University of Reading

Third-person knowledge ascriptions: a crucial experiment for contextualism

Article

Accepted Version

Grindrod, J. ORCID: https://orcid.org/0000-0001-8684-974X, Andow, J. and Hansen, N. ORCID: https://orcid.org/0000-0001-5074-1075 (2019) Third-person knowledge ascriptions: a crucial experiment for contextualism. Mind and Language, 34 (2). pp. 158-182. ISSN 1468-0017 doi: 10.1111/mila.12196 Available at https://centaur.reading.ac.uk/73232/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1111/mila.12196

Publisher: Wiley-Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

University of Reading

Central Archive at the University of Reading

Reading's research outputs online

Third-Person Knowledge Ascriptions: A Crucial Experiment for Contextualism

Jumbly Grindrod
University of Reading

James Andow
University of Reading

Nat Hansen
University of Reading

Abstract

In the past few years there has been a turn towards evaluating the empirical foundation of epistemic contextualism using formal (rather than armchair) experimental methods. By-and-large, the results of these experiments have not supported the original motivation for epistemic contextualism. That is partly because experiments have only uncovered effects of changing context on knowledge ascriptions in limited experimental circumstances (when contrast is present, for example), and partly because existing experiments have not been designed to distinguish between contextualism and one of its main competing theories, subject-sensitive invariantism. In this paper, we discuss how a particular, "third-person", experimental design is needed to provide evidence that would support contextualism over subject-sensitive invariantism. In spite of the theoretical significance of third-person knowledge ascriptions for debates surrounding contextualism, no formal experiments evaluating such ascriptions that assess contextualist claims have previously been conducted. In this paper, we conduct an experiment specifically designed to examine that central gap in contextualism's empirical foundation. The results of our experiment provide crucial support for epistemic contextualism over subject-sensitive invariantism.[1]

1. The Empirical Foundation of Contextualism

Epistemic contextualism is the view that the truth-conditional content of knowledge ascriptions can vary across different contexts of utterance.[2] There are roughly two distinct arguments in support of the view. First, epistemic contextualism is often thought to provide an attractive resolution to certain skeptical puzzles (Cohen 1988; DeRose 1995; Lewis 1996). Secondly, epistemic contextualism is supposed to provide an accurate account of how we ordinarily make and understand knowledge ascriptions. Keith DeRose, for instance, explicitly states that this constitutes the primary reason for thinking that epistemic contextualism is true. He states:

> [t]he best grounds for accepting contextualism come from how knowledge-attributing (and knowledge-denying) sentences are used in ordinary, non-philosophical talk: What ordinary speakers will count as "knowledge" in some non-philosophical contexts they will deny is such in others. (DeRose 2009, p. 47)

In support of this claim, DeRose and others employ *context-shifting experiments*. Such experiments construct a pair of contexts that differ only in terms of one or two contextual parameters. When the same sentence is used to make knowledge ascriptions in two contexts that differ in terms of

[1] This paper has been greatly improved by comments and suggestions from the following people: Mark Phelan, Mikkel Gerken, Eliot Michaelson, Emmanuel Chemla, Kathryn Francis, and several anonymous referees. This research was supported by Leverhulme Research Project Grant RPG-2016-193 and by an External Faculty Fellowship at Stanford University's Humanities Center.

[2] This is typically understood as a semantic claim that the variability in the truth conditions of knowledge ascriptions is explained in terms of the context-sensitive semantics of the word "know[s]". However, some contextualists have argued that the context-sensitivity of knowledge ascriptions is due to the pragmatic modulation of what is said (Stainton 2010; Pynn 2015). We are focusing on the empirical foundation that either understanding of contextualism relies upon, and so we remain neutral regarding these competing theoretical positions that concern the sources of context sensitivity of knowledge ascriptions.

these parameters, contextualists claim that judgments about whether the knowledge ascription is true or false will differ across the two contexts (see Table 1).

Table 1: Standard contextualist predictions for context shifting experiments

Sentence uttered	Low context	High context
P (e.g., "I know that the bank will be open tomorrow")	(MORE) TRUE	(MORE) FALSE[3]

Contextualists argue that the best explanation of the variation in truth-value judgments on display in context shifting experiments is that the truth-conditional content of knowledge ascriptions differs across the two contexts.

An alternative design for context shifting experiments (advocated by DeRose 2009, pp. 49-54), varies both certain contextual parameters (usually the stakes and mentioned possibilities of error) and whether a particular sentence or its negation is used to make a knowledge ascription.[4] In DeRose's favored design, contextualists claim that truth-value judgments remain the same in the two contexts, even though the knowledge ascriptions made in the two contexts seem to contradict each other (see Table 2).

Table 2: DeRose's predictions for the original bank case

Sentence uttered	Low context	High context
P (e.g., "I know the bank will be open tomorrow")	TRUE	
~P (e.g., "I don't know the bank will be open tomorrow")		TRUE[5]

The most discussed context-shifting experiment taken to support epistemic contextualism is DeRose's (1992) and Stanley's (2005) bank case, which employs the second design. The original version of the case has the following structure: In the "low standards" context, where nothing important is at stake, DeRose says that he knows the bank where he and his wife deposit their pay checks will be open tomorrow. In the "high standards" context, where a lot is at stake and DeRose's wife mentions the possibility that the bank may not be open tomorrow, DeRose says that he doesn't know the bank will be open tomorrow.[6] DeRose claims that the positive first-

[3] The armchair truth value judgments that were the original empirical foundation for contextualism were presented as binary truth value judgments. But once formal experiments are conducted to evaluate the contextualists' predictions, it is *tendencies* to judge claims true or false that becomes the empirical foundation of contextualist theories. The parenthetical "(more)" in Table 1 is meant to capture the fact that in formal experimental settings, evidence that conforms with contextualist predictions will register *differences* in tendencies to make truth value judgments across contexts.

[4] The significance of adopting one or other of these two experimental designs will be discussed in greater depth in §3.2.

[5] DeRose's preferred design predicts that participants in a formal experiment will have responses that are significantly above the midpoint on a scalar response, or above chance on a binary response (Turri 2016, p. 143).

[6] According to DeRose, even in the high standards case, the speaker remains as confident as he was before that the bank will be open tomorrow (2009, p. 2). It is important that the confidence of the putative knower not change

person knowledge ascription in the low standards context *and* the first-person denial of knowledge in the high standards context both seem true (see Table 2). He argues that the contribution that the verb "knows" makes to the truth conditions of the knowledge ascription varies across the two contexts, which explains why the apparent contradiction is merely apparent.

For DeRose's argument about the-truth conditional contribution of "knows" to go through, it is essential (though not sufficient) that competent speakers' truth value judgments regarding the knowledge ascriptions are actually as DeRose describes them. However, subsequent experimental findings have cast some doubt on this claim. In what follows, we give an overview of the experimental work that has sought to test the context-shifting experiments that support epistemic contextualism.

2. Experimental Work on Epistemic Contextualism

There have been two waves of experimental research that bear on epistemic contextualism, with DeRose's (2011) assessment of existing experimental investigations of contextualism marking the break between the two waves. Around 2010, a cluster of papers was published that provided experimental findings regarding folk intuitions on cases similar to the bank case (Feltz and Zarpentine 2010; May et al. 2010; Buckwalter 2010). Schaffer and Knobe (2012) cite these studies as showing that the armchair judgments that up to that point had been the empirical foundation for the debate over contextualism "do not survive empirical scrutiny" (p. 31).[7]

While DeRose (2011) offers a battery of arguments for discounting the results of these three studies as presenting a serious problem for the contextualist, another consideration concerning the power of these studies to detect relevant contextual effects can be used to make a similar point. As Hansen and Chemla (2013) put it, 'an absence of difference cannot establish that the difference does not exist, unless one also proves the counterfactual claim that the experiment would have been sufficiently powerful to detect it' (p. 292). With that in mind, it is worth considering that (a) some of the studies in this first wave of experiments did find some evidence of small effects of stakes (e.g., May et al. 2010, p. 270) and (b) the statistical power of these studies to detect meaningful contextual effects was limited.[8]

Taking into consideration DeRose's methodological criticisms of the first wave of studies, Hansen and Chemla (2013) conducted an experiment that tested four different first-person knowledge scenarios, as well as a number of other context-shifting experiments to compare the epistemic case with other cases of purported context-sensitivity. Hansen and Chemla did find a significant contextual effect on truth value judgments in the knowledge scenarios (as well as all the other cases they tested) (pp. 304-5). However, the contextual effect disappeared when participants had only seen one of each type of scenario (thereby indicating that the contextual effect might be dependent on the presence of *contrast*, or be some other form of experimental

---

between the two contexts, because that would be a potential confound for the contextualist explanation of the change in truth values of the knowledge claim (Nagel, 2010, p. 421; Pinillos and Simpson 2014). It turns out to be hard to hold the confidence of the speaker fixed in context shifting experiments involving first person knowledge ascriptions (see DeRose 2009, p. 191-2 for discussion of this issue, and Turri 2016, for relevant experimental evidence), which is another reason to prefer third person knowledge ascriptions over first person knowledge ascriptions: the confidence of the ascriber can vary across the different contexts, while the confidence of the ascribee remains fixed.

[7] For a more recent review of the empirical literature which is sympathetic to this assessment, see Buckwalter (2017).

[8] Feltz and Zarpentine's (2010) four experiments had sufficient power (at the .95 level) only to detect effects of stakes with a size of d=.87, d = .82, d=.81, and d=.61 or higher respectively (where the conventional threshold for medium effects is d=.5 and for large effects d=.8). May et al.'s (2010) between-subjects experiment had sufficient power only to detect an effect with a size of d=.47 or higher (the conventional threshold for a small effect is d=.25). Buckwalter's experiment had sufficient power only to detect an effect with a size of d =.38 or higher. (Power analyses conducted using G*Power 3.1.9.2.)

artefact). Nevertheless, Hansen and Chemla's results are consistent with contextualist predictions, and some theorists have taken their results to lend positive support for contextualism (Pynn, 2016). But such a conclusion is hasty. Further work has to be done before experimental results can support epistemic contextualism over its competitors, for two reasons.

First, it is worth emphasising a key difference between epistemic contextualism and a rival view: subject-sensitive invariantism (SSI). SSI is the view that knowledge is a relation between a subject, a proposition, and some feature of the subject's (rather than the ascriber's) context, including the practical stakes that the subject has in being right about that proposition (Fantl and McGrath 2002, Hawthorne 2004, Stanley 2005).[9] As such, according to SSI, whether a subject knows a proposition can vary across contexts in which the stakes for that subject vary. For this reason, SSI can explain the contextualist intuitions in first-person cases like the bank case, because the stakes for the subject do differ across the high and low contexts. For first-personal knowledge ascriptions, like those in the bank case, the predictions made by the subject-sensitive invariantist and the epistemic contextualist are identical. With this in mind, even if evidence is found that judgments about the bank case turn out in the way the contextualist predicts they will, this fact lends equal support to the rival SSI theory. Hansen and Chemla's results therefore do not favor contextualism over SSI, because they only tested first-person knowledge scenarios. This issue for the contextualist is further exacerbated by the fact that Pinillos (2012) presents experimental data that he argues is better explained by SSI than by contextualism (p. 193).

With that issue in mind, cases involving third-person, rather than first-person, knowledge ascriptions should be used in order to evaluate the empirical foundations of epistemic contextualism over SSI. This is because epistemic contextualism and SSI make different predictions about third person cases: if the context of ascription is varied while the context of the subject to whom knowledge is being ascribed remains fixed, then epistemic contextualism predicts a contextual effect while SSI does not. In that respect, third-person context shifting experiments constitute a *crucial experiment* (in Bacon's classic sense) for contextualism versus SSI. DeRose has argued for the importance of third person context shifting experiments for this reason (DeRose 2009, pp. 60-66), and he strongly emphasizes the importance of such cases for the debate with SSI:

> [T]hese third-person cases provide a powerful objection—to my thinking, a killer objection—to SSI. (p. 65)

Until formal tests of third-person context shifting experiments are conducted, there is no non-armchair experimental evidence that favours epistemic contextualism over SSI.[10]

Turri (2016) presents another reason to prefer testing third-person knowledge ascriptions. Turri argues that classic first-person knowledge ascription cases, like the bank case, are set up in such a way that they introduce a potential confound concerning deference, namely that "people might simply defer to others' self-regarding knowledge statements, regardless of whether the stakes vary" (p. 142).[11] Turri conducted an experiment that supports this hypothesis. He constructed a pair of cases similar to the bank cases except that all that changes across the two contexts is

[9] The term "subject-sensitive invariantism" is sometimes used interchangeably with "interest-relative invariantism", the view articulated and defended in Stanley (2005). (See DeRose 2009, p. 25, for the interchangeable use, for example.) However, an anonymous referee points out that the two views are distinct: subject-sensitive invariantism is the view that some feature of the purported knower's context is a relatum of knowledge, whereas interest-relative invariantism is the more specific view that the relatum in question is the speaker's interests or stakes.

[10] It is not the case that if a contextual effect was found in a third person case, this could *only* be explained via appeal to epistemic contextualism. For instance, it might be that the contextual effect in question is actually due to certain psychological biases, as has been argued for by Williamson (2005) and Nagel (2010b).

[11] Feltz and Zarpentine (2010, p. 689) suggest something similar when they suggest that there is an *attributor effect*, in which 'people are more reluctant to agree with third-person knowledge attributions than first-person attributions'.

whether the subject states 'I know p' or 'I don't know p'—no contextual parameters (stakes or mentioned possibilities of error) were varied across contexts. Turri found that in both contexts participants tended to agree with the statement "When Keith said 'I [do/don't] know', what he said was true". This is so despite the fact that there was no change in the contextual parameters that contextualists have claimed should matter for assessing knowledge ascriptions. Given that the statements in the two contexts seem to contradict one another, this provides compelling evidence in favour of the idea that participants simply defer to the subject regarding the self-ascription of knowledge.[12]

Note, however, that third-person knowledge ascriptions not only provide a way of supporting contextualism over SSI, they also provide a way of avoiding Turri's deference confound: if participants are likely to defer to others' *self*-ascriptions of knowledge, contextualists should focus on third-person knowledge ascriptions instead.

3. Experiment 1: Testing Third Person Knowledge Ascriptions

Even though contextualism and SSI have been the target of extensive experimental investigation, and third-person knowledge ascriptions play a crucial role in the case for contextualism, there have been no previous experimental studies of such knowledge ascriptions. We designed and conducted an experiment specifically designed to examine that crucial gap in contextualism's empirical foundation.

3.1 Task

Participants were asked to read a series of hypothetical scenarios and respond to them by judging whether the claim made by certain target sentences (indicated in bold) were true or false using a sliding scale from 0 (False) to 100 (True) (see Figure 1).

[12] Turri's deference confound is another way of raising the worry first articulated in Hansen and Chemla (2013) about DeRose's preferred design for context shifting experiments, that by varying both context and sentence polarity in his high and low standards contexts, without also testing the combinations of low standards + negative polarity, and high standards + positive polarity, DeRose doesn't have the resources to trace variation in truth value judgments to changes in context (as opposed to polarity). But by testing all four possible cells in their revised design, Hansen and Chemla have the resources to distinguish effects of context and effects of sentence polarity, that is, to assess whether there exists something like Turri's deference confound.

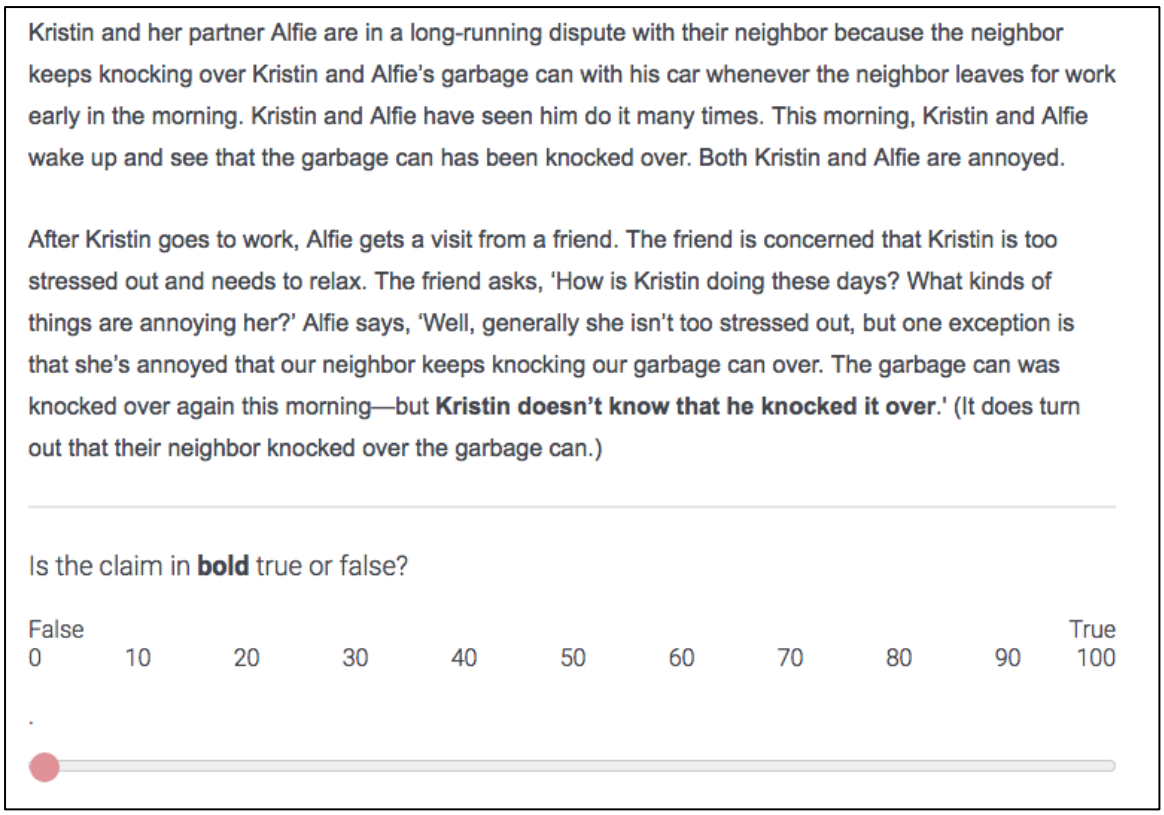


Figure 1: Sample knowledge scenario and truth value judgment task (as it appeared in the online survey)

One might worry that using a continuous scale of 0-100—rather than asking for a binary true/false judgement—is problematic because it does not exactly reproduce the truth-value judgments that are part of armchair theorizing about contextualism and because participants may be interpreting the scalar task as asking them to measure something like their confidence in the particular claim at issue rather than their assessment of its truth or falsity. But it is an open question how ordinary speakers think about the truth and falsity of what is said—whether they think of the difference as binary, or as something that can be captured in terms of a continuous scale. We agree with the defense of the continuous scale truth-value judgment task given in Hansen and Chemla (2013, p. 309):

> All methods of eliciting responses to linguistic experiments, whether they employ a binary true/false judgment task, or a Likert scale with labelled points, or the continuous true/false scale we employed, play a role in shaping the responses participants give. For example, a binary true/false judgment task demands that participants make sharp judgments, even when their responses may in fact be much more nuanced. That could obscure interesting differences between participants' responses to scenarios…. And no type of response corresponds directly to the binary, true/false (or 1/0) outputs of semantic theory, even those elicited by a binary true/false judgment task. Semantic theory has to be combined with theories of how participants will perform in response to particular experimental material and in response to particular kinds of tasks before predictions about actual participants' responses are possible.[13]

Given what we currently know, it could be the case that asking for binary true/false judgments may actually distort the way ordinary speakers think about truth and falsity. Participants are

[13] See Franke (2016) for a comparison of binary truth-value judgment tasks with Likert scale responses in experiments probing the meaning of the quantifier "some".

explicitly instructed to indicate on the scale whether the claim is true or false, and we think it is fair to assume that participants are tapping into their judgments about truth and falsity in their responses to the task. Furthermore, this is a method that has been used in previous psycholinguistic studies (Chemla 2009a,b; Chemla and Spector 2011; Chemla and Schlenker 2012; Hansen and Chemla 2013).[14]

3.2 Experimental materials: knowledge scenarios

We used the structure of a third-person context shifting experiment described by DeRose (2009, pp.4-5) as the basis for two knowledge scenarios. DeRose's own scenario is not suitable for experimental use, as it is (a) far too long and (b) the Low and High contexts are of substantially different length: 343 words for the Low context, and 494 words for the High context. Ideally, the stories participants read should be short and of roughly equal length, so that any observed differences between contexts can't plausibly be explained in terms of differences in cognitive load.

We aimed to preserve the following structural features from DeRose's original third-person context shifting experiment:

a. *The set up*: This describes the evidence available to the person to whom knowledge is ascribed (the *ascribee*). The set up remains constant across all of the different manipulations of context and polarity. (In DeRose's original third-person experiment, for example, the set up involves Lena seeing a co-worker's hat hanging outside his office, and her hearing someone asking about the co-worker.)
b. *Low standards context of ascription + positive knowledge ascription:* In this condition, in which nothing particularly important is at stake, and no one mentions any possibilities of error, another character in the story (the *ascriber*) says, "[The ascribee] knows that p". (In DeRose's original experiment, the ascriber, Thelma, hanging out in a bar, says that Lena knows that the co-worker whose hat she saw was at work that day in order to settle a small bet.)[15]
c. *High standards context of ascription + negative knowledge ascription:* In this condition, the stakes of knowing that p for the ascriber are raised, and a possibility of error is mentioned to the ascriber. The ascriber says, "[The ascribee] doesn't know that p". (In DeRose's original experiment, the ascriber, Louise, is talking to the police who are "conducting an extremely important investigation", and the possibility is raised that seeing the co-worker's hat hanging in the hall is consistent with him not being at work. When Louise is asked if Lena knows that her co-worker was at work that day, Louise says that Lena doesn't know that he was at work that day.)[16]

The evidence and confidence of the ascribee is meant to remain constant across the two contexts of ascription. That is accomplished by making it the case that the ascribee is not party to the conversation in which the stakes are raised and the possibility of error is mentioned. As knowledge ascriptions are factive, it should also be clear to the participants that the proposition in question is in fact true in all cases. This is stated explicitly at the end of each prompt.

DeRose's judgments about his third-person context shifting experiment (see Table 3) parallel his judgments about the original bank case (shown in Table 2):

---

[14] For more extensive discussion and defense of the task, see Hansen and Chemla (2013, pp. 309-311).
[15] Following DeRose (2011, pp. 89-90), we are seeking to test contextualism generally, rather than specific contextualist proposals. For this reason, we vary both the stakes for the attributor and whether a possible alternative has been raised.
[16] For DeRose's own account of how to generate context shifting experiments using third-person knowledge ascriptions, see his (2009, pp.62-63).

Table 3: DeRose's predictions for the third person context shifting scenario

Sentence uttered	Low context	High context
"Lena knows that John [her co-worker] was there."	TRUE	
"Lena doesn't know that John [her co-worker] was there."		TRUE

Notice that DeRose's design only tests two of the four available "cells" (combinations of context + sentence polarity). Hansen and Chemla (2013, p. 295) observe that looking only at those two cells is methodologically unsound:

> DeRose's design simultaneously varies *both* the target sentence used *and* the context in which the sentence is used. That will make it difficult to identify whether it is the change in context or the polarity of the sentence used that is responsible for the intuitions elicited by each cell.

Testing all four cells, in contrast, makes it possible to determine whether it is context or polarity (or the interaction of both) that is affecting truth value judgments about knowledge ascriptions.[17] For that reason we have adopted the four-cell design used in Hansen and Chemla (2013) in our evaluation of third-person context shifting experiments, yielding the following structure, in which, for each cell, *the set up* is combined with one of the four possible combinations of context + polarity (a-d):

The set up
---
a. *Low standards context + positive knowledge ascription*
b. *Low standards context + negative knowledge ascription*
c. *High standards context + positive knowledge ascription*
d. *High standards context + negative knowledge ascription*

The contextualist predictions for the modified, four-cell design go as follows, where "MORE TRUE/MORE FALSE" means that a condition should be judged to be more true/more false than the cell to its right or left (as appropriate)—indicating an effect of context on a truth value judgment (see Table 4):

Table 4: Contextualist predictions for the four cells in third person context shifting scenarios

Sentence uttered	Low context	High context

[17] One consequence of testing both polarities in each context is that it can sometimes be challenging to make a knowledge attribution of a particular polarity sound natural while keeping as much of the context fixed as possible. For instance, the knowledge attribution in the low + positive context presented below may seem quite natural, but the negative attribution in the same context may strike the reader as an odd thing to say regardless of its truth, given the conversational setting. That pragmatic oddness would be problematic if it then affected participants' truth value judgments. In response to this worry, we have, where possible, sought to make each knowledge attribution sound as natural as possible within the conversational setting - for instance by using the appropriate discourse markers. In doing so, we have tried to strike the right balance between making each attribution sound natural given the discourse, while making the positive and negative contexts as similar to one another as possible.

Lena knows that John [her co-worker] was there	(MORE) TRUE	(MORE) FALSE
Lena doesn't know that John [her co-worker] was there	(MORE) FALSE	(MORE) TRUE

It is worth being explicit about how the contextualist predictions would be realized within this experimental design. Because we vary the context *and* polarity across our stories, the contextualist would not predict a main effect of context, as this would mean that changing the context would drag scores for *both* positive and negative cells in the same direction. Instead, as Table 4 illustrates, the contextualist would predict that the effect that the context has is dependent upon the polarity of the attribution. In the analysis of results below, we will thus ask three main questions: Is there an interaction between context and polarity? Is there an effect of context on ratings for positive statements? Is there an effect of context on ratings for negative statements? The contextualist predicts affirmative answers to all three questions.

The knowledge-testing stories that participants read consist of *the set up* with one of the four conditions (a-d). To illustrate, here is a prompt that appeared in our experiment consisting of a set up and a low standards context + positive knowledge ascription (participants did not see the italicized labels; the sentence in bold is the target knowledge ascription):

> *Set up:* Kristin and her partner Alfie are in a long-running dispute with their neighbor because the neighbor keeps knocking over Kristin and Alfie's garbage can with his car whenever the neighbor leaves for work early in the morning. Kristin and Alfie have seen him do it many times. This morning, Kristin and Alfie wake up and see that the garbage can has been knocked over. Both Kristin and Alfie are annoyed.
>
> *Low standards + positive knowledge ascription:* After Kristin goes to work, Alfie gets a visit from a friend. The friend is concerned that Kristin is too stressed out and needs to relax. The friend asks, "How is Kristin doing these days? What kinds of things are annoying her?" Alfie says, "Well, generally she isn't too stressed out, but one exception is that she's annoyed that our neighbor keeps knocking our garbage can over. The garbage can was knocked over again this morning, and **Kristin knows that he knocked it over.'** (It does turn out that their neighbor knocked over the garbage can.)[18]

We developed two separate knowledge scenarios: the first involved knowledge ascriptions concerning whether a neighbor had knocked over a garbage can (the "neighbor" scenario) and the second involved knowledge ascriptions concerning whether the ascribee knew that it was going to be sunny tomorrow (the "sunshine" scenario). The knowledge scenarios we developed were substantially shorter than DeRose's original third person scenario (166 words for the above low + positive story, vs. DeRose's 343, for example) and were roughly balanced in length (166 for this low + positive vs. 172 for high + negative in this story, for example). (See the Appendix for all of the experimental materials.)

3.3 Experimental materials: color and control scenarios

In addition to the knowledge scenarios, which were our primary target of investigation, we also presented participants with a *color* scenario and a *control* scenario, both of which mirrored the

[18] Figure 1, above, shows an example of the low standards + negative knowledge cell for the "knowledge-neighbor" scenario.

four-cell design of the knowledge scenarios. The color scenario was adopted from Travis (1989) via Hansen and Chemla (2013), and involved judgments about the truth value of claims about the color of walls of an apartment. For example, the "low standards + positive polarity" and "high standards + positive polarity" cells of the color scenario read as follows (with the target sentence in bold):

> *Low standards + positive polarity:* Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. Hugo and Odile are trying to choose a rug that will go with the walls of their new apartment. Odile points at an orange rug and says, 'What do you think of this one?' Hugo says, 'I don't like it. **The walls in our apartment are brown.'**

> *High standards + positive polarity:* Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. It has recently been discovered that the walls made of brown plaster give off a poison gas. So they are being demolished and replaced. The superintendent asks Hugo to find out what sorts of walls his are. After inspecting his walls, Hugo says, **'The walls in our apartment are brown.'**

The point of including the color scenario was to evaluate the finding in Hansen and Chemla (2013) that the effects of context on truth value judgments about color statements were stronger than the effects of context on knowledge ascriptions. Since color judgments display behavior strongly indicative of context-sensitivity, they provide a benchmark against which the context sensitivity of knowledge ascriptions can be measured.

In testing color terms, we are not committed to any particular semantic treatment of color expressions, and we are certainly not committed to thinking that color terms and knowledge attributions are context-sensitive in the same way. More specifically, we are not committed to thinking that they are sensitive to the same contextual features, or map onto the same scales, or warrant similar semantic treatments. We are testing color terms to provide a clear example in our experimental setting of a successful context-shifting experiment. In doing so, we also aim to further confirm Hansen & Chemla's (2013) previous experimental evidence. We take this to be necessary, but not sufficient, for establishing that an expression is context-sensitive. So even if color terms were not context-sensitive, and instead just multiply ambiguous, this would not affect the role that they play here.

The control scenario was included as an attention check to make sure participants were paying attention and that they understood the truth value judgment task, and it too had a four-cell design (see Appendix 1). Participants who made implausible truth value judgments about the target sentences in any of the controls (that is, giving a response of less than 50 to the responses that should have been judged true, or more than 50 to the responses that should have been judged false) were prevented from completing the survey.

3.4 Experimental "Block" Design

Given the four-cell design of the two knowledge scenarios, one color and one control scenario, our participants gave truth value judgments in response to a total of 16 different stories. Using the "block" design developed in Hansen and Chemla (2013), these 16 stories were arranged in four "blocks": each block was constructed so as to contain only one cell of each of the four

scenarios (two knowledge, one color, one control) (see Figure 2). The order of the stories in each block was randomized, as was the order of the blocks themselves.

Figure 2: Block Design

Scenario Version	Block A	Block B	Block C	Block D
High+Neg	Knowledge-Neighbor	Knowledge-Sunshine	Color	Control
High+Pos	Knowledge-Sunshine	Knowledge-Neighbor	Control	Color
Low+Neg	Color	Control	Knowledge-Neighbor	Knowledge-Sunshine
Low+Pos	Control	Color	Knowledge-Sunshine	Knowledge-Neighbor

The block design allows our experiment to implement both a within-subjects design (when all of the responses are considered together), and a between-subjects design, when only the responses to the "first block" of cells are considered. That is, each block is designed so that it contains only one cell of each scenario, but the blocks are shuffled so that there are "first-block" style responses to all cells in the experiment. So, by isolating "first block" responses across the whole experiment, it is possible to compare judgments to all of the relevant cells in a between-subjects manner. (The block design was not transparent to participants.)

3.5 Participants

In total, 557 participants were recruited using Amazon MTurk. The following restrictions were placed on participation: the location of workers was limited to the United States, the HIT approval rate for requesters' HITs was 95%, and the requirement for number of HITs approved for each worker was set at 50.[19]

Participants who did not complete the survey or who failed one or more attention check were not allowed to complete the HIT. The survey remained open until 431 of participants completed the survey. This number was pre-set on the basis of a power analysis using G*Power. All the data presented and analysed is from these participants. Those who completed the survey were paid $.30.

163 (37.8%) were Male, 264 (61.3%) were Female, and 4 (0.9%) were another gender. 419 (97.2%) indicated that English was their native language. 133 (30.9%) has studied some philosophy at university level.[20] The mean age of participants was 35.94. The mean time participants took to complete the survey was 9 minutes, 32.19 seconds.

3.6 Results

[19] "HIT" stands for "Human Intelligence Task", the name for tasks posted on Amazon Mechanical Turk.
[20] The results of the analysis were not qualitatively different when these participants were excluded. Results are reported for all participants.

In this section, we analyse the results of Experiment 1. Our most important finding is that there is evidence of a contextual effect, namely, a significant interaction of context and polarity for knowledge and color scenarios in both the "global" results (when responses to all cells are considered) and for the "first block" results (when responses to only the first block of results are considered).

3.6.1 Control Cases

As described above, the control scenarios were used as attention check questions. Participants who gave implausible answers to the control scenarios were automatically prevented from answering any further questions. The number of participants who failed one of the attention checks and was filtered out in this way was 126 (leaving 431 participants).

The means and standard deviations for the responses the remaining participants gave to the control scenarios are given in Table 5.

Table 5: Mean TRUE responses for controls (Experiment 1)

Polarity	Context 1 (5 mins)		Context 2 (10 hours)	
	Mean	SD	Mean	SD
Pos	98.73	4.19	1.86	4.57
Neg	2.20	5.48	98.51	4.76

Statistical analysis is not conducted for this control scenario due to the way that participants were excluded on the basis of their answers.

3.6.2 Global Descriptive Results

The means and standard deviations for responses to all scenarios are given in Table 6. The mean responses are also plotted in Figure 3.

Table 6: Mean TRUE responses in all results (Experiment 1)

Scenario Type	Polarity	Low Context		High Context	
		Mean	SD	Mean	SD
Knowledge-neighbor	Pos	54.38	39.90	37.45	38.22
	Neg	61.75	38.97	72.23	35.91
Knowledge-sunshine	Pos	65.42	35.64	54.37	37.31
	Neg	47.61	39.98	57.34	37.31
Color	Pos	91.25	19.72	60.87	38.63
	Neg	13.27	25.72	50.52	40.29

In Figure 3, each line on the graph shows the difference in mean response between the high and low versions of a particular scenario type and sentence polarity, e.g. the high and low contexts and positive and negative sentences for "knowledge-neighbor" on the far left of the graph. The blue lines represent the responses participants gave to the first block of scenarios. The red lines represent responses for all scenarios. The red and blue lines track each other closely, illustrating the fact that the same patterns of responses were found in both within-subjects and between-subjects conditions.

This graph clearly reveals the central contextual effect from our results. Note the distinctive "V"-shape in the results for each scenario. This illustrates the *interaction* we found in each case between polarity and context. The interaction means that the effect of context is different for positive and negative polarity statements. This pattern is what the contextualist predicts, since a high standards context should result in a greater reluctance to say that 'S knows that p' is true, but a greater willingness to say that 'S does not know that p' is true, and a low standards context should produce the reverse.

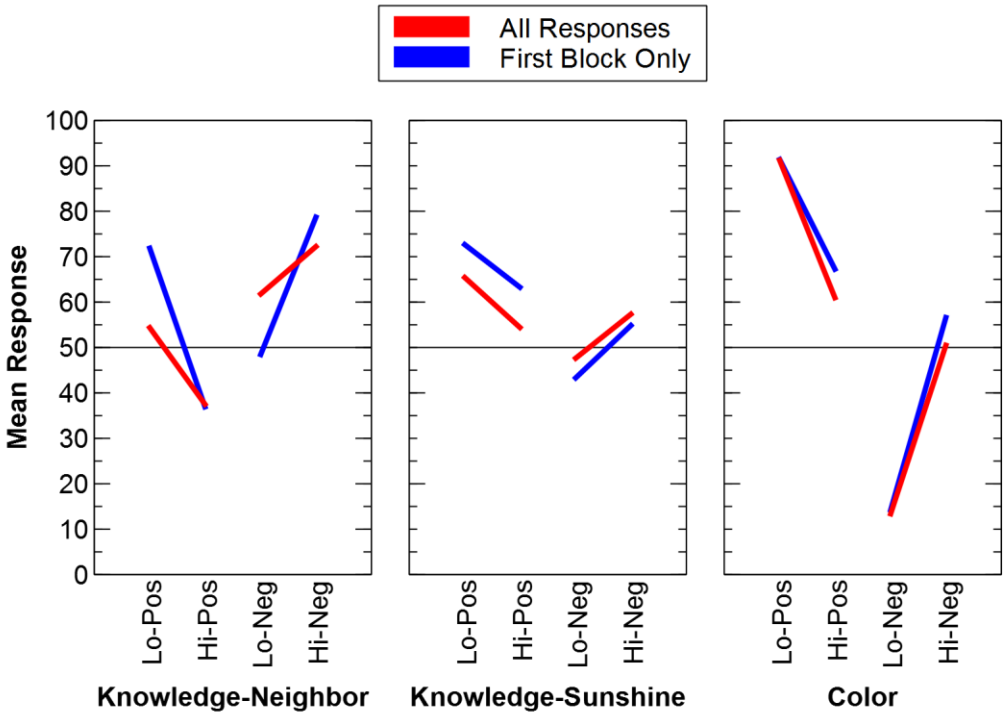


Figure 3: Graph of mean responses for all versions and all scenarios showing the contextual effect (the "V" shape in each graph) (Experiment 1)

3.6.3 Analysis of Global Results

A 2 x 2 x 3 ANOVA with three within-subjects factors (scenario type: neighbor, sunshine, color; context: hi, lo; polarity: neg, pos) revealed a significant three-way interaction ($F(1.90,817.28) = 76.30$, $p < .0005$, $\eta^2_p = 15$).[21] There were also significant two-way interactions between context and polarity ($F(1,430)=421.01$, $p < .0005$, $\eta^2_p =.50$), type and polarity ($F(1.94, 832.38)=191.26$, p

[21] Mauchly's test found that the assumption of sphericity was violated ($\chi^2(2)=25.08$, $p < .0005$, $\varepsilon = .95$). Degrees of freedom corrected using the Huyhn-Feldt correction.

$< .0005$, $\eta^2_p = .31$),[22] and type and context ($F(2,429)=8.88$, $p < .0005$, $\eta^2_p = .04$). There were also significant main effects of polarity ($F(1,430)=33.43$, $p < .0005$, $\eta^2_p = .07$) and scenario type ($F(1.95,838.34)=4.98$, $p=.008$, $\eta^2_p = .01$),[23] but not context ($p=.818$).

Three 2 x 2 ANOVAs with two within-subjects factors (context: hi, lo; polarity: neg, pos) were then conducted to consider the main effects and interaction for each scenario type. These are in turn followed up by two paired sample t-tests to examine the effect of context for each polarity.

For Neighbor, there was a significant interaction ($F(1,430) = 100.28$, $p < .0005$, $\eta^2_p = .19$). There were also significant main effects of context ($F(1,430) = 9.59$, $p = .002$, $\eta^2_p = .02$) and polarity ($F(1,430) = 49.93$, $p <.0005$, $\eta^2_p = .10$). Paired-samples t-tests reveal a significant difference between contexts for both negative ($t(430)=6.24$, $p < .0005$, $d = .30$) and positive ($t(430)=9.63$, $p < .0005$, $d = .46$).

For Sunshine, there was a significant interaction ($F(1,430) = 65.59$, $p <.0005$, $\eta^2_p = .13$). There was no significant main effect of context ($p = .593$). There was a significant main effect of polarity ($F(1,430) = 6.99$, $p = .008$, $\eta^2_p = .02$). Paired-samples t-tests reveal a significant difference between contexts for both negative ($t(430) = 6.11$, $p < .0005$, $d = .29$) and positive ($t(430) = 5.69$, $p < .0005$, $d = .27$).

For Color, there was a significant interaction ($F(1,430) = 341.64$, $p <.0005$, $\eta^2_p = .44$). There were also significant main effects of context ($F(1,430) = 9.37$, $p = .002$, $\eta^2_p =.02$) and polarity ($F(1,430) = 527.46$, $p <.0005$, $\eta^2_p = .55$). Paired-samples t-tests reveal a significant difference between contexts for both negative ($t(430) = 16.36$, $p < .0005$, $d =.78$) and positive ($t(430) = 15.14$, $p < .0005$, $d = .73$).

3.6.4 Summary of Global Results

There is a clear contextual effect. The two-way ANOVAs reveal significant two-way interactions between context and polarity for each scenario-type. For each scenario type and polarity, there is a significant difference between high and low contexts. Jacob Cohen's (1988) rules of thumb for interpreting Cohen's d effect sizes suggest that the effects of context are small for the knowledge scenarios ($.2 < d < .5$) and medium for the color scenario ($.5 < d < .8$). This suggests weaker contextual effect for the knowledge scenarios than for the color scenario.[24] This is reflected in the fact that the three-way interaction is significant in the three-way ANOVAs. In order to demonstrate this stronger effect for the color scenario, we carried out pairwise comparisons of scenario type. Significant three-way interactions remained when comparing Neighbor and Color ($F(1,430)=86.30$,$p<.0005$, $\eta^2_p =.17$) and Sunshine and Color ($F(1,430) = 115.75$, $p < .0005$, $\eta^2_p =.21$). However, there was no significant three-way interaction in the comparison of the two knowledge scenarios (Neighbor and Sunshine) ($p =.07$) (see Appendix 2 for full analyses).

3.6.5 First Block Descriptive Results

In this section, we examine the results from the first time participants saw a scenario of each type. This thus allows us to emulate a between-subjects design. Our design means that these results come from the first four scenarios a participant saw. It also means that no participant had previously seen a scenario of another type with the same combination of context and

[22] Mauchly's test found that the assumption of sphericity was violated ($\chi^2(2)=16.50$, $p < .0005$, $\varepsilon = .96$). Degrees of freedom corrected using the Huyhn-Feldt correction.
[23] Mauchly's test found that the assumption of sphericity was violated ($\chi^2(2)=13.25$, $p < .0005$, $\varepsilon = .97$). Degrees of freedom corrected using the Huyhn-Feldt correction.
[24] This is consistent with the finding of greater strength of contextual effects in color scenarios than in knowledge scenarios reported in Hansen and Chemla (2013).

polarity. The first block means and standard deviations can be seen in Table 7.). The means from this table are included in Figure 3 (above).

Table 7: Mean TRUE responses in the first block of results (Experiment 1)

Scenario Type	Polarity	Context					
		Low			High		
		N	Mean	SD	N	Mean	SD
Knowledge-Neighbor	Pos	105	71.90	33.69	107	36.83	38.49
	Neg	121	48.36	41.17	98	78.73	31.29
Knowledge-Sunshine	Pos	121	72.67	34.15	98	63.23	35.56
	Neg	105	43.27	40.94	107	54.86	36.82
Color	Pos	107	91.41	20.72	105	67.17	36.02
	Neg	98	14.09	24.63	121	56.64	37.18

3.6.6 Analysis of First Block Results

The results for each scenario type were examined separately as each participant saw only one cell of each scenario in their first block. A series of 2 x 2 ANOVAs were conducted with two between-subjects factors, context (high, low) and polarity (negative, positive). These are in turn followed up by two t-tests to examine the effect of context for each polarity.

For Neighbor, there was a significant interaction ($F(1,427) = 85.41$, $p < .0005$, $\eta^2_p = .17$), no main effect of context ($p = .509$), and a significant main effect of polarity ($F(1,427) = 6.73$, $p = .01$, $\eta^2_p = .02$).[25] Independent samples-tests reveal a significant difference between contexts for both negative ($t(216.18) = 6.20$, $p < .0005$, $d = .82$) and positive ($t(207.32) = 7.06$, $p < .0005$, $d = .97$).

For Sunshine, there was a significant interaction ($F(1,427) = 8.71$, $p = .003$, $\eta^2_p = .02$), no main effect of context ($p = .762$), and a significant main effect of polarity ($F(1,427) = 28.11$, $p < .0005$, $\eta^2_p = .06$).[26] Independent samples-tests reveal a significant difference between contexts for both negative ($t(206.79) = 2.17$, $p = .031$, $d = .30$) and positive ($t(217) = 2.00$, $p = .047$, $d = .27$).

For Color, there was a significant interaction ($F(1.427) = 125.89$, $p < .0005$, $\eta^2_p = .23$), and significant main effects of both context ($F(1,427) = 9.46$, $p = .002$, $\eta^2_p = .022$) and polarity ($F(1,427) = 217.85$, $p < .0005$, $\eta^2_p = .34$).[27] Independent samples-tests reveal a significant difference between contexts for both negative ($t(209.27) = 10.14$, $p < .0005$, $d = 1.32$) and positive ($t(165.39) = 5.99$, $p < .0005$, $d = .83$).

3.6.7 Summary of First Block Results

[25] Levene's test found that the assumption of homogeneity of variances was violated, $F(3,427) = 11.21$, $p < .0005$.
[26] Levene's test founds that the assumption of homogeneity of variances was violated, $F(3,427) = 5.17$, $p = .002$.
[27] Levene's test founds that the assumption of homogeneity of variances was violated, $F(3,427) = 31.94$, $p < .0005$.

Overall, we see a very similar pattern of results in the first block as when all results are considered. There is a clear contextual effect. The two-way ANOVAs reveal significant two-way interactions between context and polarity for each scenario-type. For each scenario type and polarity, there is a significant difference between high and low contexts. Cohen's rules of thumb for interpreting Cohen's d effect sizes suggest that the effects of context are small for the knowledge-sunshine ($.2 < d < .5$) and large for the knowledge-neighbor and color scenarios ($.8 < d$).

3.7 Discussion

In this section, we discuss the findings of our first experiment. Overall, our findings serve to support the contextualist predictions because we found a significant interaction between context and polarity that was further supported by the existence of a contextual effect on each polarity type.

3.7.1 Contextual Effects in the Knowledge Scenarios

Our most important finding for the epistemic contextualist debate is that there is a consistent interaction between context and polarity for the third-person knowledge scenarios, in both the global and first block results. This is positive evidence in favour of epistemic contextualism that doesn't also support SSI. As stated earlier, SSI does not predict a contextual effect in third-person knowledge ascriptions because the stakes of the knower (the *ascribee*) remain invariant across the contexts in question, whereas epistemic contextualism does allow variation in third person scenarios, because the context that matters is the context of the *ascriber*, which does vary in the scenarios.[28,29] In addition, the fact that we find an effect of context even in the first block of results should eliminate any worries that the contextual effect is driven by a contrast effect (which would result from seeing more than one "cell" of each scenario), or another form of experimental artefact. Our finding of a contextual effect is therefore continuous with, but goes substantially beyond, Hansen and Chemla's (2013) findings of a contextual effect in first person knowledge ascriptions in their global results only.

Furthermore, our findings constitute evidence in favour of epistemic contextualism that is immune to Turri's (2016) hypothesis that there is a general phenomenon of deference to others' self-ascriptions of knowledge. As our third-person cases do not involve any self-ascriptions, such deference cannot be at work in our scenarios.[30]

---

[28] A defender of SSI might object to this as follows: In the "sunshine" scenario, the potential closure of the hospital arguably raises the stakes for everyone, not just the ascriber. Thanks to an anonymous referee for raising this issue. But this worry is less pressing in the "neighbor" scenario. While the stakes are raised for Alfie in his conversation with the police, it's unclear why the stakes would also be raised for Kristin: she is not in the conversation with the police, and there is no suggestion that she is in danger. What we have aimed to do—in both cases—is to *clearly* raise the stakes for the attributor.

[29] Recently, some theories of the effects of changing stakes on knowledge have been proposed that don't clearly fit into the standard contextualism/SSI debate. For instance, both Grimm (2015) and Hannon (2017) have argued that whether a subject knows something is sensitive to the interests of third parties other than the subject. That would put their views at odds with subject-sensitive invariantism. It might be possible, on these views, to explain the results of our third-person knowledge ascriptions in terms of the fact that non-subject or communal stakes vary between the low and high contexts, but this will depend on how the *community of inquiry* is understood (see the discussion of bank cases in Hannon 2017, p. 616, for example).

[30] One might object by expanding the scope of the deference claim, and argue that people defer to knowledge ascriptions in general, rather than just to self-ascriptions of knowledge. But what, we wonder, would justify such deference? While there is reason to expect that first-person avowals are believed, because "we assume that people tend to be right about their own mental states" (Turri, forthcoming, p. 12 n. 6), there isn't any obvious reason to accord third person ascriptions any special weight. Invoking a completely general "agreement bias", "whereby people tend to endorse assertions" (Turri, forthcoming, p. 15), can't explain the differences that we observed—a general agreement bias, if it exists, would apply to all ascriptions equally.

3.7.2 Other findings: strength of contextual effect, acquiescence bias

In our global results, we found a larger contextual effect in the Color scenario than in the knowledge scenarios. This is consistent with the findings in Hansen and Chemla (2013), which also compared the strength of contextual effects on judgments about color and knowledge ascriptions. Finding different strengths of contextual effect is interesting because it introduces a new potential explanandum for theories of communication: what would explain not just the fact that context affects truth value judgments, but the differences in the *degree* to which context affects those judgments? As far as we are aware, no existing theories of the way context and truth conditions interact, whether radical contextualist, moderate contextualist, or minimalist, have even raised the possibility that certain ways that context affects communicated content may be stronger or weaker than others.[31]

Note that in nearly all contexts, the scores for the negative polarity and positive polarity sentences sum to over 100. One might think that if statement p in context c gets a score of n, then ¬p in c should receive a score of (100 – n). Why didn't our results fit that expectation? We think that the results we observed are due to 'acquiescence bias', wherein the experimental participant is inclined to agree with or find true any statement that they are presented with in an experimental setting (Podsakoff et al. 2003, Schaeffer and Presser 2003). This will have the effect of lifting all scores to some extent.[32] The presence of this bias doesn't problematize our experimental results, because our key findings are based upon the fact that there is an interaction between context and polarity: average truth-value judgment scores for positive sentences will be greater in *low* contexts than in *high* contexts, and average truth-value judgment scores for negative sentences will be greater in *high* contexts than in *low* contexts. The difference in average scores cannot be explained by the presence of an acquiescence effect.

We also found some results that are surprising from a contextualist point of view. For instance, looking at all responses to the Knowledge-Neighbour Low context, negative sentences were judged to be more true than positive sentences, and in the results for the first block of the Knowledge-Sunshine case, positive sentences were rated as more true than negative sentences in the High context. We're not sure how to explain this surprising pattern of results. Nevertheless, the central prediction of contextualism, that certain changes in the context of ascription can have an effect on truth-value judgments about knowledge ascriptions, is still supported by our findings.

3.7.3 A worry about the Neighbor scenario

One distinctive aspect of the high standard context in the Neighbor scenario is that Alfie – the knowledge ascriber – admits that he is unsure about who knocked the garbage can over. In presenting these cases, we have encountered the worry that the variation in context in the neighbor case is such that contextualists and non-contextualists alike would predict an effect on the truth value of knowledge attributions.[33] In particular, if knowing that p entails p (i.e. if knowledge is factive), then it would be odd for Alfie to say that he is unsure about p (in this case, that the neighbor knocked over the garbage can), and then subsequently claim that Kristen knows p, just because this would entail what he just claimed he is unsure of.

We think that this is a legitimate concern. It is worth pointing out that we found significant interaction effects between context and polarity in the Sunshine scenario where there is no room

[31] Of course, this may simply be an experimental artefact, explained by the relative difficulty of reading different scenarios. That is the explanation floated in Hansen and Chemla (2013, p. 308).
[32] Thanks to an anonymous referee for pressing this question.
[33] Thank you to [names omitted] and an anonymous reviewer for raising this worry.

for a similar worry, and so our results still provide clear evidence of an effect of context on third-person knowledge ascriptions. But we expected that the Neighbor scenario would be an effective context-shifting experiment even with the problematic statement of uncertainty removed, so we ran a follow-up experiment with the problematic statement of uncertainty removed.

4. Experiment 2: Revised Neighbor

4.1 Task, Materials and Design

The set up of the revised Neighbor scenario remained the same as in the first experiment. The positive and negative versions of the high context read as follows (for all of the materials used in Experiment 2, see the Appendix):

> ***High + Positive***: After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says, 'Sir, the garbage can could have been knocked over by a gust of wind or a raccoon, and it's really important for our investigation that we are clear on this. Does Kristin know that your neighbor knocked over the garbage can?' Alfie replies **'Yes, Kristin knows that he knocked it over.'**
>
> (It does turn out that the neighbor knocked over the garbage can).

> ***High + Negative:*** After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says, 'Sir, the garbage can could have been knocked over by a gust of wind or a raccoon, and it's really important for our investigation that we are clear on this. Does Kristin know that your neighbor knocked over the garbage can?' Alfie replies, **'No, Kristin doesn't know that he knocked it over'**.
>
> (It does turn out that the neighbor knocked over the garbage can).

Experiment 2 used the same four controls as in Experiment 1, and the same continuous scale truth-value judgment task.

The survey was designed to allow for both within- and between-subjects conditions in the same experiment (as in Experiment 1). Participants were randomly assigned to one of four groups. The first scenario for each group was one version of the revised Neighbor scenario. In each group, following the first scenario, the remaining Neighbor scenarios and the Control scenarios were presented in a random order.

4.2 Participants

We recruited 596 participants on Amazon's Mechanical Turk, who were paid $.30 each upon completion of the task. The same restrictions as in the first experiment were placed on participation: the location of workers was limited to the United States, the HIT approval rate for requesters' HITs was 95%, and the requirement for number of HITs approved for each worker was set at 50.

Participants who failed to respond appropriately to any of the controls were not allowed to complete the HIT. The survey remained open until 402 participants completed the survey on the basis of a pre-set limit of 400.[34]

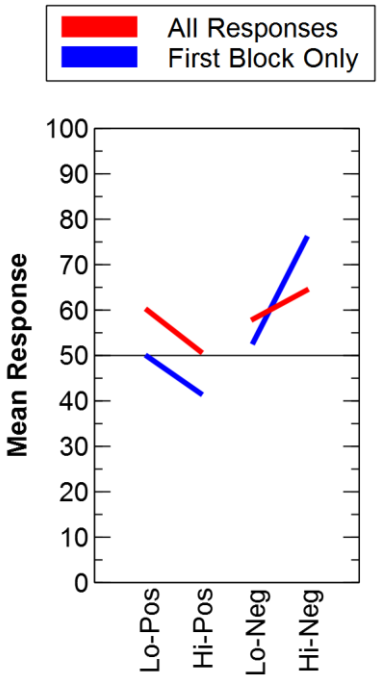
165 (41%) were Male, 236 (58.7%) were Female, and 1 (0.2%) was another gender. 397 (98.8%) indicated that English was their native language. 104 (25.9%) had studied some philosophy at university level.[35] The mean age of participants was 35.87. The mean time participants took to complete the survey was 5 minutes, 1.96 seconds.

4.3 Descriptive Results

The means and standard deviations for all Neighbor scenarios, both for all participants, and for participants who saw the relevant scenario first, are presented in Table 8, and Figure 4.

Table 8: Mean TRUE responses (Experiment 2)

Context	Polarity	Overall			When presented first		
		N	Mean	SD	N	Mean	SD
Low	Positive	402	59.95	40.10	107	49.77	42.48
Low	Negative	402	58.03	40.33	102	52.91	41.29
High	Positive	402	50.88	41.71	96	41.66	40.22
High	Negative	402	64.33	39.42	97	75.07	31.30



[34] The total number of participants thus represents an accidental over-recruitment of 2.
[35] The primary results of the analysis were not qualitatively different when these participants were excluded. Thus, results are reported for all participants.

Figure 4: Graph of mean responses for all versions and all scenarios showing the contextual effect (the "V" shape in each graph) (Experiment 2)

4.4 Within-subjects Analysis (all responses)

A 2 x 2 ANOVA was conducted with two within-subjects factors, context (high, low) and polarity (negative, positive). These are in turn followed up by two t-tests to examine the effect of context for each polarity. There was a significant interaction ($F(1,401) = 37.34$, $p < .0005$, $\eta^2_p = .09$), but no significant effects of either context ($p=.18$) or polarity ($p=.08$).[36] Paired samples t-tests reveal a significant difference between contexts for both positive ($t(401)=5.73$, $p<.0005$, $d = .29$) and negative polarities ($t(401)=3.74$, $p < .0005$, $d = .19$).

4.5 Between-subjects Analysis (first "block" only)

A 2 x 2 ANOVA was conducted with two between-subjects factors, context (high, low) and polarity (negative, positive). These are in turn followed up by two t-tests to examine the effect of context for each polarity. There was a significant interaction ($F(1,398) = 14.96$, $p < .0005$, $\eta^2_p = .04$), but no significant effects of either context ($p=.72$) or polarity ($p=.44$). Independent samples t-tests reveal a significant difference between contexts for negative ($t(187.81)=4.28$, $p<.0005$, $d=.93$) but not positive ($p=.165$) polarities.[37]

4.6 Summary

The results for Experiment 2 display a similar pattern as observed in Experiment 1. The two-way ANOVAs reveal significant two-way interactions between context and polarity for each scenario-type. Looking at all participants' responses there is a small significant difference between high and low contexts for both positive and negative polarities (although for negative polarity, Cohen's d falls just short of $d < .02$, the conventional threshold for a 'small' effect). Looking at responses to the first "block" only, the results are slightly different. There is a very large effect of context for the negative polarity cases, but no significant effect for the positive cases.

4.7 Discussion

The main takeaway from the follow-up experiment is that responses to the revised Neighbor scenario show the interaction of context and polarity that supports contextualism. Overall, the effect in the follow-up experiment is weaker than the effect we found in the original Neighbor scenario (the distinctive "V" shape in Figure 4 is less pronounced than the "V" shape for the Neighbor scenario in Figure 3), which could be due to the absence of the statement of uncertainty that appeared in the original Neighbor scenario. But the contextual effect remains after that statement is removed, so it wasn't responsible for all of the contextual effect we observed in the first experiment. With that in mind, we take this experiment to further support the contextualist hypothesis that there is a significant interaction between context and polarity by

[36] The only part of the analysis which is qualitatively different once participants with philosophical experience are excluded concerns the main effect of polarity in this within-subjects ANOVA. With such participants excluded, this main effect is significant. However, it is associated with an effect size which falls short of the standard rules of thumb for a 'small' effect, and is of little theoretical significance. So we have chosen to report results with such participants included.

[37] Levene's test founds that the assumption of homogeneity of variances was violated for the negative comparison, $F = 25.29$, $p < .0005$.

ruling out the possibility that the statement of uncertainty that was present in the original Neighbor case was responsible for the effect of context.[38,39]

5. Conclusion

The crucial experiments we conducted support epistemic contextualism over the rival SSI view. We found an effect of context on the truth values of knowledge attributions in the form of an interaction between context and polarity. This result, combined with previous findings in the experimental literature, means that the state of debate is now at something of an impasse. On the one hand, our findings provide evidence in favour of epistemic contextualism over SSI. On the other hand, there have been experimental findings that provide support for SSI over epistemic contextualism: Pinillos (2012) and Pinillos and Simpson (2014) found that participants would require stronger or weaker evidence in order to say that the subject (ascribee) knows something, depending upon the stakes of being wrong for the subject (but not the ascriber).

Of course, either side can appeal to cognitive biases, pragmatic explanations, or experimental artefacts in order to explain away the problematic evidence. We will conclude, however, on a more ecumenical note by emphasizing that epistemic contextualism and SSI are not actually inconsistent views. One claims that the truth of knowledge ascriptions can be sensitive to certain parameters in the context of ascription; the other claims that knowledge depends partly upon the stakes of getting a proposition wrong for the knower. The two views have been viewed as competing theories because they have provided competing explanations of the same cases—namely, first person knowledge ascriptions (e.g. the bank case).[40] But now that positive evidence exists for both views, it is possible to maintain that both views give us partial, but compatible, accounts of the complexities of how we assess whether someone knows something.

[38] One additional difference between results in the revised Neighbor scenario and the original is that in the Revised Neighbor scenario, we did not find a significant effect of context in the between-subjects (first block only) condition for positive sentences. Because we only found a small effect in the within-subjects design, it is possible that there is a small effect of context on positive sentences in the between-subjects condition that we simply lack the power to detect.

[39] One surprising result from the second experiment is that in the first block results and in the low context, negative sentences received a higher score than positive sentences. We do not have an explanation for this result, and so we group this with the surprising results found in the first experiment that we discuss at the end of section 3.72.

[40] Pinillos and Simpson (2014, p.12) and Weatherson (2012) observe that SSI is committed to existential claims about the effect of practical stakes on knowledge: *sometimes* context affects our truth value judgments about knowledge ascriptions. Contextualism plausibly involves a similar existential claim: *sometimes* certain contextual parameters (stakes and mentioned possibilities of error) can affect the truth conditions of knowledge ascriptions. The truth of neither view excludes the truth of the other. Pinillos (2012, p.194 n.7) notes that a "hybrid" view is possible, according to which "'knowledge' may express different relations (as some contextualists urge), but some of these relations are themselves sensitive to stakes".

Appendix 1: Experimental Materials

KNOWLEDGE-NEIGHBOR

Setup: Kristin and her partner Alfie are in a long-running dispute with their neighbor because the neighbor keeps knocking over Kristin and Alfie's garbage can with his car whenever the neighbor leaves for work early in the morning. Kristin and Alfie have seen him do it many times. This morning, Kristin and Alfie wake up and see that the garbage can has been knocked over. Both Kristin and Alfie are annoyed.

Low + positive: After Kristin goes to work, Alfie gets a visit from a friend. The friend is concerned that Kristin is too stressed out and needs to relax. The friend asks, "How is Kristin doing these days? What kinds of things are annoying her?" Alfie says, "Well, generally she isn't too stressed out, but one exception is that she's annoyed that our neighbor keeps knocking our garbage can over. The garbage can was knocked over again this morning, and Kristin knows that he knocked it over.'
(It does turn out that their neighbor knocked over the garbage can.)

Low + negative: After Kristin goes to work, Alfie gets a visit from a friend. The friend is concerned that Kristin is too stressed out and needs to relax. The friend asks, 'How is Kristin doing these days? What kinds of things are annoying her?' Alfie says, 'Well, generally she isn't too stressed out, but one exception is that she's annoyed that our neighbor keeps knocking our garbage can over. The garbage can was knocked over again this morning—but Kristin doesn't know that he knocked it over.'
(It does turn out that their neighbor knocked over the garbage can.)

High + negative: After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says 'But the garbage can could have been knocked over by someone else. Are you sure your neighbor did it?"
'No, I'm not sure' Alfie replies.
'Does Kristin know that he knocked over the garbage can?' the policeman asks.
Alfie replies, 'No, Kristin doesn't know that he knocked it over.'
(It does turn out that their neighbor knocked over the garbage can.)

High + positive: After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says 'But the garbage can could have been knocked over by someone else. Are you sure your neighbor did it?"
'No, I'm not sure' Alfie replies.
'Does Kristin know that he knocked over the garbage can?' the policeman asks.
Alfie replies, 'Yes, Kristin knows that he knocked it over.'
(It does turn out that their neighbor knocked over the garbage can.)

KNOWLEDGE-SUNSHINE

Set-up: It's Monday and housemates Terence and Bob are discussing their plans for the week. Terence, looking at the weather forecast on his computer, says 'The online forecast is that it's going to be sunny tomorrow'. 'Ok, thanks' says Bob.

Low + Positive: Later, Bob meets his friend Jackie and they talk about what they should do on their day off together. After thinking about it for a little bit, Jackie suggests that they go to the beach tomorrow (Tuesday). They both have been working hard recently, and the beach is a nice spot to spend the day at this time of year, as long as it is sunny. Bob says, 'Terence checked the weather online and it said it is going to be sunny. **Terence knows that it's going to be sunny tomorrow**.'
(On Tuesday, it is indeed sunny).

Low + Negative: Later, Bob meets his friend Jackie and they talk about what they should do on their day off together. After thinking about it for a little bit, Jackie suggests that they go to the beach tomorrow (Tuesday). They both have been working hard recently, and the beach is a nice spot to spend the day at this time of year, as long as it is sunny. Bob says, 'Terence checked the weather online and it said it is going to be sunny. But **Terence doesn't know that it's going to be sunny tomorrow**.'
(On Tuesday, it is indeed sunny).

High + Positive: Later, Bob meets his friend Rocco. Rocco is anxious because he is responsible for a big outdoor charity fundraiser for the local children's hospital. If he doesn't raise a million dollars at the fundraiser, the hospital will have to close down. He will only raise a million dollars if the fundraiser takes place on a sunny day. He has just had a call from Terence telling him that it will be sunny tomorrow. Hearing this, Bob says 'He read online that it will be sunny. The website he looked at isn't always right about the weather. **But Terence does know that it's going to be sunny tomorrow**.'
(On Tuesday, it is indeed sunny).

High + Negative: Later, Bob meets his friend Rocco. Rocco is anxious because he is responsible for a big outdoor charity fundraiser for the local children's hospital. If he doesn't raise a million dollars at the fundraiser, the hospital will have to close down. He will only raise a million dollars if the fundraiser takes place on a sunny day. He has just had a call from Terence telling him that it will be sunny tomorrow. Hearing this, Bob says 'He read online that it will be sunny. The website he looked at isn't always right about the weather. **Terence doesn't know that it's going to be sunny tomorrow**.'
(On Tuesday, it is indeed sunny).

COLOR

Context 1 [low] + positive polarity

Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. Hugo and Odile are trying to choose a rug that will go with the walls of their new apartment. Odile points at an orange rug and says, 'What do you think of this one?' Hugo says, 'I don't like it. **The walls in our apartment are brown.'**

Context 1 [low] + negative polarity

Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. Hugo and Odile are trying to choose a rug that will go with the walls of their new apartment. Odile points at an orange rug and says, 'What do you think of this one?'

Hugo says, 'I don't like it. **The walls in our apartment aren't brown.'**

Context 2 [high] + positive polarity

Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. It has recently been discovered that the walls made of brown plaster give off a poison gas. So they are being demolished and replaced. The superintendent asks Hugo to find out what sorts of walls his are. After inspecting his walls, Hugo says, **'The walls in our apartment are brown.'**

Context 2 [high] + negative polarity

Hugo and Odile have a new apartment. When their building was built, two sorts of walls were put in: ones made of white plaster and ones made of brown plaster. The walls of their apartment are painted brown, but are made of white plaster. It has recently been discovered that the walls made of brown plaster give off a poison gas. So they are being demolished and replaced. The superintendent asks Hugo to find out what sorts of walls his are. After inspecting his walls, Hugo says, **'The walls in our apartment aren't brown.'**

CONTROLS

Context 1 + Positive polarity
Parag and Ayesha are deciding what movie to watch. One movie they could watch, *Agent Zero*, is 5 minutes long. Parag suggests they watch it. He says 'It's a long movie'.

Context 1 + Negative polarity
Parag and Ayesha are deciding what movie to watch. One movie they could watch, *Agent Zero*, is 5 minutes long. Parag suggests they watch it. He says 'It's not a long movie'.

Context 2 + Positive polarity
Parag and Ayesha are deciding what movie to watch. One movie they could watch, *Chunnel,* is 10 hours long. Parag suggests they watch it. He says 'It's a long movie'.

Context 2 + Negative polarity
Parag and Ayesha are deciding what movie to watch. One movie they could watch, *Chunnel*, is 10 hours long. Parag suggests they watch it. He says 'It's not a long movie'.

REVISED NEIGHBOR

Setup: Kristin and her partner Alfie are in a long-running dispute with their neighbor because the neighbor keeps knocking over Kristin and Alfie's garbage can with his car whenever the neighbor leaves for work early in the morning. Kristin and Alfie have seen him do it many times. This morning, Kristin and Alfie wake up and see that the garbage can has been knocked over. Both Kristin and Alfie are annoyed.

Low + Positive: After Kristin goes to work, Alfie gets a visit from a friend. The friend is concerned that Kristin is too stressed out and needs to relax. The friend asks, "How is Kristin doing these days? What kinds of things are annoying her?" Alfie says, "Well, generally she isn't

too stressed out, but one exception is that she's annoyed that our neighbor keeps knocking our garbage can over. The garbage can was knocked over again this morning, and **Kristin knows that the neighbor knocked it over**".

(It does turn out that their neighbor knocked over the garbage can.)

Low + Negative: After Kristin goes to work, Alfie gets a visit from a friend. The friend is concerned that Kristin is too stressed out and needs to relax. The friend asks, 'How is Kristin doing these days? What kinds of things are annoying her?' Alfie says, 'Well, generally she isn't too stressed out, but one exception is that she's annoyed that our neighbor keeps knocking our garbage can over. The garbage can was knocked over again this morning—but **Kristin doesn't know that the neighbor knocked it over**".

(It does turn out that their neighbor knocked over the garbage can.)

High + Positive: After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says, 'Sir, the garbage can could have been knocked over by a gust of wind or a raccoon, and it's really important for our investigation that we are clear on this. Does Kristin know that your neighbor knocked over the garbage can?' Alfie replies **'Yes, Kristin knows that he knocked it over.'**

(It does turn out that the neighbor knocked over the garbage can).

High + Negative: After Kristin goes to work Alfie gets a visit from the police. As part of a kidnapping investigation, the police are trying to establish where their neighbor was last night. After hearing the garbage can story, one of the policeman says, 'Sir, the garbage can could have been knocked over by a gust of wind or a raccoon, and it's really important for our investigation that we are clear on this. Does Kristin know that your neighbor knocked over the garbage can?' Alfie replies, **'No, Kristin doesn't know that he knocked it over'**.

(It does turn out that the neighbor knocked over the garbage can).

Appendix 2: Supplementary Analyses

Experiment 1: Global Results: Pairwise Comparison of Scenario Types

To allow a pairwise comparison of scenario types, a series of 2x2x2 ANOVAs with three within-subjects factors (scenario type, context, polarity) were conducted.

In the first, Neighbor and Sunshine were compared. The three-way interaction was not significant ($p = .07$). There were significant two way interactions between context and polarity ($F(1,430) = 151.57$, $p < .0005$, $\eta^2_p = .26$), and type and polarity ($F(1,430) = 89$, $p < .0005$, $\eta^2_p = .17$), but not type and context ($p=.108$). There were significant main effects of context ($F(1,430) = 5.71$, $p = .017$, $\eta^2_p = .01$), and polarity ($F(1,430) = 7.65$, $p = .006$, $\eta^2_p = .02$), but not type ($p = .765$).

In the second, Neighbor and Color were compared. The three-way interaction was significant ($F(1,430)=86.30$, $p<.0005$, $\eta^2_p =.17$). There were significant two-way interactions between context and polarity ($F(1,430)=392.54$, $p < .0005$, $\eta^2_p =.48$), type and polarity ($F(1,430)=332.26$, $p < .0005$, $\eta^2_p =.44$), and type and context ($F(1,430) = 17.68$, $p < .0005$, $\eta^2_p = .04$). There were significant main effects of polarity ($F(1,430) = 43.03$, $p < .0005$, $\eta^2_p = .09$) and type ($F(1,430) = 10.00$, $p = .002$, $\eta^2_p = .02$), but not context ($p = .886$).

In the third, Sunshine and Color were compared. The three-way interaction was significant ($F(1,430) = 115.75$, $p < .0005$, $\eta^2_p = .21$). There were significant two way interactions between context and polarity ($F(1,430) = 372.32$, $p < .0005$, $\eta^2_p = .46$), type and polarity ($F(1,430) = 116.14$, $p < .0005$, $\eta^2_p = .21$), and type and context ($F(1,430) = 6.09$, $p = .014$, $\eta^2_p = .01$. There were significant main effects of polarity ($F(1,430) = 231.26$, $p < .0005$, $\eta^2_p = .35$) and type ($F(1,430) = 6.13$, $p = .014$, $\eta^2_p = .01$), but not context ($p = .097$).

References

Buckwalter, W. (2010). Knowledge isn't closed on Saturday: a study in ordinary language. *Review of Philosophy and Psychology, 1*(3), 295-406.

Buckwalter, W. (2017). Epistemic Contextualism and Linguistic Behavior. In J.J Ichikawa (ed.), *The Routledge Handbook of Epistemic Contextualism.* New York: Routledge. 44-56.

Cappelen, H., and Lepore, E. (2005). *Insensitive Semantics: A Defence of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.

Chemla, E. (2009a). Presuppositions of Quantified Sentences: experimental data. *Natural Language Semantics, 17*(4), 299-340, doi:10.1007/s11050-009-9043-9.

Chemla, E. (2009b). Universal Implicatures and Free Choice Effects: Experimental Data. *Semantics & Pragmatics, 2*(2), 1-33.

Chemla, E., and Schlenker, P. (2012). Incremental vs. Symmetric Accounts of Presupposition Projection: An Experimental Approach. *Natural Language Semantics, 20*(2), 177-226.

Chemla, E., and Spector, B. (2011). Experimental Evidence for Embedded Scalar Implicatures. *Journal of Semantics*, doi:10.1093/jos/ffq023.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

Cohen, S. (1988). How to Be a Fallibilist. *Philosophical Perspectives, 2*, 91-123.

Cohen, S. (1999). Contextualism, Skepticism, and the Structure of Reasons. *Noûs, 33*(s13), 57-89.

DeRose, K. (1992). Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research, 52*(4), 913-929.

DeRose, K. (1995). Solving the Skeptical Problem. *Philosophical Review, 104*(1), 1-52.

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.

Fantl, J., and McGrath, M. (2002). Evidence, Pragmatics, and Justification. *Philosophical Review, 111*(1), 67-94.

Feltz, A., and Zarpentine, C. (2010). Do You Know More When It Matters Less? *Philosophical Psychology, 23*(5), 683-706.

Franke, M. (2016). Task Types, Link Functions and Probabilistic Modeling in Experimental Pragmatics. In F. Salfner and U. Sauerland (eds.), *Preproceedings of Trends in Experimental Pragmatics*, 56-73.

Grimm, Stephen R. (2015). Knowledge, Practical Interests, and Rising Tides. In *Epistemic Evaluation: Point and Purpose in Epistemology*, edited by John Greco and David Henderson. Oxford: Oxford University Press.

Hannon, Michael. (2017). A Solution to Knowledge's Threshold Problem. *Philosophical Studies* 174 (3):607-629.

Hansen, N. (2012). On an Alleged Truth/Falsity Asymmetry in Context Shifting Experiments. *Philosophical Quarterly, 62*(248), 530 - 545.

Hansen, N., and Chemla, E. (2013). Experimenting on Contextualism. *Mind and Language, 28*(3), 286-321.

Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press.

Kennedy, Christopher, and Louise McNally. 2010. Color, Context, and Compositionality. *Synthese* 174 (1):79-98.
Lewis, D. (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic, 8*(1), 339-359.
Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy, 74*(4), 549-567.
May, J., Sinnott-Armstrong, W., Hull, J. G., & Zimmerman, A. (2010). Practical Interests, Relevant Alternatives, and Knowledge Attributions: An Empirical Study. *Review of Philosophy and Psychology, 1*(2), 265-273.
Nagel, J. (2010a). Epistemic Anxiety and Adaptive Invariantism. *Philosophical Perspectives, 24*(1), 407-435.
Nagel, J. (2010b). Knowledge Ascriptions and the Psychological Consequences of Thinking about Error. *Philosophical Quarterly, 60*(239), 286-306.
Pinillos, N. Á. (2012). Knowledge, experiments, and practical interests. In J. Brown, & M. Gerken (Eds.), *Knowledge Ascriptions*. Oxford: Oxford University Press.
Pinillos, N. Á., and Simpson, S. (2014). Experimental Evidence Supporting Anti-Intellectualism about Knowledge. In J. Beebe (Ed.), *Advances in Experimental Epistemology*. London: Bloomsbury.
Podsakoff, Philip M., MacKenzie, Scott B., Lee, Jeong-Yeon, and Podsakoff, Nathan P. 2003. 'Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies.' *Journal of Applied Psychology*, 88(5): 897–903.

Pynn, G. (2015). Pragmatic Contextualism. *Metaphilosophy,* 46(1), 26-51.
Pynn, G. (2016). Contextualism in Epistemology. *Oxford Handbooks Online*.
Schaeffer, N.C, and S. Presser. 2003. The Science of Asking Questions. *Annual Review of Sociology* 29 (1):65-88. doi: 10.1146/annurev.soc.29.110702.110112.
Sripada, C. S., and Stanley, J. (2012). Empirical tests of Interest-Relative Invariantism. *Episteme, 9*(1), 3-26.
Stainton, R. J. (2010). Contextualism in Epistemology and the Context Sensitivity of 'Knows'. In M. O'Rourke, & H. S. Silverstein (Eds.), *Knowledge and Skepticism* (Vol. 5, pp. 113-139). Cambridge, MA: MIT Press.
Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford: Oxford University Press.
Turri, J. (2016). Epistemic Contextualism: An Idle Hypothesis. *Australasian Journal of Philosophy 95*(1): 141-156.
Weatherson, B. (2012). Knowledge, Bets, and Interests. In J. Brown, & M. Gerken (Eds.), *Knowledge Ascriptions* (pp. 75-103). Oxford: Oxford University Press.
Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
Williamson, T. (2005). Contextualism, Subject-Sensitive Invariantism, and Knowledge of Knowledge. *Philosophical Quarterly, 55*(219), 213-235.