
THE STAKES EFFECT: NEW EVIDENCE FROM A RETRACTION-BASED EXPERIMENTAL DESIGN

Nikolai Shurakov

ABSTRACT

The paper examines the influence of stakes on knowledge attributions, building on the 'retraction-based' experimental design introduced by Dinges & Zakkou (2021). Experiment 1 replicates Dinges & Zakkou's original findings and extends the investigation to third-person knowledge ascriptions. The results show that raising the stakes increases the likelihood of retraction in both first- and third-person scenarios. Experiment 2 addresses potential concerns about the retraction-based design, specifically whether participants genuinely endorse the initial knowledge claim. Experiment 2 introduced a modification to the initial design by adding a knowledge-ascribing question. This addition made the act of retraction more realistic and helped to exclude scenario sceptics—participants who disagree with the knowledge claim. The results confirm that the stakes effect persists even when participants have explicitly endorsed the initial claim. Experiment 3 explores the generalizability of the findings to two additional scenarios but fails to replicate the stakes effect. The paper discusses the findings in relation to two competing theories about the context-sensitivity of knowledge ascriptions: epistemic contextualism (EC) and subject-sensitive invariantism (SSI). EC readily accounts for the results of all three experiments, suggesting that higher stakes raise epistemic standards, leading to more retractions. SSI, which ties epistemic standards to the subject of the knowledge ascription, struggles to explain the stakes effect in third-person cases unless it appeals to a detailed theory of how attributors project their own stakes onto the subject. Overall, the paper provides further evidence for the influence of stakes on knowledge attributions, particularly in third-person contexts, and suggests that the retraction-based experimental design is a valuable tool for investigating the complexities of knowledge ascription practices. The findings also lend support to EC as a more plausible explanation for the observed patterns of retraction.

INTRODUCTION

Contrary to the traditional view that knowledge depends solely on truth-related factors (such as evidence or belief), the pragmatic encroachment thesis in epistemology posits that practical factors—often, the stakes involved in a situation—can influence whether a person knows something. In high-stakes situations, for example, the epistemic standards required for knowledge attributions are higher compared to those in low-stakes contexts, even when the evidence available is fixed. Epistemic contextualism supports this view, arguing that the epistemic standards for knowledge attributions shift according to relevant conversational context. Keith DeRose, a central proponent of contextualism, identifies the best type of evidence for this view:

The best grounds for accepting contextualism concerning knowledge attributions 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 provides not only 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 is context-sensitive. (DeRose, 2005, p. 172)

Experimental philosophy is perfectly suited to explore this evidence "of the very best type," and many empirical studies have investigated whether stakes indeed influence ordinary knowledge attributions (see (Pinillos, 2016) for an overview and more recent experiments of (Dinges & Zakkou, 2021; Francis et al., 2019a; Grindrod et al., 2019; Porter et al., 2024; Rose et al., 2019; Turri, 2017; Wu, 2023)). However, the experimental data in this area are inconclusive at best. Two primary experimental approaches—evidence-seeking and evidence-fixed—appear to measure distinct sides of the stakes-knowledge relationship (Porter et al., 2024), while an alternative retraction-based method, which has found a significant stakes effect, remains underexplored, and the present paper addresses this gap.

The stakes effect observed by (Dinges & Zakkou, 2021) is replicated (Experiment 1), withstanding modification that enhances the ecological validity of retraction and accounts for scenario-sceptics (Experiment 2). Furthermore, the stakes effect remains robust in the third-person version of the Bank case (Experiment 1) but diminishes when alternative scenarios are tested (Experiment 3). This series of retraction-based experimental results appear to support epistemic contextualism, offering a new reason to prefer it over its main competitor, subject-sensitive invariantism.

The paper first introduces the retraction-based experimental design and then successively presents three experiments, concluding with a general discussion of the findings.

ON RETRACTION-BASED EXPERIMENTAL DESIGN

This section overviews the experimental designs employed in debates over the effect of stakes¹ on knowledge ascriptions, subsequently motivating the choice of the retraction-based design.

Only some experiments have found evidence of the stakes effect and most of them have employed the evidence-seeking experimental design. Two dominant experimental designs—"evidence-seeking" and "evidence-fixed"(see (Pinillos, 2012) where this distinction was introduced) critically

¹ By "the stakes effect," I refer to the phenomenon where knowledge claims are influenced by practical considerations (both broadly understood)

differ in how the crucial question is formulated: in evidence-seeking experiments, participants are asked how much evidence one has to obtain before they know (or do not know) that p , whereas in the more traditional evidence-fixed experiments, the crucial question is whether a person knows (or whether it is true that a person knows) that p . The latter design has been employed by many (e.g. (Buckwalter, 2010; Hansen & Chemla, 2013; May et al., 2010)) but with varying success. Thus, Francis and colleagues argue that “while the “evidence-fixed” experimental design is capable of uncovering stakes effects on judgments about knowledge, those effects are hard to find and, if found, are small.”(Francis et al., 2019a, p. 439) A more recent cross-cultural study by Porter et al., (2024) concludes that the results obtained with these designs are not inconsistent, but rather, two methods test different things: "People's belief about the evidence required for knowledge is not aligned with their own practice of knowledge ascription."(Porter et al., 2024, p. 14) The evidence-seeking prompts investigate beliefs about evidence, whereas evidence-fixed prompts probe actual practices of knowledge ascription.

However, one more experimental design has attempted to find the stakes effect in a series of experiments (Dinges & Zakkou, 2021) [hereafter just D&Z]. It might be of interest here since it employs retraction as a tool in investigating knowledge-attributing practices and does not fall into any of the two above-described groups. So far, no further studies exploring this alternative methodology have been reported in the literature. In this design, participants are asked to read a setup of a situation that ends with a knowledge (self-)ascription, e.g.: “Based on this, you respond: “I know the bank will be open tomorrow”.”(D&Z, p. 736). The follow-up part of the scenario is different in three conditions: NEUTRAL (no new information regarding knowledge claim provided), STAKES (high stakes introduced), or EVIDENCE (evidence against p is offered). The crucial question then is whether participants stand by their claim that they know the bank will be open tomorrow. Choosing between ‘I do’ and ‘I don’t’, they are allowed to retract the knowledge claim. The results are robust; the retraction rates for the Bank scenario: STAKES (48%) is higher than for NEUTRAL (9.8%, large effect), and for EVIDENCE (96.1%) is higher than for STAKES (large effect). A similar pattern holds for the Typo scenario (modification of the story employed by Pinillos 2012 where a student needs to proofread an essay), with retraction rates: NEUTRAL 5.9%, STAKES 24%, EVIDENCE 58%. Across both scenarios, D&Z consistently found a medium to large effect of stakes.

These findings set the retraction-based design apart from the two dominant experimental paradigms. The consistency and effect sizes contrast sharply with the minimal effects found using evidence-fixed prompts. Unlike evidence-seeking prompts, the retraction-based approach does not require participants to assess the protagonist's amount of evidence, thereby avoiding (Buckwalter & Schaffer, 2015) critique that stakes effects may be limited to deontic modals, such as “has to”. These considerations provide strong initial support for further exploration of this methodology. D&Z also recognize the need for additional research that should explore: “third-person cases, cases featuring knowledge denials rather than ascriptions, cases where the stakes are lowered rather than raised and cases where practical factors other than stakes (e.g., time constraints) shift.”(Dinges & Zakkou, 2021, p. 746) Recognising the value of D&Z findings, this paper investigates their experimental design further.

MOTIVATION FOR EXPERIMENT 1

In what follows, two primary motivations for conducting Experiment 1 are provided.

First, there is a need to assess the robustness of D&Z's reported findings. Although the general replication rate for experimental philosophy is relatively high—"about 70%" (Cova et al., 2021, p. 10)—direct replications remain rare in debates surrounding the stakes effect, and sometimes replications provide different results. In particular, a preregistered replication of (Sripada & Stanley, 2012) experiment by Francis et al., (2019) yielded a reverse pattern: while Sripada and Stanley failed to find the stakes effect for the Basic vignette alone, Francis et al. observed the stakes effect in the Basic but not in the Implicit/Explicit or Ignorant vignette pairs. So, the possibility of not replicating the effect reported by D&Z should not be dismissed. Furthermore, the direct replication would provide reliable and clean data for future comparisons. This brings us to the second motivation for Experiment 1: the possibility of contrasting the 1st and third-person cases.

Testing third-person cases not only bypasses (Turri, 2017) objection but also offers finer-grained evidence for competing theories of the context-sensitivity of knowledge ascriptions. (Turri, 2017, pp. 147–149) found that participants tend to agree with protagonists' self-ascriptions of knowledge, regardless of stakes manipulation. In his experiment, participants agreed with both positive and negative knowledge claims made by protagonists, confirming contextualist predictions. The mere polarity of 'know' but not stakes produced the results that contextualists take as evidence for their view. Turri attributes this to the "deferral confound," whereby participants defer to protagonists' self-regarding claims. The scenarios tested by D&Z (Bank, Typo) involved only protagonists' self-ascriptions of knowledge. Thus, the results might be indirectly affected by this deferral confound. However, by focusing on third-person cases, we eliminate this confound, as self-ascriptions are no longer an issue.

Third-person cases are also crucial for distinguishing between epistemic contextualism (EC) and subject-sensitive invariantism (SSI). Both views permit stakes to influence knowledge ascriptions and align on predictions for 1st person cases but diverge regarding third-person. The first experimental study of third-person knowledge ascriptions (Grindrod et al., 2019) has emphasized why third-person cases are so important:

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 (Grindrod et al., 2019, p. 161).

Grindrod and colleagues tested two scenarios of third-person knowledge attributions (in addition to Colour and Control scenarios), and their findings offered overall support for contextualism. However, the stakes effect on knowledge was small and weaker than that observed in the Colour scenario. Mentioning the results of evidence-seeking design experiments, the authors refrain from concluding an outright victory for EC, suggesting instead that both views "give us partial, but compatible, accounts of the complexities of how we assess whether someone knows something" (Grindrod et al., 2019, p. 177). This thought resonates with the abovementioned conclusion of (Porter et al., 2024) who proposed that two experimental designs may each offer partial explanations of knowledge attributions. Experiments which employ a retraction-based design and test third-person cases might provide a piece of novel evidence to the debate.

It also has to be acknowledged that the retraction-based experimental design has intrinsic limitation in constituting a “crucial experiment.” Neither EC nor SSI makes explicit predictions on retraction. In the retraction-based design, participants initially make a knowledge ascription, assigning knowledge to a third party; then, participants in the STAKES condition are informed of higher stakes and asked if they would stand by their initial ascription. A contextualist might argue that high-stakes awareness prompts attributors to adopt higher epistemic standards, leading to a higher rate of retractions. Conversely, SSI proponents would not anticipate a change in stakes affecting the attributor’s decision, as the subject’s stakes remain constant. Thus, retraction-based design may provide further evidence supporting either EC or SSI. Despite this limitation, Experiment 1 is designed with two primary objectives in mind. First, it seeks to directly replicate D&Z’s findings to assess their robustness. Second, it examines whether the stakes effect extends to third-person scenarios.

EXPERIMENT 1

Preregistered² Experiment 1 (i) attempts to replicate the D&Z Bank case experiment directly and (ii) tests whether the stakes effect holds for the third-person version of the Bank case. The preregistrations, data and materials for the whole project can be found here [https://osf.io/VIEW_ONLY_LINK].

PARTICIPANTS

I recruited 922 native speakers of English from the US and UK via Prolific. The online experiment was implemented using Unipark. Each of the six conditions aimed for 100 valid responses (preregistered). Participants who failed to correctly answer attention check question based on the details of the assigned scenario were excluded. For the final analysis, I used the first 100 valid responses per condition, resulting in a total sample of 600 people (341 female, 2 preferred not to say, 1 person with expired data; mean age 40 years). Participants were paid £0.25 for approximately 2 minutes of their time (£7.5/h).

DESIGN, PROCEDURE, MATERIALS

The experiment utilized a 3x2 design comprising three types of stories (Neutral, Stakes, Evidence; between subjects) and two perspectives (first-person, third-person; between subjects). Participants were randomly assigned to one of six conditions. First, participants were asked to choose a response to the question of whether they stood by a knowledge claim (binary choice: “I do” or “I don’t”). They were then asked a confidence question, using a 7-point Likert item ranging from “very unconfident” to “very confident.” The intermediate options were “unconfident,” “somewhat unconfident,” “neither confident nor unconfident,” “somewhat confident,” and “confident.” An attention check followed on the next page to ensure that participants paid enough attention to the task and details of the scenario.

The first-person conditions were adapted from D&Z’s Experiment 2 (the Bank scenario). The third-person version was created by modifying the original story as little as possible. The scenarios are presented below, with the manipulation [first-person/third-person] indicated in square brackets. The setup of the scenario remains the same across all three conditions for each perspective:

Picture yourself in the following scenario:

² https://osf.io/5z6pu/?view_only=ee4c6fe6441546d583643f0392594b96

You are driving home from work on a Friday afternoon with a colleague, Peter. You plan to stop at the bank to deposit your paychecks. As you drive past the bank, you notice that the lines inside are very long, as they often are on Friday. [Peter asks whether you know/ You ask Peter whether he knows] whether the bank will be open tomorrow, on Saturday. If it is open tomorrow, you can come back tomorrow, when the lines are shorter. [You remember/ Peter says that he remembers] having been at the bank three weeks before on a Saturday. Based on this, you [respond/say]:

“[I know/ Oh, so you know] the bank will be open tomorrow”.

At this point, ...

NEUTRAL

...you receive a phone call from your partner. [S/he/ You mention that you are currently with Peter and tell your partner that Peter knows that the bank will be open tomorrow. S/he] tells you that one of your children has gotten sick and that they are still waiting at the doctor's office to get an appointment. S/he asks whether you can water the plants if you come home and prepare dinner. There's enough food at home so you don't have to buy anything extra. You agree. As you hang up, [Peter asks/ your partner asks] whether you stand by your previous claim that [you know / Peter knows] the bank will be open tomorrow. You respond:

STAKES

... you receive a phone call from your partner. [S/he/ You mention that you are currently with Peter and tell your partner that Peter knows that the bank will be open tomorrow. S/he] tells you that it is extremely important that your paycheck is deposited by Saturday at the latest. A very important bill is coming due, and there is too little in the account. You realize that it would be a disaster if you drove home today and found the bank closed tomorrow. As you hang up, [Peter asks/ your partner asks] whether you stand by your previous claim that [you know/ Peter knows] the bank will be open tomorrow. You respond:

EVIDENCE

... you receive a phone call from your partner. [S/he/ You mention that you are currently with Peter and tell your partner that Peter knows that the bank will be open tomorrow. S/he] tells you that s/he was at a different branch of your bank earlier today. A sign said that the branch no longer opens on Saturdays. You see a similar sign in the branch you were about to visit. You can't properly read the sign from the distance, but it seems to concern the opening hours. As you hang up, [Peter asks/ your partner asks] whether you stand by your previous claim that [you know/ Peter knows] the bank will be open tomorrow. You respond:

Each condition is followed by the questions below, comments in square brackets are added for the reader:

Please pick the response you would be more likely to give:

- I do
- I don't

How confident are you in your answer:

Very unconfident – unconfident – somewhat unconfident – neither confident nor unconfident – somewhat confident – confident – very confident [7 point Likert item]

According to the story, how long ago were you at the bank on a Saturday?

- You didn't mention being at the bank on a Saturday before [correct answer for the third- person conditions]
- Three weeks ago [correct answer for the first-person conditions]
- Two weeks ago
- Three months ago

The analysis closely followed the methodology of Dinges and Zakkou: binary responses were analyzed first, followed by an analysis of composite scores. Responses to the first question were coded as follows: “I do” as 1 and “I don't” as -1. Each participant's composite score was calculated by multiplying their response by their confidence level, which ranged from 1 (very unconfident) to 7 (very confident), resulting in composite scores ranging from -7 to 7.

The experiment aimed to answer the following questions: (i) Is the result reported by Dinges & Zakkou replicated?, (ii) Is the result reported by Dinges & Zakkou hold for the third-person conditions?, (iii) Is there a statistically significant difference between the Stakes first-person condition and the Stakes third-person condition? Accordingly, the results were first analyzed for the first-person conditions, which is also a direct replication of the original experiment. Then, I analyzed the results for third-person conditions to address the second question. Finally, I compared binary responses and composite scores for the STAKES conditions to determine whether switching to a third-person perspective affected retraction rates.

RESULTS

From the 922 collected responses, 225 failed the attention check and were excluded. The first 100 valid responses per condition from the remaining responses were included in the final analysis.

RESULTS – FIRST-PERSON (REPLICATION)

ANALYSIS OF BINARY RESULTS

The percentage of retractions ("I don't") in each condition was as follows: 7% in NEUTRAL, 53% in STAKES, and 95% in EVIDENCE.

A chi-square test for independence was conducted to determine the significance of these differences. A significant effect of story type was observed across all conditions, $\chi^2(2, N = 300) = 155.159, p < .001$, Cramer's $V = .719$, indicating a large³ effect size and prompting further pairwise comparisons. Participants were more likely to retract in STAKES than in NEUTRAL, $\chi^2(1, N = 200) = 48.214, p < .001$, Cramer's $V = .491$ (medium effect). They were also more likely to retract in EVIDENCE than in NEUTRAL, $\chi^2(1, N = 200) = 151.441, p < .001$, Cramer's $V = .870$ (large effect). Finally, participants were more likely to retract in EVIDENCE than in STAKES, $\chi^2(1, N = 200) = 43.685, p < .001$, Cramer's $V = .467$ (medium effect).

ANALYSIS OF COMPOSITE SCORES

Mean composite scores by condition, compared to the findings of the original experiment, are shown in Figure 1. A one-way ANOVA revealed a statistically significant difference between story types, $F(2, 297) = 167.720, p < .001, \eta^2 = .530$ [large effect]. A Tukey HSD post hoc test showed that participants were more inclined to retract in STAKES ($M = 0.28, SD = 5.32, p < .001$, Cohen's $d = 1.06$ [large effect]) and EVIDENCE ($M = -4.59, SD = 2.64, p < .001$, Cohen's $d = 3.55$ [large effect]) than in NEUTRAL ($M = 5.32, SD = 2.93$). Participants were also more inclined to retract in EVIDENCE than in STAKES ($p < .001$, Cohen's $d = 1.16$ [large effect]).

DISCUSSION

The original experiment was replicated almost perfectly. Retraction rates across conditions were similar: 9.8% vs. 7% in the NEUTRAL condition, 48% vs. 53% in STAKES, and 96.1% vs. 95% in EVIDENCE, suggesting consistent medium-to-large effects across both studies. Similarly, the composite scores (see Figure 1) closely matched those of the original study, demonstrating large effect sizes across all pairwise comparisons. These findings indicate that the stakes effect reported by Dinges and Zakkou remains robust.

³ Add references to effect sizes used

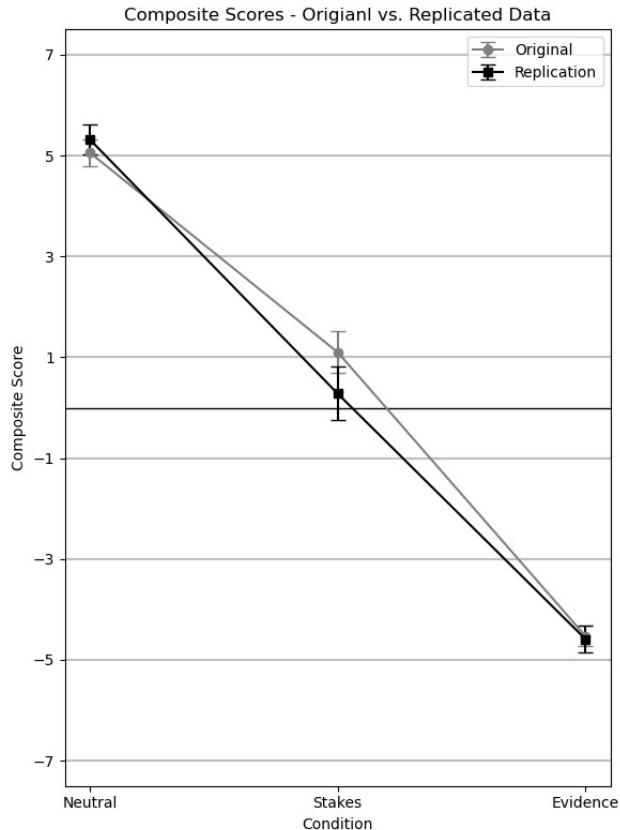


Figure 1. Mean composite score comparison between the original study – Dinges&Zakkou (2021) and the direct replication across three conditions: Neutral, Stakes, and Evidence.

RESULTS – THIRD-PERSON

ANALYSIS OF BINARY RESULTS

The percentage of retractions ("I don't") in each condition was as follows: 11% in NEUTRAL, 61% in STAKES, and 94% in EVIDENCE.

A chi-square test for independence was conducted, revealing a significant effect of story type across all conditions, $\chi^2(2, N = 300) = 141.315, p < .001$, Cramer's $V = .686$, indicating a large effect size and prompting further pairwise comparisons. Participants were more likely to retract in STAKES than in NEUTRAL, $\chi^2(1, N = 200) = 52.105, p < .001$, Cramer's $V = .510$ (medium effect). They were also more likely to retract in EVIDENCE than in NEUTRAL, $\chi^2(1, N = 200) = 134.817, p < .001$, Cramer's $V = .821$ (large effect). Finally, participants were more likely to retract in EVIDENCE than in STAKES, $\chi^2(1, N = 200) = 29.362, p < .001$, Cramer's $V = .383$ (medium effect).

ANALYSIS OF COMPOSITE SCORES

Mean composite scores by condition are shown in Figure 2. A one-way ANOVA revealed a statistically significant difference between story types, $F(2,297) = 149.771, p < .001, \eta^2 = .530$ (large effect). A Tukey HSD post hoc test revealed that participants were more inclined to retract in Stakes ($M = -0.94, SD = 5.04, p < .001$, Cohen's $d = 1.28$ [large effect]) and Evidence ($M = -4.72, SD = 2.56, p < .001$, Cohen's $d = 3.04$ [large effect]) than in Neutral ($M = 4.64, SD = 3.53$).

They were also more inclined to retract in Evidence than in Stakes ($p < .001$, Cohen's $d = 0.95$ [large effect]).

FIRST- VS THIRD-PERSON COMPARISON

No significant differences were found. The comparison of binary scores yielded a z-score of -1.143 , $p = .253$. The comparison of composite scores revealed a t-statistic of 1.664 , $p = .098$, indicating no statistically significant difference in composite scores either. This result is illustrated in Figure 3 below.

DISCUSSION

The results of the experiment for the third- person version of the bank scenario are similar to those of first-person. The same pattern—NEUTRAL < STAKES < EVIDENCE—holds for both binary responses and composite scores. The effect size is medium-to-large for binary responses and large for composite scores. There is no evidence that the third-person STAKES condition differs from the first-person version. The observed stakes effect suggests a potential advantage of epistemic contextualism over subject-sensitive invariantism.

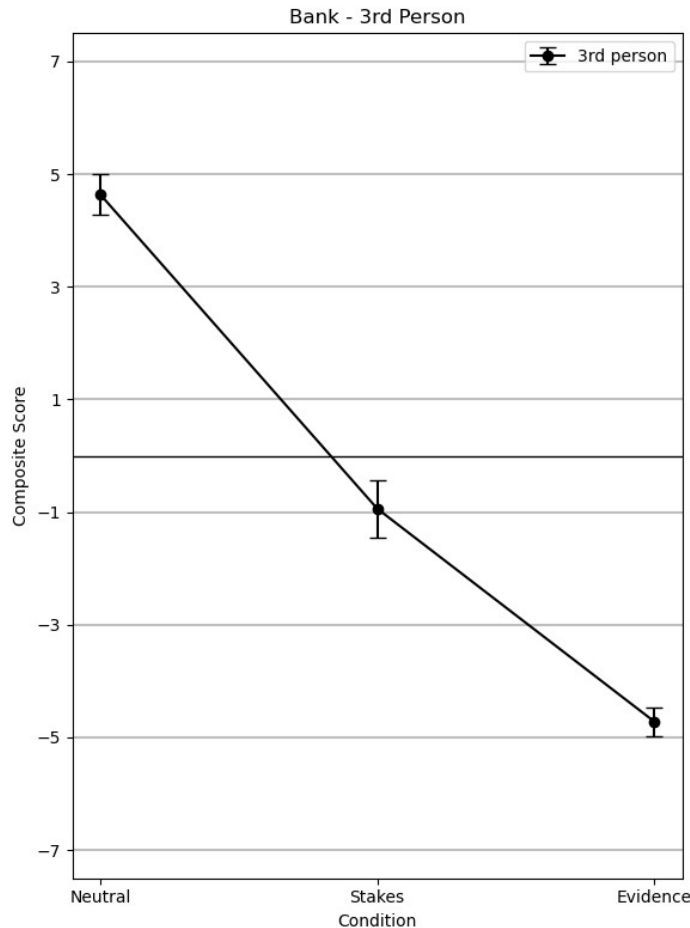


Figure 2. The mean composite scores for the third- person scenarios across three conditions: Neutral, Stakes, and Evidence.

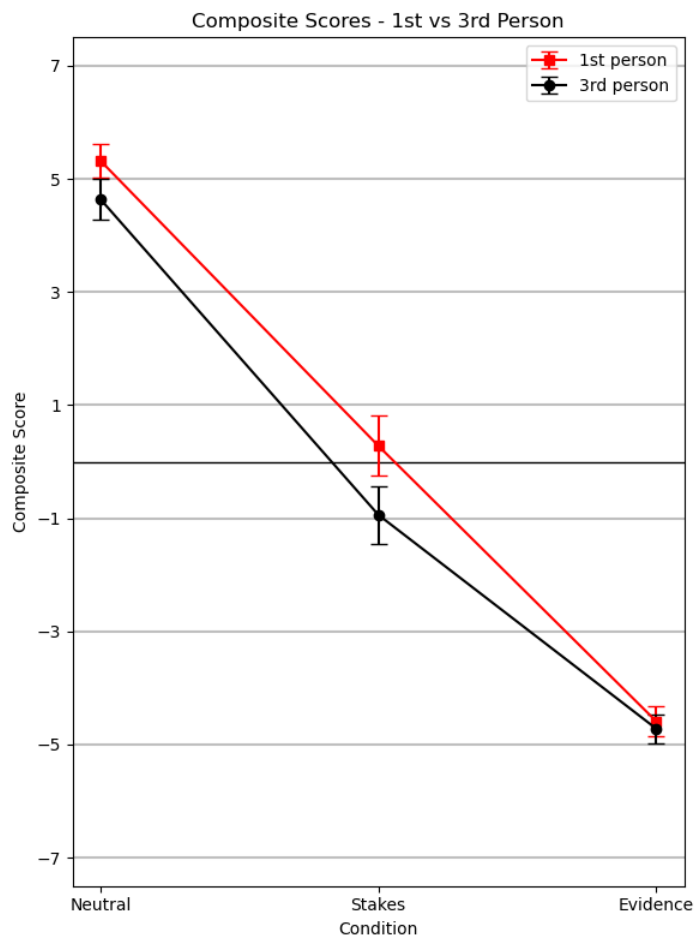


Figure 3. The mean composite scores for the first- and third-person scenarios across three conditions: Neutral, Stakes, and Evidence.

GENERAL DISCUSSION OF EXPERIMENT 1

The robust findings of Dinges and Zakkou were independently replicated. Analysis of both binary responses and composite scores revealed significant differences between story types: NEUTRAL < STAKES < EVIDENCE, with medium-to-large effect sizes. This pattern also holds for the third-person version of the Bank scenario, but between the STAKES first-person and STAKES third-person conditions no statistically significant difference was found. This consistency suggests that stakes influence third-person knowledge attributions as well, potentially offering new evidence in favour of epistemic contextualism.

Before interpreting these results in relation to broader debates, it is important to address a notable issue: 225 out of 922 participants failed the attention check. This failure rate may be partly attributable to the attention check question being presented on a separate page from the scenario, thus partially relying on participants' memory. While this design may disadvantage participants with weaker memory, it validates the responses of those who paid closer attention to details. Additionally, identifying the correct response for third-person conditions may have been more challenging, as the answer was not explicitly stated in the text provided and required some

reasoning. Nonetheless, most participants recognized that only their stakes had changed, while knowledge was attributed to another person.

Epistemic contextualism (EC) appears well-equipped to account for the results across conditions. EC holds that heightened stakes raise epistemic standards within the conversational context. In both first-person and third-person scenarios, the STAKES context led participants to adopt stricter standards for knowledge ascriptions, relevant to each conversational setting. These raised standards resulted in the retractions of knowledge claims by nearly half of the participants in both first- and third-person STAKES conditions.

While subject-sensitive invariantism (SSI) may explain the first-person results, it encounters difficulty accounting for the third-person cases, where participants in high-stakes conditions retracted knowledge claims regarding someone else. Since the stakes do not affect the knowledge subject (in this case, Peter), SSI should predict almost no retraction. Thus, the third-person findings challenge SSI. Proponents of SSI might pursue several strategies to address this issue. (Hawthorne, 2003, p. 163) suggests that people tend to "overproject our own lack of knowledge to others." A high-stakes attributor might not self-ascribe knowledge and thus project this lack of knowledge onto the low-stakes subject. Stanley (2005, p. 102) proposes that for the attributor, it is crucial whether the subject would know that p "if that person had the interests and concerns High Stakes does" (where High Stakes refers to an individual for whom p is a significant practical concern). Participants in the STAKES third-person condition may have reasoned that, if Peter were in a high-stakes situation, he would lack knowledge, leading them to retract the initial knowledge ascription.

However, (DeRose, 2009, pp. 234–238) has argued that the projectionist strategy struggles to account for cases where the attributor possesses sufficient evidence to meet even the highest epistemic standards. In such cases, the attributor would have no ignorance to project but would still be inclined to assert that the low-stakes subject lacks knowledge of p . In our case if participants imagined themselves to have at least some evidence for p , they would be less likely to project their high-stakes context onto Peter. While the projectionist strategy may offer SSI a possible means of explanation here, the existing literature provides little in terms of a detailed account of the psychological mechanisms underlying projection and how it could address DeRose's challenge. I remain neutral as to whether SSI could ultimately account for Experiment 1 results but emphasize the need for a more comprehensive account. By contrast, DeRose's contextualism already seems well-equipped to explain the results, suggesting that EC more readily accommodates the evidence than its key competitor.

Of course, these results do not provide direct evidence on how ordinary speakers use knowledge ascriptions, and the complexity of retraction as a phenomenon cautions against overly strong conclusions. Nonetheless, it appears reasonably safe to suggest that the overall evidence thus far favours epistemic contextualism.

SCENARIO SCEPTICS AND MOTIVATION FOR EXPERIMENT 2

The second experiment addresses a possible problem regarding the retraction-based experimental design. In what follows, I explicate two concerns: a worry of genuine retraction and of scenario sceptics, in the existing design and motivate one methodological change that has been implemented in Experiment 2.

The primary task for participants in the experiment is to retract—decide whether they stand by the claim. However, for a retraction to be genuine, participants must first have endorsed the knowledge claim they are asked to retract. In the original design by Dinges and Zakkou, participants are instructed to “picture yourself in the following scenario,” rather than to make the knowledge claim themselves. This instruction limits participants to imagining the scenario where the knowledge claim is treated *as if* it were their own. As a result, the act of retraction in this context may lack authenticity, as participants are merely retracting a hypothetical claim they never explicitly made or endorsed. This design flaw may make it easier for participants to retract in an imagined situation. This concern is particularly relevant in the stakes condition, where the lack of genuine endorsement may significantly affect the observed stakes effect, because standing by p becomes costly. Therefore, a modified design that allows participants to genuinely endorse and then retract the knowledge claim would be desirable to improve the ecological validity of the results.

It is also unlikely that all participants endorse the knowledge claim that “the bank will be open tomorrow.” For some, evidence of having been at the bank three weeks ago may not suffice to justify such a knowledge claim. These participants might view their evidence as insufficient for knowledge or maybe only see it as adequate for a justified belief but not for knowledge. For them—referred to here as “scenario sceptics”—the design of the D&Z experiment may be shaping responses. In NEUTRAL, such participants may maintain the knowledge claim not because they genuinely endorse it, but because there is no motivation to reconsider. In EVIDENCE conditions, however, these sceptics are likely to retract in light of the defeater that undermines the original claim. Both actions seem to be rational and align well with the results. But the presence of scenario sceptics could significantly impact the observed stakes effect, potentially leading to inflated retraction rates. Scenario sceptics are particularly prone to retract in the STAKES condition: they were already doubtful of the knowledge attribution, and the stakes raised the cost of maintaining the claim. For these participants, retraction in STAKES is very much anticipated. As the proportion of scenario sceptics increases, the actual stakes effect may be correspondingly reduced. The initial design does not account for such participants, yet it is crucial that the act of retraction be performed by participants who at least accept that this claim is regarding knowledge.

To address these concerns, the retraction-based design needs modifications to (i) allow participants a genuine opportunity to endorse the claim and (ii) to exclude scenario sceptics. Experiment 2 includes a necessary adjustment that satisfies these requirements.

EXPERIMENT 2

Experiment 2 modifies the initial design to make retraction more realistic and to screen out scenario sceptics. To accomplish this, an additional binary choice question was incorporated following the scenario setup, allowing participants to select the response they would be more likely to give in such a conversation. Scenario-sceptics then can choose not to ascribe knowledge. For participants who do ascribe knowledge, this modification makes the retraction process more realistic, as it reflects their own decision to ascribe knowledge. Only those who explicitly endorsed the claim were assigned to follow-up conditions, as participants who did not endorse it (scenario sceptics) had nothing to retract.

PARTICIPANTS

I recruited 450 native speakers of English via Prolific. The online experiment was implemented using Unipark. Each of the three conditions aimed for 100 valid responses (as preregistered)⁴. This number was estimated using G*Power based on the tests originally conducted by Dinges and Zakkou: 0.9 power, a medium effect size of 0.3 (w), 0.05 alpha for the chi-squared test for independence; 0.9 power, a medium effect size of 0.25 (f) at the standard 0.05 alpha error probability for one-way ANOVA. Participants who failed to correctly answer an attention check question based on details of the assigned scenario were excluded, as were those who disagreed with the knowledge ascription (scenario sceptics). For the final analysis, I used the first 100 valid responses per condition, resulting in a total sample of 300 participants (174 female, 2 preferred not to disclose; mean age 38 years). Participants were paid £0.25 for approximately 2 minutes of their time (£7.50 per hour).

DESIGN, PROCEDURE, MATERIALS

The experiment employed a two-level design in which all participants were initially asked a knowledge-ascribing question. Those who provided a positive answer were then randomly assigned to one of three follow-up conditions (Neutral, Stakes, Evidence; between-subjects). The scenario used was the original D&Z Bank scenario (same as the 1st person conditions of Experiment 1). Due to resource limitations, only the 1st person version of the Bank scenario was tested. The only change made is an additional binary choice question introduced after the scenario setup (response options presented in randomized order):

... You remember having been at the bank three weeks before on a Saturday. Based on this, you respond:

- “I know the bank will be open tomorrow”.
- “I don’t know the bank will be open tomorrow”.

At this point, ...

Participants who did not ascribe knowledge concluded the experiment after this question. Those who ascribed knowledge were randomly assigned to one of the three follow-up conditions (NEUTRAL, STAKES, or EVIDENCE). They were then asked the same questions as in Experiment 1: a binary retraction question (“I do” or “I don’t”), a confidence question (7-point Likert scale ranging from “very unconfident” to “very confident”), and an attention check.

The experiment aimed to answer the following questions: (i) Is the effect reported by Dinges and Zakkou replicated in the updated design? For exploratory purposes, I also checked (ii) if the retraction rate for the updated STAKES condition is lower than for the original STAKES condition. As in Experiment 1, two types of responses were analyzed: binary responses and composite scores, which were calculated exactly like in Experiment 1. Responses to the retraction question were coded as follows: “I do” as 1 and “I don’t” as -1. Each participant’s composite score was derived by multiplying their response by their confidence level (ranging from 1 for “very unconfident” to 7 for “very confident”), resulting in composite scores between -7 and 7 for each participant. To address (ii) question, I used the Stakes 1st person data from Experiment 1. Given the almost perfect replication of the original experiment in Experiment 1, this comparison data might be reliable.

⁴ Preregistered: https://osf.io/ae3pu/?view_only=ad355cofc62f47769dd5d8c7dd5519da

RESULTS

Of 450 participants, 38 failed the attention checks. Of the remaining 412, 53 (12.9%) were scenario sceptics who hadn't assigned knowledge to themselves. From the remaining 359 valid responses, only the first 100 per condition were included in the final analysis.

ANALYSIS OF BINARY RESULTS

The percentage of retractions ("I don't") in each condition was as follows: 6% in NEUTRAL, 43% in STAKES, and 95% in EVIDENCE.

A chi-square test for independence was conducted to determine the significance of these differences. A significant effect of story type was observed across all conditions, $\chi^2(2, N = 300) = 160.176, p < .001$, Cramer's $V = .731$, indicating a large effect size and prompting further pairwise comparisons. Participants were more likely to retract in STAKES than in NEUTRAL, $\chi^2(1, N = 200) = 35.032, p < .001$, Cramer's $V = .419$ (medium effect). They were also more likely to retract in EVIDENCE than in NEUTRAL, $\chi^2(1, N = 200) = 154.895, p < .001$, Cramer's $V = .880$ (large effect). Finally, participants were more likely to retract in EVIDENCE than in STAKES, $\chi^2(1, N = 200) = 60.799, p < .001$, Cramer's $V = .551$ (medium effect).

ANALYSIS OF COMPOSITE SCORES

A one-way ANOVA revealed a statistically significant difference between story types, $F(2,297) = 203.050, p < .001, \eta^2 = .582$, indicating a large effect size. A Tukey HSD post hoc test showed that participants were more inclined to retract in STAKES ($M = 1.59, SD = 5.06, p < .001$, Cohen's $d = 0.94$ [large effect]) and EVIDENCE ($M = -4.70, SD = 2.36, p < .001$, Cohen's $d = 4.04$ [large effect]) than in NEUTRAL ($M = 5.36, SD = 2.63$). Participants were also more inclined to retract in EVIDENCE than in STAKES ($p < .001$, Cohen's $d = 1.60$ [large effect]).

STAKES: ORIGINAL VS MODIFIED DESIGN

The comparison between the original STAKES condition and the modified STAKES condition did not reveal a statistically significant difference (see Figure 4). A chi-square test for independence yielded $\chi^2(1) = 1.62, p = .203$. A Mann-Whitney U test produced a U statistic of 4254.0, $p = .064$.

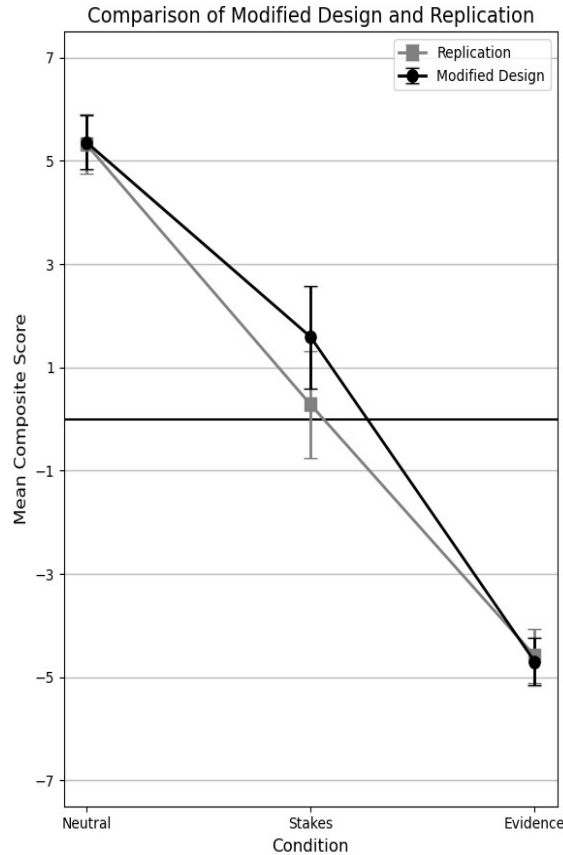


Figure 4. The mean composite scores three conditions: Neutral, Stakes, and Evidence in the replication and modified design.

DISCUSSION

The results of the experiment indicate that the worry regarding scenario sceptics is unjustified. The results are replicated in the modified design as well, implying that even when the retraction is made more realistic and scenario sceptics are excluded from further analysis, stakes do affect knowledge attribution, and a significant number of people do retract once stakes go up. The effect size again ranges from medium to large for binary responses and large for composite scores. The results possibly provide even more robust results than the initial experiment, and its replication, as the main mechanism of the experiment – retraction, is made more realistic.

By addressing the methodological concerns about the initial design, Experiment 2 strengthens the validity of the results obtained with this method. The modification implemented—allowing participants to endorse knowledge claims and excluding scenario sceptics explicitly—enhanced the ecological validity of the experiment while mostly maintaining the original structure.

EXPERIMENT 3

The results obtained with the retraction-based experimental design thus far are robust and consistent, with the stakes effect consistently proving significant. Experiment 3 seeks to explore

whether these findings hold across two additional third-person scenarios and with a design modification introduced in Experiment 2.

PARTICIPANTS

I recruited 361 native speakers of English from the US, UK, and Australia via Prolific. The online experiment was implemented using Qualtrics. The sample size⁵ was estimated using G*Power based on the statistical tests planned for this study: 0.80 power, a medium effect size of 0.3 (w), and a 0.05 alpha level for the chi-squared test for independence, as well as 0.80 power, a medium effect size of 0.25 (f), and a 0.05 alpha level for one-way ANOVA. Adjustments were made to account for potential exclusions, setting the total target sample at 320. Participants who failed to correctly answer attention check questions based on details of the assigned scenario were excluded, as were those who disagreed with the knowledge ascription (scenario sceptics). For the final analysis, I included all valid responses per condition, resulting in a total sample of 361 participants (181 female; mean age 41 years). Participants were paid £0.40 for approximately 3 minutes of their time (£8/hour).

DESIGN, PROCEDURE, MATERIALS

The experiment followed a structure similar to Experiment 2. The experiment employed a 3x2 design, comprising three types of stories (Neutral, Stakes, Evidence; between-subjects) and two scenarios (Train, Flight; between-subjects). Participants were randomly assigned to one of the two scenarios.

All participants were initially asked a knowledge-ascribing question. Those who attributed knowledge were then randomly assigned to one of three follow-up conditions (NEUTRAL, STAKES, or EVIDENCE). Participants who did not ascribe knowledge concluded the experiment after the initial question. Those who continued were asked the same follow-up questions as in Experiments 1 and 2: a binary retraction question (“I do” or “I don’t”), a confidence question (7-point Likert scale ranging from “very unconfident” to “very confident”), and an attention check.

Data analysis and statistical tests followed the method used in the experiments of D&Z and in the prior two experiments. Two types of responses were analyzed: binary responses and composite scores. The response to the retraction question was coded as follows: “I do” as 1 and “I don’t” as -1. Composite scores were calculated by multiplying the retraction response by the participant’s confidence level (ranging from 1 for “very unconfident” to 7 for “very confident”), resulting in scores between -7 and 7 for each participant.

The experiment aimed to answer the following question: (i) Does the stakes effect observed in the previous experiments generalize to two additional third-person scenarios?

Below is an example of one of the scenarios used in the Train condition:

Picture yourself in the following scenario:

You are at the railway station and want to take a train to Kensington. You ask a guy with a suitcase whether the train that is staying on the platform stops at Kensington. The guy takes out the printed train timetable, and says "Yeah, it stops at Kensington". Based on this, you conclude:

- The guy knows that the train stops at Kensington.

⁵ Preregistered: https://osf.io/vuj9q/?view_only=1ca26db238ed4a0e9cd41dbeb2a5c8cc

- The guy does not know that the train stops at Kensington

NEUTRAL

You board the train and receive a phone call from your partner. After exchanging some pleasantries, you mention that you met a guy who knows that the train stops at Kensington. S/he says that there will be a small party of your mutual friend next week, and the friend asks if you would like to join. Of course, you agree. She says, "I am sure it will be fun! I wish you a safe journey. Just to check, do you stand by your previous claim that the guy you met knows the train stops at Kensington?"

STAKES

You board the train and receive a phone call from your partner. After exchanging some pleasantries, you mention that you met a guy who knows the train stops at Kensington. S/he responds, "That's cool because I really need you here. Could you buy Drugpills for me? I accidentally ate a soup containing shellfish, and I'm not feeling well. I may need resuscitation if you don't arrive in an hour." You decide to hurry and assure her that it should only take 20 minutes to reach Kensington. She replies, "Okay, I'll be waiting for you. Just to check, do you stand by your previous claim that the guy you met knows the train stops at Kensington?"

EVIDENCE

You board the train and receive a phone call from your partner. After some small talk, you mention that you met a guy who knows the train stops at Kensington. S/he responds, "That's cool because I saw the news, and there was a huge accident on the railroad. Many trains to Kensington and to other destinations are rerouted today. S/he says, "I'll be waiting for you. Just to check, do you stand by your previous claim that the guy you met knows the train stops at Kensington?"

The second scenario involved a flight from London to Tokyo, where the higher stakes were introduced by a friend for whom it is crucial that the flight takes exactly 14 hours (see the full version in the Appendix).

RESULTS

Of 361 participants, none failed the attention checks. A total of 181 participants were randomly assigned to the Train scenario, and 180 to the Flight scenario. Of these, 20 participants in the Train condition (11.05%) and 4 participants in the Flight condition (2.22%) were scenario sceptics who did not assign knowledge. The remaining valid responses were included in the final analysis.

ANALYSIS OF BINARY RESULTS

The percentage of retractions ("I don't") in each condition for the Train scenario was as follows: 3.7% in NEUTRAL, 7.55% in STAKES, and 62.96% in EVIDENCE. For the Flight scenario, the percentages were 5.08% in NEUTRAL, 11.86% in STAKES, and 84.48% in EVIDENCE.

A chi-square test for independence was conducted to determine the significance of these differences. A significant effect of story type was observed across all conditions for both scenarios: Train, $\chi^2(2, N = 181) = 63.439, p < .001$, Cramer's $V = .628$ (large effect); Flight, $\chi^2(2, N = 180) = 101.419, p < .001$, Cramer's $V = .759$ (large effect). Pairwise comparisons revealed the following:

Train: Participants were more likely to retract in EVIDENCE than in NEUTRAL, $\chi^2(1, N = 108) = 40.042, p < .001$, Cramer's $V = .61$ (large effect), and in EVIDENCE than in STAKES, $\chi^2(1, N = 107) = 33.488, p < .001$, Cramer's $V = .56$ (medium effect). No statistically significant difference was found between STAKES and NEUTRAL, $\chi^2(1, N = 107) = 0.197, p = .66$ (not significant).

Flight: Participants were more likely to retract in EVIDENCE than in NEUTRAL, $\chi^2(1, N = 117) = 71.493, p < .001$, Cramer's $V = .78$ (large effect), and in EVIDENCE than in STAKES, $\chi^2(1, N = 117) = 58.932, p < .001$, Cramer's $V = .71$ (large effect). No statistically significant difference was found between STAKES and NEUTRAL, $\chi^2(1, N = 118) = 0.983, p = .32$ (not significant).

ANALYSIS OF COMPOSITE SCORES

Results for both scenarios are summarized in Figure 5.

Train: A one-way ANOVA revealed a statistically significant difference between story types, $F(2, 158) = 54.61, p < .001, \eta^2 = .409$ (large effect). A Tukey HSD post hoc test showed that participants were more likely to retract in EVIDENCE ($M = -1.13, SD = 5.04, p < .001$, Cohen's $d = 1.67$ [large effect]) than in NEUTRAL ($M = 5.33, SD = 2.17$), and in EVIDENCE than in STAKES ($M = 4.94, SD = 2.95, p < .001$, Cohen's $d = 1.47$ [large effect]). No statistically significant difference was found between STAKES and NEUTRAL ($p = .841$).

Flight: A one-way ANOVA revealed a statistically significant difference between story types, $F(2, 173) = 123.36, p < .001, \eta^2 = .588$ (large effect). A Tukey HSD post hoc test showed that participants were more likely to retract in EVIDENCE ($M = -2.84, SD = 3.49, p < .001$, Cohen's $d = 2.66$ [large effect]) than in NEUTRAL ($M = 4.92, SD = 2.21$), and in EVIDENCE than in STAKES ($M = 4.42, SD = 3.13, p < .001$, Cohen's $d = 2.19$ [large effect]). No statistically significant difference was found between STAKES and NEUTRAL ($p = .645$).

DISCUSSION

The main finding of this experiment is the lack of evidence for the stakes effect. In both scenarios, retraction rates in the STAKES condition were not statistically significantly different from those in the NEUTRAL condition. As in previous experiments, a significant difference was observed between the EVIDENCE condition and the other two conditions, with a large effect size. These results suggest that there may be meaningful differences between the Bank scenario used in prior experiments and the Train and Flight scenarios tested here.

Although the sample size in this experiment is smaller than in Experiments 1 and 2, it is still comparable to the original experiments by Dinges and Zakkou. Considering the medium-to-large stakes effect observed in previous experiments, it is unlikely that the absence of the stakes effect in Experiment 3 is due to insufficient statistical power. It is worth noting that this experiment combined both the third-person perspective and the exclusion of scenario sceptics modification that had not previously been tested together. While Experiments 1 and 2 strongly suggest that each of these factors individually does not reduce the stakes effect, the findings from Experiment 3 indicate that the combination of both may contribute to the stakes effect disappearance.

One plausible explanation for this disappearance is that the manipulation of stakes in the Train and Flight scenarios may not have been as effective as in the Bank scenario. However, it is hard to see why travelling to help a poisoned person who may need resuscitation would be perceived as less important than experiencing difficulties due to a missed deposit. The allergy-related stakes used in the Train scenario were partially inspired by the scenarios tested by (Sripada & Stanley, 2012, pp. 13–14), which involved a protagonist with an allergy to “Mongolian pine nuts”. As such, the operationalisation of stakes in the Train scenario appears at least comparable to those in Sripada and Stanley’s experiments.

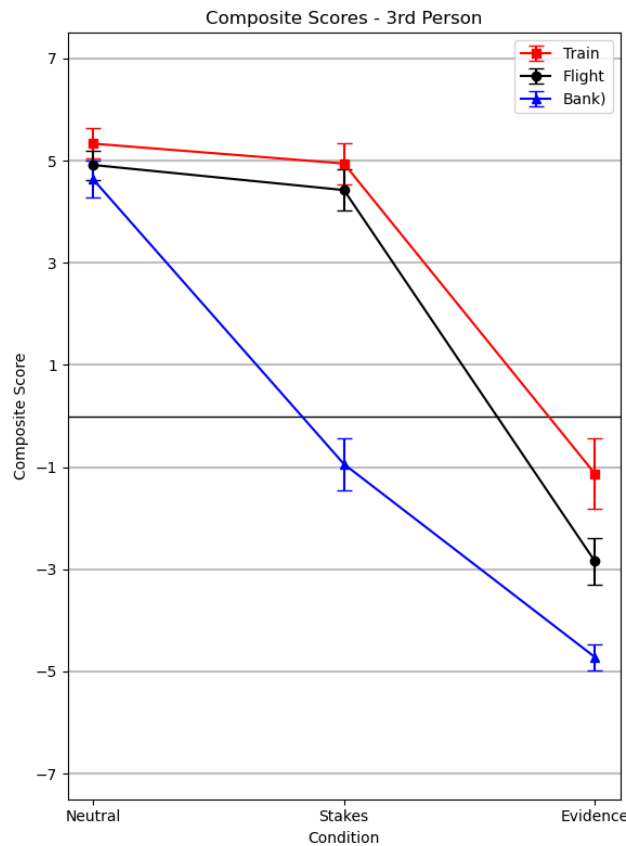


Figure 5. The mean composite scores three conditions: Neutral, Stakes, and Evidence in the Experiment 3 scenarios: Train, Flight (Experiment 3) versus the Bank (third- person; Experiment 1).

GENERAL DISCUSSION

If we temporarily set aside Experiment 3, the competing explanations can be reconstructed as follows. Epistemic contextualism (EC) accommodates the results of Experiments 1 and 2 without significant challenges. Contextualists may argue that the STAKES condition in these experiments introduced higher epistemic standards, which rendered participants unable to stand by knowledge ascriptions initially made in low-stakes settings. Subject-sensitive invariantism (SSI), while sharing the idea of heightened epistemic standards, ties these standards to the subject. This limitation creates difficulties in explaining the stakes effect in third-person cases, as seen in

Experiment 1. Proponents of SSI might appeal to the projectionist strategy, but as discussed in the context of Experiment 1, this approach lacks the details.

The position of SSI becomes even more unstable when the results of Experiment 3 are included. To explain cases in which a high-stakes attributor denies knowledge to a low-stakes subject, SSI advocates like Hawthorne and Stanley have argued that the attributor projects their own stakes/ignorance/bias onto the subject—evaluating the subject's epistemic status as if the subject were in the same high-stakes situation. However, the results of Experiment 3 suggest that this projection mechanism worked in Experiment 1 but failed in the Train and Flight scenarios. This inconsistency demands further elaboration of what projection involves and how it functions. A potential defence for SSI might involve the notion of proximity between the attributor and the subject. In the Bank case, the subject (Peter, the colleague) shares physical proximity with the attributor, being in the same car and possibly being a friend, whereas in the Train and Flight scenarios, the knowledge is attributed to strangers. It is plausible that projection is reduced entirely when the psychological distance between the subject and attributor increases. While this strategy could provide a way to defend SSI, it requires further development and empirical validation.

In contrast, EC provides a straightforward explanation for the results of all three experiments without requiring additional mechanisms. As DeRose (2009) aptly observes:

there is nothing in contextualism to prevent a speaker's context from selecting epistemic standards appropriate to the practical situation of the subject being talked about, even when the subject being discussed is no party to the speaker's conversation—which is good, because speakers often do select such standards when their conversational purposes call for it! (DeRose, 2009, p. 240)

Applying this perspective, contextualists can argue that high epistemic standards influenced the attributor in the Bank case (Experiment 1), while the low standards of the subjects (strangers) were more relevant in Experiment 3. This unified explanation elegantly accounts for the results across all three experiments without necessitating the introduction of additional variables.

Of course, these findings invite further exploration and discussion, particularly regarding other views that incorporate pragmatic encroachment on knowledge claims, such as adaptive invariantism (Nagel, 2008, 2010) or contrastivism (Schaffer & Knobe, 2012). While this paper focuses on EC and SSI, other perspectives may yield additional insights into the significance of these results.

This paper started with the methodological debate between evidence-seeking and evidence-fixed experimental designs, which have been argued to test different aspects of the stakes-knowledge relationship with varying levels of success. The retraction-based design was chosen as a promising and underexplored alternative, with the potential to reveal robust stakes effects. By implementing a single key modification aimed at improving ecological validity, this study demonstrated the design's potential for exploring third-person cases. While the design does not definitively prove the stakes effect, it has shown high potential for generating thought-provoking results and may require greater attention from experimental philosophers, especially in light of retraction becoming a subject of experimental scrutiny in other x-phi papers (see (Almagro et al., 2023; Khoo, 2015; Kneer, 2021; Knobe & Yalcin, 2014; Marques, 2018)).

CONCLUSION

Overall, the retraction-based design is confirmed to be capable of uncovering the stakes effect. While this paper explicitly addressed only the third-person cases, future research should explore other dynamics and factors (knowledge denial, polarity, stakes being lowered, etc.). Experiment 1 successfully replicated the findings of Dinges and Zakkou, also demonstrating that the stakes effect persists in the third-person version of the Bank case. Experiment 2 introduced a modification to the initial design by adding a knowledge-ascribing question. This addition made the act of retraction more realistic and helped to exclude scenario sceptics—participants who disagree with the knowledge claim. The modified design was tested using the first-person version of the Bank case and was found not to reduce or significantly influence the stakes effect. Experiment 3 attempted to explore two additional third-person scenarios—Train and Flight—using the modified design but did not observe the stakes effect.

Among the competing explanations, none explicitly predict retraction behavior, nor are they required to provide such an explanation. However, epistemic contextualism appears to offer a more uniform account of the obtained results. Contextualists can readily explain the stakes effect observed in the third-person conditions of Experiment 1 and may argue that participants in Experiment 3 simply adopted the epistemic standards relevant to their conversational context. In contrast, subject-sensitive invariantism faces greater challenges in explaining the results unless it is supported by a developed theory of projection. Even with such a theory, it would need to explain why projection occurred only in Experiment 1 and not in Experiment 3. For invariantists, this might involve alleged proximity between the subject and the knowledge attributor. However, if such a variable exists, it would require additional justification.

REFERENCES

- Almagro, M., Bordonaba-Plou, D., & Villanueva, N. (2023). Retraction in public settings. *Synthese*, 202(5), 1–25. <https://doi.org/10.1007/S11229-023-04348-3/TABLES/2>
- Buckwalter, W. (2010). Knowledge Isn't Closed on Saturday: A Study in Ordinary Language. *Review of Philosophy and Psychology*, 1(3), 395–406. <https://doi.org/10.1007/S13164-010-0030-3>
- Buckwalter, W., & Schaffer, J. (2015). Knowledge, stakes, and mistakes. *Nous*, 49(2), 201–234. <https://doi.org/10.1111/NOUS.12017>
- DeRose, K. (2005). The ordinary language basis for contextualism, and the new invariantism. *Philosophical Quarterly*, 55(219), 172–198. <https://doi.org/10.1111/J.0031-8094.2005.00394.X>
- DeRose, K. (2009). The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1. *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*, 1–304. <https://doi.org/10.1093/ACPROF:OSO/9780199564460.001.0001>
- Dinges, A., & Zakkou, J. (2021). Much at stake in knowledge. *Mind & Language*, 36(5), 729–749. <https://doi.org/10.1111/MILA.12300>
- Francis, K. B., Beaman, P., & Hansen, N. (2019a). Stakes, Scales, and Skepticism. *Ergo, an Open Access Journal of Philosophy*, 6(20201214). <https://doi.org/10.3998/ERGO.12405314.0006.016>
- Francis, K. B., Beaman, P., & Hansen, N. (2019b). Stakes, Scales, and Skepticism. *Ergo, an Open Access Journal of Philosophy*, 6(20201214). <https://doi.org/10.3998/ergo.12405314.0006.016>

- Grindrod, J., Andow, J., & Hansen, N. (2019). Third-person knowledge ascriptions: A crucial experiment for contextualism. *Mind and Language*, 34(2), 158–182. <https://doi.org/10.1111/MILA.12196>
- Hansen, N., & Chemla, E. (2013). *Experimenting on Contextualism*.
- Hawthorne, J. (2003). Knowledge and Lotteries. In *Knowledge and Lotteries*. Oxford University Press. <https://doi.org/10.1093/0199269556.001.0001>
- Khoo, J. (2015). Modal disagreements. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 5(1), 1–24. <https://doi.org/10.1080/0020174x.2015.1033005>
- Kneer, M. (2021). Predicates of personal taste: Empirical data. *Synthese*, 199(3–4), 6455–6471. <https://doi.org/10.1007/s11229-021-03077-9>
- Knobe, J., & Yalcin, S. (2014). Epistemic modals and context: Experimental data. *Semantics and Pragmatics*, 7, 1–21. <https://doi.org/10.3765/sp.7.10>
- Marques, T. (2018). Retractions. *Retractions Synthese*, 195(8), 3335–3359. <https://doi.org/10.1007/s11229-015-0852-8>
- 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. <https://doi.org/10.1007/S13164-009-0014-3>
- Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy*, 86(2), 279–294. <https://doi.org/10.1080/00048400801886397>
- Nagel, J. (2010). EPISTEMIC ANXIETY AND ADAPTIVE INVARIANTISM. *Philosophical Perspectives*, 24(1), 407–435. <https://doi.org/10.1111/J.1520-8583.2010.00198.X>
- Pinillos, Á. (2012). Knowledge, Experiments, and Practical Interests. *Knowledge Ascriptions*. <https://doi.org/10.1093/ACPROF:OSO/9780199693702.003.0009>
- Pinillos, Á. (2016). Experiments on contextualism and interest relative invariantism. *A Companion to Experimental Philosophy*, 349–358. <https://doi.org/10.1002/9781118661666.CH24>
- Porter, B., Barr, K., Bencherifa, A., Buckwalter, W., Deguchi, Y., Fabiano, E., Hashimoto, T., Halamova, J., Homan, J., Karasawa, K., Kanovsky, M., Kim, H., Kiper, J., Lee, M., Liu, X., Mitova, V., Bhaya, R., Pantovic, L., Quintanilla, P., ... Machery, E. (2024). A puzzle about knowledge ascriptions. *Nous*. <https://doi.org/10.1111/NOUS.12515>
- Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E. E., Chatterjee, A., Cheon, H., Cho, I. R., Cohnitz, D., Cova, F., Dranseika, V., Lagos, Á. E., Ghadakpour, L., Grinberg, M., Hannikainen, I., Hashimoto, T., Horowitz, A., ... Zhu, J. (2019). Nothing at Stake in Knowledge*. *Nous*, 53(1), 224–247. <https://doi.org/10.1111/NOUS.12211>
- Schaffer, J., & Knobe, J. (2012). Contrastive knowledge surveyed. *Nous*, 46(4), 675–708. <https://doi.org/10.1111/J.1468-0068.2010.00795.X>
- Sripada, C. S., & Stanley, J. (2012). Empirical tests of interest-relative invariantism. *Episteme*, 9(1), 3–26. <https://doi.org/10.1017/EPI.2011.2>

Turri, J. (2017). Epistemic Contextualism: An Idle Hypothesis. *Australasian Journal of Philosophy*, 95(1), 141–156. <https://doi.org/10.1080/00048402.2016.1153684>

Wu, S. (2023). Are Folks Purists or Pragmatic Encroachers? New Discoveries of Relation between Knowledge and Action from Experimental Philosophy. *Episteme*, 1–29. <https://doi.org/10.1017/EPI.2023.56>

APPENDIX

FLIGHT scenario, Experiment 3

You are at the airport to take a direct flight to Tokyo. Although you know the departure time is 17:45 and the arrival time is 16:50, you have no idea how many hours a direct flight to Tokyo takes. You ask a woman standing next to you in a queue to security check whether she knows how many hours a direct flight to Tokyo takes. She says: 'I've been to Japan a couple of times; it takes 14 hours to get to Tokyo if you have a direct flight'. Based on this, you conclude:

- (1) The woman knows that the flight takes 14 hours
- (2) The woman does not know that the flight takes 14 hours

NEUTRAL

You meet a close friend who happens to be on the same flight. After exchanging some pleasantries, you mention that you met a woman who knows that the flight takes 14 hours. Your friend is very happy about it and says, "I worked a lot last week, so now I am glad that I can get a proper 8-hour-long sleep during the flight. Just to check, do you stand by your previous claim that the woman knows that the flight takes 14 hours?"

STAKES

You meet a close friend who happens to be taking the same flight. After exchanging some pleasantries, you mention that you met a woman who knows that the flight takes 14 hours. Your friend is a huge Star Wars fan and says that today is the Day of Star Wars: each fan must rewatch the two original trilogies, which takes 13 hours and 38 minutes. Your friend says: "I am glad that I can have a movie marathon during the flight. It is very important to me, and I would feel betrayed if you fooled me. Just to check, do you stand by your previous claim that the woman knows that the flight takes 14 hours?"

EVIDENCE

You meet a close friend who happens to take the same flight. After exchanging some pleasantries, you mention that you met a woman who knows that the flight takes 14 hours. Your friend now works for a Japanese company and has taken this flight already. He says: "I flew several times to Tokyo with this company, and all of my flights took exactly 15 hours. But just to check, do you stand by your previous claim that the woman knows that the flight takes 14 hours?"