Synthese (2023) 202:30
https://doi.org/10.1007/s11229-023-04236-w

ORIGINAL RESEARCH

The role of confidence in knowledge ascriptions: an evidence-seeking approach

Kathryn B. Francis[1] · **C. Philip Beaman**[2]

Received: 28 September 2022 / Accepted: 17 June 2023 / Published online: 17 July 2023

Abstract

Two methods have been used in the investigation of the stakes-sensitivity of knowledge as it occurs in ordinary language: (a) asking participants about the truth or acceptability of knowledge ascriptions and (b) asking participants how much evidence someone needs to gather before they know that something is the case. This second, "evidence-seeking", method has reliably found effects of stakes-sensitivity while the method of asking about knowledge ascriptions has not. Consistent with this pattern, in Francis et al. (Ergo, 2019), we found evidence of scalar stakes effects using an evidence-seeking approach. Whilst we found this evidence across several cases using both negative ("don't know") and positive ("know") polarities, there remain questions about the directness of the relationship between stakes and knowledge ascriptions; it is possible that stakes are affecting knowledge by affecting the confidence of the attributor. For example, Bach (in: Peter & Preyer, 2005) has argued that knowledge attributions do not track truth attributions but rather thresholds for doxastic confidence. To investigate the role of confidence in knowledge ascriptions, we use our existing paradigm (Francis et al., 2019) but include measures of both participant and protagonist confidence. As far as we are aware, this is the first empirical investigation of the role of confidence in stakes effects on knowledge that incorporates an evidence-seeking approach using several scenarios. Overall, across both positive ("know") and negative ("don't know") polarity conditions, we find further evidence of a stakes effect on knowledge using an evidence-seeking paradigm. However, and importantly, we do not find evidence that changes in participant confidence partially or fully mediate the stakes effect on knowledge.

Keywords Stakes-sensitivity of knowledge · Confidence · Evidence-seeking · Knowledge ascriptions

✉ Kathryn B. Francis
k.francis@keele.ac.uk; kathrynbfrancis@gmail.com

[1] School of Psychology, Faculty of Natural Sciences, Keele University, Keele, UK

[2] School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

1 Stakes sensitivity of knowledge

Advocates of the stakes-sensitivity of knowledge hold that whether a subject (*s*) knows some proposition (p) can depend on how much is at stake for *s* in being either right or wrong about *p*. Depending on the view taken, the explanation for this varies. For example, *epistemic contextualists* hold that "to know" is sensitive to context and so truth conditions of knowledge ascriptions vary across conversational contexts (e.g., DeRose, 2009) while *interest-relative invariantists* see a unique epistemic role for practical factors (e.g., stakes) in knowledge ascriptions (e.g., Stanley, 2005). Opponents of the stakes-sensitivity of knowledge view on the other hand, hold that the truth-conditions of knowledge ascriptions are independent of practical stakes. Experimental work has attempted to test these perspectives among ordinary language users and has produced mixed findings with some failing to find evidence of a stakes effect (e.g., Buckwalter, 2010, 2014; Buckwalter & Schaffer, 2015; Feltz & Zarpentine, 2010; Francis et al., 2019; Rose et al., 2019) and others supporting the existence of stakes effects on knowledge (e.g., Buckwalter & Schaffer, 2015; Dinges & Zakkou, 2021; Francis et al., 2019; Pinillos, 2012; Pinillos & Simpson, 2014).

Experimental research in this area has largely incorporated two experimental paradigms[1] to test these effects: the *canonic* (traditional), or "evidence-fixed" paradigm versus the *evidence-seeking* paradigm. In the evidence-fixed paradigm, participants read a vignette describing a protagonist in a low or high stakes situation in which the evidence available to the protagonist is held constant across cases. Participants are then asked about the truth or acceptability of knowledge ascriptions concerning the protagonist. For example, participants might read the following scenario (Francis et al., 2019):

> Elaine is a medical researcher. Her task is to create a vaccine for a virus. Elaine has done this before, and she has a check list that specifies all of the steps she needs to take to make the vaccine. Elaine is following all of the steps correctly. Elaine's assistant has informed her that there is one human research participant who has volunteered to trial the vaccine before it is distributed more widely. If Elaine does not follow the steps correctly, it will produce an ineffective combination that when administered to the research participant will give them mild cold-like symptoms.

Following this, participants are asked to rate the extent to which they agree or disagree with the following statement:

> Elaine **knows** that she is making the vaccine correctly (*positive polarity*)

Or:

> Elaine **doesn't know** that she is making the vaccine correctly (*negative polarity*)

In the evidence-seeking paradigm on the other hand, evidence is no longer fixed in the vignette and participants are asked how much evidence the protagonist needs to

[1] Note that Dinges and Zakkou (2021) incorporated a third paradigm.

gather before they know that something is the case. For example, participants might read the same scenario as above but will be instead asked:

> How many times does Elaine need to consult her check list before she **knows** that she is making the vaccine correctly? (*positive polarity*)

Or:

> How many times can Elaine consult her check list and still **not know** that she is making the vaccine correctly (*negative polarity*)

This second, evidence-seeking method of investigating stakes-sensitivity has reliably found that participants offer higher evidence ratings in the high stakes cases compared to the low stakes cases (e.g., Francis et al., 2019; Pinillos, 2012). However, the evidence-fixed paradigm of asking about knowledge ascriptions has produced little evidence of differences between low stakes and high stakes cases (Buckwalter, 2010, 2014; Feltz & Zarpentine, 2010; Francis et al., 2019). For example in recent work, we investigated the scalar nature of stakes effects on knowledge ascriptions (Francis et al., 2019) extending empirical work beyond the use of binary paradigms (low versus high stakes). Incorporating both evidence-fixed and evidence-seeking approaches, we found evidence of a scalar stakes effect in the latter paradigm using both positive ("know") and negative ("don't know") polarities but no evidence of this effect in the evidence-fixed paradigm.

2 The role of confidence in stakes sensitivity

Evidence-seeking approaches have been criticised, however. Sceptics have questioned whether responses to these prompts truly concern the nature of knowledge. For example, Gerken (2017) argues that the apparent stakes effect found in these paradigms results from an "Epistemic Actionability-Proxy" whereby the participant is considering how much evidence is needed by the protagonist *to act* rather than *to know.* Further, Bach (2005) has argued that knowledge attributions in these experiments do not track truth attributions but rather thresholds for doxastic confidence. In high stakes cases, the threshold for confidence is raised and so the attributor will demand more evidence, but this does not mean that *knowledge* requires the same level of evidence. Bach argues that the standard for the truth of a knowledge attribution stays the same in these cases when compared to lower stakes cases. For example, consider the following high-stakes case (Francis et al., 2019):

> Tracy is taking part in a game show that involves answering general knowledge trivia questions. The game show host has asked Tracy, "What is the capital of Tanzania?". Tracy has recently read a list of the most obscure[2] world capitals and the city "Dodoma" pops into her head. In fact, Tracy is right: the capital of

[2] This example was created using a list of commonly mistaken world capitals. In the present study and previous studies using these materials, the participant samples have been collected from the US and UK and we believed that this capital would be unknown to most participants.

> Tanzania is Dodoma. As this is the final round of the game show, $1,000,000 is at stake: answering this question correctly will result in Tracy winning $1,000,000 and answering incorrectly will result in her losing $1,000,000.
>
> How many minutes does Tracy need to spend considering her answer before she knows that the capital of Tanzania is Dodoma?

In this case, we might respond by stating that Tracy needs to spend several minutes considering her answer before she *knows* the answer. However, if we lower the stakes to $1 instead, a majority of participants would likely decrease the number of minutes that Tracy needs to consider her answer before she *knows* it (see Francis et al., 2019). While a contextualist might argue that this reflects a true stakes effect, opponents might argue that the protagonist's and indeed the participant's own doxastic situation drives the effect (Bach, 2010). In this case, the attributor may not confidently *believe* that Tracy knows the answer and will subsequently want more evidence before asserting that she *knows* the answer. This view that raising the stakes can weaken the ascriber's own confidence in a proposition (e.g., Bach, 2010; Nagel, 2008) has been supported in some empirical studies where raising the stakes affects the level of confidence in attributions of knowledge but not knowledge attributions directly (e.g., May et al. (2010)[3]). As such, there remain questions about the directness of the relationship between stakes and knowledge ascriptions given this possibility that stakes are affecting knowledge by affecting the *confidence* of the attributor.

3 Experiment: stakes, confidence, and knowledge

To investigate the role of confidence in knowledge ascriptions, we use our existing paradigm (Francis et al., 2019) but include measures of both participant *and* protagonist (subject) confidence. As far as we are aware, this is the first empirical investigation of the role of confidence in stakes effects on knowledge that incorporates an evidence-seeking approach using several scenarios.

3.1 Hypothesis

If knowledge is sensitive to stakes, we should observe differences in responses to low versus high stakes cases and across polarities (know versus don't know) as seen in existing research (see Francis et al., 2019). However, if this effect is driven by thresholds for doxastic confidence rather than truth attributions, this effect will be mediated by the attributor's confidence.

[3] Although note that May et al. (2010) did not explicitly measure attributor confidence and instead assumed that degree of agreement reflected level of confidence in the truth of that claim.

3.2 Participants

Participants (N = 249) were recruited from MTurk and paid $1.75 each for participating.[4] We ensured that the Human Intelligence Task (HIT) approval for HITs was 95% and the requirement for number of HITs approved for each participant was set at 50. The study took approximately 10 min to complete (data were collected in 2019). This research received ethical approval from the University of Reading, UK and informed consent was obtained from all participants. The sample size was reduced after removing incomplete responses (N = 30) and following ex post procedures to identify suspicious or low-quality responses[5] in the dataset (N = 9). As an additional screening procedure, participants who responded incorrectly to one of two control prompts were removed from further data analysis (see Francis et al., 2019) (N = 23), leaving a final sample of 187 participants (69 females, 118 males) between 19 and 66 years old (M = 34.70 years, SD = 9.92 years).

3.3 Measures and procedure

Participants were randomly assigned to the positive polarity condition (N = 97) or the negative polarity condition (N = 90) following the paradigm developed in Francis et al. (2019). The highest- and lowest-stakes versions of six scenarios developed in Francis et al. (2019) were presented to each participant. These scenarios involved the manipulation of different types of stakes (lives; physical injury; embarrassment; money; damage to objects of personal value) (see OSF project page for all scenarios and low versus high stakes versions of each).[6] For example, the 'paramedic' scenario involves changing the severity of an injury that will be left untreated (stakes) should a paramedic make a wrong turn on the way to an accident:

> *Low Stakes*
> Megan, a paramedic, has just been called to an accident involving a collision between two cars. Megan is driving an ambulance carrying her team of paramedics to the scene of the accident. Megan is familiar with the surrounding

[4] A power analysis using the ShinyApp Monte Carlo Power Analysis for Indirect Effects (Schoemann et al., 2017) was performed using a 95% confidence level, 80–90% target power, assuming a medium stakes effect as seen in previous research using the evidence-seeking paradigm and a small-medium effect between x and m, and m and y in mediation analysis, and based on 1000 replications. This determined that a sample size of N = 180 to N = 200 would provide 85%-90% certainty of detecting a true effect. N = 249 participants were recruited to account for attrition via incomplete responses, low-quality responses, and failed control prompts.

[5] Screening procedures were performed by identifying identical GPS locations with unique IP addresses, determining whether IP addresses derived from an Internet Service Provider (ISP) or data center, and evaluating open-ended responses against a set of criteria (for full details regarding this procedure see Dennis et al., 2018).

[6] Note that hypothetical versions of high stakes scenarios are not likely to produce responses that would occur in real (or more realistic) versions of those situations where the stakes are tangible for the participant (e.g., FeldmanHall et al., 2012; Francis et al., 2016, 2017). However, there is evidence that participants will often project their own cognitive state on to the protagonist in a scenario and will treat their concerns as shared (e.g., Dinges 2018; Nagel, 2008).

area, she has GPS on her phone that she can check if necessary, and she is traveling on the right route to get to the accident. Over the radio, Megan is told that there is **one person at the scene of the accident with a broken arm, which is not life-threatening.** If Megan makes a wrong turn on the way to the accident, the injured person **will be inconvenienced, but nothing terrible will happen**. Megan thinks to herself, "I am completely confident that I can get to the accident without taking a wrong turn".

In the above 'low' stakes version of the scenario, one individual will be inconvenienced if the protagonist arrives late. These stakes are raised significantly in the high stakes version of the scenario:

High Stakes

…Over the radio, Ottoline is told that one of the trucks was carrying flammable fuel and **several vehicles, including a school bus carrying 50 children, are on fire**. If Ottoline makes a wrong turn on the way to the accident, she will be too late and **the children will die**. Ottoline thinks to herself, "I am completely confident that I can get to the accident without taking a wrong turn".

Importantly, the final sentence of each scenario was added to the current study stating that *s* thought or said aloud that they were confident that *p*, to ensure that protagonist confidence was held constant across scenarios regardless of what was at stake. After reading each scenario, participants were asked to respond to three prompts:

Evidence-seeking prompt. An evidence seeking prompt (positive or negative) asking how much evidence would need to be collected in order for the protagonist to *know* that *p*. For example, in the low stakes paramedic case, participants in the positive polarity condition were asked:

What is the **minimum** number of times Megan needs to check her GPS before she **knows** that she will make it to the accident without taking a wrong turn?
_____times

and participants in the negative polarity condition were asked:

What is the **maximum** number of times Megan can check her GPS and **not know** that she will make it to the accident without taking a wrong turn?
_____times

As in Francis et al. (2019), after reading the prompts, participants were asked to respond as follows:

Positive polarity
enter a whole number: 1, 2, 3 . . . etc. if you think Megan **knows** without having to check, write "0". If you think Megan will never know no matter how many times she checks, write "never"
Negative polarity
enter a whole number: 1, 2, 3 . . . etc. if you think Megan will **never know** no matter how many times she checks, write "never"

Participant confidence prompt. After the evidence-seeking prompt and to measure participant confidence, participants were asked how confident *they* were that *s* would need to gather that much evidence in order to know that *p* on a scale from 1 (Not at all confident) to 5 (Very confident). The wording of these varied depending on polarity condition:

> *Positive polarity*
> How confident are **you** that Megan needs to check a minimum of that many times in order to **know** that she will make it to the accident without taking a wrong turn?
> *Negative polarity*
> How confident are **you** that Megan can check a maximum of that many times and **not know** that she will make it to the accident without taking a wrong turn?

Protagonist confidence prompt.[7] Participants were also asked to rate how confident they thought the protagonist was on a scale from 1 (Not at all confident) to 5 (Very confident). For example, participants were asked "How confident is Megan that she will make it to the accident without taking a wrong turn?". The same question was asked regardless of polarity condition assignment.[8]

Prompts were arranged in two blocks after each scenario so that the evidence-seeking prompt was followed by the protagonist confidence prompt (block one). The participant confidence prompt was presented separately (block two). Block one (evidence-seeking prompt followed by the participant confidence prompt) and block two (the protagonist confidence prompt) were then presently randomly after each scenario to avoid order effects. For example, having read a scenario, participants may have been asked to respond to the participant confidence prompt first (block two) followed by the evidence-seeking and protagonist confidence prompts next (block one) or vice versa. After responding to all six scenarios, an average (1) evidence-seeking score, (2) participant confidence score, and (3) protagonist confidence score was calculated for each participant and included in further analyses. All measures used in the present study and all data collected are available on the project's OSF page: https://osf.io/n8x7q/.

4 Results

One participant reported extreme values in the negative polarity condition (see OSF project page). To account for this and the subsequent skew in the data, this was removed prior to further analysis. Across both positive and negative polarity conditions, participants had higher average evidence-seeking scores for high stakes cases compared to

[7] We chose to avoid using the term "subject confidence" here given the regular use of "subject" *and* "participant" to refer to the same individuals in psychological research.

[8] We assess the role of confidence in knowledge ascriptions under the following assumption; that the ascriber projects something of their own hypothetical doubt on to the protagonist, which affects the amount of evidence they think the protagonist should get, even if the ascriber is told that *p* is true. Thus, the methodological approach outlined here, allows the measurement of the doxastic status of the ascriber (participant) while holding the doxastic status of the ascribee (protagonist). This was the main aim of the study.

Table 1 Means and standard deviations (SDs) of evidence-seeking, participant confidence, and protagonist confidence scores across stakes and polarity conditions

	Protagonist confidence M(SD)		Knowledge (evidence-seeking) M(SD)		Participant confidence M(SD)	
	Low	High	Low	High	Low	High
Positive ("Know")	4.51 (0.57)	4.56 (0.59)	2.34 (2.01)	4.19 (3.72)	4.11 (0.68)	4.01 (0.65)
Negative ("Don't Know")	4.58 (0.58)	4.54 (0.62)	3.31 (3.95)	2.43 (3.02)	3.76 (0.93)	3.69 (0.92)

low stakes cases, consistent with previous evidence-seeking approaches (see Table 1). Participants also reported that they were less confident that *s* would need to gather that much evidence in order to know that *p* (or negative equivalent) in high stakes cases compared to low stakes cases (see Table 1). Participants' rating of subject/protagonist confidence appeared to be similar across low and high stakes cases as expected (see Table 1).

4.1 Protagonist confidence

Prior to investigating the role of protagonist confidence on knowledge, we analysed protagonist confidence to ensure that it was constant across all scenarios and stakes (see Sect. 3.3). Analyses comparing participants' judgments about each subject's confidence showed no significant differences between low stakes and high stakes cases across the majority of scenarios in the positive polarity group ('know') and negative polarity ('don't know') ($ps > 0.106$). However, in the Vaccine scenario, there was a significant difference between participants' judgments about each subject's confidence between cases for the positive polarity group, with participants stating that the subject was more confident in the high stakes case, $t(96) = -4.68$, $p < 0.001$. Additionally, in the Personal Value scenario, there was a significant difference between participants' judgments about each subject's confidence between cases for the negative polarity group, with participants stating that the subject was more confident in the low stakes case, $t(88) = 2.50$, $p = 0.014$ (see Fig. 1).

4.2 Mediation analyses

Mediation analysis allows us to determine whether there is a sequence of relations whereby an antecedent variable (x) influences a mediating variable (m), which then affects an outcome variable (y). Essentially, a mediating variable transmits the effect of an independent or predictor variable on a dependent or outcome variable. This analysis is appropriate in the given context as we are able to identify and examine whether there is a mediating role of confidence in the relationship between stakes

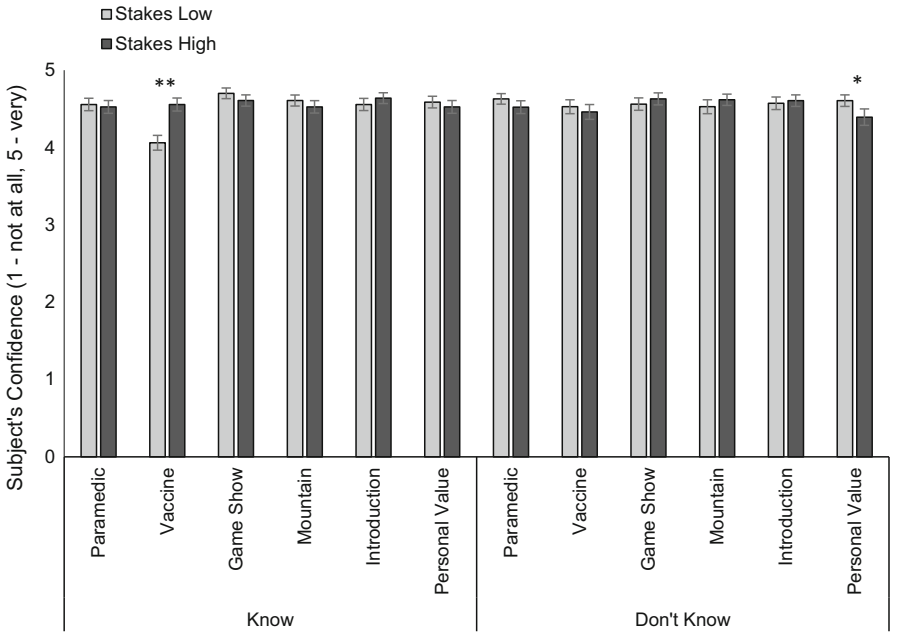

Fig. 1 Mean protagonist confidence ratings across scenarios, stakes, and polarity conditions. Protagonist confidence was constant across the majority of cases. It significantly differed in the Vaccine scenario in the positive polarity and in the Personal Value scenario in the negative polarity. Error bars represent ± 1 Standard Error (SE). $^{**}p < 0.001$, $^{*}p < 0.05$

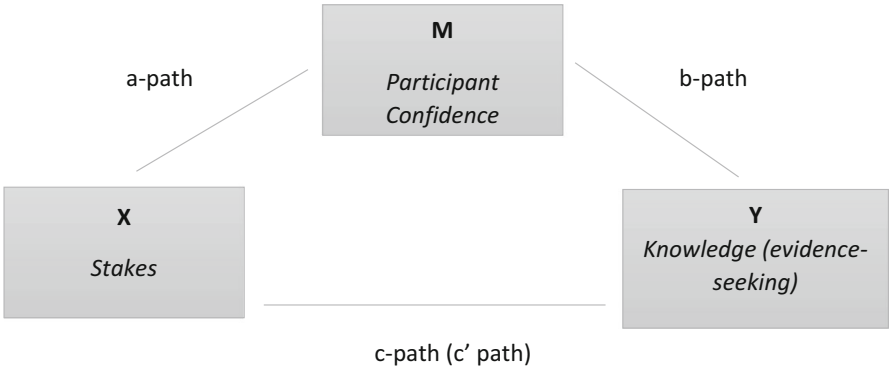

Fig. 2 Model of mediation. Stakes (X) predict changes in knowledge (Y) and this is mediated by participant confidence (M)

and knowledge. As such and to investigate this, we conducted a mediation analysis to determine if participant (attributor) confidence (M) mediated the effect of stakes (X) on knowledge (Y) (see Fig. 2).

Fig. 3 Average evidence-seeking (knowledge) scores and participant's confidence in both low and high stakes cases in the positive polarity condition. There was a stakes effect on knowledge and a smaller effect on participant confidence. Error bars represent ± 1 SE

For mediation analyses, we used the R-MEMORE package (and MEMORE macro for replication) which supports within-subjects statistical mediation analysis (Montoya & Hayes, 2017). Subsequently, we assessed the following paths in the model to establish mediation[9] (see Fig. 2):

(1) Do stakes (X) predict knowledge (Y)? *Path c*
(2) Do stakes (X) predict participant confidence (M)? *Path a*
(3) Do stakes (X) and participant confidence (M) together predict knowledge (Y)?:
    a. Does participant confidence (M) predict knowledge (Y)? *Path b*
    b. Do stakes (X) no longer predict knowledge (Y) or is this effect lessened when accounting for participant confidence (M)? *Path c'*

4.3 Positive polarity

For the positive polarity condition, mediation analyses found a statistically significant effect of stakes on knowledge via the evidence-seeking path (path *c*), $t(96) = -5.90$, $p < 0.001$, 95% CI [− 2.47, − 1.23] and a smaller but statistically significant effect of stakes on participant confidence (path *a*), $t(96) = 2.17$, $p = 0.032$, 95% CI [0.001, 0.21] (see Fig. 3). There was no statistically significant effect of participant confidence on knowledge (path *b*), 95% CI [− 1.58, 0.99]. The direct effect of stakes on knowledge (evidence-seeking) (path *c'*) was statistically significant, $t(96) = -5.62$, $p < 0.001$, 95% CI [− 2.46, − 1.18]. The indirect effect of stakes on knowledge through

[9] Note that these steps are listed according to the causal steps traditional to mediation analysis and that these do not indicate that sequential analytical steps were performed using the MEMORE package. They are simply included to explain the paths in the conceptual model (Fig. 2).

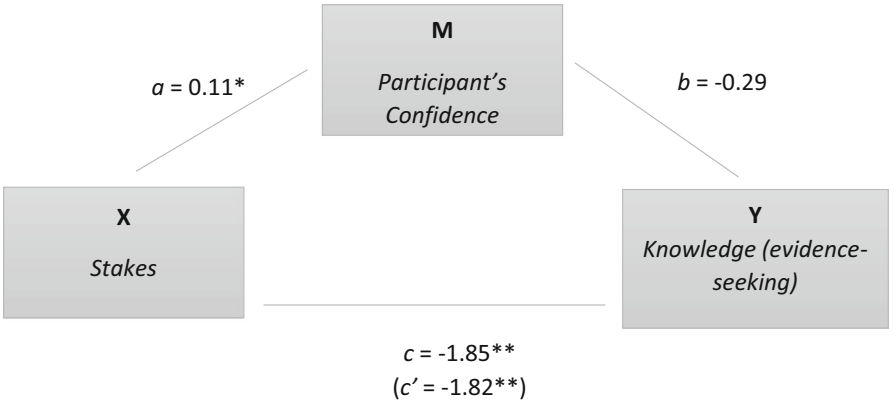

Fig. 4 Stakes (X) significantly predicted changes in knowledge (Y) and participant confidence. Participant confidence did not affect knowledge. There was no evidence of a mediation effect

participant confidence was not statistically significant, 95% CI [− 0.22, 0.11]. This indicates that while stakes had effects on both participant confidence and knowledge (evidence-seeking scores), the effect of stakes on knowledge was not found to be partially nor fully mediated by the effect of participant confidence (see Fig. 4).

Given that protagonist confidence was statistically significantly different between the low and high stakes cases in the Vaccine scenario in the positive polarity condition (see Sect. 4.1), analyses were repeated with this case removed. All effects were replicated and again indicated that while stakes had effects on both participant confidence and knowledge (evidence-seeking scores), the effect of stakes on knowledge was neither partially nor fully mediated by the effect of participant's confidence.

4.3.1 Negative polarity

Mediation analyses found a statistically significant effect of stakes on knowledge (evidence-seeking) (path *c*), $t(88) = 2.25$, $p = 0.027$, 95% CI [0.10, 1.65] (see Fig. 5). There was no statistically significant effect of stakes on the participant confidence (path *a*), 95% CI [− 0.02, 0.15] and no statistically significant effect of participant confidence on knowledge (path *b*), 95% CI [− 2.05, 1.83]. The direct effect of stakes on knowledge (evidence-seeking) (path *c'*) was statistically significant, $t(88) = 2.22$, $p = 0.029$, 95% CI [0.09, 1.67] and the indirect effect of stakes on knowledge through participant confidence was not statistically significant, 95% CI [-0.13, 0.12]. This indicated a stakes effect on knowledge (evidence-seeking scores) only. There was no effect of stakes on participant's confidence and the effect on knowledge was neither partially nor fully mediated by the effect of participant confidence (see Fig. 6).

Given that protagonist confidence was statistically significantly different between the low and high stakes cases in the Personal Value scenario in the negative polarity condition (see Sect. 4.1), analyses were repeated with this case removed. All effects were replicated indicating that stakes had an effect on knowledge (evidence-seeking scores) but not participant's confidence. There was no evidence that the effect of

Fig. 5 Average evidence-seeking (knowledge) scores and participant's confidence in both low and high stakes cases in the negative polarity condition. There was a stakes effect on knowledge only. Error bars represent ± 1 SE

stakes on knowledge was partially or fully mediated by the effect of participant's confidence.[10]

5 Discussion

To investigate the role of confidence in knowledge ascriptions, we used an adapted evidence-seeking paradigm (Francis et al., 2019) to examine stakes effects on knowledge ascriptions and the potentially mediating effect of participant's confidence. Overall, across both positive ("know") and negative ("don't know") polarity conditions, we find further evidence of a stakes effect on knowledge using an evidence-seeking paradigm. Participants' ratings of the confidence of the protagonist in their knowledge was largely unaffected by any of the manipulations (either polarity or stakes). In the positive polarity condition only, we find evidence of a small stakes effect on participants' confidence in the protagonists' knowledge. However, and importantly, we do not find evidence that changes in participant confidence partially or fully mediate the stakes effect on knowledge.

Existing research has criticised both evidence-fixed and evidence-seeking paradigms, asking whether responses to these prompts truly concern the nature of knowledge. Given that Bach (2005) has argued that knowledge attributions in these

[10] Power analyses were originally performed on the assumption that both negative and positive polarity conditions would be analysed together. However, given the substantial differences between the prompts, the data were not merged for mediation analysis. If datasets are combined following power analyses, the effect of stakes on participant confidence and on knowledge (evidence-seeking) is replicated. There is no effect of participant confidence on knowledge and no evidence of partial or full mediation as in the previous analyses.

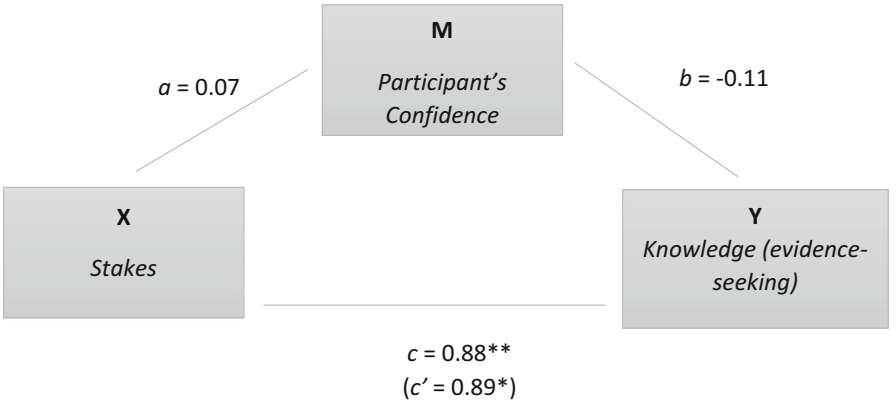

Fig. 6 Stakes (X) significantly predicted changes in knowledge (Y) only. There was no evidence of a mediation effect

experiments do not track truth attributions and instead track thresholds for doxastic confidence, we might have expected participant confidence to mediate (even fully) the stakes effects on knowledge found in these experiments. Here, we find little evidence to support this. In the positive polarity condition we did find evidence that, as the stakes increase, the threshold for confidence is raised. However, we do not find evidence to support the argument that this then explains why the attributor will demand more evidence as the stakes increase. The experimental work presented here provides initial evidence that the standard for the truth of a knowledge attribution does vary across cases when using the evidence-seeking approach, and we do not find evidence that suggests this is related to changes in doxastic confidence. In the negative polarity condition, we find evidence of a stakes effect on knowledge in the evidence-seeking paradigm and no further effects of stakes on participant confidence. This provides further evidence against Bach's claim that participant confidence is playing a mediating role here.

The data presented here represent the first attempt to empirically examine the possibility that evidence-seeking investigations of knowledge-ascription track confidence in a participant's knowledge ascription rather than a protagonist's knowledge per se. We do not find evidence that the stakes effect on knowledge (using the evidence-seeking approach) is driven, either wholly or in part, by a confidence judgement.[11] The conclusion that there is no such effect of confidence needs to be viewed with caution on statistical and methodological grounds, however.

On statistical grounds, the failure to find an effect cannot be taken as evidence for no effect—at its most fundamental level, our power analysis indicates that we have an 85–90% probability of correctly rejecting a null hypothesis of no effect in our overall data given an alternative hypothesis that an effect of a certain size or larger is correct.

[11] We do not interpret or ground the present findings within invariantist *or* epistemic contextualist theories as the current methodological approach does not allow these views to be teased apart. Existing research has argued for the inclusion of third-person (as opposed to first-person) cases which do allow these accounts to be empirically tested and compared (see Grindrod et al., 2019).

The probability of correctly rejecting a null hypothesis in our data given a smaller effect of confidence is, of course, lower than this—how much lower is dependent upon how small an effect of confidence we are willing to consider as large enough to (wholly or in part) mediate the scalar effect of stakes which we did observe. Clearly, further replications of our results are required to support our interim conclusion statistically. On the more interesting methodological grounds, our study employed a "snapshot" method of obtaining participants' confidence judgements. Alternative and more dynamic methods—for example, incrementing the number of times the protagonist seeks evidence (e.g., Megan checks her GPS) and asking for repeated confidence judgements following each increment—might give different results. While we have no a priori reasons to imagine this to be the case, we note that any single approach necessarily gives data relevant only to the way in which the question has been framed (and interpreted by the research participants) in that instance. There is a long history of different approaches returning different results and hence different interpretations in experimental psychology and this pattern looks set to be repeated in experimental philosophy (for example, the different results obtained in evidence-seeking versus evidence-fixed approaches) and for this reason multiple approaches need to be attempted to provide a more complete picture. By this same token, however, even if other approaches give different or conflicting results, the data reported here that stakes effects in knowledge do not seem to be mediated by confidence, would still stand in need of explanation.

Acknowledgements Thanks to Dr Sarah Fisher for very helpful comments on the draft of this paper. Funding from the Leverhulme Trust Research Project Grant RPG-2016-193 made this research possible.

Data availability All measures used in the present study and all data collected are available on the project's OSF page: https://osf.io/n8x7q/

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Bach, K. (2005). The emperor's new 'knows.' In G. Peter & G. Preyer (Eds.), *Contextualism in philosophy* (pp. 51–89). Clarendon Press.

Bach, K. (2010). Knowledge in and out of context. In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.), *Knowledge and skepticism* (p. 105). The MIT Press.

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. (2014). The mystery of stakes and error in ascriber intuitions. In J. R. Beebe (Ed.), *Advances in experimental epistemology* (pp. 145–173). Bloomsbury Publishing.
Buckwalter, W., & Schaffer, J. (2015). Knowledge, stakes, and mistakes. *Noûs, 49*(2), 201–234.
DeRose, K. (2009). *The case for contextualism*. Oxford University Press.
Dinges, A. (2018). Anti-intellectualism, egocentrism and bank case intuitions. *Philosophical Studies, 175*, 2841–2857.
Dinges, A., & Zakkou, J. (2021). Much at stake in knowledge. *Mind & Language, 36*(5), 729–749. https://doi.org/10.1111/mila.12300
FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition, 123*(3), 434–441.
Feltz, A., & 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., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PloS one, 11*(10).
Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports, 7*(1).
Francis, K. B., Beaman, P., & Hansen, N. (2019). Stakes, scales, and skepticism. *Ergo: An Open Access Journal of Philosophy*. https://doi.org/10.3998/ergo.12405314.0006.016
Gerken, M. (2017). *On folk epistemology: How we think and talk about knowledge*. Oxford University Press.
Grindrod, J., Andow, J., & Hansen, N. (2019). Third-person knowledge ascriptions: A crucial experiment for contextualism. *Mind & Language, 34*(2), 158–182.
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*, 265–273.
Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*(1), 6.
Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy, 86*(2), 279–294.
Pinillos, N. A. (2012). Knowledge, experiments and practical interests. In J. Brown & M. Gerken (Eds.), *Knowledge ascriptions* (pp. 192–219). Oxford University Press.
Pinillos, N. A., & Simpson, S. (2014). Experimental evidence supporting anti-intellectualism about knowledge. In J. R. Beebe (Ed.), *Advances in experimental epistemology* (pp. 9–44). Bloomsbury Publishing.
Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., et al. (2019). Nothing at stake in knowledge. *Noûs, 53*(1), 224–247. https://doi.org/10.1111/nous.12211
Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379–386.
Stanley, J. (2005). *Knowledge and practical interests*. Clarendon Press.