DOI: 10.1111/nous.12515

ARTICLE

Noûs

A puzzle about knowledge ascriptions

Brian Porter[1] | Kelli Barr[1] | Abdellatif Bencherifa[2] | Wesley Buckwalter[3] | Yasuo Deguchi[4] | Emanuele Fabiano[5,6] | Takaaki Hashimoto[7] | Julia Halamova[8] | Joshua Homan[9] | Kaori Karasawa[10] | Martin Kanovsky[11] | Hackjin Kim[12] | Jordan Kiper[13] | Minha Lee[14] | Xiaofei Liu[15] | Veli Mitova[16] | Rukmini Bhaya[17] | Ljiljana Pantovic[18] | Pablo Quintanilla[6] | Josien Reijer[16] | Pedro Romero[19] | Purmina Singh[17] | Salma Tber[20] | Daniel Wilkenfeld[21] | Stephen Stich[22] | Clark Barrett[23] | Edouard Machery[1,16]

Correspondence
Edouard Machery, University of Pittsburgh, 1117 CL Pittsburgh PA 15260 USA.
Email: machery@pitt.edu

Funding information
John Templeton Foundation

Abstract

Philosophers have argued that stakes affect knowledge: a given amount of evidence may suffice for knowledge if the stakes are low, but not if the stakes are high. By contrast, empirical work on the influence of stakes on ordinary knowledge ascriptions has been divided along methodological lines: "evidence-fixed" prompts rarely find stakes effects, while "evidence-seeking" prompts consistently find them. We present a cross-cultural study using *both* evidence-fixed and evidence-seeking prompts with a diverse sample of 17 populations in 11 countries, speaking 14 languages. Our study is the first to use an evidence-seeking prompt cross-culturally, and includes several previously untested populations (including indigenous populations). Across cultures, we *do not find* evidence of a stakes effect with our evidence-fixed prompt, but *do* with our evidence-seeking prompt.

> We argue that the divergent results reveal a tension within folk epistemology: people's beliefs about when it is appropriate to ascribe knowledge differ significantly from their actual practice in ascribing knowledge.

1 | INTRODUCTION

Is knowledge sensitive to practical considerations in addition to evidential considerations? Many philosophers believe that it is.[1] In particular, many have been convinced by bank cases that knowledge ascriptions are sensitive to *stakes* (DeRose 2005): a person has some evidence that their bank will be open on a given day; in one case, nothing bad will happen if the person is wrong; in another case, it will be bad if they turn out to be wrong. Philosophers often judge that the person knows that the bank will be open in the "low stakes" case, but not in the "high stakes" case. They therefore conclude that knowledge ascriptions and possibly knowledge itself are in some way sensitive to practical factors like stakes.

In contrast to philosophers' consensus, empirical results examining whether knowledge ascription is influenced by stakes have been mixed. Some studies presenting bank cases to non-philosophers find evidence that knowledge ascriptions are sensitive to changes in stakes;[2] others fail to find evidence.[3] While most studies have been limited to English-speaking participants, Rose et al. (2019) found that the stakes involved did not notably alter knowledge ascriptions across languages and cultures, with few exceptions.[4]

Some of the variation in results is explained by the use of two different methods: evidence-*fixed*, and evidence-*seeking*. In the former, participants are told how much evidence the character in the cases has gathered and are asked to assign knowledge to them; in the latter (introduced by Pinillos, 2012), participants are asked how much evidence the character in the vignette should obtain in order to know some proposition. Almost all the evidence against stakes effects has come from evidence-fixed methods; evidence-seeking prompts consistently find stakes effects.[5] Francis and Beaman (2023) argue that in contrast to evidence-seeking methods, evidence-fixed methods

[1] See, for example, Cohen (1999), DeRose (1992, 1995, 2005, 2009), Fantl and McGrath (2002, 2009), Jackson (2021), Nagel (2008), Rysview (2001), and Stanley (2005).

[2] See, for example, Buckwalter and Schaffer (2015), Dinges (2023), Dinges and Zakkou (2021), Francis et al. (2019), Francis and Beaman (2023), Pinillos (2012), Pinillos and Simpson (2014), Sripada and Stanley (2012), and Turri et al. (2016).

[3] See, for example, Buckwalter (2010, 2014), Buckwalter and Schaffer (2015), Feltz and Zarpentine (2010), Francis et al. (2019), Rose et al. (2019), and Turri et al. (2016).

[4] Out of 19 populations sampled, only 3 had any statistically significant effect: Spain, Japan, and the UK.

[5] In most cases, evidence-fixed prompts find little to no evidence of a stakes effect (Buckwalter 2010, 2014; Buckwalter and Schaffer 2015; Feltz and Zarpentine 2010; Francis et al. 2019; and Rose et al. 2019). But there are notable exceptions: Sripada and Stanley (2012) find a small stakes effect, although the effect does not replicate in Francis et al. (2019); Francis and Beaman (2023) find a stakes effect with their evidence-fixed prompt; and Pinillos (forthcoming) finds a stakes effect using a modified evidence-fixed prompt. Turri et al. (2016) find evidence of a stakes effect in one of their two experiments, but argue that the stakes effect is only indirectly on knowledge; stakes affect judgements about *actionability*, which in turn affect knowledge judgements.

have produced inconsistent results, and that this inconsistency may be evidence that this method is flawed.[6]

We interpret the disagreements between the two methods differently: we speculate that the different results between evidence-fixed and evidence-seeking methods reflect a tension between people's epistemic beliefs and practices. The relevance of stakes varies depending on whether people are asked to attribute knowledge or are asked to reflect on the conditions for attributing knowledge. To examine this tension, our study asked participants both to assign (or decline to assign) knowledge to someone having a given body of evidence and to decide how much evidence would be needed to know something. To address the concerns raised by Jackson (2021) and Francis et al. (2019), we also use a more realistic scenario (particularly for small-scale societies), involving a dam on a river, and participants read a "high stakes" scenario in which the stakes are life or death.

Finally, this study extends earlier cross-cultural research such as Rose et al. (2019) by considering a broader and more diverse participant base. Our sample includes individuals from 17 distinct populations across 11 countries, speaking 14 languages. Prior empirical work on stakes-sensitivity has focused primarily on highly-educated and urban populations; this study incorporates a wider range of participants, with an emphasis on including populations—such as indigenous communities—that have been underrepresented in prior work.

2 | METHODS

This study was part of a larger study on knowledge and understanding, carried out under the Geography of Philosophy project (www.geographyofphilosophy.com); the preregistration for this study can be found at https://osf.io/ym45u.

Data was collected from convenience samples across 17 sites in 11 countries, using the online survey platform Qualtrics at some sites, and paper and pencil at others (see Table 1 for details). A total of 5,573 participants were surveyed, with a minimum of 80 participants in each population.[7]

Participants were randomly and evenly assigned to one of four conditions, based on a 2 × 2 design (Evidence: weak vs. strong; Stakes: low vs. high). Each group saw a different continuation of the following vignette:

> [name]'s[8] kids want to go swimming in a local river. The water has slightly risen due to several days of light rain. So he decides to build a dam. He stacks a number of logs in the river to hold the water back and create a shallow pool where his kids can swim.

Participants were either told that the protagonist checked the logs once, or that he checked multiple times:

---

[6] Different theories have been offered to explain what exactly the flaw is. Jackson (2021) blames the failure to find evidence for the stakes-sensitivity of knowledge attributions with evidence-fixed prompts on the unrealistic nature of the scenarios used in many of these studies. Francis et al. (2019) argue that stakes are scalar—whether stakes are "high" or "low" is a matter of degree—and that at least some experiments may have failed to find evidence of stakes-sensitivity simply because the "high stakes" and "low stakes" scenarios used in those studies were too close together on the scale.

[7] All populations have at least 200 participants except for the Meitei population in India, which has 83.

[8] Names varied across populations. For each population, appropriate names were chosen by local members of the population who were native speakers of the target language.

TABLE 1 Demographic Characteristics

Sample	Language	Students	Age[a] mean	range	sd	Gender Percentage M/F/other	*N*	Passed Check	Passed Check Percentage
China	Mandarin	both	22.75	18–52	5.53	43.9/55.8/0.3	416	261	62.7
Russia	Russian	non-students	25.97	18–75	10.28	40/60/0	739	558	75.5
Slovakia	Slovak	both	26.28	18–74	9.07	54.4/44.9/0.7	713	290	40.7
Ecuador	Spanish	students	21.39	18–42	3.21	32.9/66.7/0.5	390	208	53.3
India	Hindi	both	29.98	18–62	10.23	45.3/53.8/0.9	200	100	50
India	Meitei	non-students	23.26	18–45	6.26	47.6/50/2.4	86	42	48.8
Japan	Japanese	both	30.31	19–62	10.76	53.3/46.3/0.4	544	445	81.8
Morocco	Arabic	both	32.96	18–76	11.81	44.4/55.6/0	551	323	58.6
Peru	Shipibo	non-students	34.68	18–83	13.77	53.9/46.1/0	200	102	51
Peru	Spanish	students	22.75	18–49	3.28	43.1/56.9/0	286	216	75.5
South Africa	Afrikaans	both	41.63	18–73	12.85	34.3/65.7/0	270	178	65.9
South Africa	Sepedi	non-students	29.33	18–57	7.88	36.5/62/1.5	258	122	47.3
South Africa	isiZulu	non-students	31.05	18–76	9.84	27.9/69.9/2.2	268	130	48.5
South Korea	Korean	both	35.40	18–68	10.37	41.5/58.5/0	256	193	75.4
United States	English	students	35.85	19–73	10.52	55.5/44.5/0	396	348	87.9

[a] Participants who reported an age under 18 were excluded from the study; participants who input non-numeric text or a number greater than 100 were excluded from our age calculations here, but were otherwise included in the study.

Weak evidence: Once he has finished building the dam, he checks the logs once to make sure the dam is secure.

Strong evidence: Once he has finished building the dam, he checks the logs several times to make sure the dam is secure.[9]

Participants were also either told that it was very important that the dam was secure, or that it was not at all important:

High stakes: It is very important that the dam is secure. Due to the high water level, if it breaks, his kids will be swept away by the water.

Low stakes: It is not at all important that the dam is secure. Due to the low water level, if it breaks, the water won't even rise to knee level.

For example, here is the weak evidence/high stakes version of the vignette:

[name]'s kids want to go swimming in a local river. The water has slightly risen due to several days of light rain. So he decides to build a dam. He stacks a number of logs in the river to hold the water back and create a shallow pool where his kids can swim. Once he has finished building the dam, he checks the logs once to make sure the dam is secure.

It is very important that the dam is secure. Due to the high water level, if it breaks, his kids will be swept away by the water.

Participants were first asked a comprehension question to determine whether they recognized the stakes in the vignette:

Comprehension: It is very important that the dam is secure (Y/N)

Participants were then given an evidence-fixed prompt, in which they can either attribute knowledge or decline to do so:

Evidence-fixed: [name] ____________ that the dam is secure.[10]

○ knows
○ only thinks he knows

Participants were finally given an evidence-seeking prompt, in which they are asked how much evidence is required for knowledge:

Evidence-seeking: How many times do you think [name] has to check the logs before he knows that the dam is secure? [open-text response]

[9] The preregistration for this study used the phrases "a few times" and "multiple times", rather than "several times", when discussing the strong evidence condition. However, the .qsf survey file uploaded to OSF at the time of preregistration used the phrasing presented here.

[10] This prompt is a simplified version of the "Strict Knowledge Attribution" prompt used in Rose et al (2019).

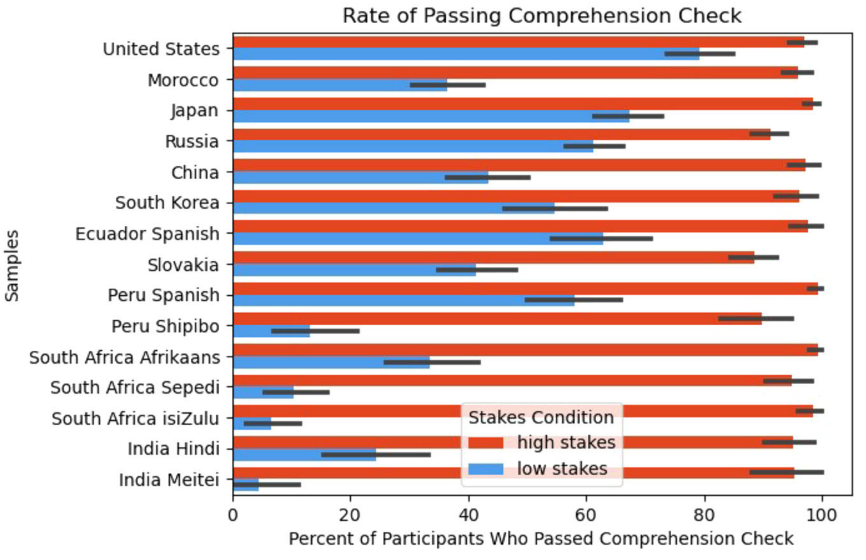

FIGURE 1 Comprehension Check Across Samples [Color figure can be viewed at wileyonlinelibrary.com]

All three prompts were presented on the same page, in the order given above. Vignettes and prompts were translated into the target languages by competent native speakers and presented in the participants' native languages.[11]

Based on Rose et al. (2019), we predicted that stakes would not influence knowledge ascriptions or the number of times the protagonist needed to check. We further hypothesized that the protagonist's amount of evidence would not influence knowledge ascription or the number of checks needed for knowledge.

3 | RESULTS[12]

2,057 participants were excluded either for failing the comprehension check, failing to complete the survey, or for reporting an age below 18.[13] Nearly all participants in the high stakes condition passed the comprehension check (Figure 1). However, in the low stakes condition, participants incorrectly answered that the dam being secure was "very important" at higher than expected rates. There was wide variation across populations; this may be due to translations issues, or to cultural variation in participants' willingness to accept the vignette's assertion that the children are not in danger in the low stakes scenario.[14]

[11] They were also backtranslated to check the accuracy of the translation.

[12] The data and the code can be found at osf.io/tyge9/files/osfstorage.

[13] Exclusion criteria were not preregistered, although the presence of a comprehension check was preregistered.

[14] Due to the high rate of failures for the comprehension check, we also reanalyzed the data without comprehension check exclusions. The results were very similar to those reported below, and can be found in our supplementary materials on OSF at https://osf.io/tyge9/files/osfstorage.

TABLE 2 Stakes Effects for the Three Models for the Evidence-Fixed Prompt

Model	Estimate	Std. Error	z value	Pr(>|z|)	Odds Ratio
M_1	0.02864	0.03758	0.762	0.4460	1.0290553
M_2	−0.10123	0.04028	−2.513	0.0120	0.9037275
M_3	0.02579	0.09105	0.283	0.7770	1.026121

TABLE 3 Model Fit of the Three Nested Models for the Evidence-Fixed Prompt

Model	npar	AIC	BIC	Log Likelihood	Chisq	Pr(>Chisq)
M_1	4	4587.5	4612.1	−2289.7		
M_2	5	4308.9	4339.8	−2149.5	280.51	< 2.2e-16
M_3	7	4299.6	4342.8	−2142.8	13.30	0.001294

3.1 | Evidence-Fixed

Removing participants who failed the comprehension check, we analyzed the remaining 3,516 responses to the evidence-fixed knowledge ascription prompt using three nested logistic regression models:[15]

1. A simple linear model M_1: $Ascription_s \sim \beta_0 + \beta_1 Stakes_s + \beta_2 Evidence_s + \beta_3 Stakes_s{*}Evidence_s + \varepsilon_S$;
2. A mixed-effects model that varies the intercept by population (P) M_2: $Ascription_{s\in P} \sim \beta_0 + \beta_{0P} + \beta_1 Stakes_s + \beta_2 Evidence_s + \beta_3 Stakes_s{*}Evidence_s + \varepsilon_S$;
3. A mixed-effects model that varies the intercept as well as the slope of the stakes factor by population (P) M_3: $Ascription_{s\in P} \sim \beta_0 + \beta_{0P} + (\beta_1 + \beta_{1P}) Stakes_s + \beta_2 Evidence_s + \beta_3 Stakes_s{*}Evidence_s + \varepsilon_S$.

The results for the stakes effect in each model can be found in the table below. Following Benjamin et al. (2018), we set the standard for statistical significance at .005. We compare the fit of the models using ANOVA, and find that the model that varies both intercept and slope provides the best fit for the data (Tables 2 and 3).

As predicted, in none of the models do we find a statistically significant stakes effect (Figure 2).[16] We do find a suggestive ($p = 0.0120$) stakes effect in M_2, but do not even find a suggestive effect in M_3, which provides a significantly better fit. However, in M_3 we do find a suggestive p-value

[15] Preregistration for this study did not specify the exact model to be used; only that we would analyze stakes, evidence, and their interaction between subjects, along with demographics, as predictors of knowledge ascription in a logistic regression. In the models presented here, we limit demographics to nationality. Effect coding was used for both the Stakes and Evidence predictor variables.

[16] We also fit versions of M_1, M_2 and M_3 that did not exclude participants who failed the comprehension check, and that replaced the stakes condition variable with participants' responses to the comprehension check. The idea is that responses to the comprehension check, in which participants report whether or not it is "very important" that the character's belief is true, can be taken to measure *perceived* stakes. We did not find the predicted stakes effect in those models; the p-value for perceived stakes is significant (0.000687) in this version of M_1, the simple logistic regression model with no random effects, but the effect is not in the expected direction: the model estimates that higher importance *increases* the likelihood of knowledge attribution. We do not find even a suggestive effect in M_2 (0.4189) or M_3 (0.3740), both of which provide a significantly better fit than M_1.

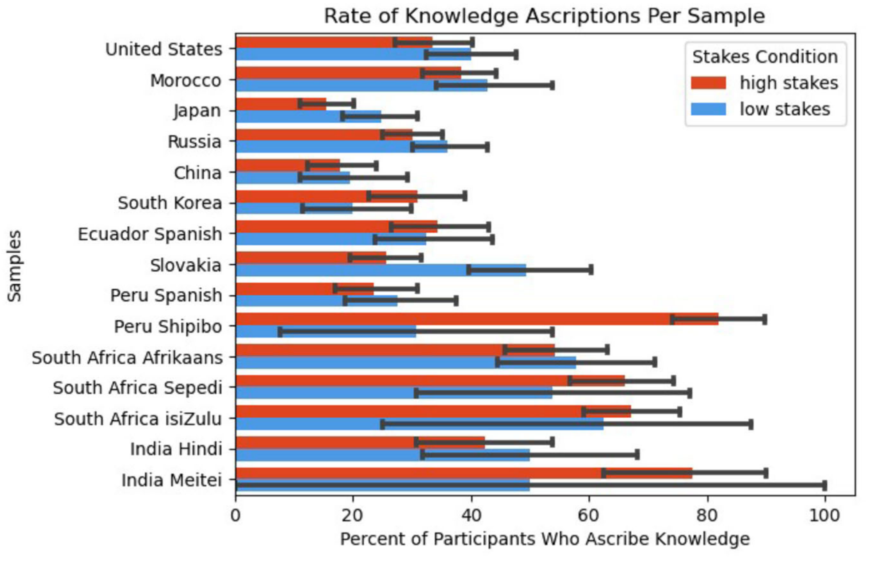

FIGURE 2 Knowledge Ascription Per Sample for the High and Low Stakes Conditions for the Evidence-Fixed Prompt [Color figure can be viewed at wileyonlinelibrary.com]

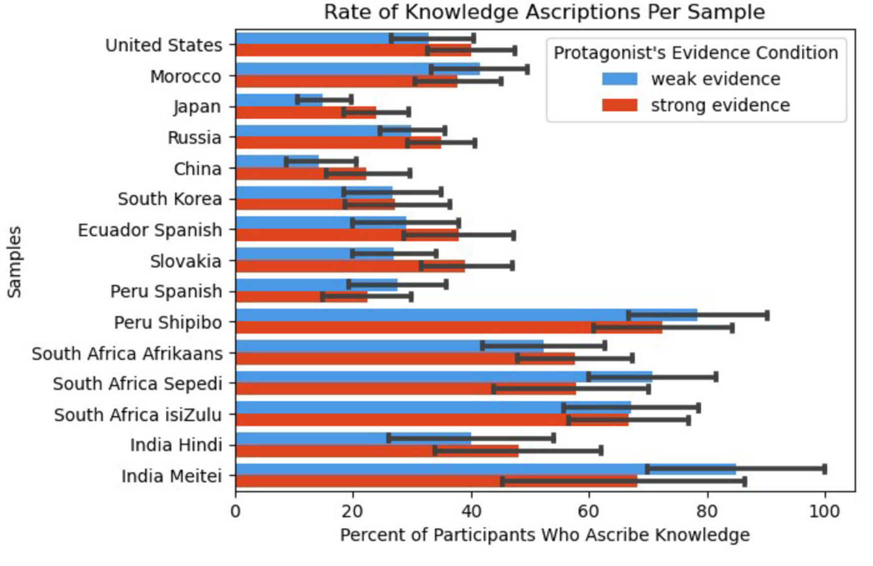

FIGURE 3 Knowledge Ascription Per Sample for the Weak and Strong Evidence Conditions for the Evidence-Fixed Prompt [Color figure can be viewed at wileyonlinelibrary.com]

($p = 0.0473$) that the protagonist's amount of evidence affects knowledge ascriptions (Figure 3). This is contrary to our preregistered predictions, but on second thought it is not surprising that stronger evidence results in higher rates of knowledge ascription.

TABLE 4 Stakes Effects Per Sample for the Evidence-Fixed Prompt

Sample	Estimate	Std. Error	z value	Pr(>|z|)	Odds Ratio
China	−0.1053	0.1763	−0.597	0.5504	0.900064
Russia	−0.12438	0.09253	−1.344	0.179	0.8830474
Slovakia	−0.53170	0.13445	−3.955	7.67e-05	0.5876024
Ecuador - Spanish	0.07867	0.15698	0.501	0.6163	1.0818479
India - Hindi	−0.1369	0.2462	−0.556	0.578	0.8720179
India - Meitei	0.4236	0.7480	0.566	0.571	1.5275252
Japan	−0.30100	0.12327	−2.442	0.0146	0.7400757
Morocco	−0.1050	0.1304	−0.805	0.42072	0.9003190
Peru - Shipibo	5.154	442.314	0.012	0.991	173.18192856
Peru - Spanish	−0.09741	0.16222	−0.600	0.548	0.9071871
South Africa - Afrikaans	−0.07561	0.17525	−0.431	0.666	0.9271746
South Africa - Sepedi	0.4357	0.3535	1.233	0.218	1.5461104
South Africa - isiZulu	0.08355	0.39391	0.212	0.832	1.0871401
South Korea	0.28943	0.17864	1.620	0.105	1.3356652
United States	−0.1457	0.1124	−1.296	0.195	0.8644485

Noteworthily, we observed some limited cultural variation (Table 4).[17] In particular, we find a significant stakes effect in the Slovakia sample ($p = 7.67e\text{-}05$) and a suggestive effect in the Japan sample ($p = 0.0146$).[18] We also find a statistically significant interaction between the stakes condition and the number of times the protagonist has checked in Morocco.[19] This suggests that within the Moroccan population sampled, there is a statistically significant stakes effect *when the protagonist has checked once*, but not when the protagonist has checked multiple times. This could be taken as evidence of a stakes effect on knowledge ascription in the Moroccan population: checking multiple times is sufficient evidence for knowledge in both the high and low stakes cases, but checking once is only sufficient evidence in the low stakes case.

However, even in those populations where we see evidence of a stakes effect, the effect size is relatively small. Even in Slovakia, the odds ratio is 1.7; which, following Chen et al. (2010), is a small effect, roughly equivalent to a Cohen's d of 0.2.

Overall, our evidence-fixed prompt finds little to no evidence of the sort of stakes effect on knowledge ascription that philosophers expect.

Before turning to the evidence-seeking prompt, we should examine whether these results are vitiated by the high number of participants failing the comprehension check (see Table 1). Since the comprehension check was binary, half of the participants answering randomly may have answered correctly by chance, and their random answers in the rest of the survey may explain why we failed to find any difference between the high and low stakes conditions.[20] There are

[17] We find large variation in the base rates of knowledge ascription between samples, but we hesitate to draw any substantive conclusions from this finding. It could be evidence of real cultural variation in knowledge ascription practices, but it could also easily be the result of variations in how the verb "to check" was translated.

[18] Note that the significance level was set at .05 in the original study. So, the Japan sample result would be significant by that standard.

[19] $p = 0.00237$, odds ratio = 1.4864723.

[20] We are grateful to an anonymous reviewer for pressing us to address this question.

two main things to say in response to this particular concern. First, failure to answer the comprehension check is not mostly due to random answering, but to participants assigned to the low states condition viewing the scenario as high stakes. Since these participants were removed for the analysis reported in the main text, concerns about random answering do not apply to our results. (Relatedly, note that including all participants and using the comprehension check as a measure of the *perceived* stakes does not change our results—see footnote 17). Second, this objection does not explain why if people answer randomly, we found systematic differences with the evidence-seeking prompt that we discuss next.

3.2 | Evidence-Seeking

Our evidence-seeking prompt was a free response question, and we had to code participants' answers. Although some answers were easily quantified ("5", or "three times"), others were harder to quantify ("many times", or "no guarantee no matter how many times") and some were straightforwardly unquantifiable ("I have no idea" or "no"). To convert responses to numeric data, we used the following rules:

1. If a response included digits ('2', '25', '100', etc.) the number was extracted from that response (so 'not more than 10 times' would be stored as 10)
2. Number words like "zero", "four", "eleven", "thousand" were converted to integers (0, 4, 11, 1000). After "twelve" we skipped straight to "hundred", "thousand", and "million". So any response that said "twenty" would be ignored.
3. Plurals were doubled. So "dozens" was stored as "24", "hundreds" was stored as 200, etc.
4. "Once" and "twice" were stored as 1 and 2 (so "at least once" and "more than once" were both stored as 1).
5. The term "couple" was stored as 2; "few" was stored as 3; "many", "several", "frequently", "lot", "lots", and "repeatedly" were all stored as 10.
6. Commas were removed, so that "1,000" and "1,000,000" were stored as 1000 and 1000000, respectively.
7. If multiple numbers could be extracted from an answer, the highest number was chosen. So "three or four times" was stored as 4.
8. All numbers extracted that were greater than 100 were stored as 100. This is to deal with outliers.[21]

Using this process, we were able to extract numbers from 2,833 responses, leaving 815 responses that could not be converted to numeric data. The results are shown in Figures 4 and 5.

Again, we fit three models:

1. A simple linear model M1: $Response_S \sim \beta_0 + \beta_1 Stakes_S + \beta_2 Evidence_S + \beta_3 Stakes_S{}^* Evidence_S + \varepsilon_S$;
2. A mixed-effects model that varies the intercept by population (P) M2: $Response_{S \in P} \sim \beta_0 + \beta_{0P} + \beta_1 Stakes_S + \beta_2 Evidence_S + \beta_3 Stakes_S{}^* Evidence_S + \varepsilon_S$;

[21] For example, the answer of one Russian participant included a "9" followed by 132 "0"s, which increased the mean of the "high stakes" condition by several orders of magnitude.

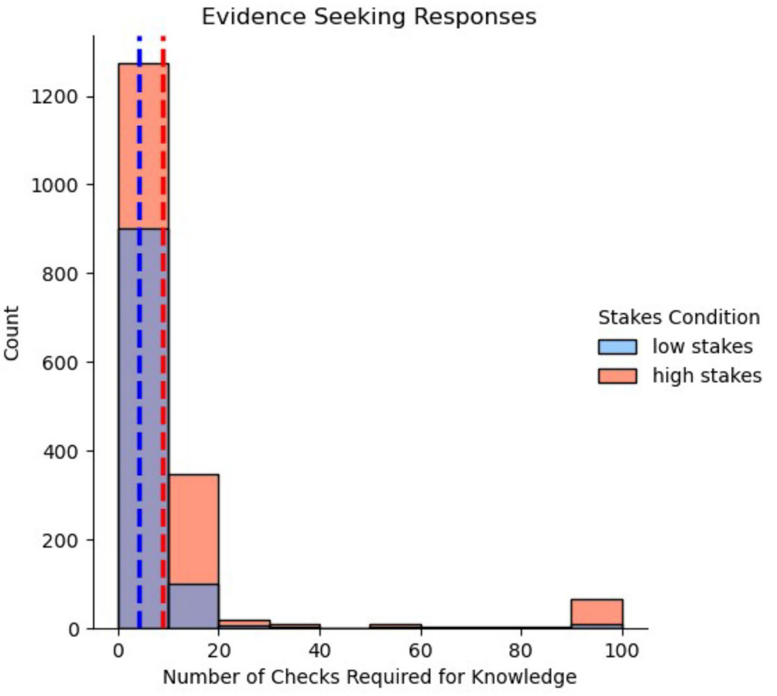

FIGURE 4 Responses from the "High Stakes" Condition in Red; Responses from the "Low Stakes" Condition in blue. Means are shown with dotted lines. The mean for "high stakes" is 8.85; the mean for "low stakes" is 4.15. [Color figure can be viewed at wileyonlinelibrary.com]

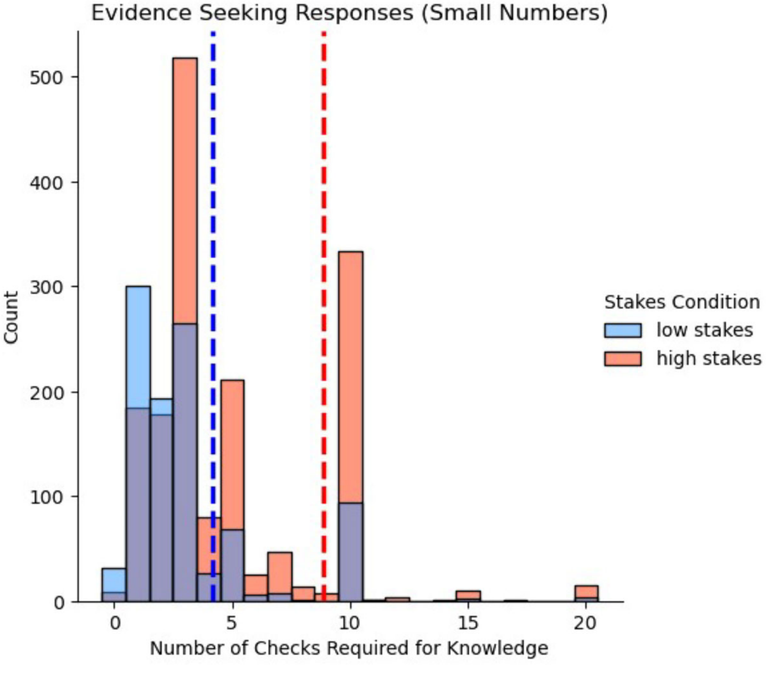

FIGURE 5 Counts are shown for each number less than or equal to 20. Each bar represents one number, and is centered on that number. Means are shown with dotted lines, and are the same as in Figure 4. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Stakes Effect for the Three Nested Models for the Evidence-Seeking Prompt

Model formula	Estimate	Std. Error	z value	Pr(>|z|)
M_1	0.391371	0.008925	43.853	< 2e-16
M_2	0.406727	0.008983	45.276	< 2e-16
M_3	0.322831	0.058406	5.527	3.25e-08

TABLE 6 Model Fit for the Three Nested Models for the Evidence-Seeking Prompt

Model formula	npar	AIC	BIC	Log Likelihood	Chisq	Pr(>Chisq)
M_1	4	43372	43396	−21682		
M_2	5	41170	41200	−20580	2203.95	< 2.2e-16
M_3	7	40820	40862	−20403	353.99	< 2.2e-16

3. A mixed-effects model that varies the intercept as well as the slope of the stakes factor by population (P) M3: $Response_{S \in P} \sim \beta_0 + \beta_{0P} + (\beta_1 + \beta_{1P})Stakes_S + \beta_2 Evidence_S + \beta_3 Stakes_S{*}Evidence_S + \varepsilon_S$.

Models were fit to a Poisson distribution, with a natural log link function. We find a statistically significant stakes effect on participant responses.[22] The results for the stakes effect in each model can be found in Table 5. [23]

Once again, an ANOVA reveals that the model with varied intercept and slope provides the best fit (Table 6).

We see similar effects across most cultures (Table 7). We find statistically significant stakes effects in 8 out of 14 samples tested, and a suggestive effect ($0.005 < p < 0.05$) in 2 samples. But it is worth noting that in one sample, South Africa – Afrikaans, we find a statistically significant interaction effect between stakes and evidence, suggesting that at least in the low evidence condition there is a statistically significant stakes effect in a surprising direction: participants report *lower* numbers of checks required when the stakes are *higher*. It is not clear how to interpret this result. However, we take the highly significant stakes effect found in the full cross-cultural sample (in the expected direction), combined with the significant or suggestive stakes effect in the expected direction found in 10 out of 14 samples, as strong evidence that there is a stakes effect on evidence-seeking responses across cultures.[24]

[22] We also see a statistically significant interaction effect between stakes and the protagonist's evidence (estimate = -0.055691, p-value = 4.95e-10 in our best-fit model). More detailed information can be found in our supplementary materials on OSF.

[23] Interestingly, we also find a statistically significant effect of the protagonist's amount of evidence (estimate = 0.222313, p< 2e-16). This suggests that the number of times the protagonist in fact checks has a significant effect on the number of times they have to check in order to know. It's not clear why this would be the case, though one possible explanation may be that participants infer that the protagonist who checks once is checking more thoroughly than the protagonist who checks several times, and thus requires fewer checks to collect the same amount of evidence.

[24] As in the evidence-fixed case, we also fit version of M_1, M_2, and M_3 without excluding anyone for failing the comprehension check, and replacing the stakes condition variable with the participant responses to the comprehension check as a measure of "perceived stakes." We find a statistically significant effect for perceived stakes in all three models, and in 9 out of the 14 tested populations; we again find the surprising negative interaction effect in the South Africa – Afrikaans sample.

TABLE 7 Stakes Effects Per Sample for the Evidence-Seeking Prompt

Sample	*N*	Estimate	Std. Error	z value	Pr(>|z|)
China	182	0.03758	0.09836	2.278	0.0227
Russia	530	0.54644	0.02207	24.758	< 2e-16
Slovakia	243	0.68343	0.03532	19.350	< 2e-16
Ecuador - Spanish	172	0.50301	0.03878	12.970	< 2e-16
India - Hindi	84	0.37698	0.06906	5.459	4.79e-08
India - Meitei	37	0.216273	0.164037	1.318	0.187
Japan	380	0.389493	0.019106	20.386	< 2e-16
Morocco	301	0.30593	0.03819	8.010	1.15e-15
Peru - Shipibo[a]	3	NA	NA	NA	NA
Peru - Spanish	126	0.11756	0.04612	2.549	0.01080
South Africa - Afrikaans	132	−0.06780	0.05164	−1.313	0.189
South Africa - Sepedi	22	0.364944	0.233666	1.562	0.118
South Africa - isiZulu	12	0.04766	0.24772	0.192	0.8474
South Korea	193	0.40194	0.02529	15.893	< 2e-16
United States	327	0.41068	0.02588	15.870	< 2e-16

[a] We were unable to test for a stakes effect in the Peru-Shipibo population, because all 3 of the responses that could be converted to numeric data were from participants in the "high stakes" condition. The numbers extracted were 5, 5, and 4.

One might worry that the fixed order of the comprehension check and the two prompts (comprehension check before the two prompts) might have influenced our results.[25] Perhaps the comprehension check ("It is very important that the dam is secure") makes the stakes salient and influences the answers to the prompts following the comprehension check. It would indeed have been good to vary the presentation order of the comprehension check and the prompts, but we made the choice to simplify the data collection process (particularly for sites where data collection was arduous) by using only one presentation order. On the other hand, it is not clear how the comprehension check is meant to influence participants' answers. Perhaps it highlights the similarities between the low and high stakes conditions (a dam might break in both conditions); but, equally plausibly, it might highlight the difference between the high (lives threatened) and low (no life threatened) stakes condition at least for those participants who answer correctly the comprehension check. Furthermore, it is not clear at all how the comprehension check is meant to explain the main result of our study, viz. the difference between evidence-fixed and evidence-seeking prompts.

4 | DISCUSSION

Our results are consistent with previous findings, in which evidence-fixed prompts have only rarely found evidence of a stakes effect, while evidence-seeking prompts consistently find evidence of a stakes effect. This difference has sometimes been taken to offer a methodological

[25] We are grateful to an anonymous reviewer for pressing us to address this question.

puzzle: two different methodologies give us different results, and the puzzle is to figure out which methodology is the reliable one, and what makes the other unreliable.

Much of the debate has therefore been centered on explaining why one of the two methods is flawed, and why the other method should be trusted. Defenders of stakes effects tend to argue that evidence-seeking prompts show that ordinary knowledge ascriptions are affected by stakes, and that evidence-fixed prompts fail to find one because the tests are flawed for one reason or another.[26] Supporters of stakes effects tend to argue that evidence-fixed prompts show that there is no stakes effect on ordinary knowledge ascriptions, and that evidence-seeking prompts seem to find one because *those* tests are flawed for one reason or another.[27]

Instead, we propose that neither methodology is flawed: both kinds of prompt are successfully testing different aspects of ordinary knowledge ascription, and neither set of results needs to be explained away. Evidence-fixed prompts ask participants to ascribe knowledge; by contrast, evidence-seeking prompts ask participants how much evidence a character would need in order to have knowledge. In answering this question, participants are not ascribing knowledge or declining to do so; they are expressing their *beliefs* about when knowledge ascriptions would be appropriate. The former prompt tests practice; the latter, beliefs about this practice.

The difference in results between the two kinds of prompts is naturally explained by the observation that the two prompts test different things. Although one might expect that these two prompts coincide, they do not: people's belief about the evidence required for knowledge is not aligned with their own practice of knowledge ascription. Misalignments between our practices and beliefs about them are of course not unusual. People believe that their assessment of others' performance is not affected by their race or gender, although it often is; people believe that they are good at detecting lies, while nearly all of us are lousy lie detectors (e.g., Bond and DePaulo, 2006).

The misalignment between people's belief about and their practice of knowledge ascription raises an interesting question. By their own lights, people must be making a mistake: if their belief about knowledge ascription is right, then they are mistaken to overlook the stakes when they ascribe knowledge; if their knowledge ascriptions in low and high stakes conditions are correct, then they are mistaken to believe that stakes should matter for knowledge ascription. Similarly, epistemologists might think that lay people's practice of knowledge ascription or their belief about it is mistaken. Classical invariantists, who argue that the truth-value of knowledge ascriptions does not shift across contexts, might argue it is the practice that is correct, and that people's belief about knowledge ascription is simply mistaken. Alternatively, contextualists such as DeRose, who hold that the truth-value of knowledge ascriptions is to be assessed relative to a context, and interest-relative invariantists such as Stanley, who hold that the truth-value of knowledge ascriptions depend on practical facts such as stakes, might argue that the belief is correct, and it is the practice that is mistaken. On this view, people's belief that stakes affect the truth conditions of knowledge ascriptions shows that there really is a stakes effect on (true) knowledge ascriptions; the fact that actual practice does not align with this belief means that people are

[26] In particular, it is often suggested that evidence-seeking prompts are finding stakes effects on something besides knowledge, such as epistemic anxiety (Nagel 2011, 2008), deontic modality (Buckwalter and Shaffer 2015), action (Gerken 2017), or estimates of knowledge-relevant factors (but not the factors themselves) (Dinges and Zakkou 2021).

[27] For example, Francis et al. (2019) suggest that participants take the evidence in the vignettes to be sufficient for knowledge in both high *and* low stakes scenarios. Pinillos (2012) and Sripada and Stanley (2012) suggest that participants assume that the protagonist in the high-stakes scenario has gathered evidence more thoroughly than the protagonist in the low stakes scenario.

making systematic errors in ascribing knowledge, either over-ascribing when stakes are high or under-ascribing when stakes are low.[28]

However, even if either people's practice or their belief about it must be incorrect, there might be no fact of the matter about which is incorrect. One relevant piece of information is what people do when presented with the inconsistency between their belief and their practice. If people judge that they were mistaken in ignoring stakes when they ascribe knowledge, then this would give us a reason to follow the lead of contextualists and interest-relative invariantists. Alternatively, they might not have any opinion on the matter. In future work, we plan to investigate this question.

5 | CONCLUSION

We presented evidence showing that in many languages and countries people do not take stakes into account when they assign knowledge in low- and high-stakes situations, using a more realistic scenario than in previous research in a more diverse group of populations. We have also shown that these same people nonetheless judge that stakes matter when asked to decide how much evidence is needed to know something. People's answers reveal an unexpected tension in folk epistemology between people's practice of knowledge ascription and their belief about it.

AFFILIATIONS

[1]Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, United States

[2]Public Policy Center, Université Internationale de Rabat, Rabat, Morocco

[3]Department of Philosophy, George Mason University, Fairfax, United States

[4]Department of Philosophy, University of Kyoto, Kyoto, Japan

[5]Centre for Social Studies, Universidade de Coimbra, Coimbra, Portugal

[6]Department of Philosophy, Pontificia Universidad Católica del Perú, Lima, Peru

[7]Department of Social Psychology, University of Tokyo, Tokyo, Japan

[8]Institute of Applied Psychology, Comenius University, Bratislava, Slovakia

[9]Department of Anthropology, University of Kansas, Lawrence, United States

[10]Department of Psychology, University of Tokyo, Tokyo, Japan

[11]Institute of Social Anthropology, Comenius University, Bratislava, Slovakia

[12]School of Psychology, Korea University, Seoul, South Korea

[13]Department of Anthropology, University of Alabama at Birmingham, Birmingham, United States

[14]Department of Psychology, Seoul National University, Seoul, South Korea

[15]School of Philosophy, Wuhan University, Hubei, China

[16]African Centre for Epistemology and Philosophy of Science, University of Johannesburg, Johannesburg, South Africa

[17]Department of Humanities & Social Sciences, Indian Institute of Technology, Delhi, India

[18]Institute for Philosophy and Social Theory, University of Belgrade, Belgrade, Serbia

[19]Department of Economics, Universidad San Francisco de Quito, Quito, Ecuador

[20]Collège des Sciences Sociales, Université Internationale de Rabat, Rabat, Morocco

[28] If one takes this line, then our results suggest that there is interesting cross-cultural variation regarding which side of the fence the error falls on: some populations have high rates of knowledge ascription in both high and low stakes cases, suggesting a possible bias towards over-ascribing knowledge; others have low rates in both cases, suggesting a possible bias towards under-ascribing.

[21]School of Nursing, University of Pittsburgh, Pittsburgh, United States

[22]Department of Philosophy, Rutgers University, New Brunswick, United States

[23]Department of Anthropology, University of California, Los Angeles, Los Angeles, United States

ORCID

Wesley Buckwalter https://orcid.org/0000-0002-6222-5083
Joshua Homan https://orcid.org/0000-0001-6010-3843
Xiaofei Liu https://orcid.org/0000-0001-6720-644X
Edouard Machery https://orcid.org/0000-0002-9944-6138

REFERENCES

Benjamin, D., Berger, J., Johannesson, M., Nosek, B., Wagenmakers, E.-J., Berk, R., Bollen, K., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C., Clyde, M., Cook, T., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., … Johnson, V. (2018). Redefine Statistical Significance. *Nature Human Behaviour*, *2*, 6–10. https://doi.org/10.1038/s41562-017-0189-z

Bond Jr, C F., & DePaulo, B M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214–234.

Buckwalter, W. (2014). The Mystery of Stakes and Error in Ascriber Intuitions. In J Beebe (ed.), *Advances in Experimental Epistemology*. Bloomsbury Academic.

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

Buckwalter, W., & Schaffer, J. (2015). Knowledge, Stakes, and Mistakes. *Noûs*, *49*(2), 201–234.

Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, *39*, 860–864. https://doi.org/10.1080/03610911003650383

Cohen, S. (1999). Contextualism, skepticism, and the structure of reasons. *Philosophical Perspectives*, *13*, 57–89.

DeRose, K. (2009). *The Case for Contextualism*. Oxford University Press.

DeRose, K. (2005). The ordinary language basis for contextualism, and the new invariantism. *Philosophical Quarterly*, *55*(219), 172–198.

DeRose, K. (1995). Solving the skeptical problem. *Philosophical Review*, *104*(1), 1–52.

DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, *52*, 913–29.

Dinges, A. (2023). Knowledge and Asymmetric Loss. *Review of Philosophy and Psychology*, *14*(3), 1055–1076.

Dinges, A., & Zakkou, J. (2021). Much at stake in knowledge. *Mind and Language*, *36*(5), 729–749.

Fantl, J., & McGrath, M. (2009). *Knowledge in an uncertain world*. New York: Oxford University Press. Edited by M McGrath.

Fantl, J., & McGrath, M. (2002). Evidence, pragmatics, and justification. *Philosophical Review*, *111*(1), 67–94.

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

Francis, K B., & Beaman, C. P (2023). The role of confidence in knowledge ascriptions: an evidence-seeking approach. *Synthese*, *202*(2), 1–15.

Francis, K.; Beaman, P., & Hansen, N. (2019). Stakes, Scales, and Skepticism. *Ergo: An Open Access Journal of Philosophy*, *6*, 427–487.

Gerken, M. (2017). *On Folk Epistemology. How we think and talk about knowledge*. Oxford: Oxford University Press.

Jackson, A. (2021). Are knowledge ascriptions sensitive to social context? *Synthese*, *199*(3), 8579–8610.

Nagel, J. (2011). The Psychological Basis of the Harman-Vogel Paradox. *Philosophers' Imprint*, *11*, 1–28.

Nagel, J. (2008). Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy*, *86*(2), 279–294.

Pinillos, Á. (forthcoming). Bank Cases, Stakes and Normative Facts. In S Nichols, & J Knobe (eds.), *Oxford Studies in Experimental Philosophy, vol*. 5. Oxford, UK: Oxford University Press.

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

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

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E E., Chatterjee, A., Cheon, H., Cho, In-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. *Noûs*, *53*(1), 224–247.

Rysiew, P. (2001). The context-sensitivity of knowledge attributions. *Noûs*, *35*(4), 477–514.

Sripada, C. S., & 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.; Buckwalter, W. & Rose, D. (2016). Actionability Judgments Cause Knowledge Judgments. *Thought: A Journal of Philosophy*, *5*(3), 212–222.

How to cite this article: 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. *Noûs*, 1–17. https://doi.org/10.1111/nous.12515