



PSYCHOLOGY AND PSYCHIATRY  
NEGATIVE-RESULT

# Uncertain veracity judgments: Does the framing of the question influence deception detection?

Mircea Zloteanu 

Department of Psychology, Kingston University, Kingston, United Kingdom

Email: [m.zloteanu@kingston.ac.uk](mailto:m.zloteanu@kingston.ac.uk)

Portions of these findings were presented as a poster at the 2021 International Behavior and Law Congress on Non-Verbal Communication.

(Received 20 December 2022; Revised 27 March 2023; Accepted 29 March 2023)

## Abstract

In deception research, little consideration is given to how the framing of the question might impact the decision-making process used to reach a veracity judgment. People use terms such as “sure” to describe their uncertainty about an event (i.e., aleatory) and terms such as “chance” to describe their uncertainty about the world (i.e., epistemic). Presently, the effect of such uncertainty framing on veracity judgments was considered. By manipulating the veracity question wording the effect of uncertainty framing on deception detection was measured. The data show no difference in veracity judgments between the two uncertainty framing conditions, suggesting that these may operate on a robust and invariant cognitive process.

**Keywords:** Deception detection; framing; judgment; uncertainty; veracity

## Introduction

People struggle to detect lies (Zloteanu et al., 2021), being barely better than chance (Bond & DePaulo, 2008), overconfident (DePaulo et al., 1997), and truth-biased (Levine et al., 1999). Most deception research centers on methods of improving deception detection (Hauch et al., 2014), but little on how people make veracity judgments.

Deception scholars have long debated how deception should be conceptualized and how deception detection (i.e., veracity judgments) should be measured. Some suggest a binary scale (Levine, 2001), a continuum of deceptiveness (Burgoon et al., 1994), or several dimensions (McCornack, 1992). Levine (2001) and Levine et al. (2010) argue that deception is a binary construct and, as such, deception judgments suit a binary decision (lie/truth), and that using a continuous scale confounds various constructs (e.g., moral condemnation and lie importance), and question if gradation of deception is even conceptually sound. Burgoon et al. (1994, 1995, 1996), conversely, assert that confounding veracity with other such message features is crucial to the judgment process, and, thus, a continuous scale must be used. Continuous scales capture more variability in responding, and alleviate social norms impinging on expressing strict opinions, although, at the expense of decision clarity (Levine, 2001). From an operational perspective, however, judgment accuracy and bias are stable regardless of the approach employed (Levine et al., 2010).

Yet, while the measurement debate continues, few consider how the framing of the event itself may impact judgment, especially if we assume that “detection” per se is difficult if not impossible for judges (Zloteanu et al., 2021). The scarce research on question framing has explored unprompted versus prompted veracity evaluations (Clare & Levine, 2019), phrasing the task as detecting “deception” or “truthfulness” (Masip et al., 2009), asking indirect questions (ten Brinke et al., 2014), and, as mentioned,

using a continuous versus binary scale (Levine et al., 2010). However, the framing of uncertainty in such questions has not been explored.

The psychology of judgment and decision-making has focused on understanding how people make judgments under uncertainty (Fischhoff & Broomell, 2020; Koehler & Harvey, 2004; Tversky & Kahneman, 1974). Verbal and written probabilities can be indicated in many ways (Honda & Yamagishi, 2006; Teigen & Løhre, 2017), and there exist multiple uncertainty frameworks (Fox & Ülkümen, 2017; Kahneman & Tversky, 1982; Teigen & Løhre, 2017; Ülkümen et al., 2016). While the terminology may differ, two forms of uncertainty are typically recognized: aleatory (i.e., randomness in an outcome) and epistemic (i.e., missing knowledge) (Fox & Ülkümen, 2011).

People tend to use phrases like “I am 80% sure...” to describe uncertainty about their own knowledge (i.e., epistemic) and phrases like “There is 80% chance...” to describe their uncertainty about external events (i.e., aleatory). This difference in framing is not simply a linguistic feature but is argued to reflect (and affect) the underlying cognitive (Fox & Ülkümen, 2017; Ülkümen et al., 2016) and/or motivational (Heath & Tversky, 1991) process used to make a judgment. Thus, the semantic choice may affect the way people interpret and answer the (veracity) question asked.

### Objectives

The study compared how framing a question in terms of certainty or likelihood can impact veracity judgments, adopting the framework of Ülkümen et al. (2016). It was predicted that the different framing would influence how judges perceive and answer the task.

**Hypothesis 1:** The framing of uncertainty—in terms of certainty or likelihood—will affect uncertainty estimates.

**Hypothesis 2:** The framing manipulation will shift response bias, judgment confidence, and deception detection accuracy.

### Method

The study was preregistered. The preregistration document, data, and analysis output can be accessed from <https://osf.io/cqjvb>.

### Participants

The sample size was determined a priori, using MorePower 6.0.4 (Campbell & Thompson, 2012). A sample of  $N = 96$  was estimated for a  $2 \times 2$  interaction, at  $\alpha = .05$ , power = 80%, and Cohen’s  $f = .30$ , and  $N = 192$  for the pairwise comparisons of interest, for a Cohen’s  $d = 0.40$  or  $\eta_p^2 = .04$ . These estimates would also produce evidence at a Bayes Factor (BF)  $> 10$  for both interaction and pairwise comparisons.

A total of 261 participants were recruited using the University’s SONA system. After removing incomplete cases ( $n = 51$ ), the final sample contains  $N = 210$  ( $F = 184$ ,  $M = 25$ ,  $O = 1$ ;  $M_{Age} = 26.4$ ,  $SD = 10.58$ ); Certainty ( $n = 103$ ) and Likelihood ( $n = 107$ ). A two-way mixed design was used, with Framing (Certainty vs. Likelihood) as the between-subjects factor, and Veracity (Truth and Lies) as the within-subjects factor. The dependent measures were the uncertainty level, deception detection accuracy, judgment confidence, and response bias.

### Stimuli

Twenty videos (10 lies and 10 truths) from the Bloomsbury Deception Set (Street et al., 2011) were used. They are approximately 30-s long. Senders are describing a holiday, and half are lying. The videos are gender-matched per veracity, and no sender was used twice.

### Procedure

The study was conducted using the Qualtrics platform. Participants were randomly allocated to either the Likelihood or Certainty condition. This affected the framing of the veracity question. In the Likelihood condition, the question was “How *likely* is it this person is telling a lie?” whereas in the Certainty condition, it was “How *sure* are you this person is telling a lie?” They responded by entering a percentage (0–100%). For example, in the “likely” condition, they were asked, “Using a percentage from 0 to 100%, how likely is it this person is telling a lie?” They also rated their confidence using a 5-point scale (1 = *Not at all* to 5 = *Completely*). The videos were presented in random order. The task took  $Med = 18$  min ( $MAD = 3$  min).

### Ethics

The study was approved by the University’s ethics board and was carried out in accordance with the General Data Protection Regulation, the Data Protection Act (2018), and the Declaration of Helsinki. All participants provided informed consent.

### Results

Data coding and cleaning were performed as per the preregistered protocol. Data were analyzed using JASP (JASP Team, 2022). Inferences are made using frequentist and Bayesian hypothesis testing.

#### Uncertainty framing

After removing outliers ( $n = 3$ ), a 2(Framing: Certainty vs. Likelihood)  $\times$  2(Veracity: Lies and Truths)-mixed ANOVA was conducted on the uncertainty responses. No effect of Framing was found,  $F(1, 205) = 2.28, p = .132, \eta_p^2 = .01, 90\% \text{ CI } [.00, .05], BF_{inc} = 0.50$  (anecdotal; Jeffreys, 1983), with responses in the Certainty ( $M = 42.86, SD = 10.92$ ) and Likelihood ( $M = 40.51, SD = 11.43$ ) conditions being similar;  $M_{diff} = -2.35 (SE = 1.56), d_s = -0.18, 95\% \text{ CI } [-.42, .06], JZS BF_{10} = 0.49$  (anecdotal). The Veracity effect was significant,  $F(1, 205) = 59.29, p < .001, \eta_p^2 = .22, 90\% \text{ CI } [.15, .30], BF_{inc} = 1.03e^{10}$  (decisive), with Lies ( $M = 45.28, SD = 13.83$ ) being given a higher percentage than Truths ( $M = 38.06, SD = 12.33$ );  $M_{diff} = 7.24 (SE = 0.94), d_{av} = 0.55, 95\% \text{ CI } [.40, .71], JZS BF_{10} = 1.04e^{10}$  (decisive). The interaction was not significant,  $F(1, 205) = 1.27, p = .262, \eta_p^2 = .01, 90\% \text{ CI } [.00, .04], BF_{inc} = 0.27$  (substantial). See Figure 1 for details.

A two one-sided test (TOST) of equivalence (Lakens, 2017) was conducted to discern if the Framing result reflects no effect or insufficient data, given the unclear Bayes Factor. The TOST, using equivalence bounds of  $\pm 4.36\%$  ( $d = \pm 0.39$ ), yielded a significant result for the lower bound,  $\Delta_L, t(205) = -4.32, p < .001$ , and a nonsignificant result for the upper bound,  $\Delta_U, t(205) = 1.29, p = .099$ . This is interpreted as *not equivalent and not different* (i.e., insufficient data; Lakens, 2017). However, a Bayesian equivalence test (Morey & Rouder, 2011) finds strong evidence for the nonoverlapping hypothesis that the effect lies within the equivalence interval,  $BF_{\notin} = 22.73$ .

#### Veracity judgments

The uncertainty responses were recoded to create an accuracy score. Values of  $>50\%$  were taken as a “lie” judgment,  $<50\%$  as a “truth” judgment, and  $50\%$  as “unsure (incorrect).” The judgments were compared to the veracity of each video. If the judgment matched the video veracity, it was taken as “correct” (1); if not, it was taken as “incorrect” (0) and summed per condition.

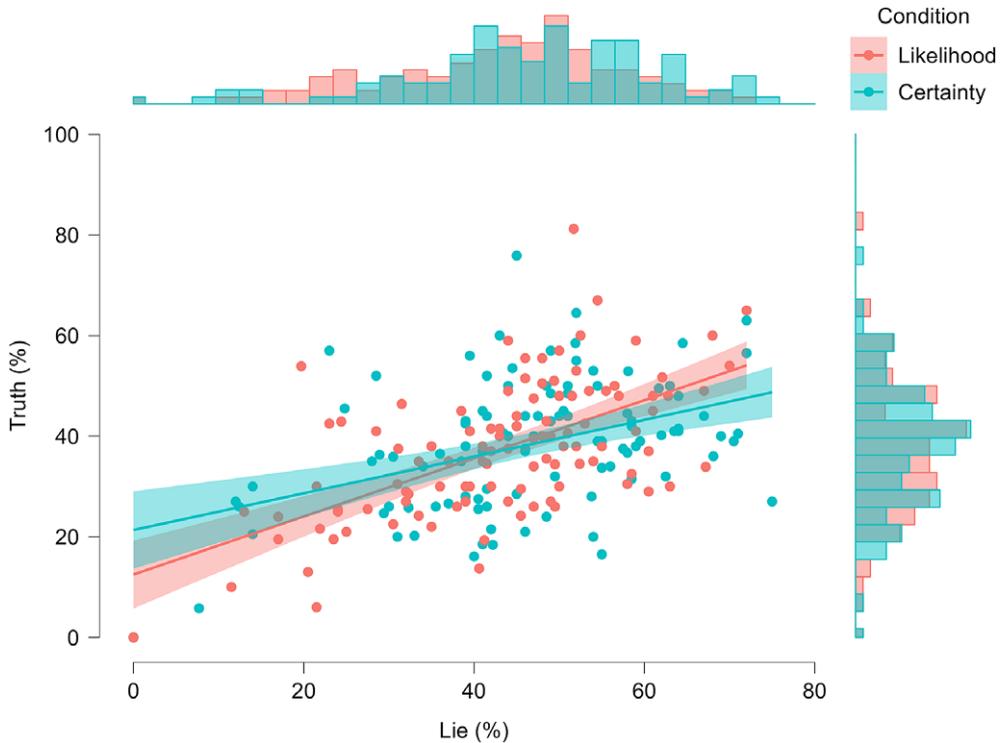


Figure 1. Scatter Plot with Histograms for the Uncertainty Ratings as a Function of Framing and Veracity.

### Deception detection

A mixed ANOVA tested deception detection accuracy (%). The Framing effect was nonsignificant,  $F(1, 208) = 3.12, p = .079, \eta_p^2 = .02, 90\% \text{ CI } [.00, .05], \text{BF}_{\text{inc}} = 0.27$  (substantial), with similar accuracy in the Certainty ( $M = 51.65\%, SD = 11.53\%$ ) and Likelihood ( $M = 48.93\%, SD = 10.82\%$ ) conditions;  $M_{\text{diff}} = -2.73\% (SE = 1.54\%), d_s = -0.14, 95\% \text{ CI } [-.29, .02], \text{JZS BF}_{10} = 0.23$  (substantial). The Veracity effect was significant,  $F(1, 208) = 77.10, p < .001, \eta_p^2 = .27, 90\% \text{ CI } [.19, .35], \text{BF}_{\text{inc}} = 9.79e^{19}$  (decisive) with lies ( $M = 40.10\%, SD = 20.57\%$ ) being harder to detect than truths ( $M = 60.43\%, SD = 19.79\%$ );  $M_{\text{diff}} = -20.27\% (SE = 2.31\%), d_{\text{av}} = -1.01, 95\% \text{ CI } [-1.25, -.76], \text{JZS BF}_{10} = 9.71e^{12}$  (decisive). The interaction was nonsignificant,  $F(1, 208) = 2.14, p = .145, \eta_p^2 = .01, 90\% \text{ CI } [.00, .04], \text{BF}_{\text{inc}} = 0.64$  (anecdotal).

### Bias

An independent samples  $t$ -test compared judges' response bias between the framing conditions. Bias was computed by classifying each lie response as  $-1$ , each true response as  $+1$ , and  $0$  for unsure, which were then summed across trials. A positive total indicates a truth bias, a negative total indicates a lie bias, and  $0$  represents unbiased.

Bias appears higher (i.e., stronger truth bias) in the Likelihood condition ( $M = 4.58, SD = 6.65$ ) than the Certainty condition ( $M = 3.35, SD = 6.07$ ); however, the test was nonsignificant with high uncertainty,  $t(208) = 1.40, p = .163, M_{\text{diff}} = 1.23 (SE = 0.88), d_s = 0.14, 95\% \text{ CI } [-.08, .46], \text{JZS BF}_{10} = 0.38$  (anecdotal). A one-sample  $t$ -test revealed that judges were overall truth-biased,  $M = 3.98 (SD = 6.38), t(209) = -2.10, p < .001, d_z = 0.62, 95\% \text{ CI } [.47, .77], \text{JZS BF}_{10} = 4.48e^{13}$  (decisive).

An equivalence test, using equivalence bounds of  $\pm 2.48 (d = \pm 0.39)$ , yielded a significant result for  $\Delta_U, t(208) = 4.22, p < .001$ , and a nonsignificant result for  $\Delta_L, t(208) = -1.42, p = .078$ , thus *not*

*equivalent and not different*. The Bayesian equivalence test, conversely, yielded very strong evidence for the effect lying within the equivalence interval,  $BF_{\neq} = 32.73$ .

### Signal detection theory

The data were also analyzed using signal detection theory (SDT), as this allows the separation of response bias (criterion;  $c$ ) from accuracy (discriminability;  $d'$ ). The SDT analysis of discriminability ( $d'$ ) revealed a similar pattern of (nonsignificant) results for accuracy, but with higher uncertainty,  $t(208) = -1.88$ ,  $p = .061$ ,  $M_{diff} = -0.16$  ( $SE = 0.08$ ),  $d_s = -0.26$ , 95% CI  $[-.53, .01]$ , JZS  $BF_{10} = 1.27$  (anecdotal), with similar scores in Certainty ( $M = 0.08$ ,  $SD = 0.62$ ) and Likelihood ( $M = -0.08$ ,  $SD = 0.60$ ). An equivalence test, using equivalence bounds of  $\pm 0.24$  ( $d = \pm 0.39$ ), yielded a nonsignificant result for  $\Delta_U$ ,  $t(208) = 0.941$ ,  $p = .174$ , and a significant result for  $\Delta_L$ ,  $t(208) = -4.71$ ,  $p < .001$ , thus *not equivalent and not different*. The Bayesian equivalence test yielded strong evidence for the effect posterior falling within the equivalence interval,  $BF_{\neq} = 12.86$ .

The SDT analysis of bias ( $c$ ) revealed a more uncertain pattern of results that the raw bias results, as the difference between the two framing conditions was significant,  $t_{Welch}(207.833) = -2.15$ ,  $p = .033$ ,  $M_{diff} = -0.17$  ( $SE = 0.08$ ),  $d_s = -0.30$ , 95% CI  $[-.57, -.02]$ , JZS  $BF_{10} = 1.29$  (anecdotal), with bias being more pronounced (i.e., a more liberal threshold for classifying a video as truthful) in the Likelihood condition ( $M = -0.37$ ,  $SD = 0.61$ ) than the Certainty condition ( $M = -0.20$ ,  $SD = 0.57$ ), yet the BF indicates insufficient evidence to make a decisive claim. Indeed, a Bayesian equivalence test finds more of the effect posterior falling within the equivalence interval than outside it,  $BF_{\neq} = 8.11$  (moderate).

### Confidence

A mixed ANOVA tested differences in judgment confidence. The framing effect was nonsignificant,  $F(1, 208) = 0.86$ ,  $p = .356$ ,  $\eta_p^2 < .01$ , 90% CI  $[.00, .02]$ ,  $BF_{inc} = 0.39$  (anecdotal), with similar confidence in the Certainty ( $M = 3.40$ ,  $SD = 0.57$ ) and Likelihood ( $M = 3.33$ ,  $SD = 0.47$ ) conditions;  $M_{diff} = -0.67$  ( $SE = 0.72$ ),  $d_s = -0.12$ , 95% CI  $[-.38, .14]$ , JZS  $BF_{10} = 0.23$  (substantial). The veracity effect was nonsignificant,  $F(1, 208) = 1.80$ ,  $p = .181$ ,  $\eta_p^2 = .01$ , 90% CI  $[.00, .04]$ ,  $BF_{inc} = 0.25$  (substantial). The interaction was nonsignificant,  $F(1, 208) = 0.05$ ,  $p = .825$ ,  $\eta_p^2 < .01$ , 90% CI  $[.00, .01]$ ,  $BF_{inc} = 0.16$  (substantial).

A TOST for the framing effect, using equivalence bounds of  $\pm 0.20$  ( $d = \pm 0.39$ ), yielded a significant result for  $\Delta_U$ ,  $t(208) = 1.89$ ,  $p = .030$ , and for  $\Delta_L$ ,  $t(208) = -3.74$ ,  $p < .001$ . This is interpreted as *statistically equivalent and not different*. Likewise, the Bayesian equivalence test finds very strong evidence for the effect lying within the equivalence interval,  $BF_{\neq} = 93.23$ .

### Discussion

Hypotheses 1 and 2 were not supported, with the data showing stronger evidence for no effect of uncertainty framing on veracity judgments. The results contradict Ülkümen et al.'s (2016) cognitive processes argument, as framing in terms of epistemic or aleatory uncertainty did not affect judges' cognition and judgment (cf. Heath & Tversky, 1991). This interpretation is further backed by the average uncertainty level, 42%, being below the expected 50%, especially in the aleatory ("chance") framing condition (Fox & Ülkümen, 2011). Thus, the current framing does not seem to shift probabilistic beliefs from a focus on subjective/internal information to stochastic/external information.

The limitation of the current study is the ambiguity of the data and strength of evidence for several results (e.g., bias). Future research must increase the sample size and investigate other operationalizations of uncertainty, for instance, rephrasing the question to focus on truthfulness (Masip et al., 2009), or changing the base rate to match real-world probabilities (Levine et al., 2014). Additionally, although the veracity framing seems not to affect judges, it may affect listeners, as these seemingly prefer speakers who

use certainty instead of likelihood statements (Ülkümen et al., 2016). However, listeners also misinterpret probability statements based on the perceived information severity and speaker tactfulness (Bonnefon & Villejoubert, 2006). Researchers should investigate the perceived evidentiary value of probabilistic veracity statements if made using certainty or likelihood phrasing. Lastly, the similarity in the confidence ratings and the uncertainty judgments should be explored. These ought to reflect separate constructs (e.g., confidence is a meta-judgment regarding one's decision), especially in the aleatory condition, yet the substantial overlap suggests that judges view the two as parallel.

## Conclusion

Contrary to the focal prediction, uncertainty framing—aleatory or epistemic—did not affect veracity judgments. Either the manipulation was too subtle to be captured by the current data, or the process by which people judge veracity is unaffected by uncertainty framing.

**Open peer review.** To view the open peer review materials for this article, please visit <http://doi.org/10.1017/exp.2023.9>.

**Data availability statement.** The data are available at <https://osf.io/cqjvb>.

**Author contribution.** Conceptualization: M.Z.; Data curation: M.Z.; Formal analysis: M.Z.; Investigation: M.Z.; Methodology: M.Z.; Project administration: M.Z.; Writing – original draft: M.Z. designed the study, collected the data, performed the analyses, and wrote the article.

**Funding statement.** This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

**Competing interest.** The author declares none.

## References

- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, *134*, 477–492. <https://doi.org/10/bn7z8h>
- Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*, 747–751. <https://doi.org/10.1111/j.1467-9280.2006.01776.x>
- Burgoon, J. K., Buller, D. B., Dillman, L., & Walther, J. B. (1995). Interpersonal deception: IV. Effects of suspicion on perceived communication and nonverbal behavior dynamics. *Human Communication Research*, *22*, 163–196. <https://doi.org/10/bq3265>
- Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal deception: V. Accuracy in deception detection. *Communications Monographs*, *61*, 303–325. <https://doi.org/10/cgr3h9>
- Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs*, *63*, 50–69. <https://doi.org/10.1080/03637759609376374>
- Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*, 1255–1265. <https://doi.org/10.3758/s13428-012-0186-0>
- Clare, D. D., & Levine, T. R. (2019). Documenting the truth-default: The low frequency of spontaneous unprompted veracity assessments in deception detection. *Human Communication Research*, *45*, 286–308. <https://doi.org/10/gjpm62>
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, *1*, 346–357. <https://doi.org/10/br3pdr>
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and decision making. *Annual Review of Psychology*, *71*, 331–355. <https://doi.org/10.1146/annurev-psych-010419-050747>
- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkeboen, & H. Montgomery (Eds.), *Perspectives on thinking, judging, and decision making* (pp. 1–14). Universitetsforlaget.
- Fox, C. R., & Ülkümen, G. (2017). Comment on Löhre & Teigen (2016). “There is a 60% probability, but I am 70% certain: Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, *23*, 483–491. <https://doi.org/10.1080/13546783.2017.1314939>
- Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*, *43*, 283–343. <https://doi.org/10/f8h5hq>
- Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*, 5–28. <https://doi.org/10.1007/BF00057884>

- Honda, H., & Yamagishi, K. (2006). Directional verbal probabilities: Inconsistencies between preferential judgments and numerical meanings. *Experimental Psychology*, *53*, 161–170. <https://doi.org/10.1027/1618-3169.53.3.161>
- JASP Team. (2022). *JASP (version 0.16.4) [Computer software]*. <https://jasp-stats.org/>
- Jeffreys, H. (1983). *Theory of probability* (3rd ed.). Clarendon Press/Oxford University Press.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*, 143–157. [https://doi.org/10.1016/0010-0277\(82\)90023-3](https://doi.org/10.1016/0010-0277(82)90023-3)
- Koehler, D. J., & Harvey, N. (Eds.) (2004). *Blackwell handbook of judgment and decision making*. Blackwell. <https://doi.org/10.1002/9780470752937>
- Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*, 355–362. <https://doi.org/10.1177/1948550617697177>
- Levine, T. R. (2001). Dichotomous and continuous views of deception: A reexamination of deception ratings in information manipulation theory. *Communication Research Reports*, *18*, 230–240. <https://doi.org/10/bn9jf2>
- Levine, T. R., Clare, D. D., Green, T., Serota, K. B., & Park, H. S. (2014). The effects of truth–lie base rate on interactive deception detection accuracy. *Human Communication Research*, *40*, 350–372. <https://doi.org/10/f57jxw>
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect”. *Communication Monographs*, *66*, 125–144. <https://doi.org/10/dfswnk>
- Levine, T. R., Shaw, A. S., & Shulman, H. (2010). Assessing deception detection accuracy with dichotomous truth–lie judgments and continuous scaling: Are people really more accurate when honesty is scaled?. *Communication Research Reports*, *27*, 112–122. <https://doi.org/10/bsgxv6>
- Masip, J., Alonso, H., Garrido, E., & Herrero, C. (2009). Training to detect what? The biasing effects of training on veracity judgments. *Applied Cognitive Psychology*, *23*, 1282–1296. <https://doi.org/10/fwshqd>
- McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, *59*, 1–16. <https://doi.org/10.1080/03637759209376245>
- Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*, 406–419. <https://doi.org/10.1037/a0024377>
- Street, C. N., Tbaily, L., Baron, S., Khalil-Marzouk, P., Hanby, B., Wright, K., & Richardson, D. C. (2011). Bloomsbury deception set. In *British psychological society division of forensic psychology conference*. Division of Forensic Psychology Conference. <https://www.researchgate.net/publication/220027496>
- Teigen, K. H., & Løhre, E. (2017). Expressing (un)certainly in no uncertain terms: Reply to Fox and Ülkümen. *Thinking and Reasoning*, *23*, 492–496. <https://doi.org/10.1080/13546783.2017.1314965>
- ten Brinke, L. M., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychological Science*, *25*, 1098–1105. <https://doi.org/10/f54swx>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. <https://doi.org/10/gwh>
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology*, *145*, 1280–1297. <https://doi.org/10.1037/xge0000202>
- Zloteanu, M., Bull, P., Krumhuber, E. G., & Richardson, D. C. (2021). Veracity judgement, not accuracy: Reconsidering the role of facial expressions, empathy, and emotion recognition training on deception detection. *Quarterly Journal of Experimental Psychology*, *74*, 910–927. <https://doi.org/10/ghk9h4>

---

Cite this article: Zloteanu M (2023). Uncertain veracity judgments: Does the framing of the question influence deception detection?. *Experimental Results*, *4*, e8, 1–9. <https://doi.org/10.1017/exp.2023.9>

# Peer Reviews

**Reviewing editor:** Dr. Massimo Grassi

Universita degli Studi di Padova Scuola di Psicologia, Department of General Psychology, via Venezia 8, Padova, Italy, 35131

Minor revisions requested.

doi:10.1017/exp.2023.9.pr1

## Review 1: Uncertain Veracity Judgements: Does the Framing of the Question Influence Deception Detection?

**Reviewer:** Dr. Glynis Bogaard 

Maastricht University, Maastricht, Netherlands, 6200 MD

Date of review: 16 January 2023

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

**Conflict of interest statement.** Reviewer declares none.

### Comment

Comments to the Author: Please find below my comments on the current version of the ms.

-In the introduction, briefly explain the effects of question framing on veracity evaluations in addition to existing manipulations.

-Please include a description of sample size determination of the preregistration, as this is much clearer than what is included in ms.

-Either include more information about the list of seven deception cues or remove from ms. Currently, it is not clear what these cues add to your message.

-Please include how accuracy scores were calculated under veracity judgments, you do this nicely in the pre-registration.

-I believe that the bias analysis should be conducted following the recommendations of Stanislaw and Todorov (1999). Specifically, see p. 142 on how to calculate response bias ( $c$ ). " $c$  is found by averaging the  $z$  score that corresponds to the hit rate and the  $z$  score that corresponds to the false alarm rate, then multiplying the result by negative one." Consequently, you must exclude the 50% judgements from the analyses, as these participants fail to make a choice. You can additionally investigate whether the number of unsure judgments differ between framing conditions.

-The confidence analysis seems redundant. You ask participants to provide a percentage, this percentage already shows their confidence. The more extreme, the more confident the participant was. This also shows in your confidence findings, which are identical to the uncertainty framing analyses.

### Score Card

#### Presentation



Is the article written in clear and proper English? (30%)

4/5

Is the data presented in the most useful manner? (40%)

4/5

Does the paper cite relevant and related articles appropriately? (30%)

4/5

## Context



Does the title suitably represent the article? (25%)

4/5

Does the abstract correctly embody the content of the article? (25%)

4/5

Does the introduction give appropriate context? (25%)

3/5

Is the objective of the experiment clearly defined? (25%)

4/5

## Analysis



Does the discussion adequately interpret the results presented? (40%)

4/5

Is the conclusion consistent with the results and discussion? (40%)

4/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

4/5