



Contrasting cases

Book or Report Section

Accepted Version

Hansen, N. (2014) *Contrasting cases*. In: Beebe, J. R. (ed.) *Advances in experimental epistemology*. Bloomsbury, London, pp. 71-95. ISBN 9781472505316 Available at <http://centaur.reading.ac.uk/37843/>

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

Publisher: Bloomsbury

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

Central Archive at the University of Reading

Reading's research outputs online

Contrasting Cases¹

Nat Hansen

Keywords: contextualism, anti-intellectualism, experimental design, contrast, evidence, knowledge

Abstract

This paper concerns the philosophical significance of a choice about how to design the context shifting experiments used by contextualists and anti-intellectualists: Should contexts be judged *jointly*, with contrast, or *separately*, without contrast? Findings in experimental psychology suggest (1) that certain contextual features are more difficult to evaluate when considered separately, and there are reasons to think that one feature—*stakes* or *importance*—that interests contextualists and anti-intellectualists is such a difficult to evaluate attribute, and (2) that joint evaluation of contexts can yield judgments that are more reflective and rational in certain respects. With those two points in mind, a question is raised about what source of evidence provides better support for philosophical theories of how contextual features affect knowledge ascriptions and evidence: Should we prefer evidence consisting of “ordinary” judgments, or more reflective, perhaps more rational judgments? That question is answered in relation to different accounts of what such theories aim to explain, and it is concluded that evidence from contexts evaluated jointly should be an important source of evidence for contextualist and anti-intellectualist theories, a conclusion that is at odds with the methodology of some recent studies in experimental epistemology.

¹ Thanks to Zed Adams, Jonas Åkerman, James Beebe, Gunnar Björnsson, Mikkel Gerken, Chauncey Maher and Eliot Michaelson for helpful comments. Special thanks to Mark Phelan for comments and discussion.

1 Background: Experiments & Context

The empirical foundation of the debate over the nature and extent of context sensitivity in natural language rests in large part on data generated primarily by experiments of a certain kind: *context shifting experiments*.² Context shifting experiments are devised to isolate the effects of some particular feature of context on particular kinds of judgments about specified features of the context. So, for example, a context shifting experiment might vary *what's at stake* for participants in a conversational context, or whether some *possibility of error* has been mentioned, and elicit metalinguistic judgments concerning some semantic or pragmatic property of the use of a target expression when those features are varied: what some particular use of a sentence *says*; whether it says something *true* or *false* (or neither); how *acceptable* the use of the expression in each context is, and so on.³ As long as there aren't more plausible non-linguistic explanations of those judgments, they are evidence of underlying semantic and pragmatic phenomena that linguistic theories aim to explain (Ludlow, 2011, ch. 3).

Alternatively, instead of eliciting judgments about linguistic features of the context (e.g., whether what is said is true or acceptable), one might elicit judgments about some non-linguistic aspect of the context, such as whether some character in the story *knows* something, or how *confident* she should be that something is the case (rather than whether what is said by a use of a sentence

² “Context shifting experiments” are a part of (and the name is derived from) what Cappelen and Lepore (2005, p. 10) call “Context Shifting Arguments”. A Context Shifting Argument takes the data generated by a context shifting experiment as a premise.

³ For discussion of metalinguistic judgments, see Birdsong (1989) and Schütze (1996, Ch. 3).

ascribing knowledge or confidence is true or false).⁴

Some of the experimental philosophers who have investigated the claims of anti-intellectualism—the view that whether one counts as knowing a proposition, or the quality of one’s evidence in favor of the proposition, partly depends on the “stakes” or practical costs of getting it wrong—employ this kind of context shifting experiment (May et al., 2010; Phelan 2013).

The goal of context shifting experiments is to set up conditions so that the effects (if there are any) of changing specific features of the relevant context (the independent variable) on judgments (the dependent variable) can be observed. Contextualists and their opponents then go on to try to explain those observed effects using their preferred theoretical resources: indexicality, free enrichment, occasion-sensitivity, conversational implicature, focal bias, and so on.⁵

Many context shifting experiments have been conducted informally, from the theorist’s armchair. But with increasing frequency, formal versions of context shifting experiments have

⁴ Hazlett (2010, pp. 497-8) distinguishes “two competing methods of theorizing in epistemology—one based on intuitions about knowledge, and the other based on intuitions about language”. DeRose argues that only meta-linguistic context shifting experiments yield data that can confirm or disconfirm predictions made by his particular variety of contextualism. For his argument, see DeRose (2009, p. 49 n. 2) and (2011, pp. 84–85). Sosa (2000, p. 1) characterizes contextualism as engaging in “metalinguistic ascent”, whereby it “replaces a given question with a related but different question. About words that formulate one’s original question, the contextualist asks when those words are correctly applicable”. Sosa goes on to say that there are questions, like the nature of justification, that the epistemologist can discuss “with no metalinguistic detour” (p. 6).

⁵ I say “relevant” context because theorists with different theoretical commitments vary elements of different kinds of contexts: the context of *utterance* (DeRose 1992, 2009, e.g.), or the context of the *subject who is said to know* (Stanley 2005, e.g.).

been conducted with all the apparatus of contemporary psychology at their disposal. The turn to formal versions of context shifting experiments is motivated on one hand by a general skepticism about the reliability of philosophers' intuitions and on the other by the need to respond to such skepticism (see Hansen and Chemla, forthcoming, for discussion of such skepticism as well as vindications of certain armchair judgments about context shifting experiments). One side effect of the turn to more formal experiments is that it has drawn attention to subtle but important elements of the design of context shifting experiments that have been largely overlooked in their informal use. As an illustration of the features of a context shifting experiment that are brought into relief when they are adopted for use in formal experiments, consider the highlighted features of the following much-discussed context shifting experiment introduced by Keith DeRose (1992, 2009):

Bank Case A. My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paychecks. But as we drive past the bank, we notice that the lines inside are very long, as they often are on Friday afternoons. Although we generally like to deposit our paychecks as soon as possible, it is not especially important in this case that they be deposited right away, so I suggest that we drive straight home and deposit our paychecks on Saturday morning. My wife says, 'Maybe the bank won't be open tomorrow. Lots of banks are closed on Saturdays'. I reply, 'No, **I know it'll be open. I was just there two weeks ago on Saturday. It's open until noon.**' [The bank is open on Saturday.]

Bank Case B. My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paychecks on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and discovered that it was open until noon. But in this case, we have just written a very large and very important check. If our paychecks are not deposited into our checking account before Monday morning, the important check we wrote will bounce, leaving us in a *very* bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, 'Do you know the bank will be open tomorrow?' Remaining as confident as I was before that the bank will be open then, still, I reply, 'Well, no, **I don't know.** I'd better go

in and make sure'. [The bank is open on Saturday.]⁶

The metalinguistic judgments DeRose expects us to make in response to the “bank” context shifting experiment—truth value judgments about the sentences in boldface—are supposed to provide evidence of the context sensitivity of the word “know”.

But there are two asymmetries between the two contexts DeRose describes that make it difficult to isolate the effect that changes in the context of utterance have on metalinguistic judgments about the target sentences.

First, in addition to varying specific features of the contexts of utterance, DeRose also varies the *sentences* that are supposed to be evaluated in each context (those that I have marked in boldface). He varies the *polarity* of the sentences (“I know...” vs. “I don’t know. . .”), whether there is anaphoric reference to the bank (“it”) and what linguistic material is elided (“I know it’ll be open [tomorrow]” vs. “I don’t know [the bank will be open tomorrow]”), and whether the discourse marker “Well,…” is present.⁷ Varying all of those linguistic elements makes it harder to defend the idea that it is the change in the *context of utterance* that is affecting our judgments about the uses of the sentences, rather than the changes DeRose makes in the sentences that are used (or

⁶ I have added boldface to pick out the sentences we’re supposed to evaluate, and I have underlined the sentences where the character in the stories who claims to know or denies that he knows gives evidence in support of the proposition that the bank will be open tomorrow.

⁷ For a discussion of the pragmatic significance of the discourse marker “well”, see Jucker (1993). Thanks to Emma Borg for bringing this paper to my attention.

some combination of both factors).

Second, the underlined sentences are where the character in the story who claims to know the bank will be open tomorrow states evidence in support of the proposition that the bank will be open tomorrow. But those statements differ subtly in how they are worded, occur in different places in the story, and the statement in Case A is in direct discourse, while the statement in Case B is in indirect discourse. The statement of evidence is arguably more salient in Case A, where DeRose's judgment is that he knows that the bank will be open, while it is less salient in Case B, where DeRose's judgment is that he does not know the bank will be open. It is possible that simply locating that statement in different places in the story affects our judgment of whether or not the character's statement that he knows the bank will be open is true.

This is not to argue that these factors *do* affect our judgments in these cases, only that they make it more difficult to isolate the effect that changing the context has on our judgments. Anyone interested in identifying those effects should revise their context shifting experiments accordingly, so that as little as possible is varied between contexts except for the relevant features of the context of utterance (in DeRose's investigation of "know", those features are the stakes and whether a possibility of error is mentioned).⁸

Even once the unnecessary asymmetries between the contexts being evaluated are eliminated, there remain questions about how subtle features of experimental design affect judgments. For example, there is evidence that the order in which scenarios are presented (Schwitzgebel and Cushman, 2012), whether the

⁸ More recent context shifting experiments avoid these asymmetries. See, e.g., Sripada and Stanley (forthcoming) and the context shifting experiment discussed below, taken from Phelan (2013).

sentences participants are asked to judge are positive or negative (Hansen and Chemla, forthcoming), and whether participants only see contexts separately (without contrast) or jointly (with contrast) (Phelan, 2013) can significantly affect judgments about them. In this paper, I will consider this final feature of the design of context shifting experiments—whether to employ separate or joint evaluation of contexts—in detail. I will first describe reasons to think that separate evaluation (involved in experiments with a between-subjects design) is the better design for context shifting experiments because it more closely resembles the structure of ordinary judgments (which do not involve explicit comparisons between contexts). I will then draw on findings in experimental psychology to argue that joint evaluation of contexts can yield judgments that are more “rational” in certain respects. With those two arguments in place, it is then possible to raise a question about which experimental design generates better evidence for contextualist and anti-intellectualist theories: Should the evidence consist of “ordinary” judgments, or more reflective, perhaps more “rational” judgments? How one answers that question depends on what one understands the explanatory project of contextualist and anti-intellectualist theories to be. In the final section of the paper, I’ll describe two different ways of understanding those explanatory projects and how they bear on the question of what kinds of experiments provide the best evidence for such theories.

2 DeRose on Joint vs. Separate Evaluation of Contexts

DeRose says that when his contextualist scenarios (like the bank scenario discussed above) are considered separately, the intuitions that they generate are “fairly strong” (DeRose 2005, p. 175/2009, p. 49) “fairly clear” (DeRose, 2005, p. 193), or “quite strong” (DeRose 1999, p. 196, 2009, p. 55 n. 7).⁹ But he worries that if the two contexts that make up a context shifting experiment are

⁹ See also DeRose (2009, p. 2).

considered jointly, we may become less certain of our intuitions about the two contexts:

Of course, we may begin to doubt the intuitions above when we consider [the contexts] together, wondering whether the claim to know in the first case and the admission that I don't know in the second can really both be true (DeRose 2002, p. 195 n. 6/2009 p. 55 n. 7).

One interesting feature of DeRose's remarks is that he doesn't say whether he finds joint or separate evaluation of contexts (if either) preferable. His *practice* favors joint evaluation: The informal presentation of DeRose's context shifting experiments (indeed, of all the informal context shifting experiments in the contextualist debate) require judgments about contexts that are presented jointly.¹⁰ But I get the feeling that DeRose would prefer that the contexts be considered individually, since that would, by his own account, produce intuitions that more strongly aligned with his predictions. And DeRose is committed to a view about what constitutes the best evidence for contextualist theories which lends support to the practice of using context shifting experiments that present contexts separately:

The 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.

This type of basis in ordinary language not only provides the best grounds we have for accepting contextualism concerning knowledge attributions, but, I believe, is evidence of the very best type one can have for concluding that any piece of ordinary language has context-sensitive truth-conditions (DeRose 2005, p. 172/2009, pp. 47–48).

Given that DeRose thinks that the “best grounds for accepting contextualism come from how...sentences are used in ordinary,

¹⁰ It would be awkward (though not impossible) to craft a paper in which readers only saw one or the other context by itself.

non-philosophical talk”, and given that, as Daniel Kahneman puts it, “We normally experience life in the between-subjects mode, in which contrasting alternatives are absent” (Kahneman, 2011, p. 354), it seems plausible that DeRose should think that context shifting experiments that present contexts separately generate better grounds for contextualism than context shifting experiments that present contexts jointly.¹¹

Further support for this idea can be found in recent experimental philosophy, where it has been explicitly argued that evidence gathered from context shifting experiments that evaluate contexts separately is preferable to evidence gathered from joint evaluation of contexts.

3 Experimenting with Separate and Joint Evaluation

Phelan (2013) conducted a series of experiments that revealed significant effects of a feature of context invoked in certain context shifting experiments, namely *practical importance*, or *what is at stake*, in contexts evaluated jointly. But those effects disappeared when each of the contexts making up the context shifting experiment was considered separately, in a “non-juxtaposed” experimental design. Phelan’s finding of no significant difference between responses to contexts when those contexts are evaluated separately lines up with other recent experimental results concerning anti-intellectualism about knowledge (Feltz and Zarpentine, 2010) and contextualism about knowledge ascriptions (Buckwalter, 2010), both of which relied exclusively on separate evaluation of contexts.

In this section, I will describe Phelan’s findings. Later, I will argue that while Phelan’s findings may suggest a problem for using

¹¹ For other examples of the claim that everyday life resembles a between-subjects experiment, see Kahneman (2000, p. 682) and Shafir (1998, p. 72).

contrasting cases in the design of context shifting experiments, it isn't at all obvious whether that problem is genuine.¹²

Phelan takes as his target the “anti-intellectualist” view that the practical importance, or “cost” or “stakes”, of being right or wrong about a proposition has an effect on one’s evidence supporting the proposition (p. 3).¹³ Anti-intellectualism about evidence is motivated in part by judgments about context shifting experiments in which only the practical importance (or “stakes”, or “costs”) of being right about a proposition is varied between contexts. For example (given certain assumptions¹⁴) the anti-intellectualist view targeted by Phelan would predict that judgments about how confident the character Kate is in the following two contexts should vary in the following way: In the *Unimportant* context, Kate should be more confident that she is on Main Street than she is in the *Important* context. (The italicized material in the contexts that follow is not present in the version given to participants. I follow Phelan in underlining material that varies in the two contexts; the material that follows is the same in both contexts.)

Unimportant (Passerby): Kate is ambling down the street, out on a walk for no

¹² An early, unpublished (but often cited) version of Phelan’s study (Neta and Phelan, Ms.) contains the claim that their studies “obviously suggest a problem for the philosophical strategy of [using] contrasting cases to elicit intuitions in support of one position or another” (p. 24).

¹³ Phelan discusses two subtly different versions of this view, “Anti-intellectualism about Evidence”, given in Stanley (2005) and Stanley (2007).

¹⁴ In order for anti-intellectualism about evidence to make testable predictions about ordinary judgments, Phelan introduces what he calls the “Bridge from Rational Confidence to Evidence (BRCE): People’s implicit commitments about an agent’s evidence set *or* quality of evidence are reflected in their explicit intuitive judgments about how confident that agent ought to be in various propositions supported by that evidence” (p. 7). The BRCE allows Phelan to draw conclusions about people’s commitments about *evidence* from their judgments about how *confident* subjects ought to be.

particular reason and with no particular place to go.

Important (Passerby): Kate needs to get to Main Street by noon: her life depends on it.

She comes to an intersection and asks a passerby the name of the street. “Main street”, the passerby says. Kate looks at her watch, and it reads 11:45 AM. Kate’s eyesight is perfectly normal, and she sees her watch clearly. Kate’s hearing is perfectly normal, and she hears the passerby quite well. She has no special reason to believe that the passerby is inaccurate. She also has no special reason to believe that her watch is inaccurate. Kate could gather further evidence that she is on Main Street (she could, for instance, find a map), but she doesn’t do so, since, on the basis of what the passerby tells her, she already thinks that she is on Main Street.

Phelan goes about attempting to verify the prediction by asking participants in his experiment to rate, on a 7-point Likert scale (anchored at 1 with “not confident” and at 7 with “very confident”), how confident the character Kate should be that she is on Main Street. He found no significant difference between judgments about Kate’s confidence in the two contexts when each participant was asked to judge only one of the two contexts.¹⁵

But, interestingly, Phelan found that changing the stakes had a significant effect on judgments of confidence in “juxtaposed cases”, when participants were allowed to jointly evaluate both the *Unimportant* and *Important* contexts.¹⁶ Phelan then ran two

¹⁵ The usual caveats about drawing conclusions from null results apply here.

¹⁶ Phelan reports that the mean responses to the important and unimportant contexts were 4.5 and 5.32, respectively, with $p < .001$. Emmanuel Chemla and I (Hansen and Chemla forthcoming) uncovered a similar result with truth value judgments about knowledge ascriptions using several different context shifting experiments based on DeRose’s bank scenario. We found a significant effect of changing contexts on truth value judgments about bank-style scenarios only when participants had the chance to make judgments about multiple contexts. In our experiment, unlike Phelan’s, participants never saw two contexts simultaneously. Instead, over the course of the experiment, participants in our experiment made judgments about knowledge ascriptions in response to 16 bank-style contexts. Hsee et al. (1999, p. 576 n. 1) says the kind of evaluation mode we used

additional context shifting experiments testing for the effects of changing stakes, but which differed from the scenario described above in terms of the *reliability of the information source* that supplies Kate with the information that she's on Main Street. In the second version, it is a pair of "drunks" who tell Kate that she is on Main Street, while in the third version, Kate gets her information about what street she's on from a street sign. In each experiment, there was a significant difference in responses to the important and unimportant contexts when participants saw them "juxtaposed", but that difference disappeared when they saw them separately.

As Phelan points out, his findings are interesting because the context shifting experiments that involve joint evaluation of contexts more closely mirror the standard, informal set up of context shifting experiments. Those reading a philosophy paper, for example, form their judgments while having multiple contexts simultaneously in view.¹⁷ One might conclude that philosophers who unreflectively employ informal context shifting experiments with joint evaluation of contexts are mistakenly offering theories that aim to explain what turns out to be merely an artifact of their particular experimental design, rather than a fact about judgments made in ordinary circumstances.¹⁸

4 Why is Contrast a Problem?

Here is a schematic representation of central results of Phelan's

"involve[s] a JE [Joint Evaluation] flavor because individuals evaluating a later option may recall the previous option and make a comparison".

¹⁷ Stanley's (2005) bank context shifting experiment involves considering *five* related contexts.

¹⁸ As mentioned above, Neta and Phelan (Ms.) draw just such a conclusion from observations about the role played by joint evaluation in judgments about the effect of stakes on confidence.

experimental study:

- Changing stakes *do not* have a significant effect on judgments of confidence about contexts when participants see those contexts separately, without contrast.
- Changing stakes *do* have a significant effect on judgments of confidence about contexts when participants see those contexts jointly, with contrast.

Phelan infers that it is problematic for philosophers to cite the effect of changing stakes on judgments of confidence seen in jointly considered contrasting cases in support of a theory like anti-intellectualism about evidence. But that inference is only reasonable given a commitment to the idea that effects that only show up in “juxtaposed” contrasting cases *do not* reveal genuine effects of stakes on judgments of confidence. Why accept that commitment?

Phelan considers two arguments that defend the importance of effects that show up only in contexts considered jointly, and he criticizes and rejects both. I’ll briefly sketch both arguments and his responses before developing a third argument in favor of embracing effects that show up only in contexts considered jointly that avoids Phelan’s criticisms.

First, one might argue that the effect of changing stakes on judgments of confidence emerges only in contexts considered jointly because only then are stakes *salient*. When contexts are evaluated separately, the stakes are not a particularly prominent feature of the context and so do not end up affecting judgments of confidence.¹⁹ Second, one might argue that when evaluating

¹⁹ Sripada and Stanley (forthcoming, pp. 6–7) make an argument along these lines, defending anti-intellectualism against experimental results indicating that stakes do not affect judgments about knowledge based only on separate evaluation of contexts.

contexts separately, participants are uncertain how to respond, and so make judgments that “land, more or less arbitrarily, somewhere in the middle of the scale” (p. 11).²⁰ But when evaluating contexts jointly, they have “more guidance”, and so better represent the role that stakes play in affecting judgments of confidence.²¹

Phelan responds to both of these arguments by comparing responses to the non-juxtaposed contexts in the three versions of his context shifting experiment that differ in terms of the reliability of the information source that provides Kate the information that she is on Main Street. He observes that, even in contexts considered separately, the mean responses of how confident Kate should be that she is on Main Street track the reliability of the source of her information that she is on Main Street: “[T]he mean value of participants’ answers for the non-juxtaposed cases involving the highly reliable street sign (5.7) was higher than that for cases involving the moderately reliable passerby (5.02), which was higher than that for the unreliable drunks (4.56)” (p. 12). Phelan found that there was a significant effect of the reliability of the information source on responses in “non-juxtaposed” cases, but no significant effect of importance. He then takes that finding to support the denial of the consequent in the following conditional:

[I]f participants’ responses to a single case do not properly reflect the extent to which stakes matter, then they should also not properly reflect the extent to which

²⁰ DeRose (2011, p. 94) hilariously calls this kind of response the “WTF?! neutral response”.

²¹ Ludlow (2011, p. 75) gives an example of how joint evaluation can improve subjects’ understanding of an experimental task: “As reported in Spencer (1973), Hill (1961) notes that sentences drawn from *Syntactic Structures* drew mixed results from experimental subjects. ‘The child seems sleeping’ was accepted by 4 of the 10 subjects *until* it was paired with ‘The child seems to be sleeping’ at which point all 10 subjects vote negatively. Establishing the contrast helped the subjects to see what the task demand was”.

other, equally salient, factors matter (p. 12).

But, because both the antecedent and consequent of the conditional involve negations, it is easier to see what's going on here if you take the experiment to affirm the antecedent of the conditional's contrapositive:

If participants' responses to a single case properly reflect the extent to which factors that are equally salient to stakes matter, then they should also properly reflect the extent to which stakes matter.

Participants in Phelan's experiments had significantly different responses about how confident a character should be when she received information about what street she was on from sources of varying reliability (a drunk, a normal passerby, and a street sign), and they did so in contexts presented *separately*. If reliability of the information source in a context is as salient as what is at stake, then Phelan has good reason to affirm the antecedent of the (rewritten) conditional and conclude that participants' responses to a single case properly reflect the extent to which stakes matter. Put another way, without some reason to think that participants' responses to stakes and reliability of information source differ systematically, "it would be *ad hoc* to claim that they do not . . . notice, or do not properly respond to, the stakes in the single cases" (p. 13).

A key part of Phelan's argument is the assumption that the reliability of information sources is equally as salient as what is at stake. If there is reason to reject that assumption, then his argument against the idea that judgments about contexts presented separately do not properly reflect the extent to which stakes matter is not convincing. I will present some reasons to reject that assumption in

the following section.

5 Further Case Studies on Separate and Joint Evaluation

Hsee et al. (1999, pp. 583–4) discuss several experiments in which switching from separate to joint evaluation corresponds not just with a significant difference in judgments, but with a *reversal* in the judgments of participants. So, for example, when participants in an experiment (conducted in Hsee 1998) were asked to judge how much they would be willing to pay for each of the two sets of dinnerware in Table 1, they judged set J to be more valuable when the sets were presented jointly.

Table 1: Judging the Value of Sets of Tableware

	Set J (includes 40 pcs)	Set S (includes 24 pcs)
Dinner Plates	8, in good condition	8, in good condition
Soup/salad bowls	8, in good condition	8, in good condition
Dessert plates	8, in good condition	8, in good condition
Cups	8, 2 of which are broken	—
Saucers	8, 7 of which are broken	—

But when participants only saw one or the other set of tableware and asked to judge how much they would be willing to pay for them, judgments were reversed: Participants were willing to pay more for Set S than for Set J (Hsee, 1998; Hsee et al., 1999; Kahneman, 2011). Hsee et al. (1999, p. 584) notes that even though Set J contains all the pieces in Set S plus six additional intact cups and one more intact saucer, participants were willing to pay more for Set S when the sets were considered separately, “although it was the inferior option”.

Or consider another experiment from Hsee (1998), which “asked students to imagine that they were relaxing on a beach by Lake

Michigan and were in the mood for some ice cream” (Hsee, et al. 1999, p. 583). Like the Tableware experiment, some participants were asked to judge how much they were willing to pay for each of two ice cream servings offered by two vendors presented jointly, while others were asked to judge how much they were willing to pay for one or the other serving option, presented separately (see Table 2).

Table 2: Choosing Ice Cream

Vendor J	Vendor S
10 oz. cup with 8 oz. ice cream	5 oz. cup with 7 oz. ice cream

Both serving options were accompanied by a drawing depicting the serving. Hsee et al. (1999, p. 583) reports the findings of the earlier study as follows:

Note that, objectively speaking, Vendor J’s serving dominated Vendor S’s, because it had more ice cream (and also offered a larger cup). However, J’s serving was underfilled, and S’s serving was overfilled. The results revealed a JE/SE [Joint Evaluation/Separate Evaluation] reversal: In JE [Joint Evaluation], people were willing to pay more for Vendor J’s serving, but in SE [Separate Evaluation], they were willing to pay more for Vendor S’s serving.

What accounts for this (and many other) reversals in judgment between separate and joint evaluation of cases? The answer given in (Hsee et al., 1999, p. 578) turns on the fact that “some attributes. . . are easy to evaluate independently, whereas other attributes. . . are more difficult to evaluate independently”. For example, whether a particular set of table-ware *has broken pieces* or whether an ice cream cup is *overfilled* is easy to evaluate independently, while the *significance of the total number of pieces* in a set of tableware, or “the desirability of a given amount of ice cream”, is more difficult to evaluate independently.

Whether an attribute is easy or difficult to evaluate, according to

Hsee, et al., “depends on the type and the amount of information the evaluators have about the attribute”. Relevant information includes which value for the attribute would be evaluatively neutral, what the *best* and *worst* values for the attribute would be, and “any other information that helps the evaluator map a given value of the attribute onto the evaluation scale” (p. 578). An extremely difficult attribute to evaluate would be one where the judge has no information about the upper and lower values the attribute can have, or what the average value of the attribute would be. So, for example, suppose you were asked to judge how suitable a candidate is for entry into philosophy B.A. program based solely on her score of 15 on her French *baccalauréat général*.²² Unfortunately you don’t know what a good or bad score on the *bac* would be, or even what the average is. You only know that higher scores are better. Suppose also that you also don’t get to *compare* the candidate with any others—she’s the only French applicant to the program. In this situation, any judgment would be a stab in the dark—there’s no grounds on which to give the candidate either a positive or a negative evaluation.

Your job is easier if you know what the average, neutral value for the attribute is, even if you don’t know what the highest and lowest values for the attribute would be. Given a particular score, you can then easily judge whether it falls above or below the average, and correspondingly give it a positive or negative evaluation. So suppose you know that the average score on the *bac* is 11. Now you can evaluate the student’s score of 15 positively, but you have no way to judge *how* positively it should be evaluated.

Still easier is a situation in which you know not only the average

²² This example is based on an experiment conducted in Hsee et al (1999), concerning evaluations of a foreign applicant to a University who has taken an “Academic Potential Exam” in her home country.

score, but scores on the high and low end of what is possible:

In the *baccalauréat général*, ten out of twenty is a pass. . . 16 is a *très bien* (*summa cum laude*), a big bouquet of starred As in the British system. Cambridge expects 17 from a French *bachelier* (Harding, 2012).

Now you are in a position to make a much more nuanced evaluation of the applicant's score. It's quite good—not fantastic, but good enough for this program (it's not Cambridge, after all).

With a more concrete sense of the kind of information that makes an attribute easy or difficult to evaluate, we can then ask whether there is any reason to think that *what's at stake* in a context is more difficult to evaluate than the *reliability* of an information source. I think the answer is that it *is* more difficult to evaluate what's at stake. First of all, the reliability of an information source has a clear upper and lower bound: a source can be 100% reliable, or completely unreliable, never producing the correct answer. Given a particular information source (a drunk, an ordinary passerby, a street sign), it is possible to make an informed (if rough) judgment about where that information source falls on the (upper and lower bounded) scale of reliability, even without comparing it to the reliability of other information sources. In contrast, there is no clear upper bound to *what can be at stake* in a context. It seems that there is a lower bound: *Nothing* might turn on whether a proposition turns out to be true or false. That seems to be an element of the “Unimportant” context Phelan describes. But, on the other end of the scale, what's the most important thing that could turn on whether or not a proposition is true or false? Certainly whether someone lives or dies is important, but there's always something more important (two people's lives, a million, the fate of the country, the planet, the universe, all possible universes. . .). Since there's no clear upper bound, there's also no clear sense of what something of *average* importance would be. So when a participant in a survey is asked to make a judgment about a single context in which *what's at stake* is mentioned, that attribute

counts as *difficult to evaluate*, in contrast with the reliability of an information source, which is (comparatively) easy to evaluate.²³

Phelan wants to defend the idea that responses to contexts considered separately provide better evidence for anti-intellectualism than cases considered jointly. He responds to the idea that joint evaluation might make subjects better equipped to evaluate what's at stake in a context as follows (this is my reconstruction of his response):

1. If participants' responses to a single case do not properly reflect the extent to which stakes matter, then they should also not properly reflect the extent to which other, equally salient, factors matter (p. 18).
2. The reliability of an information source is as salient as what is at stake in a context.
3. Participants' responses to a single case *do* properly reflect the reliability of a relevant information source.

Conclusion: Participants' responses to a single case *do* properly reflect the extent to which stakes matter.

The upshot of the discussion of what makes an attribute easy or difficult to evaluate in this section is that premise (2) in Phelan's argument is false, assuming that the ease or difficulty of evaluating an attribute is a suitable construal of Phelan's notion of "salience".

²³ Hsee et al. (1999, p. 580) observe that the fact that an attribute is difficult to evaluate does not mean that subjects do not *understand* what the attribute means: "For example, everybody knows what money is and how much a dollar is worth, but the monetary attribute of an option can be difficult to evaluate if the decision maker does not know the evaluability information for that attribute in the given context. Suppose, for instance, that a person on a trip to a foreign country has learned that a particular hotel room costs \$50 a night and needs to judge the desirability of this price. If the person is not familiar with the hotel prices of that country, it will be difficult for him to evaluate whether \$50 is a good or bad price".

The reliability of an information source is easier to evaluate than what is at stake. That explains why the effect of changing the reliability of the relevant information source shows up in separate evaluation, while the effects of changing stakes only show up in joint evaluation.²⁴ So Phelan's argument that responses to contexts considered separately *do* properly reflect the extent to which stakes matter (in contrast with responses elicited in contexts considered jointly) should be resisted. But that's only to say that there isn't yet a convincing argument that separate evaluation should be favored over joint evaluation—so far, it's still an open question whether data gathered using separate or joint evaluation is better evidence for contextualism and anti-intellectualism.

6 Which Type of Evaluation Generates Better Evidence for Contextualism and Anti-Intellectualism?

Phelan observed that changing stakes only seemed to have an effect on judgments about confidence when contexts were evaluated jointly. He then argued that the effect of stakes observed in contexts evaluated separately does genuinely reflect the effect of what's at stake on judgments about confidence. In the last section I challenged that argument. Now, in this section, I will consider another argument that tries to show that effects that show up in contexts considered separately are better evidence for contextualist and anti-intellectualist theories than effects that show up only in contexts evaluated jointly.

²⁴ Hsee et al (1999) conducted an experiment that tested for effects of different types of evaluability information that subjects might have, corresponding to the three situations described above: no information, information about average scores, and best and worst score information. Their flat (no significant difference between scores) result for the *no information* situation parallels Phelan's result for evaluations of contexts involving different stakes considered separately, whereas the significant differences they observed between evaluations of different scores in the situation where participants had information about best and worst scores parallels Phelan's result for separate evaluation of contexts involving sources of information of varying reliability.

Here is my reconstruction of the argument, which is implicit in DeRose's remarks concerning "the best grounds for accepting contextualism" and his attitude towards contexts considered separately and jointly (introduced in §2, above):

1. The best grounds for accepting contextualism come from how knowledge-attributing (and knowledge-denying) sentences are used in ordinary, non-philosophical talk (DeRose 2005, p. 172/2009, p. 47).
2. Contexts evaluated separately (and not contexts evaluated jointly) accurately represent how subjects use ordinary, non-philosophical talk.
3. So data gathered from contexts considered separately (and not contexts considered jointly) provides the best grounds for accepting contextualism.

DeRose does not explicitly commit himself to premise 2., but as discussed above, I think there is reason to think he implicitly accepts it.

Embracing this argument would mean that the proper design of context shifting experiments (both informal and formal) should involve separate evaluation of contexts, and not joint evaluation.

I now want to challenge premise (1) in (my reconstruction of) DeRose's argument by giving reasons to think that, for certain purposes, data generated by joint evaluation of contexts should be at least on the same footing as (if not considered superior to) data generated by separate evaluation of contexts. The essential move in my argument can be summarized by the following remark from Kahneman (2011, p. 361):

...rationality is generally served by broader and more comprehensive frames, and

joint evaluation is obviously broader than single evaluation.²⁵

Subjects tend to make better, more informed, more “rational” judgments about contexts when they are given more than one context to evaluate. This idea was present in the earlier discussion of judgments about the value of the two sets of tableware and the different ice cream options: When considered side by side, ice cream option J is obviously preferable, and participants select it, but when considered separately, subjects do not choose the dominant option, they choose the “objectively inferior option” (Hsee et al. 1999, p. 588). That is a clear illustration of how being able to evaluate options jointly can lead to improved judgments.²⁶

²⁵ In Kahneman’s Nobel Prize lecture, he makes a claim that can seem like it’s in tension with this idea. He says,

...intuitive judgments and preferences are best studied in between-subjects designs. . . The difficulties of [within-subjects] designs were noted long ago by Kahneman and Tversky (1982), who pointed out that ‘within-subjects designs are associated with significant problems of interpretation in several areas of psychological research (Poulton 1975)’” (Kahneman 2002, pp. 473–474).

But the apparent tension is resolved when it is pointed out that “intuitive judgments” for Kahneman are rapid and automatic, and contrast with “deliberate thought processes”, which are slow and involve reflection. Separate evaluation may be the right way to study intuitive judgments in Kahneman’s sense, but the question under consideration in this section is whether it is better to employ “intuitive judgments” or “deliberate thought processes” as evidence for contextualism and anti-intellectualism. It is possible to both think that “deliberate thought processes” are more rational than “intuitive judgments”, and therefore provide better evidence, and also that separate evaluation is the best way to study “intuitive judgments”. For further discussion of the distinction between “intuitive” and “deliberate” (or type 1 and type 2 processes) in relation to the contextualist debate, see Gerken (2012). Thanks to Mikkel Gerken for pointing out the passage in Kahneman.

²⁶ Additional reflection on this idea can be found in Pinillos et al. (2011). Pinillos et al. conducted a study of the Knobe Effect, which, unlike Knobe’s original study, allowed joint evaluation of scenarios, and found that participants were “less likely to give the asymmetric ‘Knobe’ response” (p. 129). Discussing this result, Pinillos

Another illustration of how joint evaluation can produce improved judgments is given in Kahneman and Tversky (1996) in relation to the “conjunction fallacy”. The “conjunction fallacy” is the tendency of subjects, in certain conditions, to judge that $p \& q$ is more probable than p alone. So, for example, consider the following vignette and response options (Kahneman and Tversky, 1996, p. 587):

Linda is in her early thirties. She is single, outspoken, and very bright. As a student she majored in philosophy and was deeply concerned with issues of discrimination and social justice. Suppose there are 1,000 women who fit this description. How many of them are

- (a) high school teachers?
- (b) bank tellers? or
- (c) bank tellers and active feminists?

Kahneman and Tversky report that when participants were allowed to see options (a), (b) and (c), 64% conformed to the conjunction rule, which holds that that conjunctions must be less probable (or equally probable) than either conjunct. But in an experiment with a between-subjects design (that is, one where subjects consider the relevant responses separately), when participants saw only either options (a) and (b) or (a) and (c), “the estimates for feminist bank tellers (median category: ‘more than 50’) were significantly higher than the estimates for bank tellers (median category: ‘13-20,’ $p <$

et al. say “we believe that presenting agents with both vignettes (and letting them see the range of multiple choice answers) pushes them to think more carefully before giving the final judgment. If we compare this with the original Knobe experiments (where subjects were given only one vignette followed by just two answer options), it is plausible that subjects there were less careful in their reasoning” (p. 133).

.01 by a Mann-Whitney test)” (p. 587). That is, in the between-subjects design, when participants were asked to evaluate the probability of (b) and (c) separately, they tended to violate the conjunction rule, while in the within-subjects design, when they were allowed to see both objects jointly, they tended to adhere to the rule.

So there is an argument that supports the idea that we should favor data generated by contexts considered jointly over data generated by contexts considered separately. And we’re now in a position to be able to challenge DeRose’s assumption that

The best grounds for accepting contextualism come from how knowledge-attributing (and knowledge-denying) sentences are used in ordinary, non-philosophical talk (DeRose 2005, p. 172/2009, p. 47).

There is now a competing conception of what might be considered “better” grounds for accepting contextualism, namely more informed judgments, based on joint evaluation of contexts. Pinillos et al. (2011, p. 127) put the idea this way: “In general, giving subjects further relevant information will allow them to make a more informed judgment. In short, it will put them in a better epistemic situation”.

7 Conclusion: Two Explanatory Projects

One explanatory project that contextualists and anti-intellectualists might be engaged in is a branch of cognitive science. In the case of contextualism, this project is closely related to the explanatory projects of empirical semantics and pragmatics: The goal is to build up a linguistic theory that explains and predicts certain linguistic phenomena. Evidence of those phenomena can be uncovered by eliciting judgments in linguistic experiments, looking at linguistic corpora, and recording and transcribing linguistic use “in the wild”. While the immediate goal of this project is to explain a domain of specifically linguistic phenomena, evidence for and against competing theories also comes from how

well theories mesh with neighboring areas of empirical investigation. The ultimate goal is a satisfactory explanation of “the total speech act in the total speech situation”—how linguistic capacities interact with other forms of cognition to produce the richly textured conversational understanding we enjoy. This explanatory project is essentially focused on *language* and linguistic activity. I think it is uncontroversial that *both* evidence collected from separate and joint evaluation of contexts is relevant to this explanatory project. Those engaged in this type of project want to know, among other things, why linguistic judgments differ in separate and joint evaluation (when they do), and to know that, we obviously need both kinds of evidence.²⁷

The second explanatory project is not essentially focused on psychological explanation. It seeks answers to *metaphysical* questions: What is *knowledge*? What is *evidence*? We might approach those metaphysical questions by way of answers to linguistic questions: How do we use the word “know”? Or by way of questions about judgments involving the relevant concepts: How do people make judgments about how *confident* someone should be? These routes to the nature of knowledge or evidence depend on controversial assumptions about the relation between our linguistic behavior with “know” or our judgments about confidence and the nature of knowledge and evidence. I won’t engage here in disputes over the best way to understand that relation.²⁸ Instead, I only want

²⁷ For example, Kahneman and Tversky (1996, p. 587) say that “the between-subjects design is appropriate when we want to understand ‘pure’ heuristic reasoning; the within-subjects design is appropriate when we wish to understand how conflicts between rules and heuristics are resolved”, and Stanovich (2011, pp. 124–5) discusses the way that within- and between-subjects designs may interact differently with individual differences in rational thinking dispositions.

²⁸ There are many views about the relation between linguistic facts about “know” and the nature of knowledge. Ludlow (2005, p. 13) claims that “any investigation into the nature of knowledge which did not conform to some significant degree with the semantics of the term “knows” would simply be missing the point. . .

to suggest that insofar as one is engaged in the project of getting at the nature of knowledge and evidence via linguistic or psychological investigations, it makes sense to be interested in the *best* judgments that subjects make about knowledge ascriptions or how confident subjects should be, and not exclusively in “ordinary” judgments, subject as they are to known forms of bias and distortion. If subjects’ judgments are taken to be a mirror of reality, that mirror should be as polished as possible.

So, insofar as contextualists are interested in getting at the nature of knowledge, or anti-intellectualists are interested in getting at the nature of evidence, in addition to being engaged in an aspect of the (extremely worthwhile) project of empirical linguistics and psychology, they should drop the commitment to the idea that the best grounds for contextualism are offered by ordinary uses of knowledge-ascribing (and knowledge-denying) sentences in ordinary talk. Better grounds for contextualism and anti-intellectualism, understood as theories concerning the nature of knowledge and evidence, are how speakers use knowledge-ascribing and knowledge-denying sentences, or make judgments about confidence, in situations where all the necessary work has been done to eliminate avoidable sources of bias. Employing context shifting experiments that ask for joint evaluation of contexts is a step towards generating that kind of improved

epistemological theories might be rejected if they are in serious conflict with the lexical semantics of ‘knows’”. And DeRose (2009, p. 19) says that “It’s essential to a credible epistemology, as well as to a responsible account of the semantics of the relevant epistemologically important sentences, that what’s proposed about knowledge and one’s claims about the semantics of ‘know(s)’ work plausibly together. . . .”. In contrast, Sosa (2000, p. 3) argues that epistemic contextualism as a “a thesis in linguistics or in philosophy of language” is plausible, but its interest as a theory of *knowledge* “is limited in certain ways” (p. 8), and for an argument in favor of a “divorce for the linguistic theory of knowledge attributions and traditional epistemology”, see (Hazlett, 2010, p. 500)—though see Stokke (2012) for a criticism of the reasons Hazlett offers in favor of the divorce.

evidence.

In summary, whether contextualists and anti-intellectualists take themselves to be engaged in the *cognitive scientific* or the *metaphysical* explanatory project (or both), they should be interested in—and cannot dismiss as mere experimental artifacts—responses to contexts evaluated *jointly*. Moreover, experimental results that show no significant effect of changing stakes on judgments when those contexts are evaluated separately (e.g. Buckwalter 2010, Buckwalter and Schaffer forthcoming, Feltz and Zarpentine 2010, Phelan 2013) don't pose a serious challenge to anti-intellectualism, since there is reason to think that *what's at stake* in a context is a difficult-to-evaluate attribute the effects of which emerge most clearly in joint evaluation of contexts.

References

- Birdsong, D. (1989). *Metalinguistic Performance and Interlinguistic Competence*. Springer-Verlag, Berlin.
- Buckwalter, W. (2010). Knowledge isn't closed on Saturdays. *Review of Philosophy and Psychology*, 1(3):395–406.
- Buckwalter, W. and Schaffer, J. (2012). Knowledge, stakes and mistakes. Forthcoming in *Noûs*.
- Cappelen, H. and Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell, Oxford.
- DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, LII(4):913–929.
- DeRose, K. (1999). Contextualism: An explanation and defense. In *The Blackwell Guide to Epistemology*. Blackwell, Oxford.

- DeRose, K. (2002). Assertion, knowledge, and context. *The Philosophical Review*, 111(2):167–203.
- DeRose, K. (2005). The ordinary language basis for contextualism, and the new invariantism. *The Philosophical Quarterly*, 55(219):172–198.
- DeRose, K. (2009). *The Case for Contextualism*. Oxford University Press, Oxford.
- DeRose, K. (2011). Contextualism, contrastivism, and x-phi surveys. *Philosophical Studies*, 156(1):81–110.
- Feltz, A. and Zarpentine, C. (2010). Do you know more when it matters less? *Philosophical Psychology*, 23(5):683–706.
- Gerken, M. (2012). On the cognitive bases of knowledge ascriptions. In Brown, J. and Gerken, M., editors, *Knowledge Ascriptions*, pages 140–170. Oxford University Press, Oxford.
- Hansen, N. and Chemla, E. (forthcoming). Experimenting on contextualism. Forthcoming in *Mind & Language*.
- Harding, J. (2012). Short cuts. *London Review of Books*, 34(6):38.
- Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, LXXX(3):497–522.
- Hill, A. (1961). Grammaticality. *Word*, 17:1–10.
- Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, 11:107–121.
- Hsee, C. K., Loewenstein, G. F., Blount, S., and Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5):576–590.

- Jucker, A. H. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics*, 19(5):435–453.
- Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, 23(5):681–683.
- Kahneman, D. (2003). Maps of bounded rationality. In Frañgsmyr, T., editor, *Les Prix Nobel. The Nobel Prizes 2002*. Nobel Foundation, Stockholm.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar Straus & Giroux, New York.
- Kahneman, D. and Tversky, A. (1982). On the study of statistical intuitions. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment under Uncertainty: Heuristics and Biases*, pages 493–508. Cambridge University Press, New York.
- Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3):582–591.
- Ludlow, P. (2005). Contextualism and the new linguistic turn in epistemology. In Preyer, G. and Peter, G., editors, *Contextualism in Philosophy: Knowledge, Meaning and Truth*, pages 11–50. Oxford University Press, Oxford.
- Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford University Press, Oxford.
- May, J., Sinnott-Armstrong, W., Hull, J. G., and Zimmerman, A. (2010). Practical interests, relevant alternatives, and knowledge attributions: An empirical study. *Review of Philosophy and Psychology*, 1(3):265–273.
- Neta, R. and Phelan, M. (Ms.) “Evidence that stakes don’t matter for evidence”.

- Phelan, M. (2013) Evidence that stakes don't matter for evidence. *Philosophical Psychology*, iFirst, [dx.doi.org/10.1080/09515089.2012.733363](https://doi.org/10.1080/09515089.2012.733363)
- Pinillos, N. A., Smith, N., Nair, G. S., Marchetto, P., and Mun, C. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind & Language*, 26(1):115–139.
- Poulton, E. (1975). Range effects in experiments with people. *American Journal of Psychology*, 88(1):3–32.
- Schütze, C. T. (1996). *The Empirical Base of Linguistics*. The University of Chicago Press, Chicago.
- Schwitzgebel, E. and Cushman, F. (2012). Expertise in moral reasoning? order effects on moral judgments in professional philosophers and non-philosophers. *Mind & Language*, 27(2):135–153.
- Shafir, E. (1998). Philosophical intuitions and cognitive mechanisms. In DePaul, M. R. and Ramsey, W., editors, *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, pages 59–83. Rowman and Littlefield, Lanham, MD.
- Sosa, E. (2000). Skepticism and contextualism. *Philosophical Issues*, 10(1):1–18.
- Spencer, N. (1973). Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research*, 2:83–98.
- Sripada, C. S. and Stanley, J. (forthcoming). Empirical tests of interest-relative invariantism. Forthcoming in *Episteme*.
- Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford University Press, Oxford.

Stanley, J. (2007). *Precis of Knowledge and Practical Interests*. *Philosophy and Phenomenological Research*, 75(1):168–172.

Stanovich, K. (2011). *Rationality and the Reflective Mind*. Oxford University Press, Oxford.

Stokke, A. (2012). Protagonist projection. Forthcoming in *Mind & Language*.