

When (and why) expressing uncertainty affects perceptions of competence

Joel M. Levin*

Working paper, June 2026

Abstract

What do we infer about others when they express uncertainty? A broad literature on this question has yielded results that appear contradictory. In prior work, expressing more (versus less) certainty has been associated with being perceived as being both more and less competent. In this paper, I propose and test a theory to reconcile these findings. I argue that people make inferences about others on the basis of fit between the amount of certainty that those others express and their own beliefs about the epistemicness (predictability) of the environment. Fit between expressed certainty and perceived epistemicness suggests calibration and in turn competence, while a lack of fit suggests poor calibration and (relative) incompetence. I find support for this account across four experimental studies involving both real and hypothetical communication. In each study, I manipulate two factors: the amount of certainty that an advisor or teammate expresses and a characteristic that affects the epistemicness of a given prediction or estimation. Participants display a preference for others who express more certainty in more epistemic environments, but this preference attenuates (S1) or reverses (S2 - S4) in less epistemic environments. Participants' perceptions of others' calibration follow the same pattern (S3).

Imagine asking two colleagues whether your department will hire more people in the coming year. One of them says “we’ll probably hire” and the other says “we’ll definitely hire.” Whose prediction would you put more trust in? Who do you think knows more? And who would be more likely to influence your beliefs?

When evaluating others' competence within a given domain and in choosing whether to heed their advice, we often use the degree to which they express certainty to inform our judgments. However, findings on the relationship between the amount of certainty that people express and the degree to which they are perceived to be competent have been inconsistent. Expressing more (versus

*Rady School of Management, University of California San Diego. Email: joelmlevin@gmail.com.

less) certainty has appeared to have positive (Kennedy, Anderson, & Moore, 2013; Price & Stone, 2004; Sah, Moore, & MacCoun, 2013; Snizek & Van Swol, 2001; Van Swol & Snizek, 2005; Vulloud, Clément, Scott-Phillips, & Mercier, 2017) negative (Du, Budescu, Shelly, & Omer, 2011; Gaertig & Simmons, 2023) and ambiguous (Bles, Linden, Freeman, & Spiegelhalter, 2020; Gaertig & Simmons, 2018; Karmarkar & Tormala, 2010; Løhre & Teigen, 2024) impacts on perceptions of a person's competence.

I propose and test a model to explain this apparently discrepant pattern. When evaluating others' (hereafter: *senders*') competence, peoples' (hereafter: *receivers*') judgments depend on both the degree to which senders express certainty and on receivers' beliefs about the type of certainty that characterizes the outcome in question. I then test this model across four studies, investigating whether receivers' beliefs about the *type* of uncertainty that characterizes an outcome moderate the relationship between the amount of certainty that senders express and the degree to which receivers perceive them as competent. To characterize the nature of uncertainty, I draw on prior work which has distinguished between two types: *epistemic uncertainty*, which stems from a salient lack of knowledge about outcomes which are perceived to be knowable, and *aleatory uncertainty*, which stems from the belief that outcomes are inherently less knowable (e.g., C. R. Fox & Ülkümen, 2011; Ülkümen, Fox, & Malle, 2016). I find that receivers' beliefs about certainty type predict perceptions of competence in systematic ways: When participants (receivers) perceive outcomes to be relatively more epistemic, they find senders who express higher levels of certainty to be more competent (S1 – S3) and are more likely to look to them for information (S4), but this pattern attenuates (S1) or reverses (S2 - S4) when outcomes are perceived to be more aleatory.

Background and theory

Prior work has examined the relationship between the amount of certainty that a sender expresses and the way that he or she is perceived across a variety of contexts. In most empirical papers on this topic, results have suggested there is a positive relationship between the degree to which a sender expresses certainty and the degree to which they are perceived to be competent, credible, persuasive, and trustworthy (e.g., Alzahawi & Flynn, 2025; Anderson, Brion, Moore, & Kennedy, 2012; Kennedy et al., 2013; Price & Stone, 2004; Sah et al., 2013; Snizek & Van Swol, 2001;

Van Swol & Sniezek, 2005; Vulliamou et al., 2017; Yaniv & Foster, 1995). For example, Yaniv & Foster (1995) asked receivers to evaluate senders' hypothetical judgments about questions such as the year that the Sino-Japanese war began or the distance between Chicago and New York, finding that receivers tended to prefer interval judgments that were more precise (e.g., "1890-1900") to those that were less precise (e.g., "1800-1950"). In a closely related paradigm, Van Swol & Sniezek (2005) paired participants in judge-advisor dyads and tasked judges (receivers) with answering a set of factual questions about computer operations with the aid of an advisor's (sender's) recommendation. Advisors' recommendations were communicated in the form of an answer and a confidence assessment (the advisor's subjective probability that their recommendation was correct; e.g., "I'm 90% sure that the answer is X."). They found that judges were more trusting of advisors and were more likely to follow those advisors' recommendations when recommendations were accompanied by more extreme confidence assessments (closer to 100%, indicating more subjective certainty) than when they were accompanied by less extreme assessments (closer to 50%¹, indicating less subjective certainty).

However, recent work has cast doubt on the ubiquity of this relationship, finding null, negative, or ambiguous relationships between certainty and competence (Bles et al., 2020; Du et al., 2011; Gaertig & Simmons, 2018, 2023; Karmarkar & Tormala, 2010; Løhre & Teigen, 2024). For example, Du et al. (2011) asked participants to evaluate earnings forecasts that were communicated in the form of either point estimates (more certain) or range estimates (less certain). They report that receivers often prefer the latter to the former – finding that participants (in the role of receiver) rate forecasts that are made in the form of range estimates as being more trustworthy and more credible than those that are made in the form of point estimates. They also report more confidence in range-estimate forecasts relative to point estimate forecasts. Similarly, participants (receivers) in Gaertig & Simmons (2018) (Studies 1-2) made predictions about the total number of points scored in professional basketball games and evaluated advice made by hypothetical advisors (senders) based on the predictions contained in their advice. Consistent with Du et al. (2011), they find that participants tend to prefer advisors who convey some uncertainty (by communicating their predictions in the form of range estimates) to those who express no uncertainty at all (by

¹When outcomes are categorical, minimum certainty is expressed as $1/(\text{number of options})$. In the case of a recommendation being correct or not, for which the number of options is two, minimum certainty is $1/2 = 50\%$.

communicating them as point estimates). In subsequent studies using binary rather than continuous outcomes, Gaertig & Simmons also find that people appear to be indifferent between advisors who give highly-certain advice and those who express uncertainty in a variety of other ways, such as by including a numerical probability or claiming that one outcome is “more likely” than another. In these studies, the dependent variables are scales composed of Likert items assessing features such as the perceived quality of the advice and the credibility of the advisors.

Several explanations have been proposed for the empirical pattern that receivers tend to prefer senders who express higher levels of certainty, including informativeness, aversion to ambiguity, and social factors (for reviews, see C. R. Fox & Tversky, 1995; Tenney, Spellman, & MacCoun, 2008). In the domain of evaluating competence, the pattern can be well explained by the *presumption of calibration hypothesis*, which posits that senders who express certainty are perceived to be more competent because receivers assume that senders are well-calibrated (in the absence of evidence to the contrary, Tenney et al., 2008). Calibration is “skill at correctly assigning probabilities” (Yaniv, Yates, & Smith, 1991) and reflects “insight into the quality of (one’s) own knowledge” (Tenney et al., 2008). A person who is well calibrated gives N% confidence intervals that contain the correct answer N% of the time and claims to be N% certain about the outcome of an event that occurs N% of the time. As such, if senders are well-calibrated, those who express more certainty are necessarily more competent than those who express less.

Critically, the presumption of calibration only holds in the absence of evidence that the sender is not well calibrated. In a series of compelling papers, Tenney, MacCoun, Spellman, & Hastie (2007; 2011; 2008) demonstrate that expressions of certainty can backfire when senders reveal themselves to be poorly calibrated, such as when a receiver observes a sender being correct at a rate that does not match the certainty that they’ve expressed. They propose that this is because receivers correctly identify that the combination of high certainty and low accuracy suggests poor calibration. In the prototypical study from this literature, participants make inferences about calibration by comparing senders’ statements of certainty with empirical truths (i.e., observed accuracy – in the prototypical study, a revealed error). For example, a witness could demonstrate imperfect calibration by claiming to be 90% certain about a series of events and being anything but 90% accurate.

The key insight of this paper is that there exist conditions under which calibration can be inferred in

the absence of information about a sender's accuracy. Specifically, I propose that when receivers hold sufficiently strong beliefs about the degree to which it is possible to know or predict a given outcome (i.e., that outcome's *epistemicness*) they make inferences about senders' calibration by comparing their own beliefs about epistemicness to the degree of certainty that a sender has expressed. Specifically, receivers will perceive senders as being less competent when they express an amount of certainty that the receiver perceives as being unlikely to be consistent with good calibration. For example, one need not watch the evening news to infer that a sender who claims to be certain that he will win the lottery is likely to be lacking in the ability to correctly assign probabilities – i.e., that he is poorly calibrated.

Note that inferred calibration is conceptually and empirically distinct from true calibration. True calibration is a function of a sender's expressed certainty and his observed accuracy and does not vary across receivers. Inferred calibration, on the other hand, is a function of a sender's expressed certainty and a receiver's beliefs. Because beliefs *do* vary across individuals, so too can inferred calibration. As such, I propose that receivers will perceive senders as being well-calibrated when those senders express a degree of certainty that is plausible given the receiver's beliefs about the epistemicness of the outcome in question, regardless of whether or not that degree of certainty also reflects true calibration. This is because epistemicness reflects beliefs about whether an outcome is knowable, and because to the extent to which an outcome is more knowable, it is more plausible that a receiver would be able to predict it with accuracy.

The meaning of uncertainty

There is an additional mechanism that may contribute to many of the results found throughout the literature and in this manuscript. This mechanism does not rely on the concept of calibration and instead relies on the way that perceived epistemicness moderates the way that receivers interpret the meaning of speakers' expressions of (un)certainty. To illustrate this mechanism, consider the receiver's perception of the meaning of "50%" in each of the following two statements from a hypothetical sender:

1. "There's a 50% chance that the urn contains mostly blue balls."

2. “There’s a 50% chance that the next ball drawn from the urn will be blue.”

The first statement is about an outcome which is in theory knowable in advance and is therefore the domain of purely epistemic uncertainty. As such, “50%” is directly informative about the speaker’s beliefs about the quality of their own knowledge, communicating that they are minimally certain about whether the urn contains mostly blue balls. In contrast, the second statement is more ambiguous because the statement involves two stages of uncertainty: first, uncertainty about the distribution of balls in the urn (which is epistemic), and second, the outcome of a single random draw (which is aleatory.) As such, it could reflect either that the speaker is completely uncertain about the distribution of balls in the urn (as above) or that they are completely certain that the distribution of balls in the urn is exactly 50% blue. Because either interpretation is valid, and because only one of them is informative with regard to the sender’s competence, the second statement offers a weaker signal about senders’ competence than does the first statement. Note that this mechanism can produce null findings on the relationship between certainty and competence, but cannot explain negative findings.

Overview of studies

In each of the studies reported in this paper, I manipulate two factors: (1) a characteristic of an outcome that makes it more or less epistemic and (2) the degree to which a sender expresses certainty. I assess receivers’ perceptions using survey items (Studies 1 - 3) and an incentive compatible choice of advisor (Study 4). In each study, I also capture participant beliefs about epistemicness using a combination of survey measures and incentivized judgments. Studies 1, 2, and 4 were preregistered.²

Pilot

To provide initial insight into the source of these divergent findings, I conducted a pilot study to characterize the type of uncertainty surrounding the outcomes in paradigms used in prior work. I recruited 350 participants from an online subject pool and had them rate the stimuli in prior

²The preregistration for Study 3 cannot be located. I plan on replicating the study before submitting the manuscript for publication.

work using the four-item version of the Epistemic-Aleatory Rating Scale (C. R. Fox, Tannenbaum, & Ülkümen, 2016). I selected seven studies from eight different papers, so approximately 50 participants rated each study in a between-subjects design. Following the methodology from prior work (C. R. Fox et al., 2016), scores were generated by subtracting aleatory items from epistemic items, such that positive values indicate a belief that uncertainty is more epistemic and negative numbers indicate a belief that uncertainty is more aleatory. Results revealed that all of the paradigms in which participants displayed a preference for senders who expressed more certainty involved predictions about outcomes that were relatively more epistemic in nature. In contrast, both of the paradigms in which participants displayed preferences for less-certain senders involved outcomes that were relatively more aleatory. This pilot study provides initial evidence consistent with the idea that perceived epistemicness shapes the relationship between sender certainty and receiver perceptions. Table 1 summarizes the results of this pilot study. Full details are available in the supplement.

Study 1

In Study 1, I examine whether the epistemicness of an environment affects preferences for advice that is conveyed with a high or low probability.

Participants

A total 408 participants (184 female, 2 other; age: $M = 37.16$ years, $SD = 12.27$) were recruited from Amazon Mechanical Turk.

Procedure

Studies 1 - 3 adapt a paradigm developed by Gaertig & Simmons (2018) in which participants make real, incentivized predictions with the aid of hypothetical advice. In Study 1, participants predicted the high temperature in New York City. Study 1 employs a 2 (epistemicness: high [1 day in advance] / moderate [1 week in advance]) X 2 (advice certainty: high [95%] / moderate [60%]) mixed design. All participants made predictions about the temperature both one day and one week

Table 1: Selected prior work

Paper(s)	Domain	Uncertainty type	Preference	Rated epistemicness*
Anderson, Brion, Moore, & Kennedy, 2012	Placing cities on an unmarked US map	Epistemic	Certainty	2.52
Vullioud, Clément, Scott-Phillips, & Mercier, 2017	Providing the name and office location of a colleague	Epistemic	Certainty	2.31
Snizek & Van Swol, 2001; Van Swol & Snizek, 2005	Answering questions about computer operations	Epistemic	Certainty	2.28
Yaniv & Foster, 1995	Answering factual questions about history	Epistemic	Certainty	1.80
Sah, Moore, & MacCoun, 2013	Estimating peoples' weights from pictures	Epistemic/both**	Certainty	1.17
Du, Budescu, Shelly, & Omer, 2011	Forecasting company financial performance	Both	Uncertainty	0.45
Gaertig & Simmons, 2018	Predicting the outcomes of sporting events	Both	Uncertainty /Mixed	-0.72

*Mean rated epistemicness from a pilot study in which participants from an online subject pool (N = 350) responded to questions drawn from (or read vignettes based on) each of the studies reported in this table and then completed the four-item version of the Epistemic-Aleatory Rating Scale (C. R. Fox et al., 2016). The study was conducted between-subjects, with 50 participants rating each stimulus. Scores were generated by subtracting aleatory items from epistemic items and averaging. The scale has a theoretical range of [-4, 4], with positive values indicating a belief that uncertainty is more epistemic and negative values indicating a belief that uncertainty is more aleatory. Full materials are available in the supplement.

**Participants estimated weights based on a picture. Although the weight of each person is in-principle knowable (i.e., epistemic), participants have relatively limited information from which to make an estimate. Limiting information introduces a degree of measurement error, which is a fundamental source of randomness. In related work (related to the manipulation check in Study 4), we find that introducing more measurement error makes people perceive a given estimation as being less epistemic.

in advance (in random order) and were randomly assigned to receive advice with either high or moderate certainty within each round. Prior to conducting Study 1, I conducted a pre-test of the epistemicness manipulation to confirm that people perceived the task of predicting the weather one week in advance to be less epistemic than the task of predicting the weather one day in advance.

Within each round, participants received advice and made a prediction on a single page. In the high (moderate) certainty condition, the sender said: “I am 95% (60%) confident that the high temperature... will be within one degree of X”.. The advice that all participants received was “good” in that it was generated by averaging the forecasts of three major public and private weather websites. After each round of receiving advice and making a prediction, participants rated the sender who provided the advice and the quality of the advice on the following page.

Measures

Advice / advisor quality: The dependent measure used in Studies 1 - 3 is a composite of three survey items: “How competent is this advisor?”; “How much do you trust this advisor?”; “Please rate the quality of this advice.”. The first two items use 7-item Likert response scales bounded by *not at all* and *extremely* and the third item uses a 7-item Likert response scale bounded by *lowest quality* and *highest quality*. These items were selected because they were the three highest-loading items from the 7-item composite measure used in Gaertig & Simmons (2018). The dependent measure is a simple average of the three items.

Epistemicness: Participants reported their beliefs about the epistemicness of each prediction using the six-item version of the Epistemic-Aleatory Rating Scale (EARS: C. R. Fox et al., 2016). The EARS measures beliefs in two dimensions of uncertainty: epistemic (uncertainty due to factors that are in-principle knowable) and aleatory (uncertainty due to factors that are random in nature). Example items for epistemic uncertainty include (the belief that the prediction...) “is knowable in advance, given enough information”, and “is something that well-informed people would agree upon”. Example items for aleatory uncertainty include (the belief that the prediction...) “is something that has an element of randomness”, and “could play out in different ways on similar occasions”. To generate an index of epistemicness, we reverse-code the aleatory items and average the items, as in Tannenbaum, Fox, & Ülkümen (2017).

Results

Epistemicness: Participants perceived predictions made one day in advance to be more epistemic ($M = 0.27$, $SD = 1.25$, 95% C.I. = $[0.10, 0.45]$) than predictions made one week in advance ($M = -0.47$, $SD = 1.27$, 95% C.I. = $[-0.64, -0.29]$; $t(403) = 5.90$, $p < .0001$, 95% C.I. = $[0.50, 0.99]$).

Advice / advisor quality: When making predictions one day in advance, participants displayed a preference for advisors who gave high certainty advice ($M = 4.58$, $SD = 1.43$, 95% C.I. = $[4.29, 4.86]$) over those who gave moderate certainty advice ($M = 3.97$, $SD = 1.32$, 95% C.I. = $[3.71, 4.23]$; $t(199) = 3.20$, $p < .01$, 95% C.I. = $[0.23, 0.99]$). However, when making predictions one week in advance, participants were indifferent between advisors who gave high ($M = 4.27$, $SD = 1.06$, 95% C.I. = $[4.06, 4.48]$) and low certainty advice ($M = 4.19$, $SD = 1.14$, 95% C.I. = $[3.97, 4.41]$; $t(203) = 0.51$, $p > .50$, 95% C.I. = $[-0.22, 0.38]$). A two-way mixed analysis of variance confirmed a significant epistemicness / advice certainty interaction ($f(396) = 7.53$, $p = .029$).

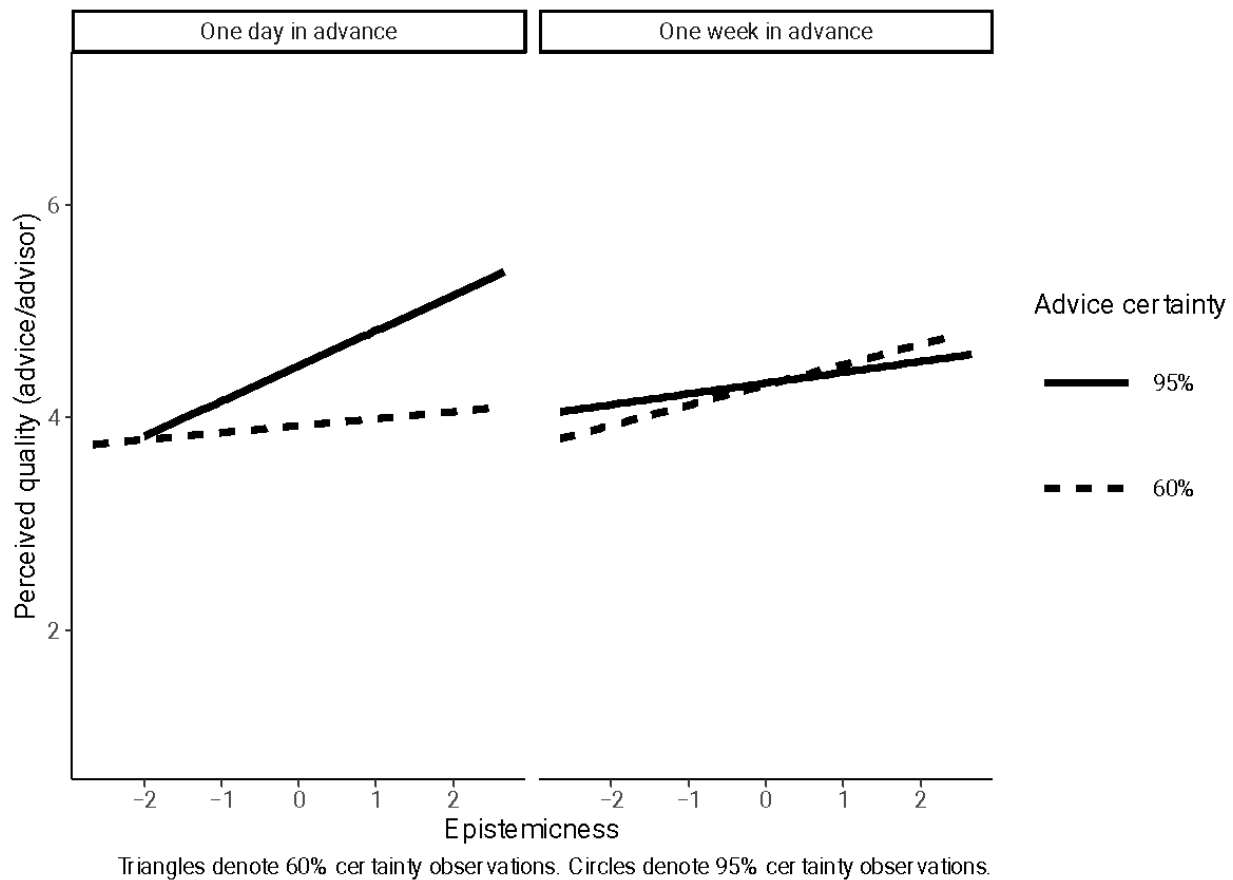


Figure 1: Study 1 epistemicness and quality

Discussion

Study 1 provided additional support for the hypothesis that the amount of certainty that people prefer is affected by the epistemicness of the environment. Participants displayed a preference for senders who expressed high levels of certainty when making predictions about the weather one day in advance and were indifferent between advisors who expressed high and moderate levels of certainty when making predictions one week in advance. Further, I found that the preference for high-certainty advisors when making predictions one day in advance was strongest among those participants who perceived the environment as being highly epistemic.

Study 2

Study 2 is intended to test the same hypothesis as Study 1 in a different context, with a between-subjects design, and using a more familiar mode of communicating uncertainty.

Participants

A total of 400 participants (184 female; age: $M = 37.12$ years, $SD = 11.10$) were recruited from Amazon Mechanical Turk. The task was advertised as being “for NFL football fans” and participants were required to answer two general football knowledge questions at the beginning of the task in order to proceed. Three participants failed both of two attention checks and were excluded from analysis, yielding a sample of 397.

Procedure

The procedure for Study 2 was similar to Study 1, with several notable differences. Participants made a binary (rather than continuous) prediction, the study used a between subjects (rather than mixed) design, and advice certainty was manipulated using verbal expressions of certainty (rather than numerical probabilities).

The study employed a 2 (epistemicness: high / low) X 2 (advice certainty: high / moderate) between-subjects design. Participants made a prediction about one of two anonymized American

Football games and received hypothetical advice about who the winner would be. The games were selected from a database of completed NFL games and were identical except for the amount of time remaining. In all conditions, the score was 21-10, the leading team had possession, and the leading and trailing teams previous-season records were 9-7 and 10-6 respectively. All participants received objectively accurate advice that the leading team would win.

To manipulate epistemicness, I varied the amount of time remaining in the game. In the high epistemicness condition, the game had two minutes remaining. In the low epistemicness condition, the game had three quarters remaining. To manipulate advice certainty, I varied whether the sender said that the leading team would “definitely win” (high certainty) or “probably win” (moderate certainty). Prior research demonstrated that people interpret the words “definitely” and “probably” as being indicative of high and moderate levels of certainty, respectively (e.g., Reagan, Mosteller, & Youtz, 1989; Wallsten, Budescu, & Zwick, 1993). A preregistration can be found at https://aspredicted.org/QKG_RHF.

Measures

All measures used in Study 2 were the same as those used in Study 1.

Results

Epistemicness: Participants in the high epistemicness condition perceived the environment as being more epistemic ($M = 1.14$, $SD = 1.38$, 95% C.I. = [0.95, 1.34]) than did participants in the low epistemicness condition ($M = -0.95$, $SD = 1.31$, 95% C.I. = [-1.14, -0.77]; $t(391) = 15.0$, $p < .0001$, 95% C.I. = [1.83, 2.36]).

Advice / advisor quality: As in Study 1, participants in the high epistemicness condition (predicting with two minutes remaining) displayed a preference for advisors who gave highly certain (“definitely”) advice ($M = 4.80$, $SD = 1.15$, 95% C.I. = [4.57, 5.03]) over those who gave moderately certain (“probably”) advice ($M = 4.15$, $SD = 1.49$, 95% C.I. = [3.85, 4.45]; $t(192) = 3.40$, $p < .0001$, 95% C.I. = [0.27, 1.03]). However, participants in the low epistemicness condition (predicting with three quarters remaining) displayed a preference for advisors who gave less certain (“probably”)

advice ($M = 4.27$, $SD = 1.18$, 95% C.I. = [4.04, 4.50]) over those who gave highly certain (“definitely”) advice ($M = 3.42$, $SD = 1.14$, 95% C.I. = [3.20, 3.65]; $t(201) = 3.40$, $p < .0001$, 95% C.I. = [0.27, 1.03]). A two-way analysis of variance confirmed a significant epistemicness / advice certainty interaction ($f(393) = 35.95$, $p < .0001$).

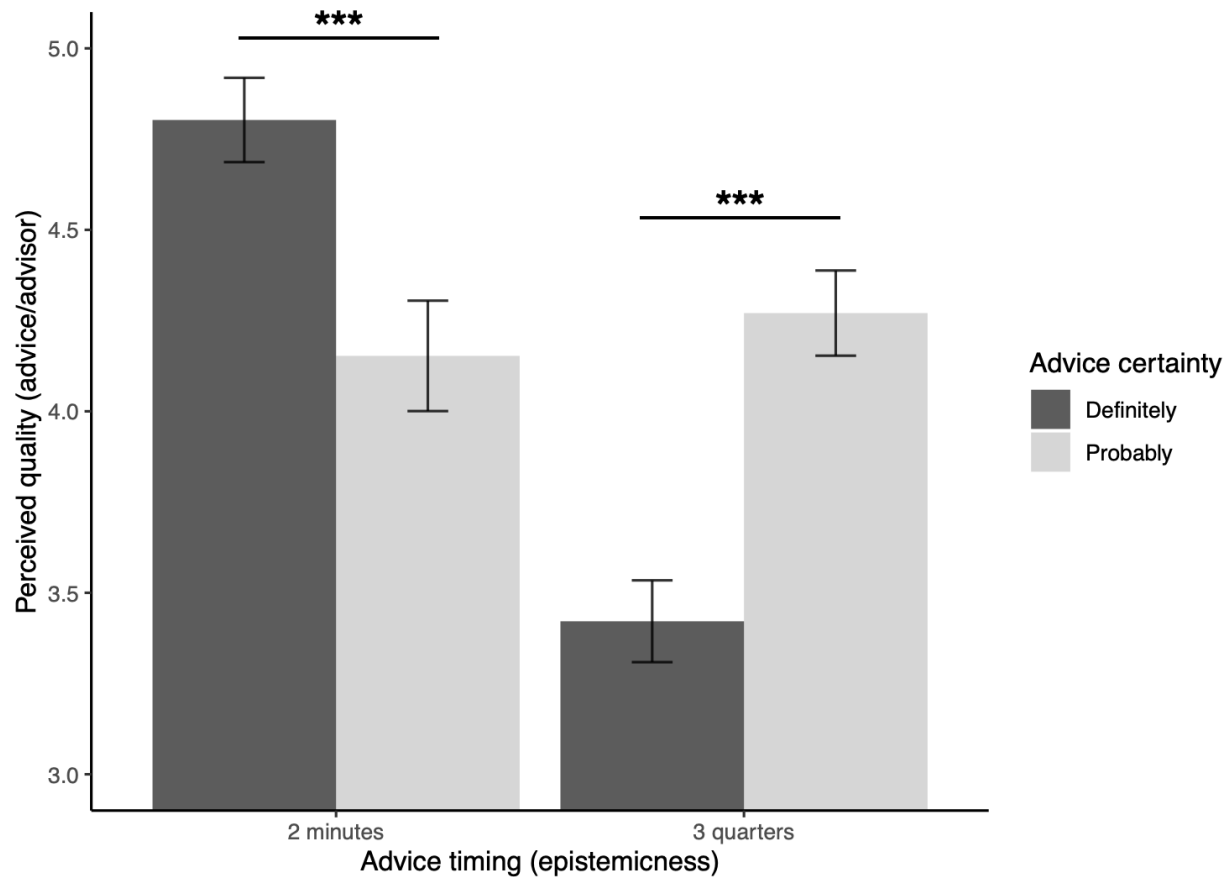


Figure 2: Study 2 epistemicness and quality

Discussion

Study 2 replicated the interaction between epistemicness and advice certainty. I found a full reversal of preferences for advice certainty as a function of epistemicness: whereas there was no condition in Study 1 in which participants preferred less-certain advice to more certain advice, participants in the low epistemicness condition in Study 2 displayed a strong preference for senders who used less certain “probably” language over those who used more certain “definitely” language. Study 2 also demonstrated that people are sensitive to the degree of certainty that people convey using natural

language, which is a particularly common way for people to communicate uncertainty (Budesu & Wallsten, 1995; Reagan et al., 1989; Wallsten et al., 1993).

Study 3

In Study 3, I sought to directly assess whether perceptions of calibration underlie the pattern that receivers prefer senders who are more (less) confident when those receivers believe that an outcome is relatively more epistemic (aleatory).

Participants

A total of 603 participants were recruited from Prolific Academic.

Procedure

The procedure for Study 3 was very similar to that of Study 1, with two key differences. First, speakers communicated their degree of certainty using either point estimates (higher certainty) or range estimates (lower certainty). Second, I included a short scale to assess beliefs about the sender's calibration.

Participants predicted the high temperature in Washington DC with the aid of hypothetical advice. Study 3 employs a 2 (epistemicity: high [1 day in advance] / low [1 month in advance]) X 2 (advice certainty: high [point estimate] / low [range estimate]) between-subjects design. Participants received advice and made a prediction on a single page. In the high certainty condition, the sender said: "The high temperature one month from today ([date]) will definitely be exactly [N] degrees.". In the low certainty condition, the sender said: "The high temperature one month from today ([date]) will probably fall somewhere between [N-10] and [N+10] degrees." The point estimate of the advice that all participants received was "good" in that it was generated by averaging the forecasts of three major public and private weather websites. After making their predictions, participants first completed the perceived calibration scale and then completed the dependent measures.

Measures

The measures are identical to those used in Studies 1-2, with two exceptions. First, to assess receivers' beliefs about senders' calibration, I developed a 3 item scale ($\alpha = 0.88$, 95% C.I. = [0.86 – 0.89]). For all items, participants were asked to respond on a 7-point Likert scale, with the lowest point (1) labeled as “not at all”, the highest point (7) labeled as “extremely” and the midpoint (4) labeled as “moderately.” I did not collect the EARS scale, to avoid prompting people to think about anything in particular before responding to the calibration scale items.

Table 2: Perceived calibration scale

Name	Text	Mean (sd)	α if dropped
Self awareness	How self-aware are they about their ability to predict the weather?	4.01 (1.45)	0.83
Understanding	How well do you think they understand how good they are at predicting the weather?	4.02 (1.44)	0.80
Assigning probabilities	When predicting the weather, how good are they at correctly assigning probabilities ?	3.87 (1.38)	0.84
<i>Calibration scale (average of items)</i>		4.06 (1.30)	

Results

Calibration: Participants in the high epistemicness condition (predicting tomorrow's temperature) rated highly certain advisors as better calibrated ($M = 3.97$, $SD = 1.44$, 95% C.I. = [3.75, 4.19]) than moderately certain advisors ($M = 3.66$, $SD = 1.06$, 95% C.I. = [3.47, 3.84]; $t(295) = 2.16$, $p = .032$). Participants in the low epistemicness condition (predicting the temperature one month out) directionally reversed this pattern, rating moderately certain advisors as better calibrated ($M = 4.22$, $SD = 0.96$, 95% C.I. = [4.08, 4.36]) than highly certain advisors ($M = 3.95$, $SD = 1.51$, 95% C.I. = [3.69, 4.20]), although this simple effect was not significant ($t(214) = 1.83$, $p = .069$). A two-way analysis of variance confirmed a significant epistemicness / advice certainty interaction ($F(1, 599) = 8.05$, $p = .005$; Figure 3).

Advice / advisor quality: Participants in the low epistemicness condition preferred moderately certain advisors ($M = 4.42$, $SD = 0.96$, 95% C.I. = [4.27, 4.56]) to highly certain advisors ($M = 3.81$,

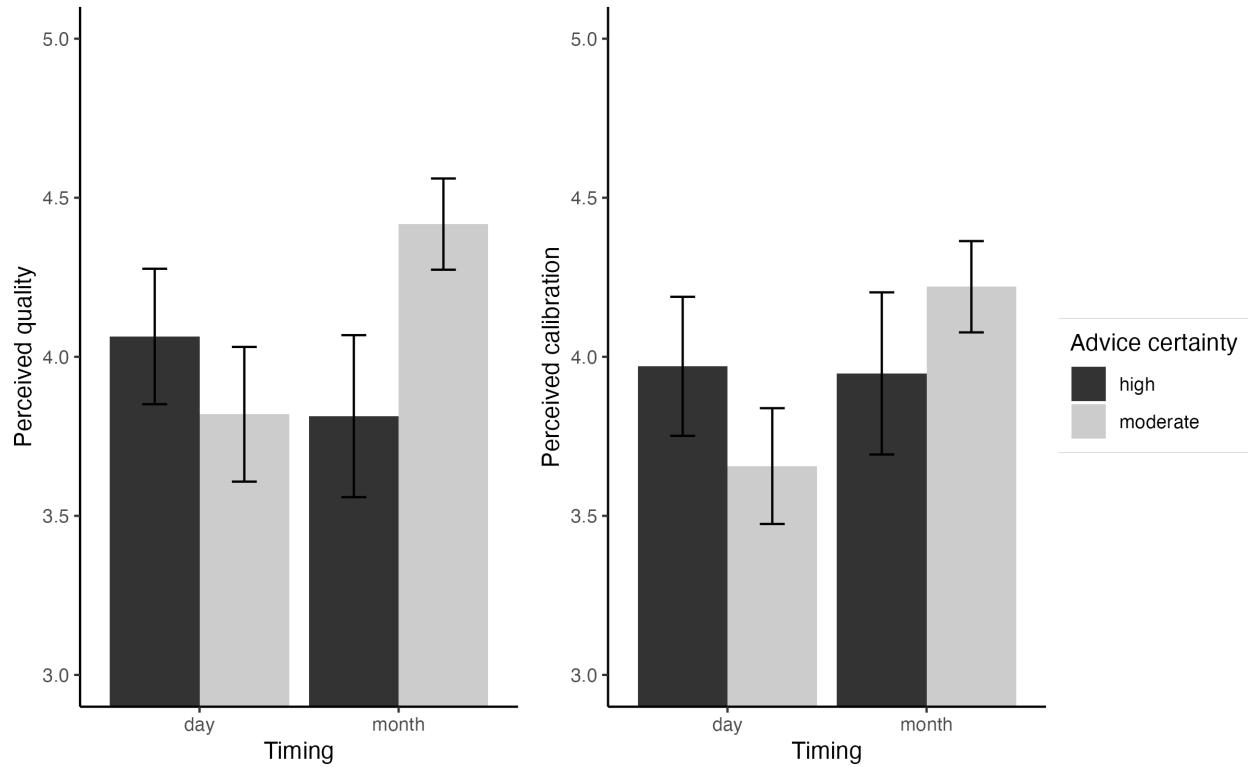


Figure 3: Calibration and dependent measures, Study 3. Vertical lines are 95% CIs

SD = 1.50, 95% C.I. = [3.56, 4.07]; $t(214) = 4.05$, $p < .001$). In the high epistemicness condition, the difference favored highly certain advisors ($M = 4.06$, SD = 1.40, 95% C.I. = [3.85, 4.28]) over moderately certain advisors ($M = 3.82$, SD = 1.24, 95% C.I. = [3.61, 4.03]) but was not significant ($t(292) = 1.60$, $p = .11$). A two-way analysis of variance confirmed a significant epistemicness / advice certainty interaction ($F(1, 599) = 16.29$, $p < .001$; Figure 3). Perceived calibration was strongly correlated with quality ratings ($r = .77$) and predicted them ($b = 0.78$, $t = 29.61$, $p < .001$) in a model that controlled for both manipulations and their interaction.

Study 4

In the prior studies reported in this paper, participants received hypothetical advice and rated advisors using survey items. In Study 4, I sought to investigate the same hypothesis in a paradigm in which participants had the opportunity to evaluate real expressions of uncertainty and to make consequential judgments about others' competence.

Participants

The focal sample contained 276 participants (150 female; age: $M = 19.66$ years, $SD = 1.16$) who were recruited through a university behavioral lab in exchange for course credit and the opportunity to earn bonuses. 25 participants were excluded based on a set of preregistered criteria (described in the supplementary materials) yielding a sample of 251. Prior to conducting the focal study, we recruited a sample of 180 prospective teammates (84 female; age: $M = 38.02$ years, $SD = 11.81$) from Amazon Mechanical Turk. From these 180, I manually selected 20 to serve as teammates for participants in the focal study.

Procedure

Participants made a series of estimates about the number of objects in an image. The study employed a 2 (epistemicness: high / low) X 2 (teammate certainty: high / moderate) mixed design, with epistemicness manipulated between subjects and teammate certainty manipulated within subjects – both at the round level.

The study proceeded as follows: in each of five estimation rounds, participants made an initial estimate about the number of objects in an image. Participants submitted all estimates in the form of a point estimate and a margin of error, which was described such that it represented a 90% confidence interval. After making an initial estimate, participants chose one of two possible teammates whose estimates they would learn prior to submitting a final estimate. While making this decision, participants were informed only of each prospective teammate's margin of error, as shown in Figure 5. After selecting a teammate, participants learned the teammate's point estimate and submitted their final estimate.

$$\text{payoff} = \begin{cases} \max(10 - \text{margin of error}, 1) & \text{if correct} \\ 0 & \text{if incorrect} \end{cases}$$

Figure 4: Points payoff function, Study 4

For both estimates, participants were incentivized using a simplified scoring rule adapted from Levin, Gilbertson, & Broomell (2019). Participants earned points only if the correct answer fell

within the range bounded by $[point\ estimate - margin; point\ estimate + margin]$. Conditional on earning points for being correct, the size of participants' payoff diminished with the size of their margin of error, such that the maximum payoff of ten points was obtained with a margin of error of zero and the minimum payoff of one point was obtained with a margin of error of nine or more. Participants were informed that one of their two estimates would be enforced to determine the number of points earned in each round. Points were used to allocate bonus payments, with the top 20 points earners receiving a cash bonus payment of \$5.

On this page, you'll pick a teammate for this estimation. On the next page, you'll see the estimate of the teammate that you picked.

Teammate	Estimate	Margin of error
Teammate E	Hidden	+ / - 3 ducklings
Teammate F	Hidden	+ / - 1 ducklings

The margin of error means that they were 90% sure that their estimate was accurate to plus or minus that number.

Which teammate do you want for this estimate?

Teammate E
☐

Teammate F
☐

Figure 5: Study 4 teammate choice

Epistemicness manipulation: To manipulate perceived epistemicness, I added measurement error to an otherwise epistemic estimation task by randomly altering image blur within each estimation round.³ In an out-of-sample pretest, we find that people perceive the task of estimating the number of

³Whether or not measurement error changes the type of uncertainty that characterizes an inquiry or merely its level

objects in high blur stimulus images to be less epistemic ($M_{\sim\text{epistemicness}} = -0.21$, $SD = 0.96$, 95% C.I. = $[-0.34, -0.09]$) than estimating the number of objects in low blur images ($M_{\sim\text{epistemicness}} = 0.86$, $SD = 0.92$, 95% C.I. = $[0.75, 0.98]$; $t(490) = 12.7$, $p < .0001$, 95% C.I. = $[0.91, 1.24]$).⁴ For comparison, high-blur estimates were perceived to be *less* epistemic than predicting whether Trump would be re-elected in 2020 (-0.14 ⁵) and low blur estimates were perceived as being nearly as epistemic as estimating the time that it takes for an object to fall from a building of known height (1.10).

Teammates & certainty manipulation: The certainty manipulation was administered within subjects, with each participant choosing between one of two teammates: one who expressed “high” certainty and one who expressed “moderate” certainty. High certainty teammates were selected from the 99th percentile of certainty (*mean margin of error* = 0.3, *range* = $[0, 1]$) and moderate certainty teammates were selected from the 60th percentile of certainty (*mean margin of error* = 2.5, *range* = $[1, 4]$). For all rounds except for one, teammates in both certainty conditions provided identical point estimates, a feature which was intended to minimize the possibility that participants would develop any inferences about a relationship between certainty and accuracy.

Measures

Certainty: (Participant) certainty was operationalized using participants’ initial margin of error for each estimate.

Teammate choice: For each round, teammate choice was coded as a binary variable (high certainty / moderate certainty).

Risk: Participants indicated their risk attitudes by responding to a single item developed by Dohmen et al. (2011). The item asks “Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?” with participants responding on a 10-point scale bounded by “not at

is both a philosophical question and an empirical one. For the purposes of this manuscript, we eschew the former and present evidence on the latter. The question of why measurement error is perceived as a source of aleatory uncertainty is beyond the scope of this project.

⁴Images were sourced in unblurred format and blur was added using the *imager* package for R. <https://cran.r-project.org/web/packages/imager/imager.pdf>

⁵Estimates of the epistemicness of Trump being re-elected were made by participants in February and March of 2019.

Image	Blur	60th percentile	99th percentile
		MoE (moderate)	MoE (high)
alpacas	Low	2	0
	High	3	0
dancers	Low	2	0
	High	2	0
ducklings	Low	2	0
	High	3	1
dumplings	Low	1	0
	High	4	1
penguins	Low	3	1
	High	3	0
Average		2.5	0.3

Figure 6: Study 4 certainty manipulation

all willing to take risks” and “very willing to take risks”.

Results

Certainty^{*6}:* Participants expressed more certainty (as indicated by smaller margins of error) when making estimates about low blur images ($M = 1.75$, $SD = 2.15$, 95% C.I. = [1.58, 1.81]) than when making estimates about high blur images ($M = 3.72$, $SD = 3.06$, 95% C.I. = [3.48, 3.96]; $t(1124) = 13.0$, $p < 0.0001$, 95% C.I. = [1.69, 2.27]). These results are based on initial margins of error (reported prior to learning about teammates) but the results are similar when using final margins (captured after learning about teammates) or an average of the two.

Choice: Consistent with the hypothesized pattern, the proportion of participants choosing a high certainty teammate was higher in the low blur condition (.63, 95% C.I. = [.59, .67]) than in

⁶Because this variable contained several extreme values, it was winsorized at the 95th percentile. Prior to winsorizing, the mean and standard deviation were 81.49 and 2727.82, respectively. After winsorizing: 3.41 and 4.37. This step was not preregistered.

the high blur condition (.45, 95% C.I. = [.41, .49]). Because each participant made a series of choices over the course of the study, we conducted primary analysis using logistic regression with standard errors clustered at the participant level. The primary specification that we estimate is:

$$choice_{ir} = \alpha + \beta_1 blur_{ir} + \beta_2 risk_i + \beta_3 female_i + \beta_4 image_r + \varepsilon_i.$$

Participants were significantly less likely to select a high certainty teammate in high blur rounds than in the low blur rounds ($\beta_1 = -0.74$, $z = -6.34$, $p < 0.0001$). The odds ratio for β_1 is 0.48 (95% C.I. = [0.37, 0.60]), indicating that participants were roughly half as likely to select the high certainty teammate in high blur rounds compared to low blur rounds. There was no association between choices and either risk or gender ($ps > 0.25$ for both), and the results for blur are robust to excluding both of these variables from analysis (for the specification excluding risk and gender: $\beta_1 = -0.74$, $z = -6.32$, $p < 0.0001$; see supplemental materials for additional robustness checks).

Because we capture participants' certainty about each estimate in the form of a margin of error before they learn anything about their teammates, we can additionally ask whether participants' own subjective certainty is predictive of whether they will choose a high versus low certainty teammate. If hypothesis two is correct, participants who themselves submit larger margins of error should in turn prefer teammates also express larger margins of error. To investigate this question, we regressed choice of teammate on initial margin of error. For ease of interpretation and to reduce kurtosis in the margin of error variable, these analyses use the $\log(2)$ of margin of error, but results are robust to using raw margin of error instead.

Collapsing across images and blur levels, we find that participants who submitted larger margins of error (indicating a belief that the estimate involved more uncertainty) were less likely choose the high certainty teammate ($\beta = -1.28$, $z = -5.81$, $p < .00001$). The odds ratio for this estimate is 0.28 (95% C.I. = [0.18, 0.43]). Because the log was taken with base 2, this means that a two-fold increase in margin of error (doubling, from a mean of 3.43) was associated with a 72% reduction in likelihood of selecting the high certainty teammate. This association between subjective certainty and teammate choice is visible in Figure 7 and Figure 8.

Figure 7 plots the proportion of participants who chose the high (versus low) certainty teammate within each decile of margin of error. Participants in the highest quantile of subjective certainty (median margin of error = 1) chose the high certainty teammate 76% of the time (95% C.I. = [71%,

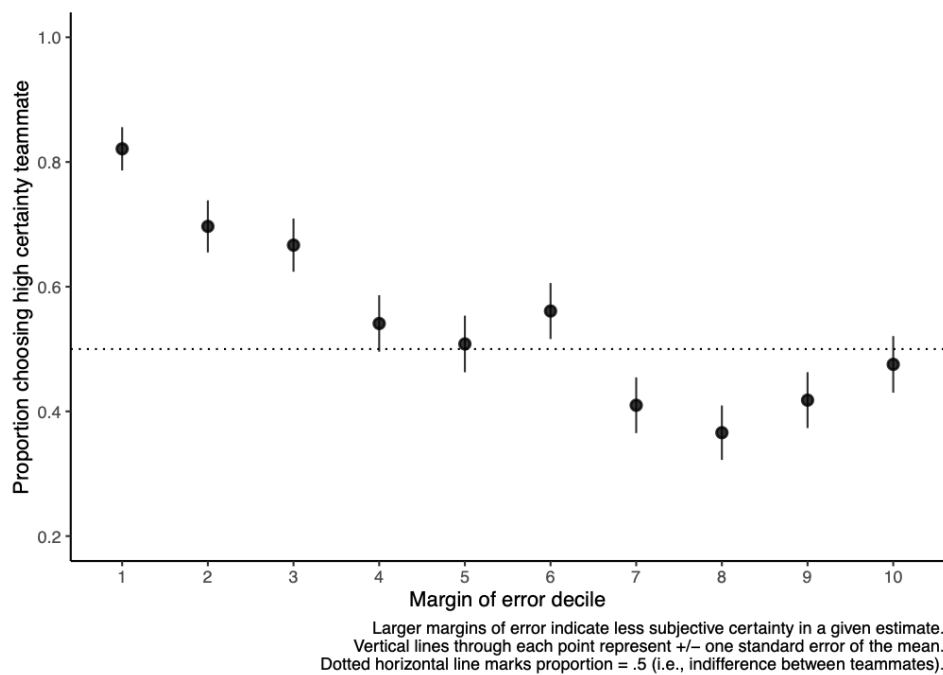


Figure 7: Study 4: proportion choosing high certainty teammate, by margin of error decile

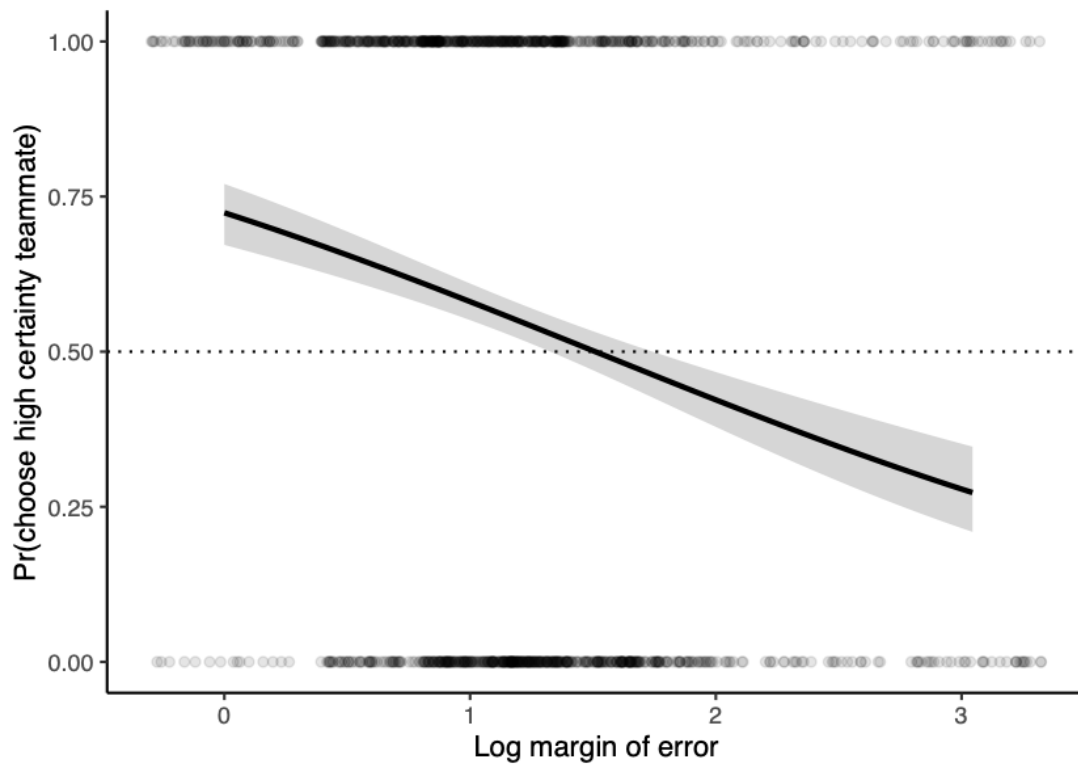


Figure 8: Study 4: fitted probability of choosing high certainty teammate, by log margin of error

81%]), whereas teammates in the lowest quantile (median margin of error = 6) chose the high certainty teammate only 45% of the time (95% C.I. = [38%, 51%]). The difference is even more pronounced when looking at participants in the second-lowest quantile (median margin of error = 3), who selected the high certainty teammate only 39% of the time (95% C.I. = 33%, 45%]). Figure 8 plots (jittered) individual participant choices along with fitted probabilities from the above specified logistic regression (without controls).

Discussion

In Study 4, I found that participants were more likely to seek information from others when those others demonstrated better calibration. Not only were participants more likely to select highly-certain teammates when they perceived estimations to be more epistemic, but they were also more likely to select less-certain teammates when they perceived estimations to be less epistemic. We identify this effect both by exogenously manipulating epistemicness (by altering blur level between subjects) and endogenously by measuring beliefs about epistemicness (by eliciting margins of error).

It should be noted that the only additional information that participants received by selecting a teammate was that teammate's point estimate, and that the only normative basis on which to prefer one point estimate over another is the belief that it is more accurate. Although it was not hypothesized, these results offer some preliminary evidence that people may perceive a direct association between calibration and accuracy.

General discussion

Across 4 studies, I find support for the hypothesis that the relationship between the amount of certainty that a sender expresses and the degree to which they are perceived as being competent depends on the receiver's beliefs about the epistemicness of the outcome in question. In each study, receivers preferred highly-certain senders in environments that were constructed to be highly-epistemic, but this pattern disappeared (S1) or reversed (S2 – S4) in less-epistemic environments. A significant interaction between epistemicness and preference for certainty was observed in all four studies.

Analysis of participants' subjective certainty allowed for further inference about the mechanism responsible for this pattern. In Study 4, participants' subjective certainty (as indicated by the margins of error of their estimates) predicted which teammate they would choose.

The findings reported in this paper offer a coherent account of why previous findings on the effects of expressing certainty on perceptions of competence have been ambiguous. These findings also have a variety of implications for contexts in which people communicate with others about uncertainty. For example: we know that people are hesitant to express uncertainty, that the failure to fully convey felt-uncertainty negatively impacts decision quality in collaborative environments, and that one factor driving this hesitation is the concern that expressing uncertainty will make one appear less-competent (R. C. Fox, 2003; Healy & Moore, 2007; Klayman, Soll, González-Vallejo, & Barlas, 1999; Simpkin & Schwartzstein, 2016; Snizek, 1992; Soll & Klayman, 2004; Soll, Palley, Klayman, & Moore, 2019). The knowledge that there are conditions in which expressing more certainty can make a person appear more competent may serve to mitigate this tendency.

Conceptually, the present work is perhaps most similar to Pescetelli & Yeung (2021), who convincingly show that receivers evaluate senders' *accuracy* by comparing the certainty that those senders express with their own internal confidence about the same outcome. However, one notable limitation of this model is that it can only describe how receivers can make well-founded inferences about senders in cases in which the sender and receiver have comparable knowledge about the outcome of interest and in which a receiver is sufficiently confident about their own estimate. As differences in knowledge across parties increase, receivers' confidence becomes a less reliable signal of senders' accuracy. A focus on perceived epistemicness allows the approach described in this paper to be more flexible and to describe a broader set of real-world phenomena. In the model described here, a receiver can themselves have low certainty in an outcome and still evaluate a highly-confident sender favorably if and only if they believe that it is possible for others to become more certain with more knowledge (i.e., for the outcome to be relatively epistemic in nature.)

There are several characteristics of these studies that limit their generalizability. First, each of the studies manipulates uncertainty across only two levels, which limits the ability to offer predictions about the *amount* of uncertainty that people will prefer as a function of their beliefs about epistemicness. Future studies should attempt to parameterize models such as the one that I have proposed

in this paper. The studies are also limited in that they employ either hypothetical expressions of certainty (Studies 1-3) or highly stylized expressions of certainty (Study 4). Future work should seek to study these questions in more natural environments. Finally, the moderator that I've tested is unlikely to be the only one that affects the relationship between senders' certainty and receivers' perceptions of their competence.

This work suggests that it is not always bad to express uncertainty when communicating with others. Instead, whether or not expressions of uncertainty are received positively depends on the receivers' beliefs about the environment. Mark Twain is reported to have said that "It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.". These results suggest that his insight may apply not only to the outcomes of one's decisions, but also to how one is perceived by others.

References

- Alzahawi, S., & Flynn, F. J. (2025). Does expressing uncertainty help or harm leaders? *The Leadership Quarterly*, 36(5), 101880. <https://doi.org/10.1016/j.leaqua.2025.101880>
- Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, 103(4), 718–735. <https://doi.org/10.1037/a0029395>
- Bles, A. M. van der, Linden, S. van der, Freeman, A. L. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, 117(14), 7672–7683. <https://doi.org/10.1073/pnas.1913678117>
- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In *Psychology of learning and motivation* (Vol. 32, pp. 275–318). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60313-8](https://doi.org/10.1016/S0079-7421(08)60313-8)
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3), 522–550. <https://doi.org/10.1111/j.1542-4774.2011.01015.x>
- Du, N., Budescu, D. V., Shelly, M. K., & Omer, T. C. (2011). The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, 114(2), 179–189. <https://doi.org/10.1016/j.obhdp.2011.05.001>

[//doi.org/10.1016/j.obhdp.2010.10.005](https://doi.org/10.1016/j.obhdp.2010.10.005)

- Fox, C. R., Tannenbaum, D., & Ülkümen, G. (2016). *The empirical case for distinguishing two dimensions of subjective uncertainty*.
- Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, 110(3), 585–603.
- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on thinking, judging, and decision making* (pp. 21–35). Oslo: Universitetsforlaget.
- Fox, R. C. (2003). Medical uncertainty revisited. In G. L. Albrecht, R. Fitzpatrick, & S. C. Scrimshaw (Eds.), *The handbook of social studies in health and medicine*. SAGE.
- Gaertig, C., & Simmons, J. P. (2018). Do people inherently dislike uncertain advice? *Psychological Science*, 29(4), 504–520. <https://doi.org/10.1177/0956797617739369>
- Gaertig, C., & Simmons, J. P. (2023). Are people more or less likely to follow advice that is accompanied by a confidence interval? *Journal of Experimental Psychology: General*, 152(7), 2008–2025. <https://doi.org/10.1037/xge0001388>
- Healy, P. J., & Moore, D. A. (2007). *The trouble with overconfidence*.
- Karmarkar, U. R., & Tormala, Z. L. (2010). Believe me, I have no idea what I’m talking about: The effects of source certainty on consumer involvement and persuasion. *Journal of Consumer Research*, 36(6), 1033–1049. <https://doi.org/10.1086/648381>
- Kennedy, J. A., Anderson, C., & Moore, D. A. (2013). When overconfidence is revealed to others: Testing the status-enhancement theory of overconfidence. *Organizational Behavior and Human Decision Processes*, 122(2), 266–279. <https://doi.org/10.1016/j.obhdp.2013.08.005>
- Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79(3), 216–247. <https://doi.org/10.1006/obhd.1999.2847>
- Levin, J. M., Gilbertson, E., & Broomell, S. B. (2019). *An incentive compatible elicitation of subjective probability density functions*.
- Løhre, E., & Teigen, K. H. (2024). When leaders disclose uncertainty: Effects of expressing internal and external uncertainty about a decision. *Quarterly Journal of Experimental Psychology*, 77(6), 1221–1237. <https://doi.org/10.1177/17470218231204350>

- Pescetelli, N., & Yeung, N. (2021). The role of decision confidence in advice-taking and trust formation. *Journal of Experimental Psychology: General*, 150(3), 507–526. <https://doi.org/10.1037/xge0000960>
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57. <https://doi.org/10.1002/bdm.460>
- Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, 74(3), 433–442.
- Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, 121(2), 246–255. <https://doi.org/10.1016/j.obhdp.2013.01.007>
- Simpkin, A. L., & Schwartzstein, R. M. (2016). Tolerating uncertainty — the next medical revolution? *New England Journal of Medicine*, 375(18), 1713–1715. <https://doi.org/10.1056/NEJMp1606402>
- Snizek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, 52(1), 124–155. [https://doi.org/10.1016/0749-5978\(92\)90048-C](https://doi.org/10.1016/0749-5978(92)90048-C)
- Snizek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307. <https://doi.org/10.1006/obhd.2000.2926>
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 299–314. <https://doi.org/10.1037/0278-7393.30.2.299>
- Soll, J. B., Palley, A. B., Klayman, J., & Moore, D. A. (2019). *Overconfidence in probability distributions: People know they don't know but they don't know what to do about it*.
- Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. Aleatory uncertainty. *Management Science*, 63(2), 497–518. <https://doi.org/10.1287/mnsc.2015.2344>
- Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, 18(1), 46–50. <https://doi.org/10.1111/j.>

1467-9280.2007.01847.x

- Tenney, E. R., Small, J. E., Kondrad, R. L., Jaswal, V. K., & Spellman, B. A. (2011). Accuracy, confidence, and calibration: How young children and adults assess credibility. *Developmental Psychology*, 47(4), 1065–1077. <https://doi.org/10.1037/a0023273>
- Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, 44(5), 1368–1375. <https://doi.org/10.1016/j.jesp.2008.04.006>
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, 145(10), 1280–1297. <https://doi.org/10.1037/xge0000202>
- Van Swol, L. M., & Snizek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology*, 44(3), 443–461. <https://doi.org/10.1348/014466604X17092>
- Vulliamd, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior*, 38(1), 9–17. <https://doi.org/10.1016/j.evolhumbehav.2016.06.002>
- Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39(2), 176–190.
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4), 424–432.
- Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110(3), 611–617.

Appendix

How much did the US federal government spend on education in 2015?

Answer in billions of dollars.

The answer to this question...

	Not at all			Very much	
is something that has an element of randomness.	☐	☐	☐	☐	☐
is determined by chance factors.	☐	☐	☐	☐	☐
is knowable in advance, given enough information.	☐	☐	☐	☐	☐
is something that well-informed people would agree on.	☐	☐	☐	☐	☐

Figure A1: Example of pilot stimulus

Imagine that you are attending an NBA basketball game and want to predict how many total points will be scored by both teams.

The answer to this question...

	Not at all				Very much
is something that has an element of randomness.	☐	☐	☐	☐	☐
is determined by chance factors.	☐	☐	☐	☐	☐
is knowable in advance, given enough information.	☐	☐	☐	☐	☐
is something that well-informed people would agree on.	☐	☐	☐	☐	☐

Figure A2: Study 2 epistemicness measure

Team A at Team B					
	1ST	2ND	3RD	4TH	SCORE
Team A (9-7)	14	-	7		21
Team B (10-6)	3	-	7		10
Team A			Team B		

**Epistemicness
(high)**

Game Status:

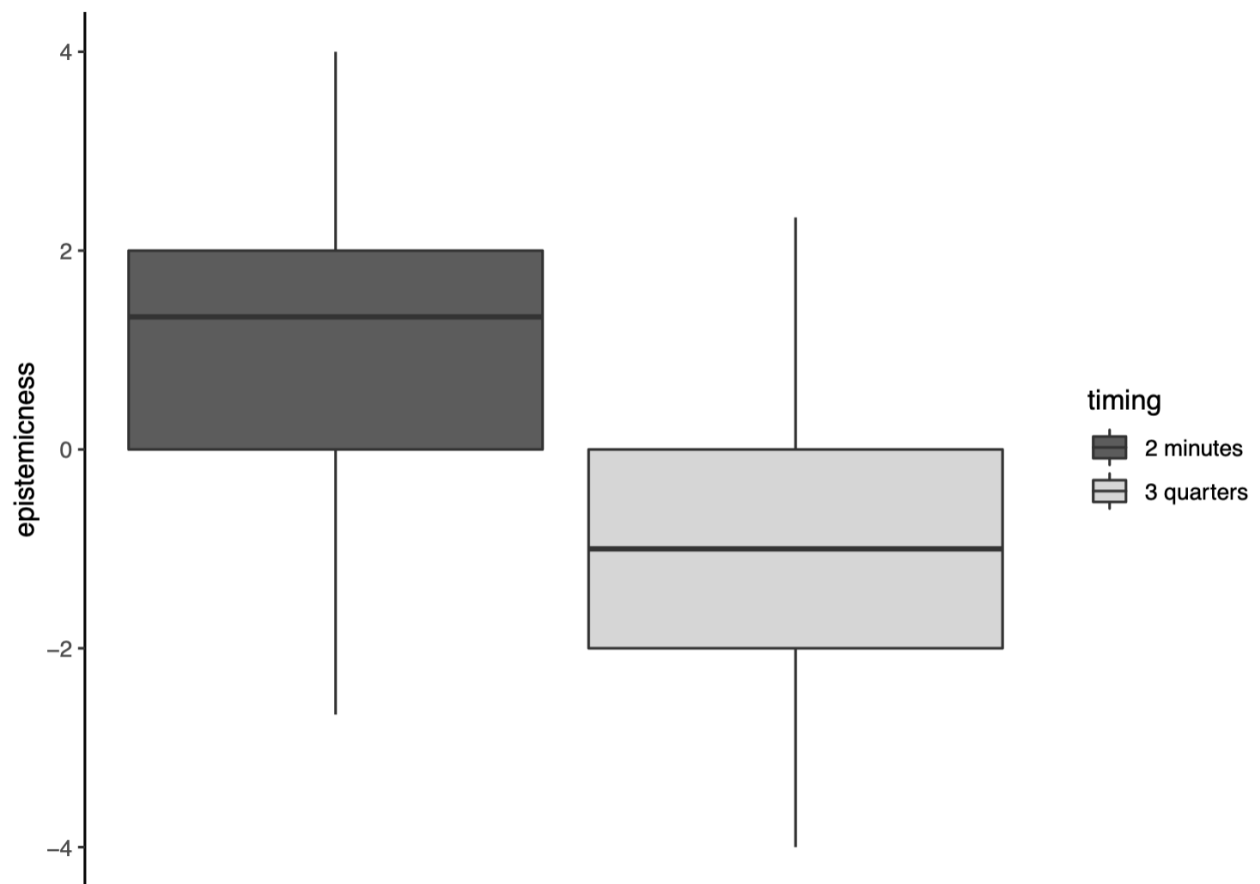
Two minute warning (2:00 remaining in regulation)
Team A has the ball on their own 27
2nd down and 8

Imagine having received the following advice:

**Certainty
(high)**

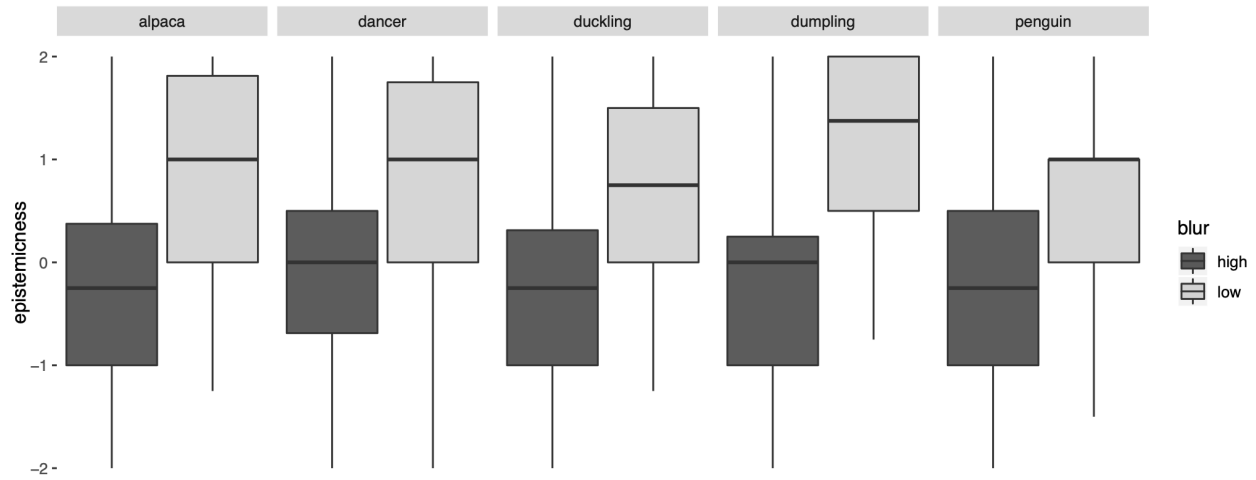
"Team A will definitely win. There's no way they can lose."

Figure A3: Study 2 stimuli



$\Delta = 2.09$; $t(391) = 15.0$, $p < .0001$, 95% C.I. = [1.83, 2.36]

Figure A4: Study 2 manipulation check



For all images:
 $\Delta = 1.07$; $t(490) = 12.7$, $p < .0001$, 95% C.I. = [0.91, 1.24]

Figure A5: Study 4 manipulation check

Chose high certainty teammate				
	(1)	(2)	(3)	(4)
Log margin of error	-1.277*** (0.220)	-1.318*** (0.221)	-1.354*** (0.228)	-1.398*** (0.230)
Constant	0.964*** (0.145)	1.275*** (0.305)	0.891*** (0.179)	1.204*** (0.323)
Individual controls	No	Yes	No	Yes
Image fixed effects	No	No	Yes	Yes

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Logistic regression with standard errors clustered at the participant level.

Standard errors in parentheses.

Figure A6: Robustness checks: choice of high-certainty teammate by log margin of error (logistic regressions)

	Chose high certainty teammate			
	(1)	(2)	(3)	(4)
Blur = high	−0.735*** (0.117)	−0.738*** (0.117)	−0.739*** (0.117)	−0.741*** (0.117)
Constant	0.549*** (0.104)	0.792*** (0.290)	0.495*** (0.142)	0.739** (0.305)
Individual controls	No	Yes	No	Yes
Image fixed effects	No	No	Yes	Yes

*p<0.1; **p<0.05; ***p<0.01

Logistic regression with standard errors clustered at the participant level.

Standard errors in parentheses.

Figure A7: Robustness checks: choice of high-certainty teammate by blur condition (logistic regressions)