Episteme (2025), 1–22
doi:10.1017/epi.2025.10060

episteme

ARTICLE

The Stakes Effect: New Evidence from a Retraction-Based Experimental Design

Nikolai Shurakov

Tartu Ulikool, Tartu, Estonia
Email: shurakovnukolay@gmail.com

(Received 27 January 2025; revised 5 June 2025; accepted 3 July 2025)

Abstract

The paper examines the influence of stakes on knowledge attributions, building on the retraction-based experimental design introduced by Dinges and Zakkou. Experiment 1 replicates Dinges and Zakkou's original findings and extends the research to third-person knowledge ascriptions. The results show that raising the stakes increases the percentage 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 claim and the worry of scenario sceptics – participants who disagree with a knowledge attribution. Experiment 2 introduces a modification to the initial design by adding a knowledge-ascribing question. This addition makes the act of retraction more realistic. The results confirm that the stakes effect persists even in an improved design. I argue that these findings constitute a serious challenge to classic invariantism and a potential challenge to subject-sensitive invariantism. Their competitors – epistemic contextualism and relativism – seem to be in a better position, even though the retraction-based design at its current stage is unlikely to distinguish between these two.

Keywords: Stakes; contextualism; knowledge attributions; invariantism; retraction

1. Introduction

Epistemic contextualism (EC) proposes that the epistemic standards for knowledge attributions shift according to conversational context. In high-stakes situations, for example, the epistemic standards required for attributing knowledge to a subject might be higher compared to those in low-stakes contexts, even when the subject's evidence remains the same. Its key rivals are *subject-sensitive invariantism* (SSI), which allows the subject's practical interests to affect knowledge attributions, and *classic invariantism,* which holds that epistemic standards for knowledge are fixed. Keith DeRose, a central proponent of contextualism, identifies the best type of evidence for his position:

> 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 well 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 and Zakkou 2021; Francis *et al.* 2019; Francis and Beaman 2023; 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, with each competing view finding at least partial support within the experimental literature. Grindrod *et al.* (2019) present empirical evidence that uniquely favors EC. Pinillos (2012), Pinillos & Simpson (2014) consistently demonstrate that subjects in high-stakes conditions require more evidence to attribute knowledge, interpreting these findings as empirical support for SSI. Nevertheless, the authors of a cross-cultural study that predominantly did not detect stakes effects "conclude that classical invariantism should be taken seriously, now that its challengers have been undermined" (Rose *et al.* 2019, p. 245). Section 2 gives an overview of existing experimental designs, demonstrating that the current experimental research struggles to resolve these theoretical debates definitively.

In response to these methodological struggles, Section 3 argues for further exploration of an alternative experimental paradigm: the retraction-based design pioneered by (Dinges and Zakkou (2021). This novel approach offers an alternative methodology for investigating a potential stakes effect, with initial findings indicating medium to large effects across two independent experiments. However, the absence of subsequent studies employing this design raises questions about the robustness and generalizability of Dinges and Zakkou's findings.

This paper presents two experimental studies. The first experiment directly replicates one of Dinges and Zakkou's experiments and also tests third-person cases. The second experiment addresses methodological concerns that could potentially undermine the findings. In Section 5, I discuss interpretations of the results from three main competing views: EC, SSI, and classic invariantism. The findings demonstrate that the stakes effect initially observed by Dinges and Zakkou (2021) is successfully replicated (Experiment 1), maintaining statistical significance even with a 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). While these retraction-based experimental results challenge classic and subject-sensitive invariantist views, they leave unresolved the question of whether EC or relativism better explains the observed effect.

2. Evidence-fixed and evidence-seeking experimental designs

In this section, I briefly present key findings from experiments employing either evidence-fixed or evidence-seeking experimental designs (the distinction introduced in (Pinillos 2012)). I aim to demonstrate that each competing view – EC, SSI, and classic invariantism (CI) – has resources to explain or at least argue against the findings. This short overview motivates the choice of an alternative retraction-based experimental design.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

To begin, let us outline the main positions in the debates over knowledge attributions.[1] A key distinction is between the *subject* of knowledge – the individual to whom knowledge is attributed – and the *attributor* – the individual making the attribution. Notably, the same person may occupy both roles. Classic invariantism denies that interests or practical factors are "knowledge-determining factors" (Nagel 2010), attributing the apparent effects of stakes to traditional epistemic factors (e.g., a lack of outright belief in high-stakes situations). SSI, by contrast, allows the stakes of the *subject* to influence whether they know. Meanwhile, EC emphasizes the role of the *attributor's* stakes (although DeRose (2009, p. 240) argues that sometimes the conversational context requires the adoption of the subject's standards and EC is compatible with it).

Each of the views has its pros and cons: "in the literature, the choice between these views is often presented as a choice of the least of three evils: given that they all have shortcomings, which shortcomings are easiest to stomach?" (MacFarlane 2014, p. 187). Rather than engaging in extensive theoretical debates between these positions, I focus on the empirical data and explore whether the findings seriously challenge any of these views.

Both SSI and EC appeal to folk intuitions regarding particular cases, with the Bank case being perhaps the most famous. In this scenario, you have a paycheck to deposit but you see huge lines to the bank office on Friday afternoon. You remember being at the bank a few weeks before, so you say, "I know the bank is open on Saturday." The high-stakes version is exactly the same, except this time, not depositing the paycheck by Monday would leave you in a very bad situation (see (DeRose 1992, 2009; Stanley 2005) for more detailed versions of this case). According to EC and SSI, you should self-ascribe knowledge in low stakes and deny it in high stakes, but your evidence in both cases is the same. However, their motivations differ: for SSI, it is your stakes *qua* subject of knowledge that do the work; for EC, it is your stakes *qua* attributor that do the work.

The experimental design in which researchers test low/high-stakes pairs of scenarios (like the Bank case just described) has been labeled "evidence-fixed". In these experiments, participants are typically asked to read a scenario and then say to what extent they agree with the target knowledge claim (e.g., (Buckwalter 2010; Feltz and Zarpentine 2010)) or whether this claim is true (e.g., (Hansen and Chemla 2013; Rose *et al.* 2019)). The results are mixed: while some experiments find an effect of stakes (e.g., (Grindrod *et al.* 2019; Hansen and Chemla 2013; Sripada and Stanley 2012)), others do not (e.g., (Buckwalter 2010; Feltz and Zarpentine 2010; May *et al.* 2010)). Furthermore, those who do find an effect report that it is small. Thus, Francis and colleagues claim 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.* 2019, p. 439). Each of the competing views can claim partial support. Proponents of EC and SSI can agree with findings that identify stakes effects, arguing that null results come from too weak stakes manipulations. Conversely, classic invariantists can highlight the lack of stakes effects in several experiments, contending that the small effect size fails to constitute robust evidence for the stakes sensitivity of knowledge attributions. The rarely observed small or nonexistent effect of stakes may be a result of particular wording of vignettes or caused by some confounds, and its size may not be enough to interfere with our knowledge-attributing practices considerably.

[1]In this article, 'knowledge ascriptions' and 'knowledge attributions' will be used interchangeably, as is common in the literature. While subtle differences between the two may be relevant for interpreting the data, the primary focus is on reporting experimental findings.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

Traditional evidence-fixed experiments have been criticized for difficulties in ensuring that the perceived evidence available to the protagonist remains constant across conditions. Pinillos (2012) introduced an alternative: the evidence-seeking experimental design. Instead of evaluating knowledge ascriptions, participants assess how much evidence a subject must gather to "know" something. For example, participants reading a scenario about a student submitting a paper might be asked, "How many times do you think Peter has to proofread his paper before he knows that there are no typos?" (Pinillos 2012, p. 199). Participants regularly reply with a higher number of checks in these experiments (Francis *et al.* 2019; Francis and Beaman 2023; Pinillos 2012; Pinillos and Simpson 2014). The experiments employing these designs do find an effect of stakes, lending support to EC and SSI (Pinillos (2012, p. 213–215) even provides some reasons why SSI offers a better explanation). Proponents of CI troubled by this may nonetheless argue that the observed effects do not reflect the stakes' influence on *knowledge* attributions. Proponents of CI can appeal to one side of the puzzle regarding knowledge ascriptions formulated by Porter and colleagues: "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). Consequently, the evidence-seeking design might be claimed to investigate folk beliefs about the evidential thresholds for knowledge rather than knowledge attributions *per se*. If so, CI can disregard the findings of evidence-seeking experiments and maintain that knowledge attributions are unaffected by practical factors.

To summarize, both dominant experimental paradigms – evidence-fixed and evidence-seeking – struggle to definitively resolve the debates concerning the stakes sensitivity of knowledge attributions. None of the competing views is seriously undermined by experimental philosophy findings. For this reason, this paper adopts and further explores an alternative retraction-based experimental design that may provide some novel evidence.

3. Retraction-based experimental design

One more experimental design has attempted to find the stakes effect in a series of experiments (Dinges and Zakkou 2021) (hereafter just D&Z). It might be of interest here since it employs retraction as a tool in investigating knowledge-attributing practices. 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.. after reading about having a paycheck to deposit and remembering being to the bank three weeks ago: "based on this, you respond: 'I know the bank will be open tomorrow'" (D&Z, p. 736). The next part of the scenario is different, depending on which of three conditions a participant has been assigned to: NEUTRAL (no new information regarding the 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 are as follows: retraction in STAKES (48%) is significantly higher than in NEUTRAL (9.8%), and retraction in EVIDENCE (96.1%) is higher than in STAKES. A similar pattern holds for the Typo scenario (wherein a student needs to proofread a paper). The retraction rates were as follows: NEUTRAL 5.9%, STAKES 24%, and EVIDENCE 58%. Across both scenarios, D&Z consistently found a medium to large effect of stakes.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

These findings set the retraction-based design apart from evidence-fixed and evidence-seeking paradigms. Its consistent results and medium-to-large effect sizes contrast sharply with the irregular small effects observed in evidence-fixed design experiments. Moreover, the retraction-based design avoids the "deferral confound" found in Turri (2017).[2] Using the evidence-fixed design (between-subjects), Turri found that participants agreed with protagonists' self-ascriptions and denials of knowledge, when stakes manipulation was not involved (Turri 2017, pp. 147–149). This tendency to defer to protagonists' epistemic claims produces a pattern that aligns with contextualist prediction, but which is not driven by the stakes. The retraction-based design skirts this issue by using a different epistemic perspective: participants are positioned not as evaluators of another person's knowledge claim but as the speakers who make the knowledge attribution themselves. As such, the resulting judgments reflect participants' own use of "know". Thus, the retraction-based design not only yields more robust evidence of a stakes effect but also avoids a confound that may undermine the findings from evidence-fixed experiments. For a similar reason, it avoids Porter *et al.*'s criticism of the evidence-seeking paradigm because the retraction-based design directly targets knowledge ascription practice (how participants use knowledge ascriptions) and not beliefs about evidence thresholds. These considerations provide 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" (D&Z, p. 746). The current paper adopts and investigates their experimental design further.

4. Experiments

4.1. Motivation for Experiment 1

In what follows, two primary motivations for conducting Experiment 1 are provided. After justifying running the direct replication of one of D&Z's experiments, I also motivate testing third-person cases.

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 and Stanley 2012) experiment by Francis *et al.* (2019) yielded a reverse pattern. For the vignette pairs where Sripada and Stanley found an effect, Francis and colleagues did not, but there was an effect for the vignette pair that was not significantly different in Sripada and Stanley's paper. Although this might be another illustration of the mixed findings associated with the evidence-fixed design, the very possibility of not replicating the effect reported by D&Z should not be dismissed. Thus, the first goal of Experiment 1 is to run a direct replication that would not only confirm the replicability of D&Z results but also would provide reliable and clean data for future comparisons. This brings us to the second motivation for Experiment 1: the possibility of contrasting the first- and third-person cases.

Third-person cases are also crucial for distinguishing between EC and SSI. Both views permit stakes to influence knowledge ascriptions and align in predictions for first-person cases but diverge regarding third-person cases. The first experimental study of third-person knowledge ascriptions has emphasized why third-person cases are so important:

[2]I thank an anonymous reviewer for bringing this point to my attention.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

> 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 Color and Control scenarios) and their findings offered overall support for contextualism.

In the retraction-based design, participants initially make a knowledge ascription, assigning knowledge to a subject; then, participants in the STAKES condition are informed of higher stakes and asked if they would stand by their initial ascription. SSI proponents should not anticipate a change in stakes to affect the attributor's decision on whether to stick to the ascription, as the subject's stakes remain the same. By contrast, EC, which prioritizes the attributor's context, predicts that stakes influence third-person knowledge ascriptions. Therefore, if a stakes effect emerges in third-person retraction-based cases, this would pose a challenge to SSI. Moreover, if the effect is both replicated in first person and shown to impact third-person scenarios, CI would face an even stronger challenge. If the results of the third-person retraction-based experiments align with those reported by Grindrod and colleagues, this would offer converging evidence for contextualist predictions across distinct experimental paradigms.

To sum up, 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 to potentially constitute a challenge to both kinds of invariantism.

4.2. Experiment 1

Preregistered 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/tys3p].

4.2.1. 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).[3] 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).

4.2.2. Design, procedure, materials

The experiment utilized a 3×2 design comprising three types of story (Neutral, Stakes, Evidence; between subjects) and two perspectives (first person, third person; between subjects). Participants were randomly assigned to one of six conditions. First,

[3]https://osf.io/5z6pu.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

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 details of the scenario and the task.

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:
>
> 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:

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

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:

o I do
o 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?

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

The experiment aimed to answer the following questions: (i) Does the result reported by Dinges and Zakkou replicate? (ii) does the result reported by Dinges and Zakkou hold for the third-person conditions? and (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.

4.2.3. 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.

4.2.3.1. Analysis of binary results. The percentage of retractions ("I don't") in first-person conditions was as follows: 7% in NEUTRAL, 53% in STAKES, and 95% in

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

EVIDENCE. In the third-person conditions, the numbers are similar: 11% in NEUTRAL, 61% in STAKES, and 94% in EVIDENCE. The detailed statistics are presented in Table 1. The overall story type and each pair-wise comparison are significantly different, with the effect ranging from medium to large.

4.2.3.2. Analysis of composite scores. The analysis followed the methodology of Dinges and Zakkou. 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. Mean composite scores by condition are shown in Figure 1.

For the first-person condition, a one-way ANOVA revealed a statistically significant difference between story types, $F(2, 297) = 167.720$, $p < .001$, $\eta^2 = .530$ (large effect). All pairwise comparisons were significant with a large effect size (see Table 2).

The same holds for third-person conditions. A one-way ANOVA revealed a statistically significant difference between story types, $F(2, 297) = 149.771$, $p < .001$, $\eta^2 = .530$ (large effect). Again, all pairwise comparisons were significant with a large effect size (see Table 2).

4.2.3.3. 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.

4.2.4. Discussion

The original experiment was replicated almost perfectly. Retraction rates across conditions were similar (D&Z vs. the replication): 9.8% vs. 7% in the NEUTRAL condition, 48% vs. 53% in STAKES, and 96.1% vs. 95% in EVIDENCE; there are consistent medium-to-large effects across both studies. Similarly, the composite scores (see Figure 2) 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.

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.

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 stated in the text provided and required some extra reasoning. In third-person vignettes, participants read that "Peter says that *he* remembers having been at the bank three weeks before on a Saturday," but the attention check asks: "How long ago were *you* at the bank?" (emphasis added). Participants had to recognize that the protagonist (i.e., themselves) was not the one who had previously been to the bank. Selecting the correct answer thus required not only paying attention to the vignette text but also some kind of perspective-taking (and most participants managed to do it).

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

Table 1. The results of the chi-squared tests for independence for Experiment 1 (first-person and third-person perspectives) and Experiment 2 (first-person: modified design).

Perspective	Comparison	χ^2 (df, N)	*p*-value	Cramer's V	Effect size
First person	Story Type (Overall)	χ^2 (2, N = 300) = 155.159	< .001	.719	Large
	STAKES vs. NEUTRAL	χ^2 (1, N = 200) = 48.214	< .001	.491	Medium
	EVIDENCE vs. NEUTRAL	χ^2 (1, N = 200) = 151.441	< .001	.870	Large
	EVIDENCE vs. STAKES	χ^2 (1, N = 200) = 43.685	< .001	.467	Medium
Third-person	Story Type (Overall)	χ^2 (2, N = 300) = 141.315	< .001	.686	Large
	STAKES vs. NEUTRAL	χ^2 (1, N = 200) = 52.105	< .001	.510	Medium
	EVIDENCE vs. NEUTRAL	χ^2 (1, N = 200) = 134.817	< .001	.821	Large
	EVIDENCE vs. STAKES	χ^2 (1, N = 200) = 29.362	< .001	.383	Medium
First person: Modified design	Story Type (Overall)	χ^2 (2, N = 300) = 160.176	< .001	.731	Large
	STAKES vs. NEUTRAL	χ^2 (1, N = 200) = 35.032	< .001	.419	Medium
	EVIDENCE vs. NEUTRAL	χ^2 (1, N = 200) = 154.895	< .001	.88	Large
	EVIDENCE vs. STAKES	χ^2 (1, N = 200) = 60.799	< .001	.551	Medium

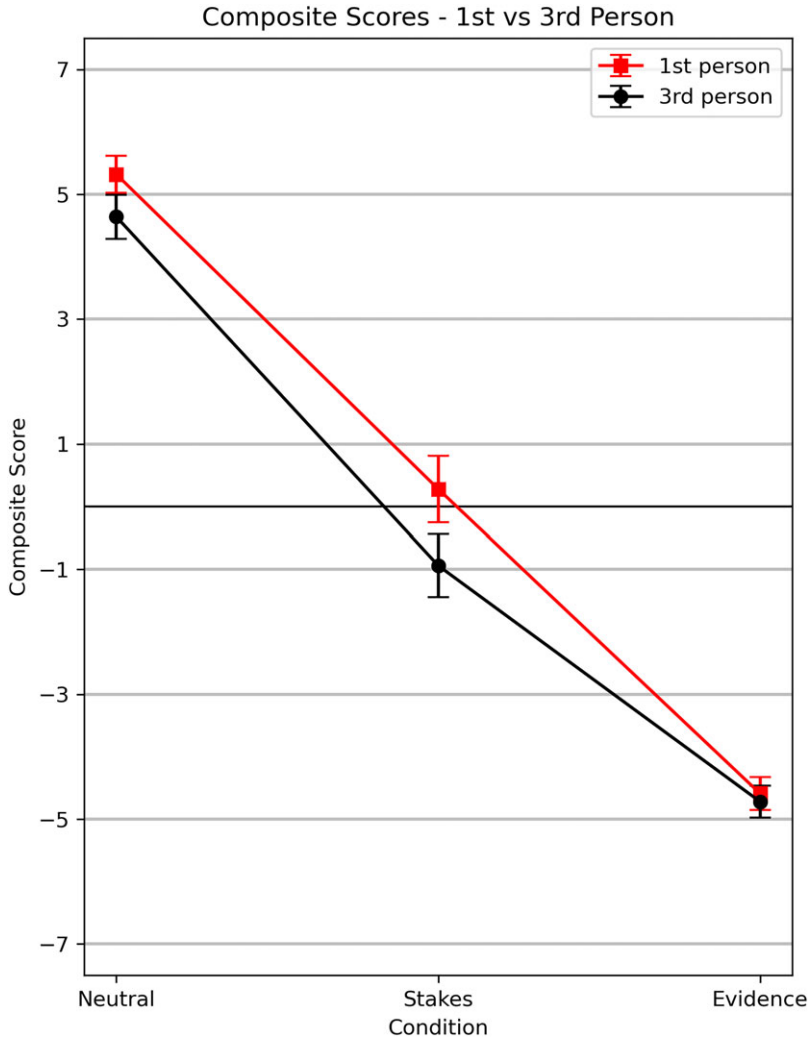

Figure 1. The mean composite scores for the first- and third-person scenarios across three conditions: Neutral, Stakes, and Evidence.

4.3. Motivation for Experiment 2

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

The primary task for participants in the experiment is to decide whether to retract a knowledge claim. However, for a retraction to be genuine, participants must first have endorsed the knowledge claim they are asked to retract. In the original D&Z design, 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

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

Table 2. The results of Tukey's HSD post hoc pairwise comparisons for Experiment 1 (first-person and third-person perspectives) and Experiment 2 (first person: modified design).

Perspective	Comparison	Means M (SD)	*p*-value	Cohen's d	Effect size
First person	STAKES vs. NEUTRAL	STAKES: 0.28 (5.32), NEUTRAL: 5.32 (2.93)	< .001	1.06	Large
	EVIDENCE vs. NEUTRAL	EVIDENCE: −4.59 (2.64), NEUTRAL: 5.32 (2.93)	< .001	3.55	Large
	EVIDENCE vs. STAKES	EVIDENCE: −4.59 (2.64), STAKES: 0.28 (5.32)	< .001	1.16	Large
Third-person	STAKES vs. NEUTRAL	STAKES: −0.94 (5.04), NEUTRAL: 4.64 (3.53)	< .001	1.28	Large
	EVIDENCE vs. NEUTRAL	EVIDENCE: −4.72 (2.56), NEUTRAL: 4.64 (3.53)	< .001	3.04	Large
	EVIDENCE vs. STAKES	EVIDENCE: −4.72 (2.56), STAKES: −0.94 (5.04)	< .001	0.95	Large
First person: Modified design	STAKES vs. NEUTRAL	STAKES: 1.59 (5.06), NEUTRAL: 5.36 (2.63)	< .001	0.94	Large
	EVIDENCE vs. NEUTRAL	EVIDENCE: −4.7 (2.36), NEUTRAL: 5.36 (2.63)	< .001	4.04	Large
	EVIDENCE vs. STAKES	EVIDENCE: −4.7 (2.36), STAKES: 1.59 (5.06)	< .001	1.6	Large

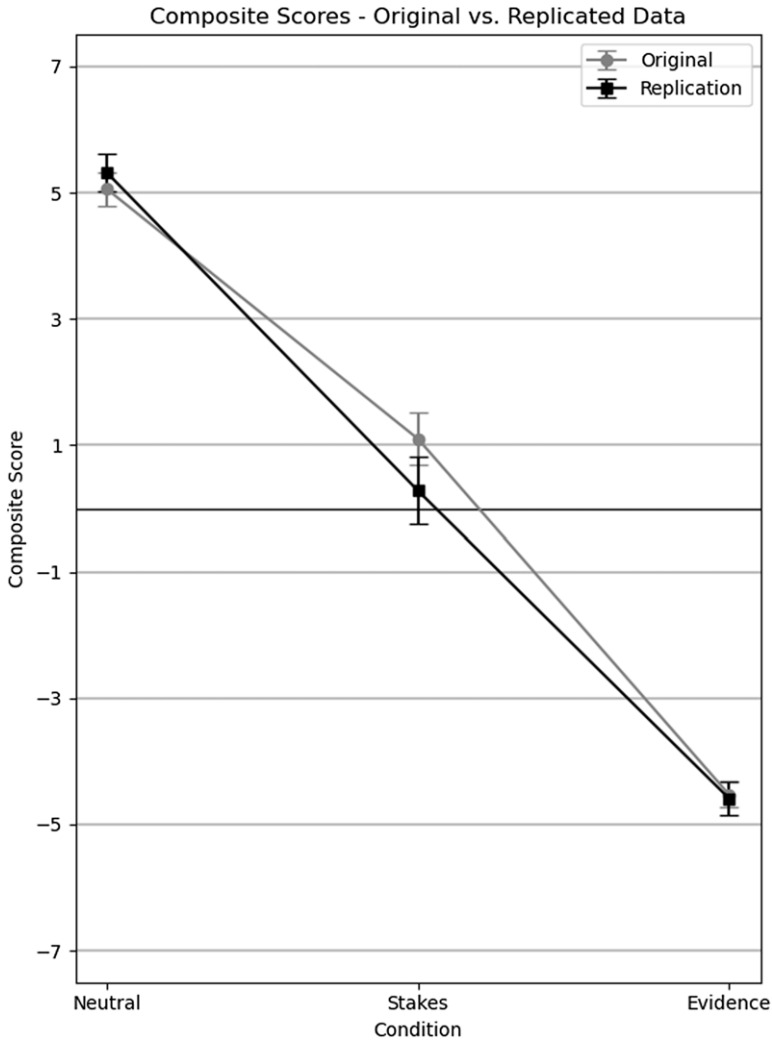

Figure 2. The mean composite score comparison between the original study – Dinges & Zakkou (2021) and the direct replication across three conditions: Neutral, Stakes, and Evidence.

retracting a hypothetical claim they never explicitly made or endorsed. There might be a significant difference between standing by the imagined claim versus retracting a claim one actually made, as the former may not correspond to any belief participants ever had. A modified design that allows participants to genuinely endorse and then retract the knowledge claim is needed to improve the ecological validity of the results.

It is also unlikely that all participants equally 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,

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

scenario 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. The presence of scenario sceptics could significantly impact the observed stakes effect because scenario sceptics seem to be 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 accept that this claim is regarding knowledge.

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

4.4. 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. 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. 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.

4.4.1. 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 say; mean age 38 years). Participants were paid £0.25 for approximately 2 minutes of their time (£7.50 per hour).

4.4.2. 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 first-person conditions of Experiment 1). Due to lack of variation between the first- and third-person conditions in Experiment 1, only the first-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):

[4]Preregistered: https://osf.io/ae3pu/.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

> . . . You remember having been at the bank three weeks before on a Saturday. Based on this, you respond:
>
> o "I know the bank will be open tomorrow".
> o "I don't know the bank will be open tomorrow".

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. To address question (ii), I used the Stakes first-person data from Experiment 1. Given the almost perfect replication of the original experiment in Experiment 1, this comparison data might be reliable.

4.4.3. Results

Of 450 participants, 53 (11.8%) were scenario sceptics and 38 failed the attention check. From the remaining 359 valid responses, only the first 100 per condition were included in the final analysis.

4.4.3.1. 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. The detailed statistics are presented in Table 1. The overall story type and each pair-wise comparison are significantly different, with the effect ranging from medium to large.

4.4.3.2. Analysis of composite scores. The analysis followed the methodology of Dinges and Zakkou. 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. Mean composite scores by condition are shown in Figure 1.

A one-way ANOVA revealed a statistically significant difference between story types, $F(2, 297) = 203.050$, $p < .001$, $\eta^2 = .582$ (large effect). All pairwise comparisons were significant with a large effect size (see Table 2).

4.4.3.3. Stakes: replication vs. modified design. The comparison between the replication STAKES condition and the modified STAKES condition did not reveal a statistically significant difference (see Figure 3). 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, $p = .064$ (both not significant).

4.4.4. Discussion

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. The results of the first

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

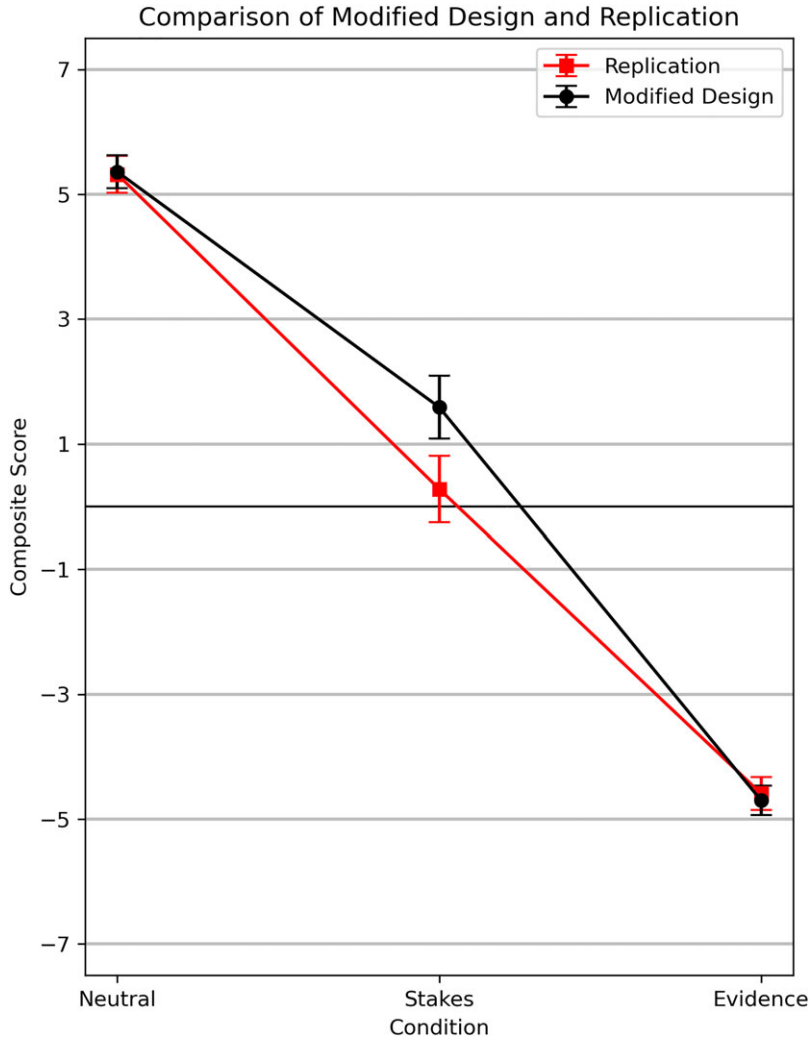

Figure 3. The mean composite scores for three conditions: Neutral, Stakes, and Evidence in the direct replication and modified design.

experiment are replicated in the modified design, implying that even when the retraction is made more realistic and scenario sceptics are excluded from further analysis, stakes do affect knowledge attributions, 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 evidence than the initial experiment, and its replication, as the main mechanism of the experiment – retraction, is made more realistic.

5. General discussion

The robust findings of Dinges and Zakkou were independently replicated. Analyses of both binary responses and composite scores revealed significant differences between

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

story types: NEUTRAL < STAKES < EVIDENCE, with medium-to-large effect sizes. This pattern held consistently across both first-person and third-person versions of the Bank scenario, with no statistically significant difference between the STAKES first-person and STAKES third-person conditions. This consistency suggests that stakes influence third-person knowledge attributions as well, potentially providing novel evidence for the stakes sensitivity of knowledge attributions. Moreover, the design modification introduced in Experiment 2 addressed several shortcomings of the retraction-based methodology, demonstrating that the robustness of the findings was unaffected by previously identified design limitations.

In what follows, I briefly evaluate whether each of the main competing views – classic invariantism (CI), subject-sensitive invariantism (SSI), and EC – can plausibly account for the observed results, inviting more detailed interpretations from any more specific position (e.g., from adaptive invariantism (Nagel 2010), contrastivism (Schaffer and Knobe 2012), or asymmetric loss (Dinges 2023).

5.1. Classic invariantism

The results of both experiments pose a serious challenge to CI. If practical factors do not influence knowledge ascriptions, as CI maintains, then the retraction rates in the STAKES and NEUTRAL conditions should be equivalent. This is not the case. The experimental outcomes – reproduced across the original experiments by Dinges and Zakkou, independent replication, and the third-person version of the Bank scenario – consistently exhibit a robust and medium-to-large stakes effect. Unlike the small effects observed in evidence-fixed experiments (e.g., Hansen and Chemla 2013; Francis *et al.* 2019), the stakes effect in the retraction-based design is both statistically significant and unlikely to result from confounding factors. The attention check question ensured that participants were reasoning explicitly about someone else's knowledge, with the vast majority answering correctly. The design modification introduced in Experiment 2 further counters potential objections to the methodology. Taken together, the results of these experiments, alongside the findings of Dinges and Zakkou, significantly undermine the orthodoxy of CI, which denies that practical factors influence knowledge attributions. Proponents of CI may still appeal to conversational implicatures (e.g., Brown 2006) or heuristic proxy (e.g., Gerken 2017), but D&Z (pp. 15–16) has already shown how their findings challenge such positions. This paper replicates and strengthens D&Z's results, putting even more pressure on CI.

5.2. Subject-sensitive invariantism

Subject-sensitive invariantism (SSI) fares better than CI in accounting for the results but encounters difficulties with the third-person conditions. SSI posits that practical factors of the subject influence knowledge attributions. However, in the STAKES third-person condition, participants retracted knowledge claims about Peter, even though his stakes remained unchanged. SSI should predict minimal retraction in such cases, as the stakes should not directly affect the epistemic status of the subject.

Proponents of SSI might appeal to projectionist explanations. Hawthorne suggests that people tend to "overproject our own lack of knowledge to others" (Hawthorne 2003, p. 163). A high-stakes attributor might not self-ascribe knowledge and thus project this lack of knowledge onto the low-stakes subject.

Against this strategy, DeRose (2009, pp. 234–238) argues that this explanation breaks down when the attributor has no ignorance (lack of knowledge) to project. To illustrate, consider an alternative: a participant has strong evidence that the bank is

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

open: perhaps they visited it recently, their partner works there, and they are well aware that the hours have not changed recently. Peter still only has a recollection of the past visit as evidence. After reading the scenario setup, the participant might think: "I know the bank is open, and Peter does too." But once stakes go up, they may still retract: "Oh, if it's *that* important, then Peter doesn't know (but I still do)." The retraction may occur even without the alleged attributor's ignorance, which is supposed to be projected. Whether this is actually the case is an empirical question: perhaps, folk intuitions are distinct from those described here, but such an objection to the projectionist strategy seems possible.

SSI also needs a detailed account of the psychological mechanisms underlying projection. Dinges (2018) demonstrates how salience sensitivity (Hawthorne's view), which appeals to *egocentric bias*, effectively deals with third-person cases. The egocentric bias is the idea that "readers project their own worry about the salient error-possibility" (Dinges 2018, p. 2848). Indeed, in STAKES condition, one can interpret "it would be a disaster if you ... found the bank closed tomorrow" as bringing a salient possibility that the bank is closed the next day. Due to the egocentric bias, STAKES third-person participants project this worry onto Peter and treat him as sharing this concern. Some participants might have reasoned that Peter's recollection of "having been at the bank three weeks before on a Saturday" is insufficient evidence to rule out this salient possibility and go for a retraction.

Interestingly, such a salience sensitivity account appears to withstand DeRose's critique when the attributor has no ignorance to project. Again, consider an attributor who has way more evidence than Peter. Such an attributor may agree that they both know that the bank is open on Saturday when both can rule out a few relevant salient alternatives. However, if the attributor is placed in a high-stakes situation that renders a specific error possibility salient, then the attributor may be able to rule it out, but Peter may not. This would lead to the attributer retracting the knowledge attribution done in the low-stakes settings. In this way, the salience sensitivity version of SSI supplemented with the egocentric bias could accommodate the results.

It is important to emphasize that the egocentric bias could not support the stakes sensitivity (see Dinges 2018, pp. 2854–2855 for argumentation), so SSI accounts that appeal to stakes may need to appeal to some other psychological mechanism (or find a way of incorporating the egocentric bias).

Nonetheless, there is a reason to think that even the salience sensitivity explanation does not account for the whole picture. D&Z already addressed an explanation that participants assigned to STAKES conditions thought of more error possibilities and, for this reason, retracted knowledge attributions. D&Z found that "the generation of error-possibilities is only *partially* responsible for stakes effects. ... Stakes effects do pose a distinctive explanatory challenge to orthodox positions" (D&Z, p. 13). Their experimental findings suggest that both salient alternatives and practical stakes influence knowledge attributions.

In brief, SSI requires a projection to accommodate the results for the STAKES third-person condition. However, one may question whether projection is a genuine psychological phenomenon. Dinges (2018) demonstrated that egocentric bias is a plausible underlying mechanism. Yet appealing to this bias only supports salience sensitivity and is not readily available for stakes sensitivity (Dinges 2018, pp. 2854–2855). The results of D&Z's third experiment indicate that salience alone does not fully explain the observed interactions. Furthermore, MacFarlane (2014, p. 185) rightly notes that "even if the projection strategy works, it is a double-edged sword," as

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

the appeal to projection would make SSI more similar to EC, and thereby subject to anti-contextualist arguments.[5] Thus, I propose that SSI faces, at least, a limited challenge in light of the results presented in this paper. I do not argue that SSI is incapable of accounting for these findings, but rather that giving such an explanation is potentially challenging to the SSI defenders.

5.3. Epistemic contextualism

EC appears well suited to explain the results across all conditions. According to EC, higher stakes raise epistemic standards within the conversational context, leading to stricter thresholds for knowledge ascriptions (DeRose 2009). In both the first-person and third-person STAKES conditions, a significant number of participants adopted these stricter standards, resulting in a substantial percentage of retractions. Participants who retracted likely judged that the available evidence – such as remembering a visit to the bank three weeks prior – was insufficient to meet the higher standards imposed in STAKES.

If the stakes are raised within a single context, participants who initially asserted that "S knows that *p*" may later retract their assertion upon realizing that higher epistemic standards are in play. This retraction occurs because the knowledge attribution does not satisfy the threshold for knowledge in a given context. Importantly, if the scenario is treated as a single context, then from a contextualist perspective, one should not assert: "I knew that the bank is open on Saturday, but now, when I see that the stakes are high, I don't." DeRose argues that "know" in the first part of the sentence must satisfy the relevant standards of the context, which are high, making the first use of "know" unassertable. Thus, contextualism not only accounts for the results but does so without endorsing contradictory "now you know it, now you don't" claims (see DeRose 2009, pp. 204–206). If the scenario involves two distinct contexts– a low-stakes setup and a high-stakes follow-up (as in the STAKES conditions) – contextualism can interpret the experiment as reflecting two claims: knowledge attribution in the low-stakes context and knowledge denial in the high-stakes context. This aligns with contextualist predictions regarding the Bank Case (DeRose 1992, 2005). I find the former explanation preferable, as the original Dinges and Zakkou experiments and Experiment 1 in this paper presented the setup and follow-up parts of the scenario as a uniform body of text. In any case, EC appears to provide a plausible explanation of the results.

However, an alternative explanation might be offered. Perhaps some participants reassessed the truth-value of the knowledge ascription in the high-stakes context. They retracted not because the epistemic standards had changed, but because the knowledge ascription was no longer true: the context of assessment had shifted. Indeed, the retraction-based design does not ask participants about the truth-value of knowledge attributions, so the results are also compatible with relativism (e.g., MacFarlane 2014). According to EC, participants who chose not to stand by the knowledge ascription would still consider the initial claim to be true (from the context of utterance), while relativists would judge it as false from the new context of assessment, thereby motivating the retraction. In its current form, the retraction-based design cannot distinguish between EC and relativism. Thus, I conclude that the data from the reported experiments are consistent with both views, and further research is required to resolve this issue.

[5]See also Langford (2025) for SSI collapsing into impurist EC, trying to address some objections.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

Notably, the rivalry between contextualism and relativism is prominent in debates over predicates of personal taste (e.g., "tasty" or "fun"). Kneer (2021) exemplifies how experimental methods can effectively contrast contextualism and relativism. Kneer gathered data regarding the retraction of assertions involving predicates of personal taste, demonstrating that "the linguistic behavior of ordinary English speakers is consistent with contextualist predictions and inconsistent with the predictions of the most widely discussed form of truth relativism advocated by John MacFarlane" (Kneer 2021, p. 6455). Future research could investigate whether similar results, which uniquely favor either of the two positions, could be obtained regarding knowledge attributions. The retraction-based design may require further modifications to achieve this.

5.4. Some limitations

Finally, I would like to acknowledge several limitations of this paper. First, the views under consideration (EC, SSI, and CI) do not make rigid predictions about retraction behavior, leaving open questions about whether the observed retraction rates undermine their views in any notable way. Nonetheless, I have argued and demonstrated how the results do challenge CI and SSI. Second, the distinction between "not standing by" a knowledge claim and actively retracting it introduces interpretive ambiguity, as only the latter involves a deliberate action of taking back what was said. It seems plausible that one may not stand by one's claim without explicitly retracting it. In this paper, I adhered to the exact wording used by Dinges and Zakkou for two reasons: to maintain experimental comparability (I preserved the crucial retraction question verbatim) and to ensure vignettes sounded natural. In my opinion, "standing by" is easier to grasp than "retract," although future studies may benefit from exploring alternative prompts. Lastly, both experiments used various versions of the Bank scenario, limiting the generalizability of the findings. It is highly probable that the same pattern would hold for other stories because Dinges and Zakkou's initial experiments demonstrated the stakes effect in both Bank and Typo scenarios. Besides, it seems unlikely that the effect of that magnitude (medium to large across all experiments) would disappear when more scenarios were tested.

6. Conclusion

Dinges & Zakkou (2021) have found the robust stakes effect on knowledge attributions, this paper explored it further. Experiment 1 successfully replicated original experiment of Dinges and Zakkou, but also showed that these results hold for third-person cases. Experiment 2 addressed some worries regarding the design and implemented the modification that made retraction more realistic. I argued that taken altogether, the results pose a significant challenge to classic invariantism and may challenge SSI. However, the retraction-based design in its current form does not distinguish between the various explanations of why retraction has occurred, leaving the results open to both contextualist and relativist interpretations.[6]

[6]I am grateful to Alex Davies and Alex Wiegmann for their feedback on earlier drafts of this paper, and to Joachim Horvath for valuable discussions of the ideas developed here. I also thank the anonymous reviewer for their comments, which significantly improved the final version of the paper. This work was supported by the Eesti Teadusagentuur [PRG2151].

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

References

Brown J. (2006). 'Contextualism and Warranted Assertibility Manoeuvres.' *Philosophical Studies* **130**(3), 407–35.

Buckwalter W. (2010). 'Knowledge isn't Closed on Saturday: A Study in Ordinary Language.' *Review of Philosophy and Psychology* **1**(3), 395–406.

Cova F., Strickland B., Abatista A., Allard A., Andow J., Attie M., Beebe J., Berniūnas R., Boudesseul J., Colombo M., Cushman F., Diaz R., van Dongen N.N., Dranseika V., Earp B.D., Torres A.G., Hannikainen I., Hernández-Conde J.V., Hu W., … **Zhou X.** (2021). 'Estimating the Reproducibility of Experimental Philosophy.' *Review of Philosophy and Psychology* **12**(1), 9–44.

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

DeRose K. (2005). 'The Ordinary Language Basis for Contextualism, and the New Invariantism.' *Philosophical Quarterly* **55**(219), 172–98.

DeRose K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*. Oxford: Oxford University Press.

Dinges A. (2018). 'Anti-Intellectualism, Egocentrism and Bank Case Intuitions.' *Philosophical Studies* **175**(11), 2841–57.

Dinges A. (2023). 'Knowledge and Asymmetric Loss.' *Review of Philosophy and Psychology* **14**(3), 1055–76.

Dinges A. and Zakkou J. (2021). 'Much at Stake in Knowledge.' *Mind & Language* 729–49. https://doi.org/10.1111/mila.12300

Feltz A. and Zarpentine C. (2010). 'Do You Know More When It Matters Less?' *Philosophical Psychology* **23**(5), 683–706. https://doi.org/10.1080/09515089.2010.514572

Francis K.B. and Beaman C.P. (2023). 'The Role of Confidence in Knowledge Ascriptions: An Evidence-Seeking Approach.' *Synthese* **202**, 3. https://doi.org/10.1007/s11229-023-04236-w

Francis K.B., Beaman P. and Hansen N. (2019). 'Stakes, Scales, and Skepticism.' *Ergo, an Open Access Journal of Philosophy* **6**(16), 427–87.

Gerken M. (2017). *On Folk Epistemology: How We Think and Talk about Knowledge*. Oxford: Oxford University Press.

Grindrod J., Andow J. and Hansen N. (2019). 'Third-Person Knowledge Ascriptions: A Crucial Experiment for Contextualism.' *Mind and Language* **34**(2), 158–82.

Hansen N. and Chemla E. (2013). 'Experimenting on Contextualism.' *Mind & Language* **28,** 286–321. https://doi.org/10.1111/mila.12019

Hawthorne J. (2003). *Knowledge and Lotteries*. Oxford: Oxford University Press.

Kneer M. (2021). 'Predicates of Personal Taste: Empirical Data.' *Synthese* **199**(3–4), 6455–71.

Langford S. (2025). 'From SSI to Contextualism.' *Episteme*, 1–14. https://doi.org/10.1017/EPI.2024.49

MacFarlane J. (2014). *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

May J., Sinnott-Armstrong W., Hull J.G. et al. (2010). 'Practical Interests, Relevant Alternatives, and Knowledge Attributions: an Empirical Study.' *Review of Philosophy and Psychology* **1**, 265–273. https://doi.org/10.1007/s13164-009-0014-3

Nagel J. (2010). 'Epistemic Anxiety and Adaptive Invariantism.' *Philosophical Perspectives* **24**(1), 407–35.

Pinillos Á. (2012). 'Knowledge, Experiments, and Practical Interests.' In J. Brown and M. Gerken (eds), *Knowledge Ascriptions*, pp. 192–219. Oxford: Oxford University Press.

Pinillos Á. (2016). 'Experiments on Contextualism and Interest Relative Invariantism.' In J. Sytsma and W. Buckwalter (eds), *A Companion to Experimental Philosophy*, pp. 349–58. Chichester: Wiley Blackwell.

Pinillos Á. and Simpson S. (2014). 'Experimental Evidence in Support of Anti-Intellectualism about Knowledge.' In J. Beebe (ed), *Advances in Experimental Epistemology*, pp. 9–44. London: Bloomsbury Academic.

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* **59**(2), 392–408.

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), 224–47.

Schaffer J. and Knobe J. (2012). 'Contrastive Knowledge Surveyed.' *Nous* **46**(4), 675–708.

Sripada C.S. and Stanley J. (2012). 'Empirical Tests of Interest-Relative Invariantism.' *Episteme* **9**(1), 3–26.

Stanley J. (2005). *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Turri J. (2017). 'Epistemic Contextualism: An Idle Hypothesis.' *Australasian Journal of Philosophy* **95**(1), 141–56.

https://doi.org/10.1017/epi.2025.10060 Published online by Cambridge University Press

Wu S. (2023). 'Are Folks Purists or Pragmatic Encroachers? New Discoveries of Relation between Knowledge and Action from Experimental Philosophy.' *Episteme* **22**(1), 1–29.

Nikolai Shurakov is a PhD student specializing in philosophy of language and experimental philosophy. He uses empirical methods to investigate context sensitivity in language.

Cite this article: Shurakov N. (2025). "The Stakes Effect: New Evidence from a Retraction-Based Experimental Design." *Episteme* 1–22. https://doi.org/10.1017/epi.2025.10060