abstract This paper defends a challenge, inspired by arguments drawn from contemporary ordinary language philosophy and grounded in experimental data, to certain forms of standard philosophical practice. The challenge is inspired by contemporary philosophers who describe themselves as practicing "ordinary language philosophy". Contemporary ordinary language philosophy can be divided into constructive and critical approaches. The critical approach to contemporary ordinary language philosophy has been forcefully developed by Avner Baz, who attempts to show that a substantial chunk of contemporary philosophy is fundamentally misguided. I describe Baz's project and argue that while there is reason to be skeptical of its radical conclusion, it conveys an important truth about discontinuities between ordinary uses of philosophically significant expressions ("know", e.g.) and their use in philosophical thought experiments. I discuss some evidence from experimental psychology and behavioral economics indicating that there is a risk of overlooking important aspects of meaning or misinterpreting experimental results by focusing only on abstract experimental scenarios, rather than employing more diverse and more ecologically valid experimental designs. I conclude by presenting a revised version of the critical argument from ordinary language.

✉ Nat Hansen
    n.d.hansen@reading.ac.uk

1   Department of Philosophy, University of Reading, Reading RG6 6AA, UK

🗍 Springer

We all use the word knowledge. *But as soon as a philosopher asks what knowledge is, we find ourselves in a terrible mess.*

Rogers Albritton

1 Constructive and critical projects in ordinary language philosophy

Ordinary language philosophy involves both *constructive* and *critical* projects. The constructive project consists of observations about how philosophically significant expressions are ordinarily used and uses those observations to support conclusions about non-linguistic aspects of the world. Austin (1957, p. 8) describes the methodology of ordinary language philosophy as follows:

> When we examine what we should say when, what words we should use in what situations, we are looking again not merely at words (or 'meanings', whatever they may be) but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena.

The constructive project is exemplified by J. L. Austin's attempt to clarify the problems of "Freedom" and "Responsibility" through an investigation of the subtly different ways we use the expressions "by mistake", "by accident", "intentionally" and "deliberately" (Austin 1957). Austin's approach to the problem of knowledge of other minds through the examination of parallels between the use of "I know" and "I promise" (Austin 1946), especially as that approach has been reconstructed by Lawlor (2013), is another example of the constructive project.

Contemporary adherents of the constructive project include both armchair and experimental philosophers. For example, contextualists about knowledge, like DeRose (2009) and Ludlow (2005), draw conclusions about the nature of knowledge (at least partly) on the basis of observations about the ordinary use of the word "knows", and experimental philosophers use empirical methods developed in the cognitive sciences to investigate philosophically significant concepts (knowledge, e.g.), and, assuming the concepts are veridically applied, the parts of reality that those concepts represent.[1] Pinillos (2012), for example, begins an experimental investigation of theories of knowledge by saying:

> The central methodological assumption I will be adopting is that information about the behavior and mental states of ordinary people, including careful observation of their deployment of the word 'knowledge', can be relevant in assessing [competing theories of knowledge].
>
> I do not believe that this is an exotic assumption.

---

[1] For a challenge to the constructive epistemological project in contemporary ordinary language philosophy, see Kukla (2015).

While the assumptions underlying the constructive project of ordinary language philosophy may not be "exotic", the *critical* project, in contrast, has not found many advocates in contemporary philosophy.[2] The critical project in ordinary language philosophy involves the charge that philosophers produce "nonsense" or are led to produce intractable philosophical problems when they depart from or ignore the way language is ordinarily used. Classic examples of the critical project include Wittgenstein's (1969, Sect. 10) remark that when one is sitting at a sick man's bedside, looking attentively into his face, neither the question "I know that a sick man is lying here?" nor the assertion "I don't know that there is a sick man lying here" makes sense and Austin's (1962, p. 15) argument that the word "directly" has been "stretched" by philosophers in discussions of perception to the point that it has become "meaningless".[3]

One of the rare contemporary advocates of the critical project is Baz (2012a, b, 2014, 2015, 2016, 2018), who argues that "the prevailing program" in contemporary analytic philosophy is fundamentally flawed, and that we don't actually understand the content of what we are being asked when confronted with philosophical thought experiments and asked to judge whether or not someone knows some proposition, or whether some knowledge ascription is true or false. Examples of such thought experiments include Gettier cases and contextualist "bank" scenarios. Because we don't understand what we are being asked in such thought experiments, any way we respond will be "unsystematic" Baz (2012b, p. 46), and will provide only an illusory foundation for philosophical theories. The strategy of this paper is to develop a less radical and more defensible version of Baz's argument from ordinary language. In the next section, I spell out Baz's radical version of the critical project of ordinary language philosophy, in Sect. 3 I raise objections to Baz's version, and in Sect. 4 I discuss experiments that support the revised argument from ordinary language.

2 Baz's challenge to "the prevailing program"

Baz criticizes a philosophical method that he says is common in "the mainstream of analytic philosophy". The method aims to develop or test philosophical theories of some subject matter by asking what Baz calls "the theorist's question", which asks for judgments whether or not "our concept of x, or [the expression] 'x', applies to some particular case, actual or imagined" (Baz 2012a, p. 1). For example, philosophers have investigated the concept of *knowledge* by asking whether or not we have intuitions that the concept applies in certain imagined situations (Gettier scenarios, driving through fake barn county, Mr. Truetemp's miraculously reliable beliefs about the temperature, and so on). Baz calls "the research program that takes answers to the theorist's question as its primary data 'the prevailing program'" (p. 1).

---

[2] For further discussion of the constructive and critical projects in contemporary ordinary language philosophy, see Hansen (2014a).

[3] For a different understanding of Austin's critical method in *Sense and Sensibilia*, see Fischer (2014), which is another example of a contemporary defense of the critical project in ordinary language philosophy.

It is controversial to describe this particular methodology as the "prevailing program", but there is little doubt that it is an influential aspect of contemporary philosophy. In particular, experimental philosophers have turned the traditional armchair method of eliciting judgments about scenarios into a branch of cognitive science by running formal experiments. These experiments ask ordinary experimental participants to make judgments about various philosophically significant expressions or concepts and using those judgments as evidence for or against philosophical theories.[4]

Baz is not alone in wanting to challenge the "prevailing program". Advocates of the "negative program" in experimental philosophy (Machery et al. 2004; Mallon et al. 2009; Weinberg et al. 2001) have criticized certain adherents of the prevailing program for assuming that the way in which a small subset of human beings apply a concept reveals something about the concept as such. And Cummins (1998) has challenged the prevailing program on the grounds that there is no way of "calibrating" the intuitions it relies on. That is, there is no independent means of determining whether or not they reliably track what they are supposed to track.

Baz's immediate target is a particular defense of the "prevailing program" against these recent challenges. Baz focuses on the defense of the "prevailing program" offered by Williamson (2004, 2005, 2007). Williamson denies that what goes on when philosophers ask whether a concept x applies to some imagined or real situation should involve eliciting *intuitions* as to whether or not the concept applies, where those intuitions are *evidence* that the concept applies or does not. That kind of approach both invites embarrassing investigations into whether or not philosophers' intuitions are widely shared and into how we could know that they are reliable indications of the subject matter under investigation, and it unnecessarily psychologizes the evidence available to philosophers. According to Williamson, the question whether a concept x applies to a particular situation can be answered by using our everyday capacity to apply concepts to actual and counterfactual situations (Williamson 2005, p. 12; Williamson 2007, p. 188). Insofar as that everyday capacity is reliable, the application of concepts to cases in philosophy should be reliable as well.[5] The prevailing program

---

[4] For surveys of just a small sample of the quickly growing experimental literature, see Alexander (2012), Hansen (2015), Knobe (2012), and Pinillos (2016).

[5] Some recent experimental work problematizes the idea that the everyday conceptual capacities are reliable when applied to certain philosophical thought experiments. Gerken and Beebe (2016), for example, propose that contrast effects that appear in knowledge scenarios are best accounted for in terms of cognitive biases that affect what participants process when reading the scenarios used in the study of contrast effects, and Fischer and Engelhardt (2016) argue that participants' willingness to make inferences characteristic of the "argument from illusion" can be explained in terms of stereotypical inferences generated by processing certain verbs of perception. These explanatory projects endorse a form of the "claim of continuity", in that they hold that the same cognitive processes are at work in philosophical knowledge ascription cases as are at work in cases of non-philosophical cognition, while at the same time denying that the continuity ensures the reliability of responses to philosophical thought experiments. Baz's radical anti-continuity argument (to be discussed below), if successful, would undercut the motivation for these explanatory projects because the questions posed in philosophical thought experiments would fail to make sense, and so there would be no way of reliably (or unreliably) responding to them. Thanks to an anonymous referee for asking about the relation between this recent experimental work and Baz's challenge to the "claim of continuity".

can then proceed to answer the theorist's question by simply reflecting on whether or not a concept of interest applies to particular actual or counterfactual situations.

Baz criticizes Williamson's "continuity defense" of the prevailing program for assuming that "what we are invited to do when we are invited (or invite ourselves) to answer the theorist's question is not essentially different from what we do when, outside philosophy, we judge that, for example, someone knows or does not know this or that" (Baz 2012a, p. 3). Focusing on "know that", and the concept *knowledge*, Baz argues that the theorist's question "is *fundamentally* different from any question to which we might need to attend as part of our everyday employment of these expressions" (Baz 2012a, p. 4). If the theorist's question is fundamentally different from everyday questions, then Williamson's defense of the prevailing program, which ties the reliability of our answers to the theorist's question to the reliability of our everyday capacity to apply concepts to encountered situations, fails. Baz takes the final sentence in the following Gettier scenario, from Weinberg et al. (2001, p. 443), as an exemplar of a "theorist's question":

> Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. *Does Bob really know that Jill drives an American car, or does he only believe it?*

Baz maintains that it is a "fundamental assumption" of the "prevailing program" that competent ordinary speakers of English (or whatever language the scenario is written in) who read this scenario understand the question that it concludes with, and are able to give it a meaningful answer. He wants to challenge that assumption, he says, "by way of a form of ordinary language philosophy" (Baz 2015, p. 4).

Baz summarizes his ordinary language procedure for challenging the "fundamental assumption" as follows (pp. 4–5):

> Take some version of the theorist's question—by which I mean, the form of words in which his question is couched—and ask how it might reasonably be understood in the course of everyday discourse, with respect to a case such as the one described by the philosopher. One thing that would then emerge is that, depending on the circumstances in which it arose, there are any number of different senses the similarly worded but non-merely-theoretical question could have—different ways the theorist's words would, or could, reasonably be understood, depending on the context in which they were uttered or considered, even though the case under consideration remained the same. That would show that, contrary to the fundamental assumption…the words and case by themselves do not suffice for fixing the theorist's question with a determinate sense, and a correct answer. In other words, it would show…that the theorist, in raising his question apart from any context that would fix his words with a determinate sense, has failed to raise a clear question.

The argumentative core of Baz's challenge to the fundamental assumption consists in five attempts to show how the theorist's question (in this case, "Does Bob really know that Jill drives an American car, or does he only believe it?", asked of the Gettier scenario from Weinberg et al. 2001 described above) might matter in a non-philosophical context (Baz 2012b, pp. 108–115). Baz argues that all of these attempts fail, leaving us without evidence that the theorist's question might naturally arise in a non-philosophical context. The burden is then on the defender of the "prevailing program" to defend the continuity of the philosophical question with ordinary questions about knowledge. I'll summarize each of the attempts and Baz's reasons for thinking that they fail.

Attempt #1: If Bob knows that Jill drives an American car, then he will be in a position to *assure* others that she drives an American car. Maybe we care about whether or not Bob is in such a position.

Reply: Given that it is stipulated in the Gettier scenario that Jill drives an American car (a Pontiac), there is no reason we, or anyone else who knows as much about the case as we do, would need assurance from Bob that Jill drives an American car. So it's not clear what point (other than the purely theoretical point of finding out what knowledge is) there would be in asking the question whether Bob really knows, or merely believes, that Jill drives an American car.

Attempt #2: Suppose some third party ("Agent") needs to know whether Jill drives an American car. Agent might wonder whether she can count on Bob's assurance that Jill does drive an American car. That would give the question "Does Bob know that Jill drives an American car, or does he merely believe it?" significance in an ordinary context.

Reply: There are two possible ways of understanding Agent's question about Bob: either Agent knows the basis for Bob's assurance and can assess it, or she does not. If she does not know the basis for Bob's assurance, or she's not in a position to assess it, then her question is not the theorist's question about Bob. The theorist's question is whether Bob's evidence is "good enough" for him to count as knowing, given that Jill does in fact drive an American car. If Agent does know the basis for Bob's assurance and can assess it, and doesn't doubt its truth, then her question is whether the fact that until recently Jill has driven a Buick gives her sufficient assurance that Jill is currently driving an American car. But that is not the same as the theorist's question about whether Bob knows that Jill drives an American car.

Attempt #3/4: Imagine that another person ("Judge") is aware of all of the facts of the Gettier scenario, and his job is to assess whether Bob was in a good enough position to assure Agent that Jill drives an American car. Imagine that Jill is an American politician, Agent is her press secretary, and Bob is Jill's personal assistant. If Jill is seen driving a foreign car, her enraged constituents will vote her out of office and Agent (the press secretary) will lose her job. One of Bob's responsibilities is to ensure that Jill is always seen driving an American car; if he fails to do so, that will have negative

consequences for both Jill and Agent.[6] Judge's question, "Does Bob really know…" is then a question about whether Bob is being sufficiently epistemically vigilant in carrying out his job, given the high stakes.

Reply: The point of Judge's question still isn't the same as the point of the theorist's question. Judge's question concerns Bob's epistemic responsibility, so "Judge must put himself in *Bob's* position if he is to judge him competently" (p. 111). But from Bob's perspective, the situation is not a Gettier scenario, so the question does not come to the same thing as the theorist's question. If the point of the question "Does Bob really know…" is instead simply whether Bob has been doing everything he should be doing with regard to keeping track of what Jill is driving, that too is a different question than the theorist's question, the point of which is just to investigate whether or not Bob *knows*.

Attempt #5: The question "Does Bob *know*…" is just the question whether Bob has a piece of information that the questioner already possesses; whether Bob is aware that Jill drives an American car. Here is an example of this kind of use of "Does [he] know…", drawn from the Corpus of Contemporary American English:

> SARA-HAINES: Does he **know** you sneak off in the middle of the night?
> SUSIE-ESSMAN: Well, when he turns around and goes like this and I'm not there. And, and you're not there? Okay. So, he, he knows now. (Inaudible).[7]

Reply: On this reading of the question "Does Bob know that Jill drives an American car", it would amount to a question about whether Bob is aware that Jill drives an American car, to which the answer is clearly yes—he would not find it informative to be told that she drives an American car. (He already knows that, in the relevant ordinary sense of "knows".) What Bob is not aware of is that Jill drives a Pontiac, not a Buick. The point of the question "Does Bob know that Jill drives an American car?", understood in this way (about what Bob is aware of) is not the same as the point of asking the theorist's question, which concerns whether Bob's justification, plus the truth of his belief that Jill drives an American car, is sufficient to count as knowledge.

Assuming that there isn't an example of the question "Does Bob really know…" in ordinary conversation that Baz has overlooked, what is the upshot of this series of failed attempts to associate the theorist's question about knowledge with various everyday questions about knowledge? Here is Baz's (2012b, pp. 115–117) account of what is going on:

> My aim is to bring out the anomalousness of her question and thereby to raise doubts about the presumed significance of the answers to it that she and others might give.…

---

[6] I am fleshing out Baz's original case in the spirit of the following remark: "To make the case more plausible, imagine that a great deal is at stake for Agent in whether or not Jill drives an American car; imagine that Judge knows this; imagine that Bob knows this as well; and imagine that Judge knows that Bob knows this" (p. 111).

[7] Date: 2015 (15/11/16); Title: SUSIE ESSMAN; "LATE NIGHT JOY"; Source: SPOK[EN]: ABC.

> In considering each of the different [everyday encounters with the question "Does Bob really know…"], we saw that the question that the person encountering Bob would naturally ask herself…is importantly different from the question that the theorist has wanted, and taken himself, to be asking. What answering the everyday question would normally involve and require, in each of the different cases, is nothing like what answering the theorist's question involves and requires.…
>
> There is good reason to suspect that no question that may naturally arise in the everyday [sic] would come to anything like the theorist's question.

Baz is not alone in observing a disconnection between the "theorist's question" and everyday questions about knowledge. For example, Bach (2005, pp. 62–63) observes that contextualists about knowledge ascriptions are not justified in treating their responses to the "theorist's question" (whether someone knows something in a particular context) as representative of ordinary uses of "knows", because

> …outside of epistemology, when we consider whether somebody knows something, we are mainly interested in whether the person has the information, not in whether the person's belief rises to the level of knowledge. Ordinarily we do not already assume that they have a true belief and just focus on whether or not their epistemic position suffices for knowing. Similarly, when we say that someone does not know something, typically we mean that they don't have the information.

(Bach is invoking the ordinary sense of "Does he know…" that appears in Baz's Attempt #5, above.)

If the "theorist's question" is indeed fundamentally different from "everyday" questions, then any answers that the philosopher receives to her question will not help answer questions about everyday uses of expressions (and vice versa). That would be a serious problem for defenders of the "claim of continuity" (like Williamson) who take responses to the theorist's question to support or undermine metaphysical theories (of knowledge, for example), as well as experimental philosophers who take answers to the theorist's question to be evidence for or against theories of the meaning of a particular expression used in ordinary thought and talk ("know", for example).[8]

In addition to arguing that the theorist's question could not arise in everyday contexts, Baz argues that we do not even know how to answer the theorist's question, or assess other people's answers to it and therefore seeking answers to it is fundamentally misguided. In order to establish that ambitious conclusion, he argues as follows:

1. "[T]he point of an everyday question guides us in answering it and in assessing our own and other people's answers".[9]

---

[8] For different worries about the differences between thought experiments as they are employed in philosophy and ordinary judgments, see Machery (2011).

[9] Deutsch (2015) criticizes Baz's notion of the "point" of a question, and Baz (2015) replies.

2. "[T]he theorist's question has no point, *in the relevant* [everyday] *sense*" (Baz 2012a, p. 327).
3. So it is not surprising that there is substantial disagreement over how to answer the theorist's question, because there is no everyday point to guide answers to the question.

Other philosophers, reflecting on the practice of asking non-philosophers to respond to versions of the theorist's question, have expressed thoughts similar to Baz's first two premises, about the way non-philosophers may have a hard time understanding the theorist's question:

> "…experimental philosophy subjects are ipso facto at a significant disadvantage since it is often a precondition of their participation that they have no idea why anyone would be interested in finding out what the folk think about Gettier scenarios, much less what a Gettier scenario actually is" (Cullen 2010, p. 281).

> "…anyone who, like me, has taken a survey when you didn't have any good feeling for why you were being asked the questions directed at you and so didn't know what to focus on should be able to appreciate how lost some ordinary person, just being asked about these strange cases on some survey, might be" (DeRose 2011, p. 93).

> "…when a person responds to a yes/no survey question (or rates assent on a Likert scale), just what is the conversational context? Who is he or she conversing with, and how do we work out what he or she assumes about the hearer's beliefs? Frankly, this is a baffling task" (Kauppinen 2007, p. 107)

There are therefore two related arguments that Baz is making against the "prevailing program". First, because the "theorist's question" (for example, "Does he know that Jill drives an American car?") lacks any practical "point" or significance, while the "point" or significance of everyday questions guides our answers to such questions, when participants in an experiment give answers to the theorist's question, we shouldn't assume that their answers tell us anything about their competence with the underlying concept that philosophers are interested in investigating. Second, Baz is arguing that because the "theorist's question" lacks an everyday point, the question lacks a determinate sense. Both of these arguments are intended to challenge Williamson's "claim of continuity".

Do those two arguments stand up to scrutiny? In the next section, I'll argue that there is experimental evidence that runs counter to the conclusion of the second argument. The first argument is more difficult to dismiss, however, and I'll show how responding to it requires rethinking how philosophers design both informal ("armchair") and formal experiments.

3 Responding to Baz's second argument against "the prevailing program"

Baz's second, more radical, criticism of "the prevailing program" alleges that there is substantial disagreement over how to answer the theorist's question, and offers a

diagnosis of the source of that disagreement in terms of the fact that the theorist's question lacks a point, in contrast with everyday questions. The most straightforward problem with this argument is that there is not evidence of substantial disagreement about how to respond to Baz's chosen "theorist's question" of a kind that would support Baz's claim that the theorist has "failed to raise a clear question" (Baz 2015, p. 5).

The central piece of empirical evidence that Baz cites in support his claim of substantial disagreement in response to the theorist's question is Weinberg et al. (2001). In that study, Weinberg et al. found that while a majority of Westerners tended to say that Bob "only believes" (and doesn't "really know") that Jill drives an American car in the Gettier scenario described above, that preference was *reversed* when East Asian participants and participants from the Indian sub-continent were asked the same question. That is a striking result, and Weinberg et al. argue that it undermines "a sizeable group of epistemological projects—a group which includes much of what has been done in epistemology in the analytic tradition" (Weinberg et al. 2001, p. 429).

The experimental evidence that has accumulated since the publication of Weinberg et al.'s study, however, has not supported the claim of substantial variation in epistemic intuitions (Turri 2016). There have been several failures to replicate the original finding of cultural variation in epistemic intuitions (Machery et al. 2017; Seyedsayamdost 2015; Turri 2013), including a study using exactly the same experimental materials as the original Weinberg et al. (2001) study but using a substantially larger sample size (Kim and Yuan 2015). And recent investigations have indicated that some variability in response to different Gettier cases is systematically related to epistemically significant features of the cases themselves, such as whether the evidence that the protagonist has for their belief is "authentic" or merely "apparent" (Starmans and Friedman 2012).

Blouw et al. (2017) and Turri et al. (2015) argue that there is in fact no epistemically unified category of "Gettier cases", but five different types of case, ranging from "Gettier-1" cases in which the agent "perceptually detects the truth, and there is a salient but failed threat to the truth of her judgment" (Goldman's (1976) fake barn county example illustrates this type of case), to "Gettier-5" cases in which "the agent fails to detect the truth, but her judgment is nevertheless made true by a state of affairs dissimilar to what she based her belief on" (p. 10) (Gettier's 1963 "Either Jones owns a Ford, or Brown is in Barcelona" case is the paradigm of this latter type) (Blouw et al. 2017, p. 9). Intermediate Gettier cases included scenarios in which:

- (Gettier-2: *detection, similar replacement*) the agent forms a true belief on the basis of "detecting" the relevant truth-maker (forming the belief that there is a pen on a table on the basis of seeing the pen), but then the truth-maker is replaced with a *similar* truth maker (another visually indistinguishable pen, for example),
- (Gettier-3: *detection, dissimilar replacement*) the agent forms a true belief on the basis of "detecting" the relevant truth-maker (she forms the belief that she has a diamond in her pocket on the basis of purchasing a genuine diamond), but the original truth-maker is replaced by a dissimilar truth-maker (a thief steals the one she bought, but there is, unbeknownst to her, another diamond stitched into her pocket),
- (Gettier-4: *no detection, similar replacement*) the agent forms a true belief but fails to "detect" the relevant truth-maker (she forms the belief that she has a diamond in

Table 1 "Really knows" dichotomous response percentages for Experiment 4 (Turri et al. 2015)

N	Control Knowledge	Gettier-1 Failed Threat	Gettier-2 Detect Similar Replace	Gettier-3 Detect Dissimilar Replace	Gettier-4 No Detect Similar Replace	Gettier-5 No Detect Dissimilar Replace	Control Ignorance
728	90%	83%	55%	39%	38%	19%	12%

her pocket on the basis of purchasing a fake diamond, which is then stolen, but her belief is made true by a genuine diamond that is slipped into her pocket without her knowledge).

There were significantly different rates of knowledge attribution in response to the different types of Gettier scenarios, ranging from knowledge attributions that do not significantly differ in rates of knowledge attribution from clear cases of knowledge in response to Goldman-style Gettier-1 scenarios (up to 83% in Turri et al. 2015), down to 19% in Gettier-5 scenarios (with the same structure as Gettier's "Barcelona" case), which do not significantly differ in rates of knowledge attribution from clear cases of non-knowledge.[10] See Table 1 for a summary of relevant results, based on Figure 1 in Turri et al. 2015; triple vertical bars indicate a significant difference in responses.

The wider pattern of responses to different types of Gettier cases reported in Blouw et al. (2017), Starmans and Friedman (2012) and Turri et al. (2015), which include responses to (theoretically) clear cases of knowledge and clear cases of non-knowledge (either cases of false belief, or true beliefs that lack justification) in fact poses a challenge to Baz's contention that the theorist's question (which, in Gettier cases is the question whether the protagonist knows that, e.g., Jill drives an American car) is not "clear" because it lacks a practical point.[11] If the theorist's question lacked a sense, as Baz claims then it should be surprising to see the consistent levels of knowledge-denial in certain kinds of Gettier cases that experimenters have found (around 80%—see Turri 2016, p. 341) as well as the consistent patterns of variation when epistemically significant features of the Gettier cases are varied (see the Appendix for details), and especially the much higher rates of knowledge attribution in theoretically clear cases of knowledge (79–90% in Starmans and Friedman 2012 and Turri et al. 2015) than in theoretically clear cases of non-knowledge (8–14% in Starmans and Friedman 2012

---

[10] Knowledge attributions are "really knows" responses when offered the dichotomous choice between "really knows" and "only believes". Nagel et al. (2013a) challenge Starmans and Friedman's results, and report significantly lower rates of knowledge attribution in Gettier cases than in "standard true belief" cases. But see Starmans and Friedman (2013) for methodological criticisms of Nagel et al.'s study, and Turri et al. (2015) for replications of Starmans and Friedman's key findings.

[11] Starmans and Friedman (2012) used 10-point Likert scale measures of confidence in combination with dichotomous knowledge judgments in all the scenarios they examined, and report consistently high confidence means across all conditions (9.1 out of 10 and 8.6 out of 10, with no significant differences across conditions). They comment: "Also, if participants had been confused in the Gettier cases, they should have given low confidence ratings to their responses, but they did not. Confidence ratings did not differ across conditions, and moreover few participants ever used the lower end of the confidence scale" (p. 280).

and Turri et al. [2015].[12] (All of these experimental studies are described in greater detail in the Appendix.)

Where does this evidence leave Baz's more ambitious argument? Even if we grant him that the theorist's question about whether the protagonist in a Gettier case knows something lacks an everyday "point", there is a substantial body of evidence that does not support the idea that participants fail to understand the content of the question they are posed. If the "theorist's question" in the Gettier cases genuinely lacked sense, then we should find a pattern of responses to versions of the "theorist's question" that indicates that participants are failing to understand the question.[13] But existing experiments do not find such a pattern.[14]

In addition to running into a body of experimental findings that challenge its conclusion, Baz's more ambitious argument also makes a deeper theoretical mistake: it assumes that there is a sharp cut-off between "everyday" questions, which are raised in contexts where there is some practical point to posing the question, and the "theorist's question", which is raised in a context that is stripped of any practical significance (for the participants attempting to answer the question). The assumption is mistaken because the distinction between the "everyday" and the "theoretical" is porous. Purely "semantic" questions come up naturally in everyday conversations, where there is no obvious point to the discussion other than sheer interest in figuring out the meaning of some expression. For example, Niedzielski and Preston ([2000]) includes a collection of 59 recordings of "everyday" or "folk" conversations pertaining to linguistic matters.

---

[12] Turri et al. ([2015], p. 387) notes: "Though comparing results from different experiments is fraught, it is still worth noting the impressive consistency of knowledge attributions in structurally analogous conditions", including the consistently high rates of knowledge attribution in knowledge controls, and low rates in non-knowledge controls.

[13] One might object to the conditional on the following grounds: Participants might not understand the "theorist's question" (because it lacks sense), and yet their responses may not indicate such a failure of understanding because they are responding to a *different* question, which they do understand and are substituting for the theorist's question (see the discussion of "attribute substitution" in Kahneman and Frederick [2002]). This is a possibility, but for it to constitute a convincing response in defense of Baz, it would have to be supplemented with some plausible account of what question is being substituted for the "theorist's question", and such a substitution account would have to be consistent with the pattern of responses observed in Starmans and Friedman ([2012]) and Turri et al. ([2015]) (see the Appendix for discussion). Baz himself ([2012b], p. 124) says that responses to Gettier cases are probably "affected by considerations that do guide us in our competent employment of 'know that'…in certain contexts (but not in others), and in this way is revelatory of an aspect of our concept of propositional knowledge". He proposes that it is the fact that we would hesitate to ascribe knowledge that someone drives an American car in an ordinary context in which it was a possibility that someone's car is stolen and replaced with a different car that explains people's reluctance to ascribe knowledge in Gettier cases. (Thanks to an anonymous referee for bringing this passage to my attention.) This explanation conflicts, however, with the results reported in Turri et al. ([2015]), in which participants are sensitive to differences in the type of evidence that subjects in Gettier cases have. For example, participants generally ascribe knowledge to subjects in Gettier-style scenarios when there is a salient, but failed threat to their perceptual relation to a truth-maker, as in Goldman's "fake barn county" thought experiment. In contrast, participants generally do not ascribe knowledge when a subject forms a belief on the basis of perceiving a truth-maker, but the truth-maker is "disrupted" and replaced with an indistinguishable back-up. If Baz's explanation were correct, participants should refuse to ascribe knowledge to subjects in Gettier cases whenever there is a salient possibility that the subject's belief is false.

[14] Thanks to Wesley Buckwalter for discussion of this point.

Those conversations include everyday discussions about the following questions of meaning:

- Is the word "maturity" associated with "closed-mindedness" or with the ability to do things "wisely" and "correctly"? (pp. 266–267)
- Does a diary consist only of "notes", or can it be "reflective" and "book-like" like a journal?
- Can a "hairdo" be correctly used to describe a man's hair? (p. 267)

These kinds of folk meta-linguistic discussion can lack a practical "point" in the same way that philosophical debates about the meaning of expressions like "knows" can lack a practical point—there may be no practical issue that turns on which way they are settled.[15] And yet the participants in these conversations can come to agree on a particular meaning for an expression. There is no principled reason why a similar conversation about the meaning of "knows" couldn't arise in an "everyday" (non-philosophical) situation.[16] Theoretical investigations of meaning are continuous with these kinds of everyday meta-linguistic conversations.

## 4 The insight in Baz's first argument: the need to diversify experimental contexts

The previous section discussed reasons to reject Baz's more ambitious second argument that the theorist's question is not "clear", and his claim that when we try to answer it we lack "orientation of the kind that is ordinarily provided by a suitable context", because it lacks an everyday "point". Experimental evidence indicates, however, that participants are not responding to the theorist's question (at least in the case of "know" and knowledge) in a way consistent with the question lacking sense.

But what about Baz's first argument, that the point of asking the theorist's question and the point of an identically worded question in an everyday context are different, so the way people respond to the question in one context doesn't necessarily tell us anything about the way they would respond to it in the other? I think that Baz's first argument is indeed an important challenge to standard experimental approaches to investigating the meaning of a term like "knows". I will raise some additional considerations in support of this argument in this section, by considering several experimental case studies, each of which lends weight to Baz's claim that when participants provide answers to the "theorist's question" about "knows", detached from features of ordinary conversation, they may be doing something substantially different than what they ordinarily do when operating with "knows" and the concept of knowledge.

---

[15] Baz (2012b, p. 118) considers these kinds of folk meta-linguistic discussions and argues that they are not genuine versions of the "theorist's question", because in the everyday situations, "a particular context of significant application is normally in place, or at least assumed or imagined". But from the transcripts in Niedzielski and Preston (2000), it looks likely that conversational participants do not always have a "particular context of significant application" in mind when they discuss questions about meaning.

[16] In the conclusion of Baz (2016), he makes a distinction between "harmless" versions of the theorist's question, which occur when "what speakers normally and ordinarily mean by the expression in question is a matter of what worldly item they mean to refer to, and if the nature of the item varies little across different contexts of speech" (p. 80). According to Baz, questions about knowledge do not fall into that category.

Fig. 1 Stimuli from the
perceptual discrimination task
used in Asch (1956, Fig. 2);
length labels did not appear on
the experimental stimuli
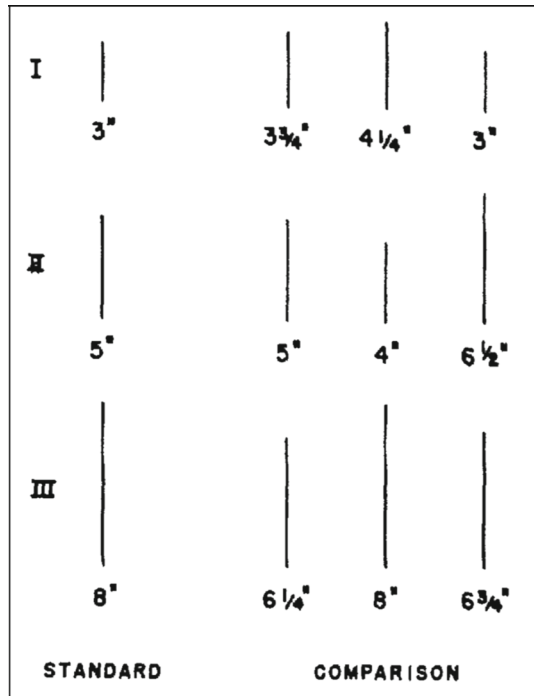
4.1 Varying the motivational context

It is possible that we are missing important dimensions of our concepts by only testing them in theoretical contexts in which participants have no stake in the outcome of their judgments. For example, a development of one of the most dramatic findings in 20th century social psychology—Asch's (1956) conformity experiments—shows that varying a participant's motivational context can affect how they perform an experimental task.

Asch's conformity experiment involves asking participants to make extremely simple perceptual judgments comparing the length of "comparison" lines with the length of a standard (see Fig. 1). The ease of the perceptual task is conveyed by the high accuracy of such comparisons (99%) when participants performed the task without any outside influence, in a control condition. The experimental manipulation involved placing the participant in a context of social influence with a group (6–8) of experimental confederates who made unanimously incorrect comparative judgments. In the social influence condition, participants' responses became significantly less accurate, conforming with the incorrect judgments of the majority in 36.8% of the trials (Asch 1955, p. 32).

Further variations indicate that other manipulations have a significant effect on rates of conformity on the perceptual judgment task. Asch (1956) provides evidence that varying the size of the majority, and the presence or absence of dissenters (both those who report accurate and inaccurate judgments) has an effect on whether participants
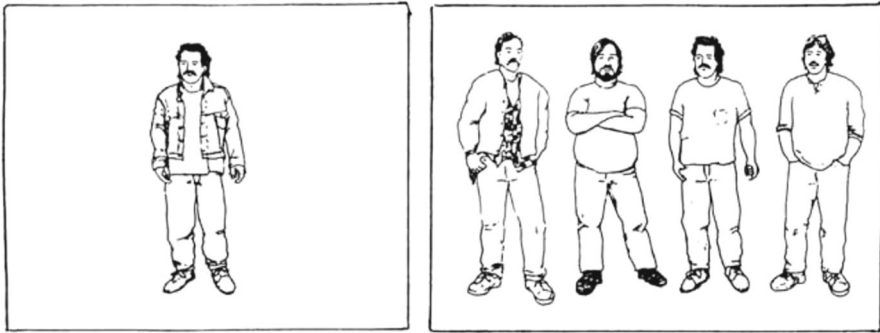
Fig. 2 "Lineup" task used in Baron et al. (1996, Fig. 1); example "perpetrator" slide is on the left, example "lineup" slide is on the right

judge in accordance with the majority. Baron et al. (1996) investigate whether the Asch conformity effect only arises because of the triviality of the perceptual task:

> One could dismiss the conformity effect as a laboratory 'hothouse' phenomenon that occurs because the potential face-to-face rejection of peers is far more important to participants than their accuracy on some unimportant 'scientific' test of perception or social judgment. (Baron et al. 1996, p. 915)

What would happen to the conformity effect if participants were given some additional motivation for performing the perceptual task accurately? To answer that question, Baron et al. used a "lineup" task, in which participants were shown a drawing of a target person and then asked to judge whether the target appeared in a lineup of four individuals in an image presented separately (see Fig. 2).

Participants were given the lineup task in four different conditions, which varied the *difficulty* of the task (low vs. high), and the *importance* of the task (low vs. high). The low-difficulty version of the task allowed participants to view the perpetrator slide and the lineup slide for five seconds each, and showed the two-slide sequence two times. In the high-difficulty version of the task, the perpetrator slide was only shown once, for 0.5 seconds. The low-importance condition involved informing participants that they were participating in a pilot study developing materials to test eyewitness testimony. In the high-importance condition, participants were told that they were calibrating an eyewitness testimony test that will soon be used by police and in courtrooms, and that if they performed in the top 12% in terms of accuracy on the test, they would receive a $20 prize.

Baron et al. found that in the low-difficulty, high-importance condition, participants were significantly less likely to be subject to the conformity effect than in the low-difficulty, low-importance condition, lending support to the idea that participants in the original Asch experiments conformed to the majority at the rates they did partly because of the low importance of the task they were asked to perform. But even more interestingly, in the high-difficulty, high-importance condition, participants were significantly *more* likely to conform to an inaccurate group consensus than in the high-difficulty, low-importance condition. Baron et al. (1996, p. 924) explain

this finding by observing that when it is difficult to "objectively" verify a particular judgment (because of the short exposure time in the high-difficulty condition), "individuals become increasingly reliant on social information to gauge the accuracy and appropriateness of their views". The Baron et al. investigation reveals that participants' responses can be affected by participants' sense of what the perceived point or importance of the experimental task is.

Embedding existing experiments on "know" and knowledge in a context where participants have some additional motivation for performing the task would require only a slight divergence from standard experimental investigations of knowledge. For example, one of the more closely studied questions in experimental epistemology is whether knowledge is sensitive to the stakes of being wrong (i.e., are people more willing to ascribe "knowledge" to an individual when the consequences of the individual being wrong are trivial than when the consequences are severe).[17] In all existing experiments probing the concept of knowledge, it is simply stated what the stakes are, and assumed that participants will take that statement at face value when asked to judge whether someone "knows" something; the actual stakes for the participants or for those who they are judging are not varied.[18]

In contrast to the methods employed by existing studies in experimental epistemology, studies in behavioral economics regularly employ methods in which the actual stakes for participants are varied. For example, stakes can be straightforwardly manipulated by varying monetary rewards (for a review of such experimental approaches see Kamenica 2012). For example, Ariely et al. (2009) found that increases in monetary stakes increased performance in simple tasks but degraded performance in complex tasks. Such a design is easily extendable to investigate the effect of stakes on judgments about knowledge and the meaning of "knows", so that participants are placed in situations where genuine financial effects of being wrong either on another or on themselves can be manipulated to determine whether self-ascription or other-ascription of knowledge is sensitive to stakes. Experiments of that form could assess whether effects similar to those observed in Baron et al. (1996) extend to assessments of knowledge.

4.2 Varying awareness of being in an experiment

The "dictator game" is used to probe whether people have a sense of "fairness" in how they allocate a monetary windfall. The game was developed to test the "unfairness" assumption in standard economic theory: "The economic agent is assumed to be law-abiding but not 'fair'—if fairness implies that some legal opportunities for gain are not exploited" (Kahneman et al. 1986, p. S286). The "dictator" receives (or is told to imagine that she receives) a certain amount of money ($20 in the original study), and is then instructed to decide how much of the windfall to offer anonymously to a recipient. Standard economic theory would predict that the dictator should keep all of the windfall.

---

[17] For a recent survey of the experimental literature on stakes sensitivity, see Pinillos (2016).

[18] This is noted by Feltz and Zarpentine (2010, fn. 17).

Kahneman et al. ([1986](#)) offered the dictator a choice between offering $2 and $10 to the recipient. The high rates of fair ($10) offers (76%) was taken as evidence against the "unfairness" assumption of standard economic theory as a model of actual human behavior (Kahneman et al. [1986](#), p. S291). Subsequent dictator game experiments which offered a wider range of response options did not reproduce the high rates of a completely fair distribution (only 22% made a 50–50 offer in the dictator experiment with actual pay in Forsythe et al. [1994](#), for example), but there has been extensive evidence from dictator game experiments that challenges the "unfairness assumption" of standard economic theory (for a summary of the results of many studies, see Camerer [2003](#), Table 2.4 and Guala and Mittone [2010](#)).

One methodological worry that has been raised about the use of dictator games to challenge the unfairness assumption is that in standard experiments participants are not *anonymous*. If the dictator's offer is not genuinely anonymous, it can't be concluded that it is purely a sense of fairness that is driving their altruistic offers—it might be, for example, the dictator's desire to protect her reputation that (partially) explains the fact that offers diverge from the predictions of standard economic theory. Hoffman et al. ([1994](#)) lent experimental weight to this worry by conducting a double-blind dictator game (in which individual participants' offers could not be known by the experimenters or the recipients of the offers, and participants knew that they could not know) which had the effect of significantly reducing the amount of the offers that the dictators made (half of the dictators offered nothing).[19]

But even in Hoffman et al. double-blind experiment, participants are still aware that they are taking part in an experiment. Winking and Mizer ([2013](#)) conducted a "natural field experiment" that removes even that residual element of the dictator's sense that her behavior is being examined (even if not *de re*) by an experimenter. Their study yielded an astonishing result: under conditions when dictators didn't realize they were participating in an experiment, they did not make *any* altruistic offers—they kept all windfalls for themselves.

Winking and Mizer's field experiment involved a pair of confederates. Confederate 1 waited at various bus stops, each of which was within one block of a casino in Las Vegas. When a potential participant also began to wait at the bus stop, Confederate 1 pretended to take a phone call on a cellular phone and "walked some distance away, facing away from the participant". Confederate 2 then walked by the participant, and "pretended to notice [casino] chips in his pocket, stopped briefly and claimed to the participant that he was late for a ride to the airport and asked the individual if he/she wanted the casino chips [$20], which he did not have time to cash in" (Winking and Mizer [2013](#), p. 290). There were three experimental conditions: In condition 1, Confederate 2 either simply walked off; in condition 2, Confederate 2 told the

---

[19] This was accomplished by giving participants an unmarked opaque envelope that contained either 20 blank slips of paper, or 10 blank slips and 10 one-dollar bills. When participants receive an envelope, each "proceeds to the back of the room, and opens the envelope inside a large cardboard box which maintains his/her strict privacy". Each participant keeps 0 to 10 of the dollar bills and 0 to 10 of the blank sheets of paper, so that the number of bills and slips of paper in the envelope add up to 10. Each envelope therefore feels equally thick. The envelopes are then put in a box, so the experimenter knows only the overall distribution of offers, not which participant made each individual offer. The contents of the box are then distributed to the recipients waiting in a separate room. (Hoffman et al. [1994](#), p. 355).

participant, when handing over the chips, "I don't know, you can split it with that guy however you want", referring to Confederate 1; condition 3 involved a set up roughly parallel to Hoffman et al. (1994), in which participants were aware they were taking part in an experiment, but the experimenter didn't see how participants allocated the \$20 in chips they received. While the results in condition 3 were consistent with laboratory dictator game results, with a mean offer of \$5.43, *no participants* in either condition 1 or condition 2 (n = 60) offered *any* chips to Confederate 1 (p. 291). Winking and Mizer's experiment indicates the dramatic effect that awareness of being in a non-ordinary (experimental) context can have on participants' behavior.

The dramatic effect of moving the dictator game out of the lab and into the wild demonstrated in the Winking and Mizer study provides a model for how to think about more naturalistic experiments investigating philosophically significant concepts (such as knowledge). With the help of confederates, it would be possible to evaluate how stakes affect the way ordinary speakers assess whether someone knows something in a covert way. For example, two confederates could play the role of parent and student at a University open day (open house). The participant would be selected from those who have volunteered to be guides for prospective students. The student confederate would ask the participant guide for directions to their next appointment (which is scheduled to take place in building B), and then walk away after receiving directions. After the student confederate walks away, the parent would then approach the participant guide and ask (condition 1, low stakes) if their child knows that their next meeting (which concerns what student clubs are available on campus) is in building B; or (condition 2, high stakes) if their child knows that their next meeting (which they *have* to be on time for because they are going to be interviewed for a full academic scholarship) is in building B. Such a design does not vary the stakes for the participant, but it creates a condition in which apparent real-world stakes (for the confederates) can vary while concealing the fact that an experiment is taking place.

4.3 Conversation versus one-off speech acts, and addressees versus overhearers

Clark (1997) observes that most experimental investigations of language employ unnatural conversational contexts, stripped of normal features of social interaction. Typically such experiments involve making judgments about one-off utterances, which participants cannot query or challenge:

> It is difficult to study understanding in the wild, so investigators have developed a variety of laboratory techniques instead. Most of these techniques are built around contrived sentences presented to people isolated from any realistic human activity. (p. 577)

Clark argues that the standard methodological assumption in experimental investigations of meaning is that understanding an utterance is "autonomous", meaning that it doesn't require any interaction beyond the passive comprehension of the speaker's utterance by the audience. Stimuli are usually written or pre-recorded spoken texts that are presented to participants, who are asked to respond to them in various ways, but querying the stimulus or asking for clarification is usually not permitted. For
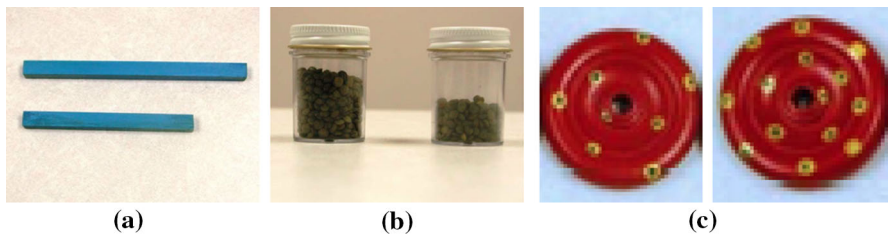
Fig. 3 Stimuli used in the presupposition assessment task, from Syrett (2007, Appendix E). **a** Please give me the long rod. **b** Please give me the full one. **c** Please give me the spotted one

example, the "presupposition assessment task" (Syrett 2007; Syrett et al. 2010; Liao and Meskin 2017; Hansen and Chemla 2017) tests whether participants are willing to accommodate the uniqueness and existence presuppositions of definite descriptions when combined with different types of adjectives.

The task involves showing participants pairs of objects with varying degrees of a particular property picked out by an adjective F, and then asking for the participant to select "the F one" (see Fig. 3). Participants are willing to accommodate both the uniqueness and existence presuppositions of the definite description when asked to select the *longer* of the two rods, but they tend to refuse both the request to hand over "the full one" (because neither jar is completely full—a failure of the existence presupposition of the definite description), and the request to hand over "the spotted one" (because both disks are spotted—a failure of the uniqueness presupposition). That pattern of responses is taken as evidence of a difference in the standards that participants associate with different types of adjective. But the task (like many experimental probes used in experimental semantics and pragmatics) is non-naturalistic in the respect that participants can't ask for clarification of the request, or confirmation that they've selected the right object.

Schober and Clark (1989) demonstrate that the ability of the audience to interact with the speaker has significant effects on successful communication. Schober and Clark provide evidence that when addressees can actively interact with speakers, they can more accurately represent what the speaker intends to communicate than "mere overhearers" who passively listen to the same conversations. In one of their experiments, a "director" was seated across from a "matcher", separated by a barrier that prevented them from seeing each other. The director has a sheet with 16 tangram figures on it, arranged in a random order (see Fig. 4). The first 12 figures on the director's sheet were numbered 1–12. The matcher had 16 cards with corresponding tangram figures on them, and ordered slots in which 12 of the cards could be placed. The primary communicative task was for the matcher to arrange 12 cards in the order in which they appeared on the director's sheet, and the director and the matcher could talk to each other as much as they wanted. Each director–matcher pair played the game six times in a row, with the order of the tangram figures randomized each time.

A secondary communicative task involved a third participant, an "overhearer", who was in the room with the director and the matcher, but who was instructed not to interact with either. The overhearer was instructed to try to match the same 12 tangram figures that the director and the matcher were trying to match. The director and the

Fig. 4 Tangram figures
(Schober and Clark 1989, Fig. 1)
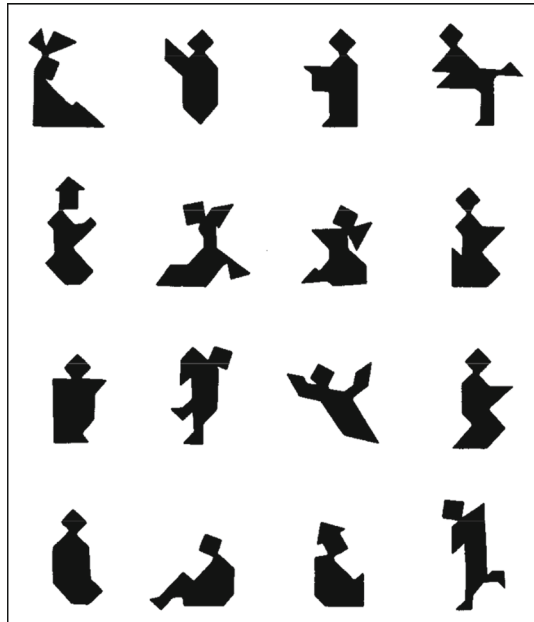
matcher were told that the overhearer was a coder who was there in order to "reduce experimental bias", in order to make sense of the presence of a silent listener (p. 222). The overhearer therefore had access to all of the same utterances as the director and the matcher, but Schober and Clark found that the matchers were significantly more accurate than the overhearers: "Matchers started out with 95% correct on Trial 1, and, by Trial 6, they all matched every reference correctly. In contrast, overhearers started out with only 78% correct and only improved to 89% by the last trial" (p. 223). That supports the idea that optimal understanding involves joint activity between speaker and addressee. Because standard experimental tasks used to probe the meaning of expressions don't involve a collaborative component, they may only be capturing a small slice of typical linguistic understanding—namely, that which is available to overhearers, rather than the optimal form of understanding that requires collaboration between speaker and addressee.

How could this conversational paradigm be applied to the investigation of "knows" and knowledge? One approach would be to adopt the interview methodology used in Niedzielski and Preston (2000), in which trained fieldworkers recorded open-ended conversations with ordinary speakers that focused on linguistic topics. It would be straightforward to prompt participants to have conversations about the meaning of "know", and steer conversation towards specific topics of theoretical interest (stakes sensitivity, what participants think of Gettier-style cases, and so on). This kind of approach would have to take steps to avoid the obvious risk of experimenter bias, but

it has the potential to reveal not just how participants apply "know" to particular cases, but also to reveal higher-level beliefs about "know" and knowledge.[20]

A different approach would be to adopt a design similar to that used in Schober and Clark (1989). Pairs of participants would be confronted jointly with standard stimuli about "know" (Gettier cases, stakes-sensitivity cases, and so on), and asked to discuss how to classify the cases. That type of design would have the advantage of yielding both "extensional" data about classification, as well as constrained contexts in which to observe meta-linguistic "intensional" data [and potentially new examples of "meta-linguistic negotiation"—see Plunkett and Sundell (2013)] about the meaning of "know".

5 A revised challenge from ordinary language

The three experimental case studies discussed above provide some empirical support to Baz's first argument that answers to the "theorist's question" may not give us an accurate picture of the concepts that speakers employ (*knowledge*, e.g.) in ordinary circumstances. With these experiments in mind, I propose a new version of the argument from ordinary language as follows:

1. Standard experimental approaches to the investigation of philosophically significant concepts assume that stripping away conversational or "pragmatic" factors from the experimental context yields a clearer picture of the underlying concepts.
2. But experimental studies in more "ecologically valid" contexts—which may include (i) motivations that go beyond just wanting to perform the experimental task, (ii) participants' awareness that they are taking part in an experiment, or (iii) an experimental task that involves active collaboration between speakers and addressees—may not interfere with or distort the application of the relevant concepts; such contexts may in fact provide *better* conditions for the application of those concepts. (At least: we don't yet have a reason to think that by stripping out standard features of ordinary situations in which a concept is applied, we get a more accurate picture of how that concept functions.)[21]
3. So drawing conclusions about philosophically significant concepts solely on the basis of answers given to the "theorist's question" in experimental contexts that lack (i–iii) is, so far, unjustified.

---

[20] In an early criticism of ordinary language philosophy, Mates (1958) argues that the standard, "extensional" method of probing meaning, in which participants classify situations as falling into the extension of an expression or as making a sentence true (e.g., "is this a case in which 'S knows that p' would apply?"), only illuminates one aspect of lexical meaning. He also recommends adopting an "intensional" approach to the study of meaning, which would involve

[asking] the subject what he means by the given word or how he uses it; [then] one proceeds in Socratic fashion to test this first answer by confronting the subject with counterexamples and borderline cases, and so on until the subject settles down more or less permanently upon a definition or account. (Mates 1958, pp. 165–166)

Mates refers to this kind of investigation as employing "Socratic questionnaires".

[21] For a parallel argument about the significance of whether or not to design experiments that include explicit contrasts, see Hansen (2014b).

The conclusion of this revised challenge from ordinary language to standard experimental ways of investigating meaning is less radical than Baz wants: it doesn't establish that there is a "fundamental" difference between the theorist's question and ordinary questions, and it could turn out that these factors (i–iii) only matter in certain cases, and that, say, the way we understand the word "know" isn't sensitive to different motivations or conversational "points", or whether people are aware that they are participating in an experiment, or whether the word is used in a collaborative conversation or just an utterance that is directed to mere overhearers. But one advantage of this revised argument is that it does not depend on any contentious (Wittgensteinian or otherwise) conceptions of meaning and understanding in general—it is a challenge grounded in experimental data and some (hopefully not overly contentious) features of non-experimental conversation.

6 Conclusion: "Nobody would really talk that way!"

The revised challenge from ordinary language can be viewed as a modest branch of the critical project in ordinary language philosophy. Endorsing the argument doesn't require saying that philosophers are speaking "nonsense" when they diverge from ordinary use (as in Malcolm 1951), or that ordinary speakers do not understand what they are being asked when confronted with Gettier scenarios, because such questions could be understood in any number of ways, and the context in which the "theorist's question" is posed doesn't provide a way of selecting among those ways (as Baz argues). But it does require some response if philosophers are going to continue to claim that formal or informal experiments illuminate the lexical meanings and concepts that ordinary speakers employ, or (more ambitiously) that such experiments tell us something about the underlying features of reality those meanings and concepts are about.

One way of responding to the revised challenge from ordinary language would involve designing experiments that probe the meaning of "know" (e.g.) while incorporating some or all of the features (i–iii) ((i) motivations that go beyond just wanting to perform the experimental task, (ii) participants' awareness that they are taking part in an experiment, and (iii) an experimental task that involves active collaboration between speakers and addressees). Such a response would require some *experimental ingenuity*. The design of such experiments that would investigate "knows" and the concept of knowledge (and possibly knowledge itself) is sketched in Sect. 4.

The quote in the title of this paper comes from a story that Keith DeRose tells about Rogers Albritton. DeRose describes his early attempts to develop pairs of examples that were supposed to illustrate the idea that knowledge ascriptions ("S knows that p") are context–sensitive. DeRose's early examples involved ascriptions that appeared to say something true, but which were conversationally inappropriate:

> My adviser, Rogers Albritton, objected, as near as I can remember, 'Nobody would really talk that way!' I replied that it didn't matter whether people would talk that way. All I needed was that such a claim would be true, and that certainly was my intuition about the truth-value of the claim. He would have

none of that, and answered, quite sternly, 'Look, if you're going to do ordinary language philosophy—and that's what you're doing here—you'd better do it right'...Albritton never explained to me why the examples should be constructed so that what's said is natural and appropriate beyond insisting that that's how ordinary language philosophy should be done. (He seemed to think it a point too obvious to require explanation, and I was not about to ask!) (DeRose 2009, p. 51)

In roughest outline, the critical project in ordinary language philosophy can be summed up as a version of Albritton's objection: It challenges standard ways of investigating the meaning of philosophically significant expressions that ignore the way people "would really talk".[22] The revised argument from ordinary language proposed in this paper, and the recommendation to enrich standard experimental investigations of "know" and knowledge is intended to focus new attention on what would be required to "do ordinary language philosophy right", at least in an experimental context.

7 Appendix: Experimental details

In this appendix, I present relevant details from the experimental studies discussed in Sect. 3. First, I describe the studies that failed to replicate the findings of cultural variations in responses to Gettier scenarios presented in Weinberg et al. (2001) (Kim and Yuan 2015; Machery et al. 2017; Seyedsayamdost 2015; Turri 2013). Second, I give details of the studies that indicate that there are different types of Gettier scenarios, which participants respond to in systematically different ways (Starmans and Friedman 2012; Turri et al. 2015).

7.1 Failures to replicate cross-cultural differences in responses to Gettier scenarios

7.1.1 Kim and Yuan (2015)

Kim and Yuan (2015) used the same Gettier "car" scenario employed in Weinberg et al. (2001), and used the same binary response option ("really knows"/"only believes"), but they failed to replicate the original finding of a significant difference in responses between East Asian and Western participants, using a larger sample size. Kim and Yuan received very similar rates of response from both East Asian (EA) and "Caucasian" (C) participants (see Table 2).

---

[22] Baz (2012b, pp. 138–139) uses DeRose's discussion of Albritton's remark to mark different ways of thinking about doing ordinary language philosophy.

Table 2 Results from Kim and Yuan (2015) in response to the Gettier car scenario

Study	N	Ethnicity	n	"Really knows" %	"Only believes" %	p-exact
		EA	82	14.6	85.4	
Kim and Yuan (2015)	140					0.89
		C	58	13.8	86.2	

7.1.2 Seyedseyamdost (2015)

This study used a slightly modified version of the "car" Gettier scenario from Weinberg et al. (2001). After reading the scenario, participants were asked to indicate whether the subject in the story REALLY KNOWS or ONLY BELIEVES the target statement ("Does Bob really know that Jill drives an American car, or does she only believe it?"). Participants were classified as being from East Asia (EA), the Indian Subcontinent (SC), or the West (W). Seyedseyamdost collected three data sets (DS) from three different groups that were used in the studies of responses to the "car" Gettier scenario: undergraduates (mainly philosophy undergraduates) at the London School of Economics (DS1), online participants registered on SurveyMonkey (DS3), and participants who voluntarily visited Harvard University's Moral Sense Test website (DS4). Whereas Weinberg et al. (2001) found significant differences between EA and W participants in responses to the car scenario, Seyedseyamdost did not find significant differences between EA and W participants (p. 103). Table 3, below, compares results from Weinberg et al. (2001) and Seyedsayamdost (2015) for the Gettier "car" scenario, first comparing EA and W participants, then (after the double horizontal line) comparing SC and W participants.[23]

7.1.3 Turri (2013)

Turri (2013) applies a novel method of presenting Gettier scenarios to participants that breaks them down into three parts, each presented on separate screens:

> Start with a belief that is well enough justified to satisfy the justification condition on knowledge. All seems well. Then introduce bad luck that would normally prevent the justified belief from being true. All seems ill. Then introduce a conspicuously distinct element of good luck that makes the belief true anyway…But not all is made well again. (p. 2)

Previous studies that found cultural differences in responses to Gettier scenarios uniformly used a one-stage method of presenting the scenarios (on a single screen), and Turri hypothesized that his tripartite structure would allow participants to keep track of all of the relevant components of the scenario, thereby bringing their responses more closely into alignment with standard philosophical judgments about Gettier

---

[23] Seyedsayamdost report that the one significant difference they found, in DS4 "may not be very meaningful" because it was based on a very small sample of SC participants (p. 103).

Table 3 Results from Seyedsayamdost ([2015](#)) and Weinberg et al. ([2001](#)) on the Gettier car scenario

Study	N	Ethnicity	n	"Really Knows" %	"Only Believes" %	p-exact
Weinberg et al. (2001)	89	EA	23	56.5	43.5	0.006
		W	66	25.8	74.2	
Seyedsayamdost (2015) DS1	120	EA	41	26.8	73.1	0.146
		W	79	15.2	84.8	
Seyedsayamdost (2015) DS3	111	EA	36	22.2	77.8	1.000
		W	75	22.7	77.3	
Weinberg et al. (2001)	89	SC	23	60.9	39.1	0.002
		W	66	25.8	74.2	
Seyedsayamdost (2015) DS1	113	SC	34	14.7	85.3	1.000
		W	79	15.2	84.8	
Seyedsayamdost (2015) DS4	208	SC	15	46.7	53.3	0.011
		W	193	17.1	82.9	

Table 4 Results from Turri ([2013](#)) and Weinberg et al. ([2001](#)) comparing Western participants and participants from the Indian sub-continent; the third row is a comparison of the "original sub-continent" (OSC) results from Weinberg et al. ([2001](#)) with Turri's results from Indian participants

Study	N	Ethnicity	n	"Really Knows" %	"Only Believes" %	p-exact
Weinberg et al. (2001)	89	OSC	23	60.9	39.1	0.002
		W	66	25.8	74.2	
Turri (2013)	53	SC	27	15	85	1.000
		W	26	11	89	
Turri (2013)	50	OSC	23	60.9	39.1	0.001
		SC	27	15	85	

scenarios. Turri's design used different Gettier scenarios than Weinberg et al. ([2001](#)). He compared responses from participants recruited on Amazon Mechanical Turk and located in India with responses from Western participants (recruited on Mechanical Turk and located in the United States), and with Weinberg et al.'s original results for participants from the Indian sub-continent. He didn't find any significant difference between the responses of Western participants and the Indian participants in his study, and he found a significantly lower rate of "really knows" responses among the Indian participants in his study compared with the original responses of participants from the Indian sub-continent in Weinberg et al. ([2001](#)) ("OSC") (see Table [4](#)).

7.1.4 Machery et al. (2017)

In this study, responses from 245 participants from Brazil, Japan, India and the US were collected in response to two Gettier scenarios ("Gettier/Hospital" and "Gettier/Trip") and two control scenarios (a clear case of knowledge, and a clear case of false, but justified belief). Participants were asked to respond to two questions about whether the

Table 5 Responses to the "knowledge 2" prompt (Machery et al. 2017)

Ethnicity	n	Scenario	"S knows that p" %	"S... doesn't actually know that p" %
Brazil	37	Gettier/Hospital	5.4	94.6
		Gettier/Trip	35.1	64.9
Japan	104	Gettier/Hospital	8.7	91.3
		Gettier/Trip	18.3	81.7
India	40	Gettier/Hospital	12.5	87.5
		Gettier/Trip	35.1	64.9
USA	64	Gettier/Hospital	14.1	85.9
		Gettier/Trip	29.7	70.3

subject in the scenario knows. The first knowledge question was "Does [the subject] know [the relevant proposition]?", followed by the following response options: "Yes, [s]he knows" and "No, [s]he doesn't know" (*knowledge 1*). The second knowledge question was "In your view, which of the following sentences better describes [the subject's] situation?" followed by the following response options: "[Subject] knows that [relevant proposition]" and "[Subject] feels like [s]he knows that [relevant proposition] but [s]he doesn't actually know [this]" (*knowledge 2*) (p. 649). The experimental materials were presented to participants in their native language (Portuguese, Japanese, Bengali, or English).

Machery et al. found various cross-cultural differences and differences within cultures in their responses to the two Gettier scenarios—for example, Brazilians and Indians were significantly more likely to ascribe knowledge in the Gettier/trip case than in the Gettier/hospital case, in response to both knowledge 1 and knowledge 2 probes, whereas Americans were not. But Machery et al. state the key finding of their cross-cultural study as follows:

> Most important, however, there was no difference in knowledge ascription between the USA, Brazil, India, and Japan in either Gettier case in response to the knowledge 2 probe. That is, Indians, Americans, Brazilians, and Japanese tend to share the Gettier intuition about Gettier cases. (p. 651)

Responses from the four groups of participants to the knowledge 2 probe are reported in Tables 5 and 6 (for statistical analyses, see Machery et al. (2017, Appendix 1).

7.2 Evidence of different patterns of responses to different types of Gettier scenarios

7.2.1 Starmans and Friedman (2012)

Starmans and Friedman (2012) investigated several different types of Getter-style scenarios alongside control scenarios that prompted clear patterns of knowledge attri-

Table 6 Responses to the "knowledge 2" prompt, organized by type of Gettier scenario (Machery 2017)

Scenario	Ethnicity	"S knows that p" %	"S… doesn't actually know that p" %
	Brazil	5.4	94.6
	Japan	8.7	91.3
Gettier/Hospital			
	India	12.5	87.5
	USA	14.1	85.9
	Brazil	35.1	64.9
	Japan	18.3	81.7
Gettier/Trip			
	India	35.1	64.9
	USA	29.7	70.3

Table 7 "Really knows" dichotomous response percentages for Experiments 1a–1c (Starmans and Friedman 2012)

Experiment	N	Clear knowledge (%)	Gettier-style (%)	False belief (%)
1a	144	88	72	11
1b	133	79	69	14
1c	46		83	

bution and denial. Starmans and Friedman's Experiments 1a and 1b each had three conditions: one control condition that described a clear case of an agent knowing that p, another control condition that involved an agent who clearly lacked knowledge that p, and a Gettier-style scenario in which the agent in the scenario forms a true belief on the basis of seeing that something is the case, but the relevant state of affairs that was the original basis of the belief is replaced, unbeknownst to the agent, with another state of affairs that still makes the belief true.

In each condition, participants were asked whether the agent "Really knows" that p, or "only thinks" that p, and they were asked to indicate their degree of confidence on a scale from 1–10. In a third experiment, 1c, participants were only asked to respond to a Gettier-style scenario (this experiment was used to evaluate whether the presence of comprehension questions was having an effect on responses, but no effect was observed). Percentage of "really knows" responses to the dichotomous prompts are given in Table 7. Starmans and Friedman did not find a significant difference between rates of knowledge attribution between the clear knowledge control scenario and the Gettier scenario in Experiments 1a and 1b, but they did find a significant difference between rates of knowledge attribution between the Gettier scenario and the case of false belief. They did not find a significant difference in responses to the Gettier scenarios across Experiments 1a–1c.

In order to evaluate whether the "lay concept of knowledge" allows beliefs that are true (but not justified) to count as knowledge, Starmans and Friedman conducted an experiment that varied the justification for the belief that p in a Gettier-style scenario and asked participants to indicate whether the agent in the scenario "really knows"

	N	Gettier-style High justification	Gettier-style Low justification
Table 8 "Really knows" dichotomous response percentages for Experiment 2 (Starmans and Friedman 2012)	51	70%	25%

	N	Gettier-style Authentic evidence	Gettier-style Merely apparent evidence
Table 9 "Really knows" dichotomous response percentages for Experiment 3 (Starmans and Friedman 2012)	43	67%	30%

or "only believes". As represented in Table 8, they found that varying the level of justification had a significant effect on whether participants attributed knowledge, and the higher level of justification was generally required for participants to attribute knowledge.

Starmans and Friedman also probed whether participants' knowledge attributions were sensitive to differences between what they called "authentic" and merely "apparent" evidence. In the merely "apparent" evidence scenarios, agents possess "evidence that only *appears* to be informative about the world, but coincidentally leads to a true belief"[24]:

> For example, consider a scenario where a student comes to believe that his professor is in her office, because the student sees a convincing hologram sitting at the professor's desk. As it turns out, the professor is in her office, but she is crouching under the desk reading philosophy. In this case, the hologram serves as the evidence for the student's belief, which turns out to be true. (p. 278)

They found that there was a significant difference between participants' rates of knowledge attribution between the "authentic" evidence Gettier-style scenarios and the merely "apparent" evidence Gettier-style scenarios (see Table 9).

Starmans and Friedman conclude that their findings "reveal a difference between two kinds of Gettier case", namely, those involving "authentic" evidence, for which "really knows" responses were in the 69–80% range (in Experiments 1a, 1b, 1c, 2, and 3), and those involving merely "apparent" evidence, for which "really knows" responses dropped to 30% (Experiment 3). They consider the possibility that participants were "confused" by the Gettier scenarios, which might explain the different types of responses, but they observe that "if participants had been confused in the Gettier cases, they should have given low confidence ratings to their responses, but they did not. Confidence ratings did not differ across conditions, and moreover few participants ever used the lower end of the confidence scale" (p. 280).

---

[24] For criticism of Starmans and Friedman's way of formulating the distinction between "authentic" and merely "apparent" evidence, see Nagel et al. (2013b).

Table 10 "Really knows" dichotomous response percentages for Experiment 1 (Turri et al. 2015)

N	Control No threat	Gettier-style (Failed) threat	Control No detection
120	81%	67%	16%

7.2.2 Turri et al. (2015)

Building on the investigation of different types of Gettier scenario in Starmans and Friedman (2012), Turri et al. (2015) present evidence that rates of knowledge attribution vary systematically depending on the presence or absence of certain epistemic features in Gettier-style scenarios. They conducted four experiments that evaluated the effects of these factors on knowledge attributions.

Turri et al.'s first experiment investigated whether "a salient, but failed threat to a perceptual judgment" blocks knowledge attributions. Participants responded to one of three scenarios: (1) An agent forms a true belief based on perceiving a "truth-maker", with no threatened disruption of that perceptual relation; (2) An agent forms a true belief based on perceiving a truth-maker, and there is a salient but failed threat to disrupt the perceptual relation (such as the presence of many fake barns in fake barn county); (3) The threat of disruption is realized, and the agent is prevented from forming a true belief. Results from the first experiment are presented in Table 10, where percentages are percentages of rates of "really knows" responses when prompted with the dichotomous choice between "really knows" and "only believes". Turri et al. found no significant difference between rates of knowledge attribution in response to the "no threat" control condition and the Gettier-style "threat" condition but did find a significant difference between the "threat" condition and the "no detection" control condition.

In their second experiment, Turri et al. tested how participants attributed knowledge in scenarios in which there is an "unnoticed change in the explanation for which an agent's belief is true". Again participants responded to three different conditions (1) An agent forms a true belief on the basis of perceiving a truth-maker, and nothing threatens to "disrupt" the truth-maker; (2) An agent forms a belief on the basis of perceiving a truth-maker, and the truth-maker is "disrupted" and replaced with a backup truth-maker (for example, an agent forms the belief that there is a pen on the table (the truth-maker), but the pen is secretly replaced with another pen, placed on the same table (disruption of the truth-maker, but the belief is still true); (3) An agent fails to detect the truth-maker for her belief, and what makes her belief true goes unnoticed (for example, the agent perceives a hologram of a professor in her office, forms a belief that the professor is in her office, a belief which is made true by the professor hiding under her desk). Results from the second experiment (rates of "really knows" responses) are presented in Table 11. There were significant differences in the rates of knowledge attributions between all three conditions.

Turri et al.'s third experiment investigated whether there is an effect of how similar the "replacement" truth-maker is to the original truth-maker on rates of knowledge attributions. Again, participants were assigned to one of three conditions (1) An agent

Table 11 "Really knows" dichotomous response percentages for Experiment 2 (Turri et al. 2015)

N	Control Normal detection	Gettier-style Replacement	Gettier-style No detection
135	88%	66%	23%

Table 12 "Really knows" dichotomous response percentages for Experiment 3 (Turri et al. 2015)

N	Gettier-style Similar	Gettier-style Dissimilar	Control No detection
558	54%	42%	8%

Table 13 "Really knows" dichotomous response percentages for Experiment 4 (Turri et al. 2015)

N	Control Knowledge	Gettier-1 Failed Threat	Gettier-2 Detect Similar	Gettier-3 Detect Dissimilar	Gettier-4 No Detect Similar	Gettier-5 No Detect Dissimilar	Control Ignorance
728	90%	83%	55%	39%	38%	19%	12%

forms a true belief on the basis of perceiving a truth-maker, but there is an unnoticed change in what makes the belief true (as in the replaced pen scenario); (2) An agent forms a true belief on the basis of perceiving a truth-maker, but there is an unnoticed change in what makes the belief true, and the replacement truth-maker is *dissimilar* in some important respect to the original truth-maker; (3) An agent fails to detect the truth and nothing makes her belief true. Results from the third experiment (rates of "really knows" responses) are presented in Table 12. There were significant differences in the rates of knowledge attributions between all three conditions.

The fourth and final experiment conducted by Turri et al. was a replication of Experiments 1–3 using different scenarios, based on Nagel et al. (2013a). In this experiment, there were seven conditions, corresponding to the seven types of conditions introduced in Experiments 1–3 (see Table 13—significant differences are marked with a triple vertical bar).

References

Alexander, J. (2012). *Experimental philosophy: An introduction*. Cambridge: Polity.

Ariely, D., Gneezy, U., Lowenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies*, *76*(2), 451–469.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70.

Austin, J. (1946). Other minds. *Aristotelian Society Supplementary*, *20*, 148–187.

Austin, J. (1956–1957). A plea for excuses. *Proceedings of the Aristotelian Society*, *57*, 1–30.

Austin, J. (1962). *Sense and sensibilia*. Oxford: Oxford University Press.

Bach, K. (2005). The emperor's new 'knows'. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning and truth* (pp. 51–89). Oxford: Oxford University Press.

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996). The forgotton variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, *71*(5), 915–927.

Baz, A. (2012a). Must philosophers rely on intuitions? *Journal of Philosophy*, *109*(4), 316–337.

Baz, A. (2012b). *When words are called for: A defense of ordinary language philosophy*. Cambridge, MA: Harvard University Press.

Baz, A. (2014). Recent attempts to defend the philosophical method of cases and the linguistic (re)turn. *Philosophy and Phenomenological Research*, *92*(1), 105–130.

Baz, A. (2015). Questioning the method of cases fundamentally-reply to Deutsch. *Inquiry*, *58*(7–8), 895–907.

Baz, A. (2016). On going (and getting) nowhere with our words: New skepticism about the philosophical method of cases. *Philosophical Psychology*, *29*(1), 64–83.

Baz, A. (2018). *The crisis of method in contemporary analytic philosophy*. Oxford: Oxford University Press.

Blouw, P., Buckwalter, W., & Turri, J. (2017). Gettier cases: A taxonomy. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining knowledge: New essays on the Gettier problem*. Oxford: Oxford University Press.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598.

Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology*, *1*(2), 275–296.

Cummins, R. (1998). Reflection on reflective equilibrium. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 113–127). Oxford: Rowman and Littlefield.

DeRose, K. (2009). *The case for contextualism*. Oxford: Oxford University Press.

DeRose, K. (2011). Contextualism, contrastivism, and x-phi surveys. *Philosophical Studies*, *156*(1), 81–110.

Deutsch, M. (2015). Avner Baz on the 'point' of a question. *Inquiry*, *58*(7–8), 1–20.

Feltz, A., & Zarpentine, C. (2010). Do you know more when it matters less? *Philosophical Psychology*, *23*(5), 683–706.

Fischer, E. (2014). Verbal fallacies and philosophical intuitions: The continuing relevance of ordinary language analysis. In B. Garvey (Ed.), *J.L. Austin on Language* (pp. 124–140). Basingstoke: Palgrave MacMillan.

Fischer, E., & Engelhardt, P. E. (2016). Intuitions' linguistic sources: Stereotypes, intuitions and illusions. *Mind and Language*, *31*(1), 67–103.

Forsythe, R., Horowitz, J. L., Savin, N., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347–369.

Gerken, M., & Beebe, J. R. (2016). Knowledge in and out of contrast. *Noûs*, *50*(1), 133–164.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, *73*(20), 771–791.

Guala, F., & Mittone, L. (2010). Paradigmatic experiments: The dictator game. *The Journal of Socio-Economics*, *39*, 578–584.

Hansen, N. (2014a). Contemporary ordinary language philosophy. *Philosophy Compass*, *9*(8), 556–569.

Hansen, N. (2014b). Contrasting cases. In J. Beebe (Ed.), *Advances in experimental epistemology* (pp. 72–96). New York: Bloomsbury.

Hansen, N. (2015). Experimental philosophy of language. *Oxford Handbooks Online*. https://doi.org/10.1093/oxfordhb/9780199935314.013.53.

Hansen, N., & Chemla, E. (2017). Color adjectives, standards and thresholds: An experimental investigation. *Linguistics and Philosophy*, *40*(3), 239–278.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, *7*, 346–380.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, *59*(4), S285–S300.

Kamenica, E. (2012). Behavioral economics and psychology of incentives. *The Annual Review of Economics*, *4*(13), 1–13.

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, *10*(2), 95–118.

Kim, M., & Yuan, Y. (2015). No cross-cultural differences in the Gettier car case intuition: A replication study of Weinberg et al. 2001. *Episteme*, *12*(3), 355–361.

Knobe, J. (2012). Experimental philosophy. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford handbook of philosophy of cognitive science*. Oxford: Oxford University Press.

Kukla, R. (2015). Delimiting the proper scope of epistemology. *Philosophical Perspectives*, *29*, 202–216.

Lawlor, K. (2013). *Assurance: An Austinian view of knowledge and knowledge claims*. Oxford: Oxford University Press.

Liao, S.-Y., & Meskin, A. (2017). Aesthetic adjectives: Experimental semantics and context-sensitivity. *Philosophy and Phenomenological Research*, *94*(2), 371–398.

Ludlow, P. (2005). Contextualism and the new linguistic turn in epistemology. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning and truth* (pp. 11–50). Oxford: Oxford University Press.

Machery, E. (2011). Thought experiments and philosophical knowledge. *Metaphilosophy*, *42*(3), 191–214.

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, *92*, B1–B12.

Machery, E., Stich, S., Rose, D., Chaterjee, A., Karasawa, K., Struchiner, N., et al. (2017). Gettier across cultures. *Noûs*, *51*(3), 645–664.

Malcolm, N. (1951). Philosophy for philosophers. *The Philosophical Review*, *60*(3), 329–340.

Mallon, R., Machery, E., Nichols, S., & Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research*, *79*(2), 332–356.

Mates, B. (1958). On the verification of statements about ordinary language. *Inquiry*, *1*(1), 161–171.

Nagel, J., Juan, V. S., & Mar, R. A. (2013a). Lay denial of knowledge for justified true belief. *Cognition*, *129*(3), 652–661.

Nagel, J., Mar, R., & Juan, V. S. (2013b). Authentic Gettier cases: A reply to starmans and friedman. *Cognition*, *129*(3), 666–669.

Niedzielski, N. A., & Preston, D. R. (2000). *Folk linguistics*. The Hague: Mouton de Gruyter.

Pinillos, A. (2012). Knowledge, experiments, and practical interests. In J. Brown & M. Gerken (Eds.), *New essays on knowledge ascriptions* (pp. 192–219). Oxford: Oxford University Press.

Pinillos, A. (2016). Experiments on contextualism and interest relative invariantism. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 349–358). Oxford: Wiley.

Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, *13*(23), 1–37.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.

Seyedsayamdost, H. (2015). On normativity and epistemic intuitions: Failure of replication. *Episteme*, *12*(1), 95–116.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124*(3), 272–283.

Starmans, C., & Friedman, O. (2013). Taking "know" for an answer: A reply to Nagel, San Juan, and Mar. *Cognition*, *129*(3), 662–665.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics*, *27*(1), 1–35.

Syrett, K. L. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. Ph.D. thesis. Evanston, IL: Northwestern University.

Turri, J. (2013). A conspicuous art: Putting Gettier to the test. *Philosophers' Imprint*, *13*(10), 1–16.

Turri, J. (2016). Knowledge judgments in "Gettier" cases. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 337–348). Oxford: Blackwell. (Chapter).

Turri, J., Buckwalter, W., & Blouw, P. (2015). Knowledge and luck. *Psychonomic Bulletin and Review*, *22*(2), 378–390.

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, *29*(1–2), 429–460.

Williamson, T. (2004). Philosophical 'intuitions' and scepticism about judgement. *Dialectica*, *58*(1), 109–153.

Williamson, T. (2005). Armchair philosophy, metaphysical modality and counterfactual thinking. *Proceedings of the Aristotelian Society*, *105*(1), 1–23.

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.

Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, *34*(4), 288–293.

Wittgenstein, L. (1969). *On certainty*. New York: Harper and Row.