



# Epistemic Contextualism: An Idle Hypothesis

John Turri

University of Waterloo

## ABSTRACT

Epistemic contextualism is one of the most hotly debated topics in contemporary epistemology. Contextualists claim that 'know' is a context-sensitive verb associated with different evidential standards in different contexts. Contextualists motivate their view based on a set of behavioural claims. In this paper, I show that several of these behavioural claims are false. I also show that contextualist test cases suffer from a critical confound, which derives from people's tendency to defer to speakers' statements about their own mental states. My evidence consists in results from several behavioural experiments. I conclude that contextualism is an idle hypothesis and I propose some general methodological lessons.

**ARTICLE HISTORY** Received 3 May 2015; Revised 19 January 2016

**KEYWORDS** knowledge; belief; evidence; philosophical method; social cognition; deferral hypothesis

## 1. Introduction

Epistemic contextualism is the view that the verb 'know' is a context-sensitive expression. More specifically, according to contextualism, in order for us to truthfully say that a person 'knows' a proposition, that person must meet the evidential standard set by our context and, critically, the standard changes across contexts. Contextualists motivate their view based on a set of empirical claims about competent speakers' linguistic behaviour in certain situations [Lewis 1996; DeRose 2009; Cohen 2013].

The most famous way of illustrating the idea involves a pair of cases about a man who wants to deposit a cheque and is deciding whether to wait in a long line at the bank on a Friday afternoon, or to come back on Saturday morning when the line would be short [DeRose 2009]. But this question arises: is this bank actually open tomorrow (Saturday) morning? The man visited this bank two Saturdays ago and it was open then, but banks do sometimes change their hours. In the 'low stakes' version of the case, nothing serious hinges on whether he deposits the cheque before the weekend is over, and the man says, 'I know that the bank is open tomorrow.' In the 'high stakes' version of the case, something very serious hinges on whether he deposits the cheque before the weekend is over, and the man says, 'I don't know that the bank is open tomorrow.' Contextualists claim that people will judge that *both* knowledge statements are true. ('Knowledge statement' includes both attributions and denials of knowledge.) That is, competent speakers will judge that the man truthfully says that he 'knows' in

the low stakes version, and that the man truthfully says that he ‘doesn’t know’ in the high stakes version.

For the sake of argument, suppose that people behave as contextualists predict. We might doubt that this is a surprising prediction, because we might expect this behavioural pattern even if contextualism is false. One possibility is that, across the pair of cases, people draw different inferences about whether the agent satisfies the (invariant) requirements of knowledge. It is widely assumed that, in order for you to know that a proposition is true, it must be that the proposition is true, that you have good evidence for the proposition’s truth, and that you think that the proposition is true. Each requirement on knowledge provides a possible explanation for why competent speakers would judge that the agent can truthfully say ‘I know’ in the low case and ‘I don’t know’ in the high case. For example, people might assume that someone who says ‘I don’t know’ is less likely to have the relevant belief (compare Bach [2005] and Nagel [2008]), or to have a *true* belief, or to have good evidence than someone who says ‘I do know.’ This could lead people to judge that the agent knows in the one case but not in the other, which in turn would lead people to judge that the one agent truthfully says ‘I know’, whereas the other agent truthfully says ‘I don’t know.’ Similarly, depending on how serious the situation is, people might draw different inferences about the truth, the agent’s evidence, and what the agent thinks.

Contrary to these suggestions, contextualists have claimed that people’s judgments about the invariant requirements of knowledge do not shift across the pair of cases. For instance, some propose that it is ‘natural’ to interpret the agent as equally confident in the low and high stakes cases [DeRose 2009: 190–3]. Some propose that it is clear that ‘nothing changes’ about the quality of the agent’s evidence across the cases (Lawlor [2013: 81]; see also DeRose [2009: 2]). And some imply that we can or will ‘assume’ that the relevant proposition is true in each case [DeRose 2009: 2].<sup>1</sup> In other words, contextualists propose that the shift in ‘knowledge’ judgments cannot be explained by appealing to widely accepted invariant conditions on knowledge (for example, truth, evidence, belief) because our underlying judgments about these factors do not shift. Thus, something else must explain the shift in ‘knowledge’ judgments, something that does shift across the cases. Contextualists propose a shifting evidential standard associated with ‘know’ to explain the behaviour.

Another possibility is that the shift in ‘knowledge’ judgments is due to a critical confound. The low and high stakes cases differ, not only in how much is at stake, but also in whether the agent says ‘I know’ or ‘I don’t know.’ People might simply defer to others’ self-regarding knowledge statements, regardless of whether the stakes vary. That is, people might assume that others are well positioned to report on their own mental states. Contextualists have not addressed this concern about a ‘deferral’ confound.

Accordingly, at least three questions are important for evaluating the motivation for contextualism. First, do ‘knowledge’ judgments shift across the pair of cases? Second, do judgments about truth, belief, or evidence shift across the pair of cases? Third, is the confound of deferral an innocent imperfection or instead a genuine problem? If ‘knowledge’ judgments shift, the other judgments do not shift, and the confound of deferral is

---

<sup>1</sup> This same assumption is not necessarily shared by previous experimental work on the issue.

an innocent imperfection, then the principal motivation offered for contextualism is supported; otherwise, it is undermined.

Prior research has not addressed the second or third of those questions and has yielded mixed results regarding the first question. Some results seemed inconsistent with contextualist predictions about shifting ‘knowledge’ judgments (for example, Buckwalter [2010, 2014] and Feltz and Zarpentine [2010]; see also May, Sinnott-Armstrong, Hull, and Zimmerman [2010]), but others seemed consistent with the predictions (for example, Hansen and Chemla [2013]; see also Alexander, Gonnerman, and Waterman [2014]). Contextualists have made several methodological objections to these studies, particularly emphasizing the importance of eliciting metalinguistic judgments about a knowledge statement’s truth value and of having the agent say ‘I know’ in the low case and ‘I don’t know’ in the high case [DeRose 2009: 49 n2, 2011]. Others express dissatisfaction at the design of previous work on the topic [Schaffer 2006]. For example, some argue that contextualist thought experiments are poorly constructed because they are supposed to produce similar results across the two cases rather than a difference, leading to the problem of interpreting a null result [Hansen and Chemla 2013: 292, 295]. But this last criticism is incorrect because contextualism does not predict a null result. Instead, contextualism predicts that people will agree with the knowledge statement in both low and high conditions. This requires a significance test against the neutral midpoint for a scaled response (or chance for a dichotomous response). If the results are significantly above midpoint in both conditions, then it vindicates contextualism’s prediction. Consistent with this, agreement might still be significantly higher in one of the conditions. (As observed below in Experiment 1, this actually happens sometimes.)

In this paper, I report a series of behavioural experiments that address all three questions. According to the results, (1) ‘knowledge’ judgments shift in the way that contextualists predict, (2) judgments about truth, belief, and evidence also shift across the pair of cases, and (3) the deferral confound is a genuine problem. Overall, these results undermine the principal motivation offered for contextualism in the literature to date.

## 2. Experiment 1

This experiment tests whether judgments about ‘knowledge’ and potential requirements of knowledge shift across a classic pair of low/high cases.

### 2.1 Method

*Participants.* Two hundred and one participants (aged 19–71 years, mean age = 31 years; 94% reporting English as a native language; 77 female) were tested. Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk and Qualtrics, and compensated \$0.40 for approximately 2–3 minutes of their time. Repeat participation was prevented within and across experiments. Participants were recruited and tested the same way as in all experiments reported here.

*Materials and Procedure.* Participants were randomly assigned to one of two conditions, Low and High, in a between-subjects design. Each participant read a single story. The stories were based on the original pair of ‘bank cases’ [DeRose 2009: 1–2]. Keith and Jane are driving home from work on Friday afternoon. Keith plans to deposit a cheque that they had just received, but then he sees that the line inside the bank is very

long. He says that he will return tomorrow instead of waiting in line. It is not important for the cheque to be cashed before Monday in the low condition, but it is very important in the high condition. Jane questions Keith calmly in the Low condition, but she questions him anxiously in the High condition. After Jane questions him, Keith claims to know that the bank is open tomorrow in the Low condition, but he claims not to know in the high condition. Here is the text for the stories:

(Low/High) Keith and his wife Jane are driving home from work on Friday afternoon. They just received a large check from a client, which Keith plans to deposit in their bank account. It is [not/very] important for him to deposit the check before Monday: [they definitely do/otherwise they won't] have enough money in their account for all their checks to clear. ¶<sup>2</sup> As they drive past the bank, they see that the lines inside are very long. Keith says, 'I hate waiting in line. I'll just come back tomorrow morning instead.' ¶ Jane responds [calmly/anxiously], 'This is really [not/very] important, [but/and] lots of banks are closed on Saturdays. Do you know that our bank is open tomorrow?' ¶ Keith answers, 'It was two Saturdays ago that I went to our bank, and it was open. So, [yes, I do/no, I don't] know that our bank is open tomorrow.'

After reading the story, participants rated their agreement with six statements.

At the end of the story:

- (1) It's true that the bank is open tomorrow.
- (2) Keith believes that the bank is open tomorrow.
- (3) Keith has good evidence that the bank is open tomorrow.
- (4) When Keith said, 'I [do/don't] know', what he said was true.
- (5) The situation is potentially very serious for Keith.
- (6) Keith should come back tomorrow morning instead.

Statement 5 was included as a manipulation check, with the expectation that participants would disagree in the Low condition and agree in the High condition. Statement 4 was included to test whether people agree that the knowledge attribution and the knowledge denial are true. The other four statements were included to test whether judgments about other potential requirements of knowledge also differ across the conditions.

Responses were collected on a standard seven-point Likert scale, 1 ('Strongly Disagree')–7 ('Strongly Agree'), left-to-right on the participant's screen. (Participants never saw numerical labels.) The statements were presented in random order and appeared on the participant's screen all at once in a matrix table, while the story remained at the top of the screen.

After rating the statements, participants proceeded to a new screen, where they answered three comprehension questions from memory (response options rotated randomly):

- (1) Keith and his wife were driving home on \_\_\_\_\_. (Friday/Saturday)
- (2) It was \_\_\_\_\_ ago that Keith was at the bank. (two Saturdays/one Saturday)
- (3) Keith said that he \_\_\_\_\_ know that the bank is open tomorrow. (does/does not)

Finally, participants went to a new screen and completed a brief demographic questionnaire. At no point could participants return to a previous screen. The same is true for all studies reported in this paper.

---

<sup>2</sup> ¶ indicates a paragraph break on the participant's screen.

**Table 1A.** Experiment 1: Mean response (standard deviation in parentheses) to the statements in the Low and High conditions, along with the results from independent samples t-tests. *d* = Cohen's *d*; MD = mean difference; CI = confidence interval.

Measure	Low	High	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	5.08 (1.23)	4.27 (1.12)	4.84	199	<.001	0.81	0.69	0.48	1.13
Belief	6.36 (0.97)	4.99 (1.37)	8.15	176 <sup>a</sup>	<.001	1.37	1.23	1.04	1.70
Evidence	5.66 (1.22)	4.75 (1.53)	4.64	187	<.001	0.91	0.68	0.52	1.30
'Know'	5.02 (1.48)	4.36 (1.81)	2.81	189	.006	0.66	0.41	0.20	1.12
Serious	2.78 (1.89)	5.85 (1.06)	-14.21	160	<.001	-3.06	2.25	-3.49	-2.64
Actionability	4.72 (1.71)	2.26 (1.28)	11.54	187	<.001	2.45	1.69	2.03	2.87

<sup>a</sup>Degrees of freedom (rounded to the nearest whole number) are adjusted lower for some dependent measures because of unequal variance on the measure across conditions.

## 2.2 Results

Participants answered the comprehension questions correctly 97% of the time, indicating that they understood and remembered the story's details extremely well. Participant response to the 'seriousness' statement shows that the experimental manipulation was extremely effective. (See Table 1A.)<sup>3</sup>

Participants in both conditions tended to agree that Keith's self-regarding knowledge statement was true: one sample t-tests, Low,  $t(101) = 6.95$ ,  $p < .001$ , MD = 1.02 [0.73, 1.31],  $d = 0.69$  (medium effect size), test proportion = 4; High,  $t(98) = 2.00$ ,  $p = .048$ , MD = 0.36 [0.01, 0.72],  $d = 0.20$  (small effect size).<sup>4</sup> (See Table 1A for descriptive statistics.) The modal response was 'Agree' (= 6) in the Low condition and 'Somewhat agree' (= 5) in the High condition. Assignment to condition affected whether participants agreed that the bank is open tomorrow, that Keith believes that the bank is open tomorrow, that Keith has good evidence that the bank is open tomorrow, and that Keith should instead come back tomorrow morning.

## 2.3 Conceptual Replication

To ensure that these results are not due to peculiarities of one particular story, I conducted a conceptual replication based on another famous low/high pair from the literature [Cohen 1999]. The experimental design and procedures were exactly the same as for Experiment 1. The basic story featured a man, Stewart, considering taking a flight to Chicago. In each case, he needs a direct flight. In the low stakes version, it is not too important whether the flight is direct: instead of arriving earlier and attending a basketball game, Stewart will arrive later and attend a hockey game. In the high stakes version,

<sup>3</sup> Preliminary tests revealed no main or interaction effects of participant gender on the dependent measures. Preliminary tests revealed two small, unpredicted, main effects of participant age. Older participants (median split = 28 years) were more likely to attribute belief overall, independent samples t-test, younger/older,  $M = 5.45/5.94$ ,  $SD = 1.48/1.20$ ,  $t(194) = -2.60$ ,  $p = .010$ , MD = -0.49 [-0.87, -0.12],  $d = 0.37$  (small effect size). Similarly, older participants were more likely to agree overall that Keith's self-regarding knowledge statement was true,  $M = 4.42/4.99$ ,  $SD = 1.80/1.50$ ,  $t(195) = -2.46$ ,  $p = .015$ , MD = -0.57 [-1.03, -0.11],  $d = 0.35$  (small effect size). I ignore these small and unpredicted demographic effects in the main text.

<sup>4</sup> Cohen's *d* is a standard measure of the size of a mean difference and is calculated by dividing the absolute mean difference by the (pooled) standard deviation. I follow Ellis's [2010] guidelines on interpreting effect sizes for *d*:  $d > 0.20$  is very small;  $0.20 > d > 0.50$  is small;  $0.50 > d > 0.80$  is medium;  $0.80 > d$  is large.

it is very important whether the flight is direct: Stewart is transporting a liver for transplantation, and a non-direct flight will cause it to spoil. Here is the text for the stories:

(Low/High) Stewart and his wife Jill are organ couriers. It's Saturday morning and they are currently [beginning their vacation/transporting a liver]. If they can get a direct flight to Chicago, then they will [go to Chicago to attend a basketball game/deliver the organ to a patient in Chicago]. But it must be a direct flight: if they encounter a layover, then [it will be too late to attend the basketball game, and they will switch their plans and attend a hockey game in Chicago/the organ will spoil before arriving, and they must board a train to Cincinnati and deliver it to a patient there] instead. ¶ As they check the airline's website, they see that it lists a direct flight to Chicago. Stewart says, 'I'll buy tickets and then we'll head straight to the airport.' ¶ Jill responds [calmly/anxiously], 'This is really [not/very] important, [but/and] sometimes the online booking information is inaccurate. Do you know that the flight is direct to Chicago?' ¶ Stewart answers, 'It was two Saturdays ago that I flew this airline to Chicago, and the flight was direct. So, [yes, I do/no, I don't] know that the flight is direct.'

After reading the story, participants answered these questions:

- (1) It's true that the flight is direct.
- (2) Stewart believes that the flight is direct.
- (3) Stewart has good evidence that the flight is direct.
- (4) When Stewart said, 'I [do/don't] know', what he said was true.
- (5) The situation is potentially very serious for Stewart.
- (6) Stewart should buy the plane tickets to Chicago.

The results replicated the findings from Experiment 1. (See Table 1B.) Participants in both conditions tended to agree that Stewart's self-regarding knowledge statement was true: one sample t-tests, Low,  $t(51) = 2.06$ ,  $p = .045$ ; High,  $t(48) = 6.14$ ,  $p < .001$ . The modal response was 'Agree' (= 6) in both conditions.

## 2.4 Discussion

Participants agreed that a knowledge attribution was true in a low stakes case, and that a knowledge denial was true in a high stakes case. However, judgments about several potential invariant requirements of knowledge also differed across the cases. In the high stakes case, participants were more likely to doubt that the relevant proposition is true, they judged the evidence to be worse, they were less willing to attribute belief, and they were more likely to deny that the agent should act.

**Table 1B.** Conceptual replication of Experiment 1: Mean response (standard deviation in parentheses) to the statements in the Low and High conditions, along with the results from independent samples t-tests.

Measure	Low	High	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	4.73 (1.19)	4.24 (1.55)	1.77	99	.079	0.49	0.36	-0.06	1.03
Belief	6.65 (0.57)	4.63 (1.70)	7.92	58	<.001	2.02	2.08	1.51	2.53
Evidence	5.13 (1.51)	4.71 (1.47)	1.42	99	.160	0.42	0.29	-0.17	1.01
'Know'	4.44 (1.55)	5.41 (1.61)	-3.07	99	.003	-0.97	0.61	-1.56	-0.34
Serious	3.25 (1.57)	6.04 (1.37)	-9.50	99	<.001	-2.79	1.91	-3.37	-2.21
Actionability	5.31 (1.25)	4.76 (1.63)	1.92	99	.057	0.55	0.39	-0.01	1.12

### 3. Experiment 2

Familiar contextualist test cases contain a potentially important confound. More specifically, the confound is that the low and high stakes cases differ not only in how much is at stake, but also in the self-regarding knowledge statement made by the agent. It could be that people tend to defer to an agent's own self-regarding knowledge statements, which leads them to agree with the self-regarding knowledge statement in the low stakes case ('I know') and high stakes case ('I don't know').<sup>5</sup> The present experiment tests this *deferral hypothesis* by comparing results from two cases differing *only* in whether the agent says 'I know' or 'I don't know'.

#### 3.1 Method

*Participants.* One hundred and ninety nine new participants (aged 18–65 years, mean age = 31 years; 97% reporting English as a native language; 80 female) were tested. Data were not collected from one person who declined to sign the consent form.

*Materials and Procedure.* Participants were randomly assigned to one of two conditions, Yes and No, in a between-subjects design. The procedures and questions were exactly the same as in Experiment 1. But this time the stories for the two conditions differed only in whether Keith attributes or denies knowledge to himself. Here is the text for the stories:

(Yes/No) Keith and his wife Jane are driving home from work on Friday afternoon. They just received a check from a client, which Keith plans to deposit in their bank account. As they drive past the bank, they see that the lines inside are very long. Keith says, 'I hate waiting in line. I'll just come back tomorrow morning instead.' ¶ Jane asks, 'Do you know that our bank is open tomorrow?' ¶ Keith answers, 'It was two Saturdays ago that I went to our bank, and it was open. So, [yes, I do/no, I don't] know that our bank is open tomorrow.'

#### 3.2 Results

Participants again answered the comprehension questions correctly 97% of the time.<sup>6</sup> In line with the deferral hypothesis, participants in each condition tended to agree with the self-regarding knowledge statement: one sample t-tests, Yes,  $t(97) = 3.44$ ,  $p < .001$ ,  $MD = 0.58$  [0.25, 0.92],  $d = 0.37$  (small effect size), test proportion = 4; No,  $t(100) = 2.74$ ,  $p = .007$ ,  $MD = 0.49$  [0.13, 0.84],  $d = 0.28$  (small effect size). (See Table 2A.) In each condition the modal response to the knowledge statement was 'Somewhat agree' (= 5).

#### 3.3 Conceptual Replication

Again I conducted a conceptual replication based on the story about the direct flight. The procedures and questions were exactly the same as in the conceptual replication for Experiment 1. Here is the text for the stories:

<sup>5</sup> This differs from the 'rule of accommodation' [Lewis 1979]. Deferral might happen because we flexibly interpret an expression's meaning to make it come out true, but a simpler explanation is just that we assume that people tend to be right about their own mental states. Deferral might be related to what researchers have called the 'attributer effect', whereby people more readily agree with first-person knowledge attributions than third-person knowledge attributions [Feltz and Zarpentine 2010: 689].

<sup>6</sup> Preliminary tests revealed no main or interaction effects of participant gender or age on the dependent measures.

**Table 2A.** Experiment 2: Mean response (standard deviation in parentheses) to the statements in the Yes and No conditions along with the results from independent samples t-tests.

Measure	Yes	No	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	5.00 (1.18)	4.64 (1.19)	2.12	197	.035	0.36	0.30	0.03	0.69
Belief	6.51 (0.63)	4.90 (1.49)	9.99	135	<.001	1.61	1.72	1.29	1.93
Evidence	5.48 (1.30)	5.14 (1.33)	1.83	197	.069	0.34	0.26	-0.02	0.71
'Know'	4.58 (1.67)	4.49 (1.78)	0.39	197	.694	0.09	0.06	-0.39	0.58
Serious	2.54 (1.27)	2.50 (1.38)	0.19	197	.849	0.04	0.02	-0.34	0.41
Actionability	4.84 (1.55)	4.41 (1.72)	1.86	197	.065	0.43	0.27	-0.02	0.89

(Yes/No) It's Saturday, and Stewart and his wife Jill are considering whether to fly to Chicago. They want a direct flight. As they check the airline's website, they see that it lists a direct flight to Chicago. Stewart says, 'I'll buy tickets and then we'll head straight to the airport.' ¶ Jill asks, 'Do you know that the flight is direct to Chicago?' ¶ Stewart answers, 'It was two Saturdays ago that I flew this airline to Chicago, and the flight was direct. So, [yes, I do/no, I don't] know that the flight is direct.'

The results replicated the findings from Experiment 2. (See Table 2B.) Participants in each condition tended to agree with the self-regarding knowledge statement: one sample t-tests, Yes,  $t(49) = 2.21$ ,  $p = .032$ ; No,  $t(50) = 2.19$ ,  $p = .033$ . The modal response to the knowledge statement was 'Somewhat agree' (= 5) in the Yes condition and 'Agree' (= 6) in the No condition.

### 3.4 Discussion

In addition to differing in how much is at stake, contextualist test cases also differ in whether the agent self-attributes knowledge or self-denies knowledge. The present experiment tested whether this difference produces the principal datum that contextualists cite as motivation for their view, even when stakes do not vary. The results showed that it did: when the only difference between conditions was that the agent said either 'I do know' or 'I don't know', participants were equally likely to agree with the statement. Participants in both conditions were equally likely to judge the situation unserious, and the differences in ratings of truth, actionability, and evidence were all greatly diminished from those observed in Experiment 1. The difference in belief attribution across conditions remained large, so deferral might go through belief attribution. Another possibility is that deferral is merely an instance of a more general agreement bias, whereby people tend to endorse assertions (Krosnick [1999: 552]; see

**Table 2B.** Conceptual replication of Experiment 2: Mean response (standard deviation in parentheses) to the statements in the Yes and No conditions along with the results from independent samples t-tests.

Measure	Yes	No	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	4.98 (1.62)	4.96 (1.51)	0.06	99	.951	0.02	0.00	-0.60	0.64
Belief	6.50 (0.79)	4.90 (1.58)	6.46	74	<.001	1.60	1.50	1.11	2.09
Evidence	4.82 (1.67)	5.18 (1.42)	-1.15	99	.252	-0.36	0.23	-0.97	0.26
'Know'	4.54 (1.73)	4.61 (1.98)	-0.18	99	.855	-0.68	0.04	-0.80	0.67
Serious	3.92 (1.40)	3.51 (1.53)	1.41	99	.163	.410	0.28	-0.17	0.99
Actionability	5.12 (1.35)	5.25 (1.41)	-0.49	99	.625	-0.14	0.10	-0.68	0.41

also Gilbert, Krull, and Malone [1990]). Further investigation into why deferral occurs is warranted but unimportant for present purposes. For present purposes, the crucial point is that the confound of deferral is a genuine problem for contextualist test cases.

## 4. Experiment 3

Instead of asking participants to rate whether an agent's self-regarding knowledge statement is true, in this experiment I asked participants to recommend whether the agent should self-attribute knowledge in order to speak truthfully. Thus participants could not simply defer to the agent's self-regarding knowledge statement, because the agent makes no such statement, thereby avoiding the deferral confound.

### 4.1 Method

*Participants.* Two hundred and two participants (aged 18–63 years, mean age = 30 years; 95% reporting English as a native language; 78 female) were tested.

*Materials and Procedure.* Participants were randomly assigned to one of two conditions, Low and High, in a between-subjects design. Each participant read a single story very similar to the ones used in Experiment 1. The difference is that this time the story did not end with Keith making a self-regarding knowledge statement. Instead, it ended when Keith says, 'It was two Saturdays ago that I went to our bank, and it was open.' After reading the story, participants were instructed, 'Keith wants to correctly answer Jane's question. In light of that, please rate your agreement with the following statement.' Here is the test statement:

In order for Keith's answer to be true, he should say, 'Yes, I do know.'

Participants then went to a new screen and rated their agreement with the same statements about truth, belief, evidence, seriousness, and actionability from Experiment 1, in the same way as in Experiment 1. The question about seriousness was included as a manipulation check, with the expectation that participants would disagree in the Low condition and agree in the High condition. After rating the statements, on a new screen participants answered a comprehension question from memory:

Keith and his wife were driving home on \_\_\_\_\_. (Friday/Saturday)

### 4.2 Results

Participants answered the comprehension question correctly 96% of the time.<sup>7</sup> Response to the 'seriousness' statement again showed that the experimental manipulation was

<sup>7</sup> Preliminary tests revealed no main effects of participant age or gender on responses to the knowledge statement. Participant gender had a small unpredicted main effect on quality of evidence, with women tending to rate the evidence better than men did: men/women,  $M = 5.25/5.64$ ,  $SD = 1.24/1.14$ ,  $t(200) = -2.24$ ,  $p = .026$ ,  $MD = -0.39 [-0.74, -0.05]$ ,  $d = 0.32$ . Participant gender and assignment to condition interacted to produce a medium unpredicted effect on response to the truth statement. Men and women rated the truth statement similarly in the High condition, but men rated it lower in the Low condition,  $M = 4.85/5.58$ ,  $t(100) = -2.76$ ,  $p = .007$ ,  $MD = -0.72 [-1.24, -0.20]$ ,  $d = 0.55$ . Participant age and assignment to condition interacted to produce an unpredicted small effect on response to the seriousness statement. Older and younger participants rated the seriousness of the situation similarly in the Low condition, but older participants rated the seriousness higher in the High condition: older/younger (median split = 27 years),  $M = 5.52/6.02$ ,  $SD = 1.35/0.96$ ,  $t(92) = -2.16$ ,  $p = .034$ ,  $MD = -0.50 [-0.96, -0.04]$ ,  $d = 0.43$ .

**Table 3A.** Experiment 3: Mean response (standard deviation in parentheses) to the statements in the Low and High conditions, along with the results from independent samples t-tests.

Measure	Low	High	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	5.14 (1.33)	4.75 (1.19)	2.18	200	.030	0.39	0.31	0.04	0.74
Belief	6.27 (0.77)	6.16 (0.88)	0.98	200	.328	0.11	0.13	-0.12	0.35
Evidence	5.59 (1.20)	5.21 (1.22)	2.23	200	.027	0.38	0.32	0.04	0.71
'Know'	4.79 (1.72)	4.64 (1.68)	0.64	200	.520	0.15	0.09	-0.32	0.63
Serious	2.48 (1.50)	5.76 (1.20)	-17.18	192	<.001	-3.28	2.48	-3.66	-2.90
Actionability	5.00 (1.49)	3.00 (1.73)	8.81	200	<.001	2.00	1.25	1.55	2.48

extremely effective. (See Table 3A.) The principal question is that of whether participants agreed more strongly in the Low condition that Keith should self-attribute knowledge in order to speak truthfully. Participants in both conditions were equally likely to agree that Keith should self-attribute knowledge in order to speak truthfully. (See Table 3A.) Mean response in each condition was significantly above the neutral midpoint (= 4): one sample t-tests, Low,  $t(101) = 4.66$ ,  $p < .001$ ,  $MD = 0.79$  [0.46, 1.13],  $d = 0.46$  (small effect size); High,  $t(99) = 3.81$ ,  $p < .001$ ,  $MD = 0.64$  [0.31, 0.97],  $d = 0.38$  (small effect size). The modal response in each condition was 'Agree' (= 6).

### 4.3 Conceptual Replication

Again I conducted a conceptual replication based on the story about the direct flight. The procedures and questions were almost the same as in the conceptual replication for Experiment 1. The difference is that this time the story did not end with Stewart making a self-regarding knowledge statement. Instead, it ended when he says, 'It was two Saturdays ago that I flew this airline to Chicago, and the flight was direct.' After reading the story, participants were instructed, 'Stewart wants to correctly answer Jill's question. Please rate your agreement with the following statement.' Here is the test statement:

In order for Stewart's answer to be true, he should say, 'Yes, I do know.'

Participants then went to a new screen and rated their agreement with the same statements about truth, belief, evidence, seriousness, and actionability from the conceptual replication of Experiment 1.

The results replicated the findings from Experiment 3. (See Table 3B.) Participants in both conditions were equally likely to agree that Stewart should self-attribute knowledge in order to speak truthfully. (See Table 3B.) Mean response in each condition was

**Table 3B.** Conceptual replication of Experiment 3: Mean response (standard deviation in parentheses) to the statements in Low and High conditions, along with the results from independent samples t-tests.

Measure	Low	High	t	df	p	MD	d	95% CI for MD	
								LLCI	ULCI
Truth	4.54 (1.01)	4.71 (1.19)	-0.75	99	.453	-0.17	0.15	-0.60	0.27
Belief	6.16 (1.00)	6.27 (0.67)	-0.68	99	.498	-0.12	0.13	-0.45	0.22
Evidence	4.86 (1.34)	5.41 (1.24)	-2.14	99	.035	-0.55	0.43	-1.06	-0.04
'Know'	4.90 (1.61)	4.76 (1.72)	0.41	99	.684	0.14	0.08	-0.52	0.79
Serious	4.02 (1.68)	6.14 (0.75)	-8.14	67	<.001	-2.12	1.99	-2.64	-1.60
Actionability	5.14 (1.25)	5.16 (1.48)	-0.06	99	.951	-0.02	0.01	-0.56	0.52

significantly above the neutral midpoint ( $= 4$ ): one sample t-tests, Low,  $t(49) = 3.96$ ,  $p < .001$ ; High  $t(50) = 3.18$ ,  $p = .003$ . The modal response was ‘Somewhat Agree’ ( $= 5$ ) in Low and ‘Agree’ ( $= 6$ ) in High.

#### 4.4 Discussion

This experiment tried to find evidence for contextualism while avoiding the confound of deferral. Instead of retrospectively rating an agent’s self-regarding knowledge statement, participants prospectively recommended whether the agent should self-attribute knowledge in order to speak truthfully. The results did not support contextualism. Regardless of whether the agent was in a high stakes or instead in a low stakes situation, participants were equally likely to recommend the self-attribution.

Before moving on, it is worth addressing one issue. In Experiment 1, participants in the high stakes condition tended to agree that the agent spoke truthfully when he said ‘I don’t know.’ But in Experiment 3 participants in the high stakes condition recommended saying ‘I do know’ in order to speak truthfully. These results are not inconsistent, because the two experiments differed in whether the agent self-denies knowledge. (Experiment 2 showed that people defer to self-regarding knowledge statements.)

### 5. Experiment 4

The present experiment attempts to detect evidence for contextualism in a new way. More specifically, I asked participants about the strength of evidence required to ‘know’.

#### 5.1 Method

*Participants.* Ninety nine new participants (aged 18–74 years, mean age = 31 years; 97% reporting English as a native language; 41 female) were tested. Data were not collected from one person who declined to sign the consent form.

*Materials and Procedure.* Participants were randomly assigned to one of two conditions, Low and High, in a between-subjects design. The procedures and stories for the conditions were very similar to Experiment 1, except that this time the story ends with Jane posing a question to Keith about evidence and knowledge. The stories were the same as in Experiment 1 through the point where Jane asks, ‘Do you know that our bank is open tomorrow?’ At that point, each story ends as follows:

Jane continues, ‘This actually raises a more general question I’ve been considering. On a scale of 1 to 10, with 10 being the highest, how strong must your evidence be in order to know that the bank is open tomorrow?’

After reading the story, participants were instructed, ‘Keith wants to correctly answer Jane’s question. In order for Keith’s answer to be true, what should he say?’ Beneath the instructions was an open sentence to complete:

‘Knowing requires evidence that rates \_\_\_\_\_ on the scale.’

Responses were collected on a ten-point scale, 1–10, left-to-right on the participant’s screen. Participants then went to a new screen and answered a comprehension question from memory:

Keith and his wife were driving home on \_\_\_\_\_. (Friday/Saturday)

## 5.2 Results

Participants answered the comprehension question correctly 92% of the time.<sup>8</sup> The critical question is that of whether assignment to condition affected response to the question about evidence. If the standards for knowledge shift across context, then mean response should be higher in the High condition. That is, participants in the High condition should think that Keith should say that knowledge requires stronger evidence. Contrary to that prediction, there was no effect of condition, independent samples t-test, High/Low,  $M = 8.62/8.31$ ,  $SD = 1.98/1.86$   $t(97) = -0.81$ ,  $p = .418$ , n.s.

## 5.3 Conceptual Replication

Again I conducted a conceptual replication with the story about the direct flight. The procedures and stories for the conditions were very similar to the conceptual replication of Experiment 1, except that this time the story ends with Jill posing a question to Stewart about evidence and knowledge. The stories were the same up through the point where Jill asks, ‘Do you know that the flight is direct to Chicago?’ Then each story ends as follows:

Jill continues, ‘This actually raises a more general question I’ve been considering. On a scale of 1 to 10, with 10 being the highest, how strong must your evidence be in order to know that the flight is direct?’

After reading the story, participants were instructed, ‘Stewart wants to correctly answer Jill’s question. In order for Stewart’s answer to be true, what should he say?’ Beneath the instructions was an open sentence to complete:

‘Knowing requires evidence that rates \_\_\_\_\_ on the scale.’

The results replicated the findings from Experiment 4. There was no effect of condition on response to the test question, independent samples t-test, High/Low,  $M = 9.08/8.59$ ,  $SD = 1.29/1.56$ ,  $t(98) = 1.69$ ,  $p = .093$ ,  $d = 0.34$ .

## 5.4 Discussion

If contextualism is true, then people should think that, in order to speak truthfully, someone in a high stakes situation must say that the evidential standard for knowledge is higher than someone in a low stakes situation will say that it is. The results did not support this prediction. Regardless of whether the agent was in a high stakes or instead in a low stakes situation, participants recommended identifying a similar evidential standard. The small numerical difference between the two situations in the conceptual replication (Experiment 4B) reached what is sometimes called a ‘marginal’ or ‘trending’ level of statistical significance. Perhaps future work could build on this to detect evidence motivating contextualism.

## 6. Conclusion

Epistemic contextualism is principally motivated by a set of empirical claims about linguistic intuitions and behaviour. It has its ‘basis in ordinary language’, which allegedly

---

<sup>8</sup> Neither participant gender nor age affected response to the test question about evidence.

provides ‘evidence of the very best type’ that the standards of ‘knowledge’ are context sensitive [DeRose 2009: 48]. The results from four experiments show that some but not all contextualist behavioural predictions were correct. Contextualists correctly predicted that people will agree when someone says ‘I know’ in a low stakes situation, and when someone says ‘I don’t know’ in a related high stakes situation. However, people also draw different inferences about (what are widely agreed to be) invariant requirements of knowledge, including whether the target proposition is true, whether the agent believes the target proposition, the quality of the agent’s evidence, and how the agent should act. Accordingly, rejecting contextualism does not force one to ‘resist the intuitions’ that lead people to endorse different self-regarding knowledge statements in the low and in the high stakes cases (cf. DeRose [2011: 89]).

Moreover, contextualist test cases contain a critical confound. The confound is that the low and high stakes cases also differ in whether the agent says ‘I know’ or ‘I don’t know.’ However, even when the stakes are held constant, people tend to agree when the agent says ‘I know’ and when the agent says ‘I don’t know’ (Experiment 2). Some contextualists have insisted that, when testing their view, it is critical to include explicit utterances of ‘I know’ and ‘I don’t know’, because this is ‘a significant part of what’s behind the intuitions [that contextualists] actually appeal to’ [ibid.: 86–7]. However, they failed to realize that this difference alone is enough to produce the relevant pattern of agreement with ‘knowledge’ statements.

In short, the principal extant motivation for contextualism fails. Contextualists still owe us a distinguishing prediction of their view, something we would confidently expect only if contextualism were true, or which contextualism seems uniquely suited to explain. In the absence of such a prediction, contextualism is an idle hypothesis and we should not accept it. Of course, it is consistent with my findings that such a prediction will be made and vindicated. The present research was not designed to test all possible such predictions, and future work exploring this possibility should not be ruled out (for a potential starting point, see the marginally significant difference observed in the conceptual replication of Experiment 4).

Some might be tempted to object that my results demonstrate only one thing: we cannot count on ordinary people to properly assess cases. Instead, we should trust the expert intuitions of trained philosophers, some of whom detect the relevant nuances and respond in line with contextualist predictions. In response, this objection is unavailable to leading contextualists, whose ‘main positive argument’ is based on ‘straightforward data concerning’ the ordinary behaviour of competent speakers [DeRose 2009: 69]. Settling for this objection would amount to a transparent and feeble bait-and-switch. We were sold a batch of ‘ordinary language’ goods, but all that we got for our money was a raincheck on some ‘expert’ intuitions.

The current findings suggest some general lessons for philosophers, including some critical improvements to standard methodology. First, for any discussion informed by patterns in judgments, including knowledge judgments, philosophers should not ignore basic tendencies in human cognition. For example, it appears that, other things being equal, we defer to others’ self-regarding mental state attributions. Accordingly, we should not be impressed if a philosophical hypothesis predicts this deferential pattern. Fulfilling this prediction does not count in the theory’s favour. It is an idle prediction.

Second, philosophers should remember that judgments about cases are influenced not only by explicitly stated details, but also by the inferences that people draw from them. For instance, suppose that increasing what is at stake makes people more likely

to deny knowledge. We should not automatically interpret this as a sign that knowledge (or ‘knowledge’, or our concept of knowledge) is partly constituted by stakes (or variable evidential standards or truth conditions), in addition to factors such as truth, belief, and evidence. For variation in underlying judgments about truth, belief, or evidence could cause the change in knowledge judgments [Turri and Buckwalter *forthcoming*]. And given how much of human cognition occurs automatically and unconsciously [Wilson 2002; Haidt 2007; Nosek, Hawkins, and Frazier 2011; Bargh, Schwader, Hailey, Dyer, and Boothby 2012], philosophers should not assume that we do not make such underlying judgments just because we do not introspect them. Relatedly, theorists proposing cases cannot, by stipulation, magically cause us to ‘hold fixed’ our underlying judgments and then legitimately proclaim that a difference must be due to some other factor.

Finally, for discussions hinging on patterns in ordinary usage, philosophers should be quick to rely on controlled experiments to separate wheat from chaff. This is especially true early in a research program, when experimental results can provide formative clues putting researchers on more promising paths. An enormous amount of time and energy has been devoted to contextualism in epistemology over the past two decades. In that time, it went from a novelty to a standard topic covered in textbooks, handbooks, anthologies, encyclopaedias, and survey courses (for example, Fumerton [2006], Sosa, Kim, Fantl, and McGrath [2008], BonJour [2010], Dancy, Sosa, and Steup [2010], Rysiew [2011], Steup, Turri, and Sosa [2013], Pritchard [2014], and Turri [2014]). Philosophers lengthily debated the finer details of contextualism, its relationship to other complicated issues in epistemology and the philosophy of language, and alternative non-contextualist explanations of alleged ‘data’. For instance, which model of context sensitivity best fits ‘knows’ (for example, Cohen [1988, 1998], Rieber [1998], Kompa [2005], Blome-Tillman [2008], DeRose [2008], Ichikawa [2011], and McKenna [2013])? Does contextualism cohere with the best theory of epistemic modality (for example, DeRose [1991] and Egan, Hawthorne, and Weatherson [2005])? What is contextualism’s relationship to views on the norm of assertion (for example, DeRose [2002] and Turri [2010])? Is contextualism preferable to non-contextualist views that relate knowledge itself (rather than ‘knowledge’) to practical factors, such as the propriety of action (for example, Hawthorne [2004], Stanley [2005], Fantl and McGrath [2009], and DeRose [2009])? However, all such questions are idle if contextualism itself is an idle hypothesis in the first place. Moving forward, some basic preliminary experiments can help to avoid similar false starts.<sup>9</sup>

## References

- Alexander, J., C. Gonnerman, and J. Waterman 2014. Salience and Epistemic Egocentrism: An Empirical Study, in *Advances in Experimental Epistemology*, ed. J.R. Beebe, London: Bloomsbury Academic: 97–117.
- Bach, K. 2005. The Emperor’s New ‘Knows’, in *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, ed. G. Preyer and G. Peter, Oxford: Oxford University Press: 51–89.

<sup>9</sup> *Acknowledgments.* For helpful feedback and discussion, I thank Michael Blome-Tillman, Wesley Buckwalter, Adam Feltz, Josh Knobe, Robin McKenna, David Rose, Patrick Rysiew, and Angelo Turri. This research was supported by the Social Sciences and Humanities Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation.

- Bargh, J.A., K.L. Schwader, S.E. Hailey, R.L. Dyer, and E.J. Boothby 2012. Automaticity in Social-Cognitive Processes, *Trends in Cognitive Sciences* 16/12: 593–605.
- Blome-Tillmann, M. 2008. The Indexicality of 'Knowledge', *Philosophical Studies* 138/1: 29–53.
- BonJour, L. 2010. *Epistemology: Classical Problems, Contemporary Responses*, 2<sup>nd</sup> edn, Lanham, MD: Rowman & Littlefield.
- Buckwalter, W. 2010. Knowledge Isn't Closed on Saturday: A Study in Ordinary Language, *Review of Philosophy and Psychology* 1/3: 395–406.
- Buckwalter, W. 2014. Salience and Epistemic Egocentrism: An Empirical Study, in *Advances in Experimental Epistemology*, ed. J.R. Beebe, London: Bloomsbury Academic: 145–73.
- Cohen, S. 1988. How to be a Fallibilist, *Philosophical Perspectives* 2: 91–123.
- Cohen, S. 1999. Contextualism, Skepticism, and the Structure of Reasons, *Philosophical Perspectives* 13: 57–89.
- Cohen, S. 2013. Contextualism Defended, in *Contemporary Debates in Epistemology*, 2<sup>nd</sup> edn, eds. M. Steup, J. Turri, and E. Sosa, Malden, MA: Wiley-Blackwell: 69–75.
- Dancy, J., E. Sosa, and M. Steup eds, 2010. *A Companion to Epistemology*, 2<sup>nd</sup> edn, Malden, MA: Wiley-Blackwell.
- DeRose, K. 1991. Epistemic Possibilities, *The Philosophical Review*, 100/4: 583–605.
- DeRose, K. 2002. Assertion, Knowledge, and Context, *The Philosophical Review*, 111/2: 167–203.
- DeRose, K. 2008. Gradable Adjectives: A Defence of Pluralism, *Australasian Journal of Philosophy*, 86/1: 141–60.
- DeRose, K. 2009. *The Case for Contextualism: Knowledge, Skepticism, and Context*, Vol. 1, Oxford: Clarendon Press.
- DeRose, K. 2011. Contextualism, Contrastivism, and X-Phi Surveys, *Philosophical Studies*, 156: 81–110.
- Egan, A., J. Hawthorne, and B. Weatherson 2005. Epistemic Modals in Context, in *Contextualism in Philosophy*, ed. G. Preyer and G. Peter, Oxford: Clarendon Press: 131–70.
- Ellis, P.D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*, Cambridge: Cambridge University Press.
- Fantl, J. and M. McGrath 2009. *Knowledge in an Uncertain World*, Oxford: Oxford University Press.
- Feltz, A. and C. Zarpentine 2010. Do You Know More When it Matters Less? *Philosophical Psychology*, 23/5: 683–706.
- Fumerton, R. 2006. *Epistemology*, Malden, MA: Blackwell.
- Gilbert, D.T., D.S. Krull, and P.S. Malone 1990. Unbelieving the Unbelievable: Some Problems in the Rejection of False Information, *Journal of Personality and Social Psychology*, 59/4: 601–13.
- Haidt, J. 2007. The New Synthesis in Moral Psychology, *Science*, 316/5827: 998–1002.
- Hansen, N., and E. Chemla 2013. Experimenting on Contextualism, *Mind & Language*, 28/3: 286–321.
- Hawthorne, J. 2004. *Knowledge and Lotteries*, Oxford: Oxford University Press.
- Ichikawa, J. 2011. Quantifiers and Epistemic Contextualism, *Philosophical Studies*, 155/3: 383–98.
- Kompa, N. 2005. The Semantics of Knowledge Attributions, *Acta Analytica*, 20/1: 16–28.
- Krosnick, J.A. 1999. Survey Research, *Annual Review of Psychology*, 50: 537–567.
- Lawlor, K. 2013. *Assurance: An Austinian View of Knowledge and Knowledge Claims*, Oxford: Oxford University Press.
- Lewis, D. 1979. Scorekeeping in a Language Game, *Journal of Philosophical Logic*, 8/1: 339–59.
- Lewis, D. 1996. Elusive Knowledge, *Australasian Journal of Philosophy*, 74/4: 549–67.
- May, J., W. Sinnott-Armstrong, J.G. Hull, and A. Zimmerman 2010. Practical Interests, Relevant Alternatives, and Knowledge Attributions: An Empirical Study, *The Review of Philosophy and Psychology*, 1/2: 265–73.
- McKenna, R. 2013. Epistemic Contextualism: A Normative Approach, *Pacific Philosophical Quarterly*, 94/1: 101–23.
- Nagel, J. 2008. Knowledge Ascriptions and the Psychological Consequences of Changing Stakes, *Australasian Journal of Philosophy*, 86/2: 279–94.
- Nosek, B.A., C.B. Hawkins, and R.S. Frazier 2011. Implicit Social Cognition: From Measures to Mechanisms, *Trends in Cognitive Sciences*, 15/4: 152–59.
- Pritchard, D. 2014. *What Is this Thing Called Knowledge?*, 3<sup>rd</sup> edn, Abingdon: Routledge.
- Rieber, S. 1998. Skepticism and Contrastive Explanation, *Noûs*, 32/2: 189–204.
- Rysiew, P. 2011. Epistemic Contextualism, *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), ed. E.N. Zalta, URL = <http://plato.stanford.edu/entries/contextualism-epistemology>

- Schaffer, J. 2006. The Irrelevance of the Subject: Against Subject-Sensitive Invariantism, *Philosophical Studies*, 127/1: 87–107.
- Sosa, E., J. Kim, J. Fantl, and M. McGrath, eds, 2008. *Epistemology: An Anthology*, 2<sup>nd</sup> edn, Malden, MA: Blackwell.
- Stanley, J. 2005. *Knowledge and Practical Interests*, Oxford: Clarendon Press.
- Steup, M., J. Turri, and E. Sosa, eds, 2013. *Contemporary Debates in Epistemology*, 2<sup>nd</sup> edn, Malden, MA: Wiley-Blackwell.
- Turri, J. 2010. Epistemic Invariantism and Speech Act Contextualism, *The Philosophical Review*, 119/1: 77–95.
- Turri, J. 2014. *Epistemology: A Guide*, Malden, MA: Wiley-Blackwell.
- Turri, J. and W. Buckwalter forthcoming. Descartes's Schism, Locke's Reunion: Completing the Pragmatic Turn in Epistemology, *American Philosophical Quarterly*.
- Wilson, T.D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Cambridge, MA: Harvard University Press.