

Maintaining credibility when communicating uncertainty: the role of directionality

Sarah C. Jenkins^a and Adam J. L. Harris^a

^a Department of Experimental Psychology, University College London. 26 Bedford Way, London, WC1H 0AP.

Sarah C. Jenkins was funded by an IMPACT studentship from UCL, half funded by the British Geological Survey (BGS).

Correspondence concerning this article should be addressed to Sarah C. Jenkins, Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, UK.

E-mail: s.jenkins.12@ucl.ac.uk

Word count (main text): 6979 (excluding abstract, tables and figure captions)

Abstract

Risk communicators often need to communicate probabilistic predictions. On occasion, an event with 10% likelihood will occur, or one with 90% likelihood will not – a probabilistically unexpected outcome. Previous research manipulating communication format has found that communicators lose more credibility and are perceived as less correct if an ‘unlikely’ event occurs, than if a ‘10–30% likelihood’ event occurs. We suggest ‘directionality–outcome congruence’ underlies the perception of predictions as ‘erroneous’. For example, the negatively directional term ‘unlikely’ led to harsher ratings because the outcome was counter to the original focus of the prediction (on the event’s non-occurrence). In the context of both probabilistically unexpected (Experiment 1) and expected (Experiment 2) outcomes, we find that communicators are perceived as less credible and less correct given ‘directionality–outcome incongruence’. Communicators should thus carefully consider the directionality implicit in their risk communications to maximise communication effectiveness.

Keywords: verbal probability expressions; directionality; risk communication; trust; expertise; credibility

Introduction

Risk communicators are often tasked with providing an accurate and informative estimate of the likelihood of an uncertain event (Fong, Rempel, & Hall, 1999). For such communications to be effective, the information must be understood as the communicator intended, and the communicator must be perceived as a credible source of information (Berry, 2004; Breakwell, 2000; Burton & Silver, 2006). Credibility is therefore one of a communicator's most precious assets (Covello & Allen, 1988). In the current paper, we consider how the communication format used to describe the (un)certainly of an event influences the maintenance of credibility in risk communications.

Much research has investigated how the communication format used to convey uncertainty information is understood by the recipient. This has highlighted a large amount of variability in the way individuals use and interpret verbal probability expressions (VPEs) such as 'unlikely' (e.g., Budescu & Wallsten, 1985) and has led to calls for VPEs to be accompanied by numerical expressions (e.g., '20% likelihood', '10–30% likelihood') to reduce misunderstandings (e.g., Budescu, Broomell, & Por, 2009; Budescu, Por, Broomell, & Smithson, 2014, though see Jenkins, Harris, & Lark, 2018 for potential downsides to such an approach). Relatively little research has examined the effect of format (e.g., VPE or numerical expression) on how the individual communicating the information is perceived (for exceptions see: Gurmankin, Baron, & Armstrong, 2004; Longman, Turner, King, & McCaffery, 2012; Peters et al., 2006). More recently, some researchers have investigated how credibility ratings of communicators change once focal event outcomes are known (Dieckmann, Mauro, & Slovic, 2010; Jenkins, Harris, & Lark, 2017, 2019). This latter question is relevant to perceptions of real-world risk communicators, who will communicate about multiple risks over time, with past prediction accuracy influencing how subsequent predictions are received (e.g., Breznitz, 1984).

Jenkins et al. (2019) found that when a communicator used the term 'unlikely' to describe the probability of an event which subsequently occurred (a probabilistically unexpected outcome), a communicator lost more credibility, and was perceived as more incorrect, than one who used a

numerical point (e.g., '20% likelihood') or range (e.g., '10–30% likelihood') format. Such an effect of communication format, however, was not consistently observed for high probability expressions. A communicator who initially described an event (which subsequently did not occur) as 'likely', suffered a similar reduction in credibility to a communicator who described the event as having an '80% likelihood' (Jenkins et al., 2017). We posit that the effect of communication format, and the disparity between findings pertaining to this effect across high and low probability VPEs, can be explained with regard to the framing of the prediction, specifically its directionality (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 2003).

Directionality.

Directionality refers to the notion that VPEs do not simply convey probability information, but also fulfil pragmatic functions by focusing attention on either the occurrence (positive directionality, e.g., 'a chance') or non-occurrence of the event (negative directionality, e.g., 'doubtful').¹ In a sentence completion task (e.g., 'It is *unlikely* that Clinton will become a good president because...'), sentences featuring positively directional VPEs tended to be completed with pro reasons (i.e., reasons why the event would occur), whereas those featuring negatively directional VPEs tended to be completed with con reasons (i.e., reasons why the event would not occur) (Teigen & Brun, 1995). Relatedly, following an event's occurrence, predictions described using negatively directional VPEs were perceived as less correct and more surprising than those described using positively directional, numerically equivalent VPEs (Teigen, 1988; Teigen & Brun, 2003). Differences in behavioural outcomes have also been observed (Honda & Yamagishi, 2006, 2009; Teigen & Brun, 1999). For instance, when a doctor describes the likelihood of side effects from a new medication as 'doubtful', a patient will be more inclined to take the medication than when there is 'some possibility' of side effects (Teigen & Brun, 1999). Thus, directionality appears integral to how VPEs are interpreted and used in subsequent decision making.

¹ See also Moxey and Sanford (2000) for a review of similar work in relation to quantifiers (e.g., 'many', 'some').

We suggest that the results in Jenkins et al. (2017, 2019), specifically the perception of a prediction as ‘erroneous’, can be explained by a ‘directionality–outcome congruence’ account, whereby a communicator is evaluated against the extent to which a phrase’s directionality and the outcome matches (c.f. Yeung, 2014). When one considers that an outcome will not occur, the possibility of its occurrence recedes into the background (Teigen & Brun, 1999). Therefore, a communicator who used the negatively directional term ‘unlikely’ suffered a greater loss of credibility than one who used ‘10–30% likelihood’ because the outcome was counter to the original focus of the prediction (i.e., on its non-occurrence). We suggest it was this ‘directionality–outcome incongruence’ which subsequently led to harsher credibility ratings of the communicator, and lower correctness ratings.

Although directionality has primarily been defined as relating to VPEs, there is evidence indicating that numerical expressions are biased towards positive interpretations (Teigen & Brun, 2000). This could explain the inconsistent effect of communication format observed for the high and low probability domains. In the low probability domain, in addition to the negative directionality of ‘unlikely’, ‘10–30% likelihood’ might be interpreted as positively directional, focussing attention on the event occurring, which would have made it directionally congruent with the outcome. In the high probability domain, however, ‘likely’ and ‘70–90% likelihood’ could be said to both be positively directional, meaning they were both incongruent with the outcome not occurring and thus suffered similar reductions in credibility.²

Overview of experiments.

The present paper aims to gain an understanding of the mechanisms behind the effects of communication format on credibility by testing the ‘directionality–outcome congruence’ account. We

² We recognise that there was a trend for a communicator who used ‘likely’ to suffer slightly more of a reduction in credibility than one who used ‘80% likelihood’, as well as a significant main effect of format on correctness ratings in Jenkins et al. (2017). This is not problematic for our account, as it has previously been noted that numerical expressions are more directionally ambiguous than VPEs (Budescu, Karelitz, & Wallsten, 2003; Teigen & Brun, 1995).

firstly test the account in the context of a probabilistically unexpected outcome (an ‘erroneous’ prediction – Experiment 1) and then in the context of a probabilistically expected outcome (Experiment 2). Having previously only focused on the VPEs ‘likely’ and ‘unlikely’ (Jenkins et al. 2017, 2019), the present research provided a chance to examine whether earlier findings would replicate with different VPEs, which share the same directionality and approximate numerical probability translation.³ Finally, it also provided the opportunity to investigate directionality in English, as much of the previous directionality research has been carried out in different languages, including Norwegian (e.g., Brun & Teigen, 1988; Teigen & Brun, 1999, 2000, 2003) and Japanese (e.g., Honda & Yamagishi, 2006, 2009). The hypotheses, methods and analyses for both of these experiments were pre-registered on the Open Science Framework (<https://osf.io/4rzfw>).

Experiment 1

As outlined above, our primary research interest was in testing the ‘directionality–outcome congruence’ account, as an explanation for credibility losses in response to ‘unlikely’ outcomes occurring. Consequently, our predictions first focused on the comparison of congruent to incongruent VPEs, before comparing these to their numerical counterparts:

Hypothesis 1.

A communicator who uses a congruent VPE (i.e. positively [negatively] directional for the low [high] probability domain) will suffer less of a reduction in credibility, will be perceived as more correct and will yield lower surprise ratings than one who uses an incongruent VPE, following a probabilistically unexpected outcome.

³ Finding VPEs of differing directionality which had similar numerical translations was difficult and thus we had to use lexical negation (i.e., use of ‘not’) to create negative directionality. Whilst ‘not entirely definite’ has not been used in directionality research before, similar use of lexical negation has been made in previous directionality studies. For instance, phrases such as ‘not quite certain’ (Teigen, 1988; Teigen & Brun, 1995, 1999, 2003), ‘not certain’, ‘not likely’ and ‘not probable’ (Brun & Teigen, 1988) have featured in such research.

Hypothesis 2_A.

As was previously found in Jenkins et al. (2017, 2019), in the low probability domain, we expect a communicator who uses an incongruent (negatively directional) VPE – ‘doubtful’ – to suffer a greater reduction in credibility and be perceived as less correct than one who uses a numerical range format, following a probabilistically unexpected outcome.

Hypothesis 2_B.

On the basis of ‘directionality-outcome congruence’ and the notion that numbers are positively directional (Teigen & Brun, 2000), in the high probability domain, we expect to find a communicator who uses a congruent VPE (negative directionality) to suffer less of a reduction in credibility compared to one who uses a numerical range, following a probabilistically unexpected outcome.

Method

Participants.

We aimed to analyse data from 300 participants to provide $\geq .98$ power to detect a medium - large effect of ‘directionality–outcome congruence’ (according to G*Power: Faul, Erdfelder, Lang, & Buchner, 2007; given an effect size of $d = 0.7$, based on comparisons of credibility difference scores for ‘unlikely’ and numerical point/range formats in Jenkins et al. 2017, 2019). As a result of the relatively high number of exclusions in our previous online studies, we increased total target sample size to 450 participants. We subsequently recruited 453 participants from Amazon Mechanical Turk via the TurkPrime platform (Litman, Robinson, & Abberbock, 2017). Seventeen participants failed the attention check (“How good are you at surviving one hour without oxygen?” c.f. Martire, Kemp, Watkins, Sayle, & Newell, 2013) and were excluded from the study, leaving a final sample of 436 participants (213 male), aged 19–74 ($Mdn = 34$). A total of 97% were native English speakers. Participants were paid \$0.40. Ethical approval was granted from the Departmental Ethics Chair for Speech, Hearing and Phonetic Sciences (University College London).

Design.

Probability (low; high) and communication format (negatively directional VPE [$n_{\text{LowProbability}}=68$, $n_{\text{HighProbability}}=79$]; positively directional VPE [$n_{\text{LowProbability}}=73$, $n_{\text{HighProbability}}=69$]; numerical range [$n_{\text{LowProbability}}=78$, $n_{\text{HighProbability}}=69$]) were manipulated in a between-participants design.

A series of pilot studies featuring 18 VPEs were carried out in order to identify four VPEs for use in Experiments 1 and 2 (a positively and negatively directional one for each probability condition). This was to ensure that the VPEs featured would have similar numerical translations, despite their opposing directionality (for results, see Supplementary Materials 1). In these pilot studies, participants were presented with an expert's prediction about the likelihood of floodwaters extending 7km and were asked to express the percentage which best represented the VPE on a slider. In total, participants saw four predictions, such that participants saw a positively and negatively directional VPE for each probability condition. Owing to the large range in interpretations given by participants in the pilot studies, we chose to use median translation when selecting VPEs. The four VPEs selected were the negatively directional 'doubtful' (median translation = 25%) and 'not entirely definite' (median translation = 70%) and the positively directional 'small chance' (median translation = 30.5%) and 'good chance' (median translation = 80%). The numerical range expressions were '10–30% likelihood' (low probability) and '70–90% likelihood' (high probability), following Jenkins et al. (2017, 2019). We opted to use a numerical range (rather than point) format, so that both the verbal and numerical communications included some imprecision (see Budescu & Wallsten, 1995 for research on the imprecision of VPEs). It is worth noting that in previous studies we have compared both numerical ranges (of different sizes) and numerical point expressions but have observed no differences between the two (Jenkins et al., 2018, 2019).

The dependent variables used were as in previous studies (Jenkins et al., 2019). Trustworthiness and expertise ratings were made pre- and post-outcome. Expertise was operationalised as: "How knowledgeable does the expert seem?" rated from 1 (*Not at all knowledgeable*) to 5 (*Extremely knowledgeable*). Trustworthiness was operationalised as: "How much

do you trust that the expert is giving you complete and unbiased information, to the best of their knowledge?” rated from 1 (*Not at all*) to 5 (*Trust a great deal*). For greater clarity of results (i.e. avoiding repetition of similar analyses), these measures were averaged to form a credibility score. As we were interested in how robust credibility perceptions were to probabilistically (un)expected outcomes, credibility difference scores ([post-outcome credibility rating] – [pre-outcome credibility rating]) were the main focus for analyses in both our experiments. Post-outcome, participants rated the communicator’s correctness, operationalised as: “How correct was the geologist’s prediction?” rated from 1 (*Not at all correct*) to 5 (*Completely correct*) (Teigen, 1988; Teigen & Brun, 2003), as well as their level of surprise: “How surprised are you by this outcome (i.e., the floodwater extended beyond 7 km/did not extend as far as 7 km)?” rated from 1 (*Not at all surprised*) to 5 (*Very surprised*) (Teigen & Brun, 2003).

Materials and procedure.

The experiment was run using Qualtrics. Participants were asked to respond to a number of demographic questions, including age and gender, before completing the attention check. Participants were randomly allocated to condition.

Participants read a ‘flood’ vignette (Jenkins et al., 2019), accompanied by a geologist’s prediction about the probability of floodwaters extending a certain distance:

“The Wayston flood plain has a history of flooding due to its flat terrain and proximity to the east side of the River Wayston. The river is currently in flood and floodwater is expected. Experts from Wayston Geological Centre are communicating information about the flood. An expert has suggested that given the river’s situation and recent weather, it is **doubtful** that the floodwater will extend 7km” (Incongruent condition, emphasis in original).

Participants were asked to provide initial ratings of the geologist's expertise and trustworthiness. On the following screen, participants were informed that the outcome did occur (low probability), or did not occur (high probability). They were then asked to provide further trustworthiness and expertise ratings, as well as rating how correct the geologist's prediction was in light of the outcome and how surprised they were (see Figure 1 for a flow chart of the procedure).

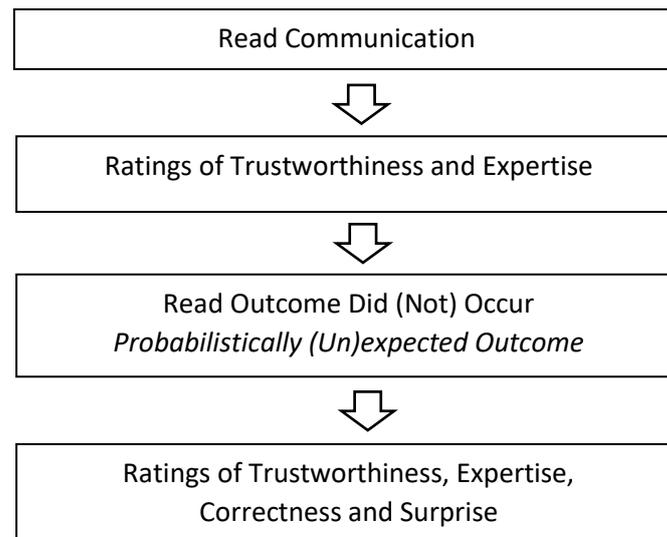


Figure 1. Procedure for Experiments 1 and 2.

Results

Planned Analyses

To test Hypothesis 1, we created a 'congruence' variable, which reflected whether the directionality of the VPE was congruent or incongruent with the actual outcome. The positively directional VPE was coded as 'congruent' in the low probability condition [where the outcome occurred] and 'incongruent' in the high probability [where the outcome did not occur] and vice versa for the negatively directional VPE (see Table 1).

Table 1.

Summary of Design Across Experiments

Experiment	Probability	Outcome	Negatively directional VPE	Positively directional VPE
1 (Probabilistically unexpected)	Low	Occurrence	'doubtful'	'small chance'*
	High	Non occurrence	'not entirely definite'*	'good chance'
2 (Probabilistically expected)	Low	Non-occurrence	'doubtful'*	'small chance'
	High	Occurrence	'not entirely definite'	'good chance'*

* Directionality of the VPE is congruent with the observed outcome

Change in credibility ratings.

Credibility was found to be highly reliable, both pre- and post-outcome ($\alpha = .85$, $\alpha = .88$, respectively).⁴ A 2 (congruence) \times 2 (probability) ANOVA⁵ revealed the predicted main effect of congruence, $F(1, 285) = 35.96$, $p < .001$, $\eta_p^2 = .11$, and a main effect of probability, $F(1, 285) = 47.05$, $p < .001$, $\eta_p^2 = .14$. Additionally, there was a significant interaction, $F(1, 285) = 14.80$, $p < .001$, $\eta_p^2 = .05$, with a greater effect of congruence for the high probability expressions than the low probability expressions (Figure 2). There was a greater loss of credibility when a communicator made a prediction using an incongruent VPE versus a congruent VPE, but this was only significant for the high probability expressions, $F(1, 285) = 49.56$, $p < .001$, $\eta_p^2 = .15$, and not for the low probability expressions, $F(1, 285) = 2.26$, $p = .13$.

⁴ Analysis of pre-outcome credibility ratings can be found in Supplementary Materials 2.

⁵ Results of 3 (format) \times 2 (probability) ANOVAs for each DV can be found in Supplementary Materials 3.

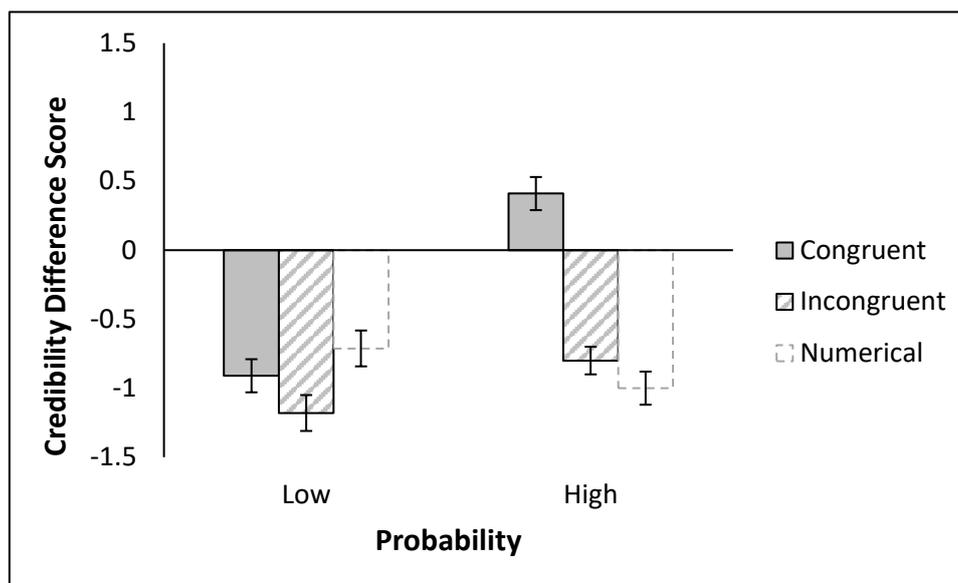


Figure 2. Mean credibility difference scores in Experiment 1 (error bars represent ± 1 standard error [SE]).

Correctness ratings.

A 2 (congruence) \times 2 (probability) ANOVA revealed a main effect of probability, $F(1, 285) = 39.20, p < .001, \eta_p^2 = .12$, and congruence, $F(1, 285) = 117.53, p < .001, \eta_p^2 = .29$. There was a significant interaction, $F(1, 285) = 9.75, p = .002, \eta_p^2 = .03$ (see Figure 3). Although the effect of congruence was larger for high probability expressions, congruent predictions were perceived as more correct than incongruent predictions across both high probability expressions, $F(1, 285) = 66.74, p < .001, \eta_p^2 = .26$, and low probability expressions, $F(1, 285) = 29.13, p < .001, \eta_p^2 = .09$.

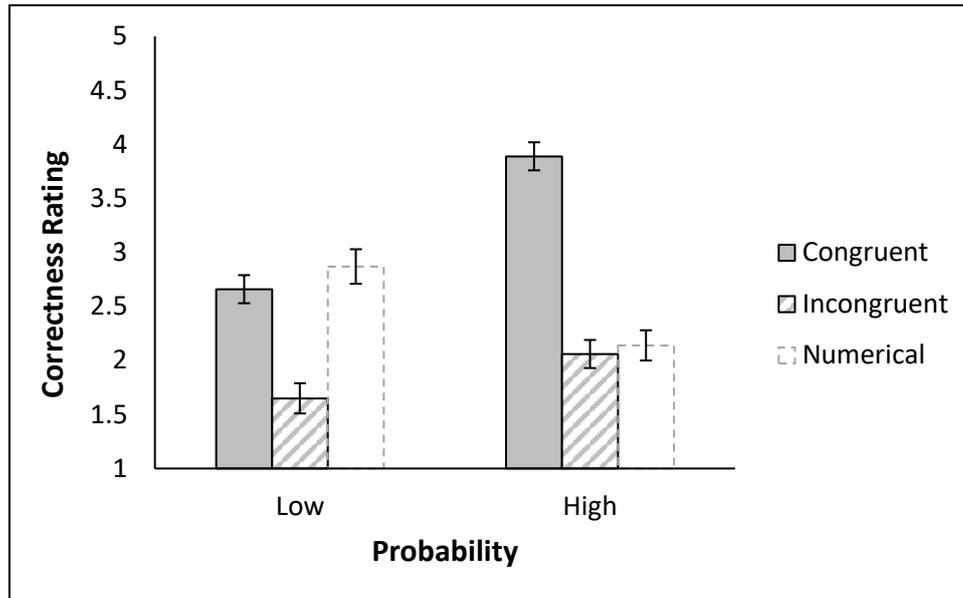


Figure 3. Mean correctness ratings in Experiment 1 (error bars represent ± 1 SE).

Surprise ratings.

A 2 (congruence) \times 2 (probability) ANOVA revealed the predicted main effect of congruence, $F(1, 285) = 4.36, p = .04, \eta_p^2 = .02$, with incongruent VPEs associated with higher levels of surprise than congruent VPEs. There was a main effect of probability, $F(1, 285) = 19.43, p < .001, \eta_p^2 = .06$ (see Figure 4), with higher levels of surprise in the low probability condition. The interaction between congruence and probability was not significant, $F(1, 285) = 1.07, p = .30$.

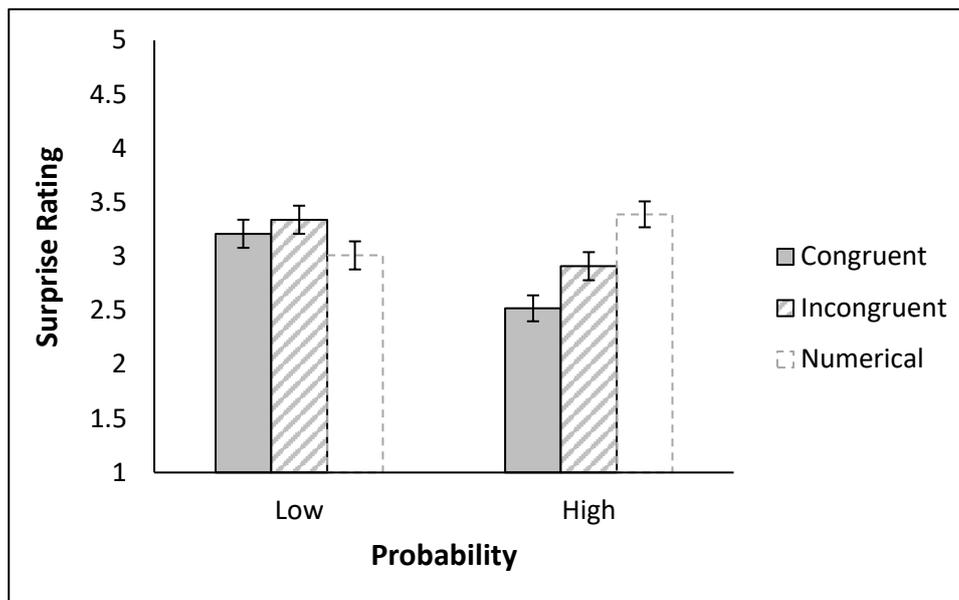


Figure 4. Mean surprise ratings in Experiment 1 (error bars represent ± 1 SE).

Comparisons to numerical ranges.

Hypothesis 2A – Low probability domain.

We compared the incongruent VPE to the numerical range using a series of independent samples *t*-tests. As seen in Figure 2, a communicator who used ‘doubtful’ suffered a greater loss of credibility after the event occurred, compared to a communicator who used ‘10–30% likelihood’, $t(144) = -2.65$, $p = .009$, $d = 0.44$. These differences were also reflected in the correctness ratings (see Figure 3), in which a prediction which featured ‘10–30% likelihood’ was perceived as more correct than one which featured ‘doubtful’, $t(138.43) = -6.08$, $p < .001$, $d = 1.13$.⁶ These results replicate findings in Jenkins et al. (2019) and demonstrate that the effect is not specific to ‘unlikely’. There was a trend for higher surprise ratings when a communicator described the event using the incongruent VPE compared to a numerical range expression, $t(140.89) = 1.96$, $p = .05$, $d = -0.32$ (see Figure 4).

⁶ Throughout the manuscript, in instances where the assumption of homogeneity of variance was violated, a Welch-Satterthwaite correction was applied (though this was not pre-registered). There is no difference in the results if the correction is not applied.

In the interests of completeness, we also compared the congruent VPE with the numerical range. No significant differences between a communicator who used ‘a small chance’ and one who used ‘10–30% likelihood’ were observed for any of the measures (all $ps > .27$).

Hypothesis 2B – High probability domain.

We compared the congruent VPE with the numerical range using a series of independent samples t -tests. As seen in Figure 2 and Figure 3, there was a significant difference between a communicator who used ‘70–90% likelihood’ and a communicator who used ‘not entirely definite’, with the latter losing less credibility, $t(146) = 7.59, p < .001, d = -1.25$, and the prediction perceived as more correct than one featuring the numerical range, $t(146) = 9.25, p < .001, -1.52$. There were higher surprise ratings when a communicator described the event using the numerical range, versus the congruent VPE, $t(144.45) = 4.79, p < .001, d = 0.78$, (see Figure 4).

As before, we also compared the incongruent VPE to the numerical range. A communicator who used ‘70–90% likelihood’ suffered a similar loss of credibility after the event did not occur to a communicator who used ‘a good chance’, $t(136) = -1.29, p = .20$. There was also no significant difference in how correct predictions were perceived to be, $t(126.60) = 0.50, p = .62$. There were higher surprise ratings when a communicator described the event using a numerical range compared to the incongruent VPE, $t(136) = 2.92, p = .004, d = -0.49$. The results slightly diverge from those found in Jenkins et al. (2017), where there was an effect of communication format on correctness ratings, although – consistent with the current results – no effect was observed for credibility ratings (using a numerical point rather than numerical range format).

Summary.

The results of the comparisons between verbal and numerical formats suggest that, following a probabilistically unexpected outcome, a communicator who used a numerical expression was more positively perceived than one who used a VPE in the low probability domain. However, the reverse was true in the high probability domain. These findings provide further support for our ‘directionality–

outcome congruence' account and for the suggestion that numerical formats are positively directional.

Unplanned Analyses

Post-outcome credibility ratings.

Although we have argued that it is the change in credibility rating that is of critical importance, given that past prediction accuracy influences how subsequent predictions are received (e.g., Breznitz, 1984), one could argue that how a communicator is ultimately perceived is also of great import. To this end, we replicate our primary analysis solely of the final credibility ratings assigned to the communicator, after the event outcome is known. The results follow those observed for the change in credibility ratings: a 2 (congruence) \times 2 (probability) ANOVA revealed a main effect of congruence, $F(1, 285) = 13.91, p < .001, \eta_p^2 = .05$, with higher credibility ratings when a communicator made a prediction using an congruent VPE, following a probabilistically unexpected outcome (see Figure 5). There was a main effect of probability, $F(1, 285) = 35.84, p < .001, \eta_p^2 = .11$, with higher credibility ratings in the high probability condition, following a probabilistically unexpected outcome. There was no significant interaction, $F(1, 285) = 0.82, p = .37$.

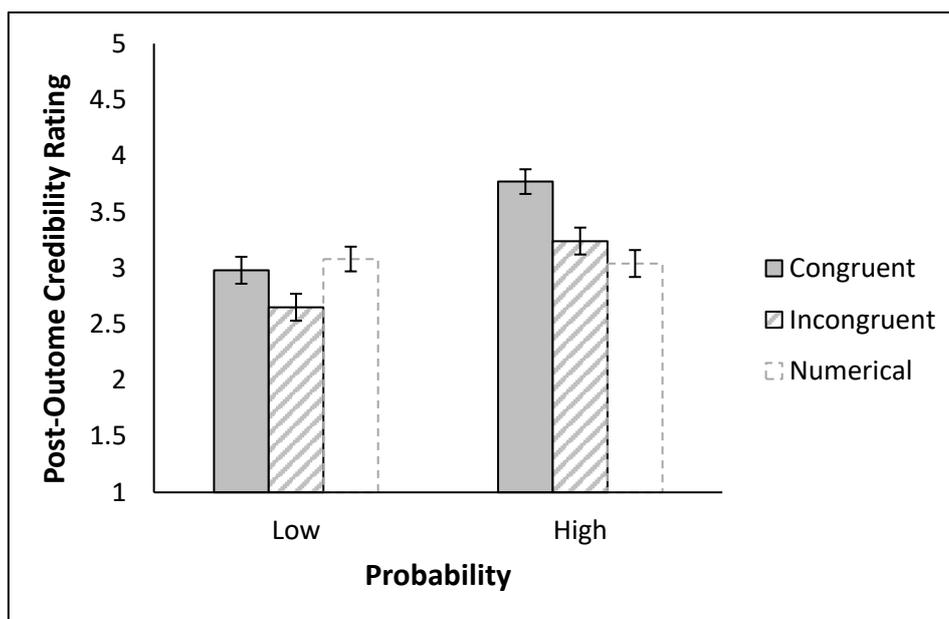


Figure 5. Mean post-outcome credibility ratings in Experiment 1 (error bars represent ± 1 SE).

Experiment 2

In Experiment 1, we found that communicators who used an incongruent VPE suffered a greater reduction in credibility and were perceived as less correct, following a probabilistically unexpected outcome (i.e., when a ‘doubtful’ event occurred, or when an event with a ‘good chance’ did not occur). Higher levels of surprise at the outcome were also observed for incongruent expressions. These results supported a ‘directionality–outcome congruence’ account, in which communicators were evaluated against the extent to which the directionality of the prediction matched the subsequent outcome.

Based on the results of Experiment 1 (as well as those previously reported in Jenkins et al. 2017, 2019), one might conclude that it is better to use positively directional VPEs for low probability events and negatively directional VPEs to describe high probability events. However, the advantage of doing so may only apply when the outcomes are probabilistically unexpected. It is therefore necessary to examine what happens in the case of probabilistically *expected* outcomes – by definition, these will occur more often than probabilistically unexpected outcomes (assuming the communicator’s forecasts are at least partially calibrated with the true state of the world). In Experiment 2, we therefore investigate perceptions of a communicator in light of a probabilistically expected outcome (i.e., a ‘doubtful’ event does not occur, or an event with a ‘good chance’ occurs). Given previous findings that negatively directional VPEs were perceived as more wrong and more surprising when the subsequent outcome was positive (versus when it was negative) (Teigen & Brun, 2003), we make the following prediction:

Hypothesis 3_A.

We expect that a communicator who uses a congruent VPE will be perceived as more correct than one who uses an incongruent VPE, following a probabilistically expected outcome.

Whether this effect of ‘directionality–outcome congruence’ will extend to ratings of surprise and credibility is an open question. Whilst a probabilistically unexpected outcome contradicts the default expectation, a probabilistically expected outcome matches it. Thus, the outcome might not

generate surprise or a *change* in credibility.⁷ However, following correctness ratings, we might expect a greater increase in credibility ratings and lower surprise ratings when a communicator uses a congruent VPE, versus an incongruent VPE (Hypothesis 3_B).

Although numerical communications do not feature in our primary research question, we continue to include a numerical range condition, enabling us to compare words versus numbers across the two experiments. This will also allow us to check that there are no unforeseen costs to credibility for a communicator who uses a numerical expression. We do not, however, make specific predictions regarding the numerical condition.

Method

Participants.

Based on the results of power calculations in Experiment 1, we recruited 450 participants from Amazon Mechanical Turk via the TurkPrime platform (Litman et al., 2017). Participants were required to be native English speakers, over 18 years of age (in order to meet ethical requirements) and have not completed any of our previous credibility experiments. Thirty-three participants failed the attention check (“How good are you at surviving one hour without oxygen?” c.f. Martire, Kemp, Watkins, Sayle, & Newell, 2013) and were excluded from the study, leaving a final sample of 417 participants (193 male), aged 19–88 (*Mdn* = 35). Participants were paid \$0.40.

Design, materials and procedure.

Design, materials and procedure were as in Experiment 1, with the exception of outcome. In this experiment, outcomes were ‘probabilistically expected’. In the low probability domain, the event did not occur, whereas in the high probability domain, the event occurred. Probability (low; high) and communication format (negatively directional VPE [$n_{\text{LowProbability}}=70$, $n_{\text{HighProbability}}= 70$]; positively

⁷ Moreover, credibility might even be said to function in a similar way to trust. Trust takes a long time to build, but can be easily lost, with actions that erode trust more impactful than those that build it (Slovic, 1993). Therefore, whilst communicators might suffer a reduction in credibility given a probabilistically unexpected outcome, an increase may not be seen given a probabilistically expected outcome.

directional VPE [$n_{\text{LowProbability}} = 63$, $n_{\text{HighProbability}} = 77$]; numerical range [$n_{\text{LowProbability}} = 76$, $n_{\text{HighProbability}} = 61$]) were manipulated in a between-participants design.

Results and Discussion

Planned Analyses

As in Experiment 1, we created a ‘congruence’ variable, which reflected whether the directionality of the VPE was congruent or incongruent with the actual outcome. Given the outcomes were probabilistically expected in this experiment, coding was reversed: the negatively directional VPE was coded as ‘congruent’ in the low probability condition (where the outcome did not occur) and ‘incongruent’ in the high probability (where the outcome did occur) and vice versa for the positively directional VPE (see Table 1).

Correctness ratings.

A 2 (congruence) \times 2 (probability) ANOVA⁸ revealed the predicted main effect of congruence, $F(1, 276) = 133.78$, $p < .001$, $\eta_p^2 = .33$, with congruent predictions perceived as more correct than incongruent predictions (see Figure 6). There was a main effect of probability, $F(1, 276) = 31.21$, $p < .001$, $\eta_p^2 = .10$, with higher correctness ratings for the low probability condition. The interaction between congruence and probability was not significant, $F(1, 276) = 0.73$, $p = .40$.

⁸ Results of the 3 (format) \times 2 (probability) ANOVAs for each DV can be found in Supplementary Materials 3.

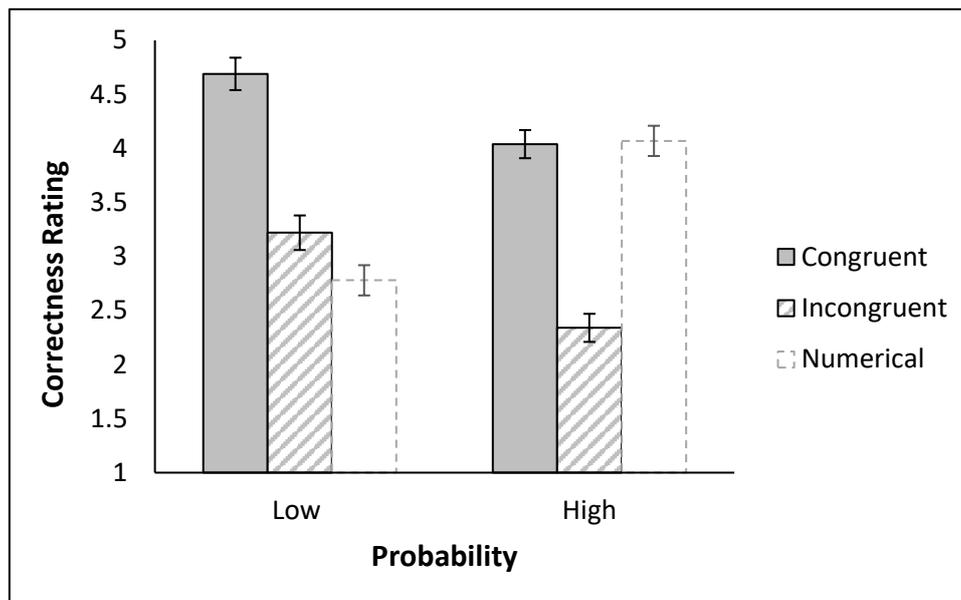


Figure 6. Mean correctness ratings in Experiment 2 (error bars represent ± 1 SE).

Change in credibility ratings.

Credibility was found to be highly reliable, both pre- and post-outcome ($\alpha = .85$, $\alpha = .90$, respectively).⁹

A 2 (congruence) \times 2 (probability) ANOVA revealed a predicted main effect of congruence, $F(1, 276) = 32.26$, $p < .001$, $\eta_p^2 = .11$, with a greater loss of credibility when a communicator made a prediction using an incongruent VPE. There was a main effect of probability, $F(1, 276) = 17.32$, $p < .001$, $\eta_p^2 = .06$ (see Figure 7), with a greater loss of credibility in the high probability condition when a communicator used an incongruent VPE. The interaction between congruence and probability was not significant, $F(1, 276) = 0.18$, $p = .67$.

⁹ Analysis of pre-outcome credibility ratings can be found in Supplementary Materials 2.

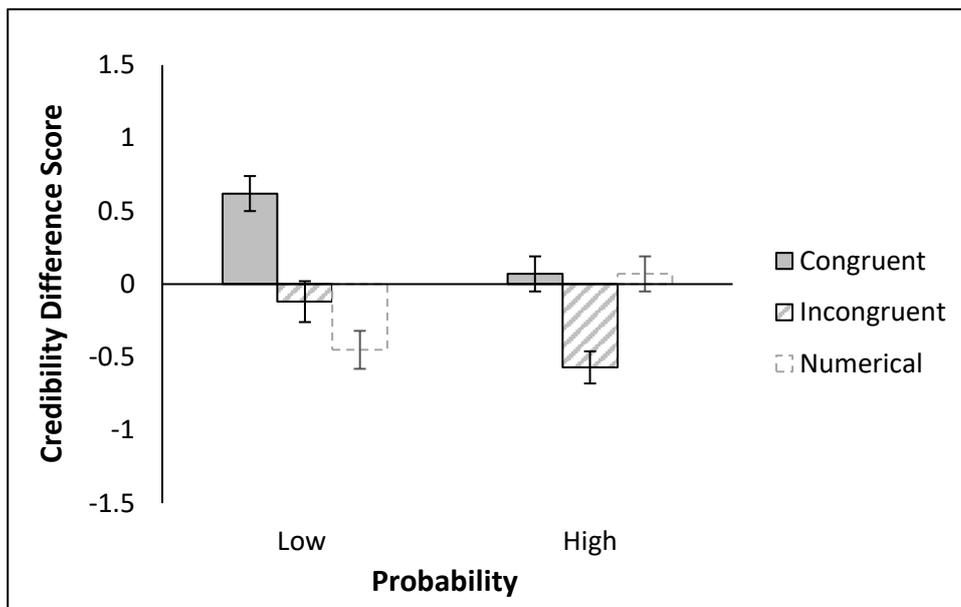


Figure 7. Mean credibility difference scores in Experiment 2 (error bars represent ± 1 SE).

Surprise ratings.

A 2 (congruence) \times 2 (probability) ANOVA showed a trend for lower levels of surprise for congruent expressions, $F(1, 276) = 3.10$, $p = .08$, $\eta_p^2 = .01$, and a non-significant effect of probability, $F(1, 276) = 0.64$, $p = .42$. There was a significant interaction between congruence and probability, $F(1, 276) = 7.72$, $p = .006$, $\eta_p^2 = .03$ (see Figure 8). No effect of congruence was observed in the low probability domain, $F(1, 276) = 0.50$, $p = .48$, but in the high probability domain, lower levels of surprise were observed when a communicator used a congruent VPE, $F(1, 276) = 10.85$, $p = .001$, $\eta_p^2 = .04$.

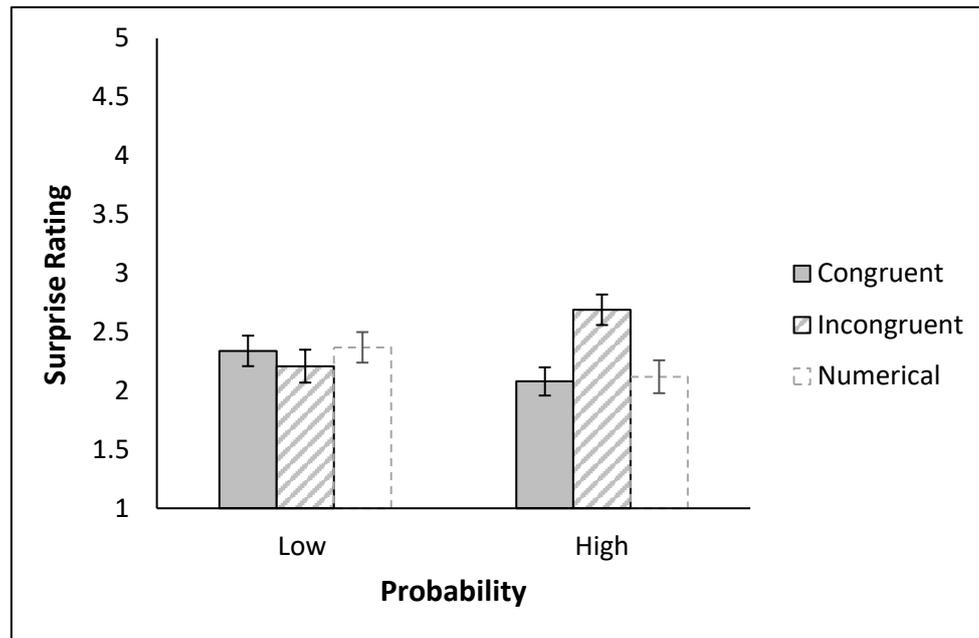


Figure 8. Mean surprise ratings in Experiment 2 (error bars represent ± 1 SE).

Comparisons to numerical ranges.

Low probability expressions.

We compared the congruent VPE to the numerical range using a series of independent samples *t*-tests. Following the probabilistically expected outcome, there was a difference in the change in credibility between the communicator who used ‘doubtful’ and the one who used ‘10–30% likelihood’ in their prediction, $t(144) = 5.84$, $p < .001$, $d = -0.97$. As seen in Figure 7, a communicator who used ‘doubtful’ experienced an increase in credibility after the probabilistically expected outcome, whereas one who used ‘10–30% likelihood’ suffered a decrease in credibility. The difference between formats was also reflected in the correctness ratings, in which a prediction featuring ‘doubtful’ was perceived as more correct than one which featured ‘10–30% likelihood’, $t(122.30) = 10.03$, $p < .001$, $d = -1.63$ (see Figure 6). There was no difference in level of surprise when a communicator described the event using the congruent VPE compared to a range expression, $t(144) = -.137$, $p = .89$ (see Figure 8).

In the interests of completeness, we also compared the incongruent VPE with the numerical range. There was little difference in credibility difference scores between a communicator who used ‘10–30% likelihood’ and one who used ‘small chance’ following the probabilistically expected

outcome, $t(137) = 1.69, p = .09, d = 0.29$. There was a trend for the prediction featuring ‘small chance’ to be perceived as more correct than which featured ‘10–30% likelihood’, $t(137) = 1.84, p = .07, d = 0.31$. There was no difference in level of surprise when a communicator described the event using the incongruent VPE compared to a range expression, $t(137) = -.83, p = .41$.

High probability expressions.

We compared the congruent VPE with the numerical range using a series of independent samples *t*-tests. There was no significant difference in credibility difference scores, correctness or surprise ratings for a communicator who used ‘70–90% likelihood’ and a communicator who used ‘good chance’ (all p s > .83).

We also compared the incongruent VPE to the numerical range. A communicator who used ‘70–90% likelihood’ experienced a small increase in credibility following the probabilistically expected outcome, whereas one who used ‘not entirely definite’ suffered a decrease in credibility, $t(117.21) = -3.95, p < .001, d = 0.67$ (see Figure 7). These differences were also reflected in the correctness ratings, in which a prediction which featured ‘70–90% likelihood’ was perceived as more correct than one which featured ‘not entirely definite’, $t(128.21) = -8.65, p < .001, d = 1.50$ (see Figure 6). Similarly, lower surprise ratings were observed when a communicator described the event using a numerical range compared to the incongruent VPE, $t(129) = -3.07, p = .003, d = 0.55$ (see Figure 8).

Summary.

The results of the comparisons between verbal and numerical formats suggest that, following a probabilistically expected outcome, a communicator who used a VPE was more positively perceived than one who used a numerical expression in the low probability domain. However, the reverse was true in the high probability domain. These findings support the notion that numerical expressions are perceived as positively directional.

Unplanned Analyses

Post-outcome credibility ratings.

Overall, the results are in line with those observed for the change in credibility ratings: a 2 (congruence) \times 2 (probability) ANOVA revealed a main effect of congruence, $F(1, 276) = 51.08$, $p < .001$, $\eta_p^2 = .16$, and a main effect of probability, $F(1, 276) = 35.84$, $p < .001$, $\eta_p^2 = .11$, (see Figure 9). Additionally, there was a significant interaction between congruence and probability (not observed in the change ratings), $F(1, 276) = 7.06$, $p = .008$, $\eta_p^2 = .03$. However, although the effect of congruence was larger for high probability expressions, congruent predictions were perceived as more credible than incongruent predictions across both high probability expressions, $F(1, 276) = 50.61$, $p < .001$, $\eta_p^2 = .16$, and low probability expressions, $F(1, 276) = 9.60$, $p = .002$, $\eta_p^2 = .03$.

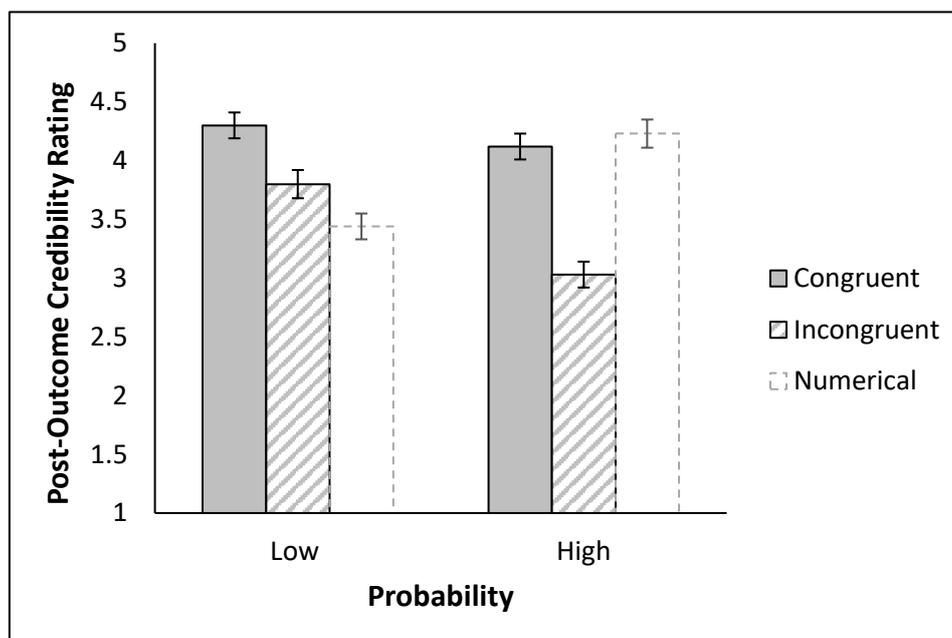


Figure 9. Post-outcome credibility ratings in Experiment 2 (error bars represent ± 1 SE).

Comparing Across Experiments (Planned Analysis).

So far, the analyses from both Experiments 1 and 2 can indicate which communication format is best to use when outcomes are probabilistically (un)expected. However, in a real-world scenario, the decision to use a specific communication format cannot be made on the basis of outcome, given this

is unknown at the time of making the prediction. That is, the communicator does not know if they are in Experiment 1 or 2. To get an initial indication of the likely credibility costs/gains associated with the different formats, we also compared results across experiments, weighing up the relative expected costs of each format for both types of outcome.

To do so, for each communication format in each probability domain, we created an expected difference score. This expected difference score incorporates the mean credibility difference scores for each potential outcome ('probabilistically unexpected' [Experiment 1]; 'probabilistically expected' [Experiment 2']), weighted by the likelihood of these alternative outcomes for a perfectly calibrated predictor (one whose 20% estimate will occur 20% of the time; see Table 2, under the simplifying assumption that the 'best estimate' of the objective probability associated with the low probability terms is 20% and with the high probability terms is 80%).¹⁰ Given the assumptions involved, and the reliance on aggregate level data, the expected difference scores will only be compared descriptively in an indicative analysis.

¹⁰ This analysis is based on the assumption that a communicator is perfectly calibrated when estimating probability, though we acknowledge that probability estimates tend to regress towards the mean (i.e., low probabilities are overestimated, and high probabilities are underestimated). We therefore provide additional (unplanned) analysis using different weights (0.3[Exp1], 0.7[Exp2], "overestimated" [regressive]), (0.1[Exp1], 0.9[Exp2], "underestimated") in Supplementary Materials 5. The results are similar to those reported in Table 2 – in the low probability domain, the lowest expected difference score was associated with use of a numerical range. In the high probability domain, the lowest expected difference score was associated with use of a negatively directional VPE. A comparison of the results with different weightings is provided in Figure S3.

Table 2. Calculation of Expected Difference Scores

Probability	Expression	Expected difference score	Mean credibility difference score		Overall 'expected difference score'
			Exp 1	Exp 2	
Low	Doubtful	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	- 1.18	0.62	0.26
	Small chance	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	- 0.91	- 0.12	- 0.28
	10–30% likelihood	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	- 0.71	- 0.45	- 0.50
High	Not entirely definite	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	0.41	- 0.57	- 0.37
	Good chance	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	- 0.80	0.06	- 0.11
	70–90% likelihood	$(0.2 \times \text{Exp 1}) + (0.8 \times \text{Exp 2})$	- 1.00	0.07	- 0.14

Note. Exp 1 and Exp 2 represent the mean credibility difference scores from each cell of Experiment 1 (probabilistically unexpected outcome) and Experiment 2 (probabilistically expected outcome) respectively.

Based on the results in Table 2, in the low probability domain, a communicator might expect to retain the most amount of credibility by using a negatively directional VPE (e.g., ‘doubtful’) in their prediction, followed by a positively directional VPE (e.g., ‘small chance’), with the lowest expected difference score associated with use of a numerical range. In the high probability domain, a communicator might expect to retain the most amount of credibility by using a positively directional VPE (e.g., ‘good chance’) in their prediction, closely followed by a numerical range, with the lowest expected difference score associated with use of a negatively directional VPE (e.g., ‘not entirely definite’). However, the choice of communication format is not as straightforward as simply relying on an expected difference score; rather, the choice depends on the goals of the communicator. For instance, whilst a communicator may wish to use ‘doubtful’ to convey the likelihood of a low probability event (on the basis that ‘doubtful’ has the largest expected difference score), such a format performs the worst in light of a probabilistically unexpected outcome. In this instance, a communicator might prefer to guard against the large reduction in credibility rather than following the expected difference score. Such a strategy might be said to be more appropriate given the literature on the nature of trust, specifically that it is more easily lost than gained (e.g., Slovic, 1993). Thus, one *might* want to consider weighting probabilistically unexpected outcomes more heavily than based solely on their probability of occurrence. The ultimate decision is, however, down to the individual communicator and their own specific goals.

General Discussion

Across two experiments, we observed consistent evidence for the role of ‘directionality–outcome congruence’ in driving evaluations of a communicator’s credibility, following either a probabilistically expected outcome, or a probabilistically unexpected outcome. In addition, Experiment 1 replicated our previous findings from the low probability domain: a communicator who uses a negatively directional VPE (such as ‘doubtful’) is most vulnerable to a probabilistically unexpected outcome, demonstrating such findings are not merely limited to the VPE ‘unlikely’.

Directionality

Verbal probability expressions.

Much research has demonstrated that VPEs fulfil a range of semantic and communicative functions, over and above simply conveying probability information (e.g., Teigen & Brun, 1995). A communicator may opt to use a specifically directional VPE to encourage a certain action, or convey a belief that an event will (not) happen. In the present studies, we suggest that participants were indeed aware that the VPE communicated more than just probability information and made inferences about the expected outcome from the directionality of the VPE. This is in line with Gricean principles (Grice, 1975), whereby the listener assumes the speaker has chosen to frame the communication in the most informative way (i.e., by highlighting the expected outcome). When the VPE was directionally incongruent with the outcome, it led to greater reductions in credibility, lower correctness ratings and higher levels of surprise. The latter replicates Teigen and Keren's (2003) findings that incongruence with the previously expected event was associated with higher ratings of surprise.

Despite the similar numerical translations given to congruent and incongruent VPEs (see Supplementary Materials 1), the different downstream effects we observed clearly support the notion that VPEs contain 'inbuilt hints' (Teigen & Brun, 1999). The current results complement previous research investigating directionality in other languages (Brun & Teigen, 1988; Honda & Yamagishi, 2006, 2009; Teigen & Brun, 1999, 2003) and add to the body of evidence demonstrating that individuals are sensitive to the pragmatic subtleties associated with VPEs. The relative weight of these 'hints' relative to the probabilistic information conveyed is yet to be determined. Given that some conditions in Experiment 2 (with probabilistically expected outcomes) yielded mean negative changes in credibility (and correctness ratings below the midpoint of the scale), could 'directionality congruence' even be psychologically *more* important than probabilistic congruence? We investigated this question in an unplanned analysis across the two experiments (see Supplementary Materials 4), coding the three formats according to whether they were directionally (in)congruent with the outcome, as well as whether they were probabilistically (un)expected. This analysis revealed

identically sized main effects for both directionality congruence and probabilistic expectedness, with no interaction. These results highlight the powerful influence of directionality – it is just as important as whether the communicator’s prediction refers to a probabilistically expected or unexpected event.

Directional ambiguity.

The fact that no effect of congruence was observed for credibility difference scores (Experiment 1) or surprise ratings (Experiment 2) in the low probability domain could be due to the ambiguous directionality of ‘small chance’. We suggest that the expression was directionally ambiguous for two reasons: the fact it was a low probability expression and that it included the word ‘small’. Low numerical probabilities have been found to be more ambiguous than high probabilities (Teigen & Brun, 2000), with lower probabilities also found to be more susceptible to contextual influences than high probabilities (Bilgin & Brenner, 2013). In addition, whilst the expression ‘a chance’ is affirmative and clearly perceived as positively directional, expressions such as ‘small’ are more ambiguous (Teigen & Brun, 1995). Indeed, Teigen and Brun (2000) found the expression ‘small chance’ was perceived by some as positively directional and by others as negatively directional. Relatedly, Sanford and Moxey (2003) suggest that ‘small risk’ puts focus on the (small) risk. Therefore, both ‘small chance’ and ‘doubtful’ may have focused expectations on the outcome’s non-occurrence, which would have meant less contrast with expectations than if the expressions focused attention in two different directions, as was the case in the high probability domain.

Inclusion of ‘small chance’ arose in part as a result of the difficulty in finding VPEs that were both similarly translated across directionalities and that fell within the numerical ranges used in our previous studies. We were further limited by the general lack of positively directional, low probability VPEs, which was perhaps inevitable given directionality and probability magnitude are typically related – positively directional, affirmative terms (e.g., ‘likely’) are typically used for higher probabilities, such as 85%, whereas negatively directional terms (e.g., ‘doubtful’) typically used to describe lower probabilities, such as 15% (Teigen & Brun, 1995).

Pragmatics associated with numerical expressions.

In both experiments, we observed no evidence for consistent differences in the way numerical expressions and positively directional VPEs were interpreted, across both credibility and correctness ratings.¹¹ Although numerical expressions have been shown to be more directionally ambiguous than their verbal equivalents, they are not directionally neutral, but rather biased towards positive interpretations (Teigen & Brun, 2000). Indeed, the positive directionality of numerical expressions can also explain why a communicator who used ‘10–30% likelihood’ suffered a loss in credibility despite a probabilistically expected outcome. In this case, the expression was directionally incongruent with the event not occurring.

In addition to the choice of which communication format to use, communicators also have a choice over how to frame their prediction, that is, in terms of the most likely (80% likelihood of occurrence) or least likely (20% likelihood of non-occurrence) outcome. Despite the mathematically equivalent nature of these frames, the attribute framing literature demonstrates that the choice of frame has downstream effects on perceptions and decisions (e.g., Levin & Gaeth, 1988; Levin, Schneider, & Gaeth, 1998; Sanford, Fay, Stewart, & Moxey, 2002). An ‘information leakage’ account has been proposed to underlie attribute framing effects, in which additional information beyond what is literally present is conveyed, which influences the inferences recipients make when presented with different frames (McKenzie & Nelson, 2003; Sher & McKenzie, 2006; for an application of this account to VPEs, see Honda & Yamagishi, 2016; Juanchich, Teigen, & Villejoubert, 2010). In the current experiments, we suggest that recipients inferred pragmatic information from both the highlighted focal outcome as well as the directionality of the expression. Whilst such inferences might be more commonly associated with VPEs, they have also been shown to occur with numerical probabilities (Sirota & Juanchich, 2012, see also Yeung, 2014). Although the addition of numbers to create a mixed

¹¹ We do, however, acknowledge that in Experiment 1, surprise ratings were significantly different in the high probability domain. We also note that for credibility difference scores and correctness ratings in Experiment 2, the effect of communication format was close to (but did not reach) conventional levels of significance.

format expression (e.g., 'unlikely [20%]') has been proposed as a solution to the problem of miscommunications associated with use of VPEs (e.g., Budescu, Broomell, & Por, 2009), communicators should be reminded that numerical expressions are not completely devoid of pragmatic implicatures.

Further considerations.

The results in Jenkins et al. (2017, 2019), specifically the different effects of communication format observed in the low and high probability domains, have been attributed to the positive directionality of numbers and relatedly, 'directionality–outcome congruency' (see Collins and Mandel [2019] for further support for this account). That is not to say that the 'directionality–outcome congruency' of numbers is the only difference given probabilistically unexpected outcomes across the low and high probability domains. Whilst the results in the current paper support a 'directionality–outcome congruency' account, we recognise that there are other factors which might also contribute to this difference, such as the differing costs of an 'erroneous' prediction in the low and high probability domains. Following an asymmetric loss function account (Harris, Corner, & Hahn, 2009; Weber, 1994), for the natural hazards we have considered (volcanic lava flows, floodwater extension, etc.), it is conceivably more costly to predict that the event will not occur when it subsequently does, than predict the event will occur when it subsequently does not. It is possible participants perceived the costs of an 'erroneous' prediction as greater in the low probability domain and thus judged the communicator more harshly. In addition, although low probability events generally evoke less concern than warranted by their probability, when they do occur, they generate far more concern than deserved (Weber, 2006), which could have further magnified the harshness of subsequent credibility ratings in Experiment 1. The asymmetric loss function account can also be applied to the reverse situation. Recognising the costs of an 'erroneous' prediction, the account suggests that there could be a greater benefit for communicators who make a 'correct' prediction in the low probability domain versus in the high probability domain. This is indeed the case in Experiment 2, seen for both credibility

and correctness ratings. The two accounts are not mutually exclusive, but rather highlight the intricacies associated with natural hazard communications.

Implications for communicating uncertainty.

The potential for misunderstandings associated with VPE usage is well established in the literature. Our findings clearly highlight, once again, that VPEs communicate additional information, over and above that regarding probability. Our work also adds to previous research on directionality by investigating its influence in the English language. Communicators should be aware of the inferences an individual might draw from the use of a particular VPE and recognise the possible implications their use could have on perceived credibility. Although previous research has indicated that numerical formats held advantages over verbal and mixed formats (Jenkins et al., 2017, 2018, 2019), the results in the current paper highlight that numerical formats are also vulnerable to pragmatic influences. Thus, it is not a straightforward choice of verbal versus numerical formats. Rather, when deciding how to frame a communication, communicators should use insights from research such as that presented here to choose a communication format appropriate for their purposes.

References

- Berry, D. (2004). *Risk, Communication and Health Psychology*. Maidenhead, England: Open University Press.
- Bilgin, B., & Brenner, L. (2013). Context affects the interpretation of low but not high numerical probabilities: A hypothesis testing account of subjective probability. *Organizational Behavior and Human Decision Processes*, *121*(1), 118–128. <https://doi.org/10.1016/j.obhdp.2013.01.004>
- Breakwell, G. M. (2000). Risk communication: factors affecting impact. *British Medical Bulletin*, *56*(1), 110–120. <https://doi.org/10.1258/0007142001902824>
- Breznitz, S. (1984). *The Psychology of False Alarms*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, *41*(3), 390–404. [https://doi.org/10.1016/0749-5978\(88\)90036-2](https://doi.org/10.1016/0749-5978(88)90036-2)
- Budescu, D. V., Broomell, S. B., & Por, H.-H. H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, *20*(3), 299–308. <https://doi.org/10.1111/j.1467-9280.2009.02284.x>
- Budescu, D. V., Karelitz, T. M., & Wallsten, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making*, *16*(3), 159–180. <https://doi.org/10.1002/bdm.440>
- Budescu, D. V., Por, H. H., Broomell, S. B., & Smithson, M. (2014). The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, *4*(6), 508–512. <https://doi.org/10.1038/nclimate2194>
- Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, *36*(3), 391–405. [https://doi.org/10.1016/0749-5978\(85\)90007-X](https://doi.org/10.1016/0749-5978(85)90007-X)

- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation*, 32(2), 275–318. [https://doi.org/http://doi.org/10.1016/S0079-7421\(08\)60313-8](https://doi.org/http://doi.org/10.1016/S0079-7421(08)60313-8)
- Burton, P. W., & Silver, B. D. (2006). The credibility of disaster evacuation warnings: the effects of issue framing and trust in government. *Annual Meeting of the American Political Science Association*. Retrieved from <https://bit.ly/2CeBKVn>
- Collins, R. N., & Mandel, D. R. (2019). Cultivating credibility with probability words and numbers. *Judgement and Decision Making*, 14(6), 683–695.
- Covello, V. T., & Allen, F. (1988). *Seven cardinal rules of risk communication*. U.S. Environmental Protection Agency. Washington, DC.
- Dieckmann, N., Mauro, R., & Slovic, P. (2010). The effects of presenting imprecise probabilities in intelligence forecasts. *Risk Analysis*, 30(6), 987–1001. <https://doi.org/10.1111/j.1539-6924.2010.01384.x>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fong, G. T., Rempel, L. A., & Hall, P. A. (1999). Challenges to improving health risk communication in the 21st century: A discussion. *Journal of the National Cancer Institute Monographs*, (25), 173–176. <https://doi.org/10.1093/oxfordjournals.jncimonographs.a024196>
- Grice, P. H. (1975). Logic and Conversation. In *Syntax and Semantics 3: Speech Acts* (pp. 41–58). <https://doi.org/10.1111/j.1365-2664.2006.01229.x>
- Gurmankin, A. D., Baron, J., & Armstrong, K. (2004). The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Medical Decision Making*, 24(3), 265–271. <https://doi.org/10.1177/0272989X04265482>
- Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition*, 110(1), 51–64. <https://doi.org/10.1016/j.cognition.2008.10.006>

- Honda, H., & Yamagishi, K. (2006). Directional verbal probabilities: Inconsistencies between preferential judgments and numerical meanings. *Experimental Psychology*, *53*(3), 161–170. <https://doi.org/10.1027/1618-3169.53.3.161>
- Honda, H., & Yamagishi, K. (2009). Perceived certainty based on verbal probability phrases: Effect of directionality and its dependence on method. *Japanese Psychological Research*, *51*(4), 266–273. <https://doi.org/10.1111/j.1468-5884.2009.00409.x>
- Honda, H., & Yamagishi, K. (2016). Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points. *The Quarterly Journal of Experimental Psychology*, *70*(10), 1–65. <https://doi.org/10.1080/17470218.2016.1225779>
- Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2017). Maintaining credibility when communicating uncertainty: The role of communication format. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 582–587). Austin, TX: Cognitive Science Society.
- Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2018). Understanding 'unlikely (20% likelihood)' or '20% likelihood (unlikely)' outcomes: The robustness of the extremity effect. *Journal of Behavioral Decision Making*, *31*(4), 572–586. <https://doi.org/10.1002/bdm.2072>
- Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2019). When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. *Journal of Risk Research*, *22*(5), 537–554. <https://doi.org/10.1080/13669877.2018.1440415>
- Juanchich, M., Teigen, K. H., & Villejoubert, G. (2010). Is guilt 'likely' or 'not certain'? Contrast with previous probabilities determines choice of verbal terms. *Acta Psychologica*, *135*(3), 267–277. <https://doi.org/10.1016/j.actpsy.2010.04.016>
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, *15*(3), 374–378. <https://doi.org/10.1086/209174>
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and

- critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149–188. <https://doi.org/10.1006/obhd.1998.2804>
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433–442. <https://doi.org/10.3758/s13428-016-0727-z>
- Longman, T., Turner, R., King, M., & McCaffery, K. J. (2012). The effects of communicating uncertainty in quantitative health risk estimates. *Patient Education and Counseling*, 89(2), 252–259. <https://doi.org/10.1016/j.pec.2012.07.010>
- Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (2013). The expression and interpretation of uncertain forensic science evidence: verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behavior*, 37(3), 197–207. <https://doi.org/10.1037/lhb0000027>
- McKenzie, C., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, 10(3), 596–602. <https://doi.org/10.3758/BF03196520>
- Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, 14(3), 237–255. [https://doi.org/10.1002/\(SICI\)1099-0720\(200005/06\)14:3<237::AID-ACP641>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<237::AID-ACP641>3.0.CO;2-R)
- Peters, E. M., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2006.01720.x>
- Sanford, A. J., Fay, N., Stewart, A., & Moxey, L. (2002). Perspective in statements of quantity, with implications for consumer psychology. *Psychological Science*, 13(2), 130–134. <https://doi.org/10.1111/1467-9280.00424>
- Sanford, A. J., & Moxey, L. (2003). New perspectives on the expression of quantity. *Current Directions in Psychological Science*, 12(6), 240–243. <https://doi.org/10.1046/j.0963-7214.2003.01270.x>

- Sher, S., & McKenzie, C. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494. <https://doi.org/10.1016/j.cognition.2005.11.001>
- Sirota, M., & Juanchich, M. (2012). To what extent do politeness expectations shape risk perception? Even numerical probabilities are under their spell! *Acta Psychologica*, *141*(3), 391–399. <https://doi.org/10.1016/j.actpsy.2012.09.004>
- Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, *13*(6), 675–682. <https://doi.org/10.1111/j.1539-6924.1993.tb01329.x>
- Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*(1), 27–38. [https://doi.org/10.1016/0001-6918\(88\)90043-1](https://doi.org/10.1016/0001-6918(88)90043-1)
- Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, *88*(3), 233–258. [https://doi.org/10.1016/0001-6918\(93\)E0071-9](https://doi.org/10.1016/0001-6918(93)E0071-9)
- Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, *80*(2), 155–190. <https://doi.org/10.1006/obhd.1999.2857>
- Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: When does $p = 0.3$ reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making*, *13*(3), 345–362. [https://doi.org/10.1002/1099-0771\(200007/09\)13:3<345::AID-BDM358>3.0.CO;2-U](https://doi.org/10.1002/1099-0771(200007/09)13:3<345::AID-BDM358>3.0.CO;2-U)
- Teigen, K. H., & Brun, W. (2003). Verbal probabilities: A question of frame? *Journal of Behavioral Decision Making*, *16*(1), 53–72. <https://doi.org/10.1002/bdm.432>
- Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts. *Cognition*, *25*(4), 291–293. <https://doi.org/10.1016/S0>
- Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, *115*(2), 228–242. <https://doi.org/10.1037/0033-2909.115.2.228>
- Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why

global warming does not scare us (yet). *Climatic Change*, 77(1–2), 103–120.

<https://doi.org/10.1007/s10584-006-9060-3>

Yeung, S. (2014). Framing effect in evaluation of others' predictions. *Judgment and Decision Making*, 9(5), 445–464.

Supplementary Materials

1. Translation of VPEs From Pilot Studies	40
Table S1. High Probability Expressions	40
Table S2. Low Probability Expressions	41
2. Pre-Outcome Credibility Ratings – Unplanned Analyses	42
Experiment 1 (Figure S1).....	42
Experiment 2 (Figure S2).....	43
3. Comparing VPEs and Numerical Expressions – Unplanned Analyses	44
Experiment 1.....	44
Experiment 2.....	45
4. Weighing up Directionality Congruence and Probabilistic Expectedness	46
5. Comparing Across Experiments (Unplanned Analyses)	47
Table S3. Calculation of Expected Difference Scores (Underestimation)	47
Table S4. Calculation of Expected Difference Scores (Overestimation)	48
Figure S3.....	49

1. Translation of VPEs From Pilot Studies

Table S1. High Probability Expressions

Directionality	VPE	<i>n</i>	Mean	SD	Median	Range
Negative	A small doubt	50	49.44	29.83	48.5	94
	Not certain	41	50.43	23.46	50	87
	Not quite certain	46	54.15	21.51	52.5	95
	Some doubt	48	60.00	20.72	62	86
	Not completely certain	50	64.00	18.84	65.5	75
	Not entirely certain	47	63.96	18.77	66	82
	Not completely definite	53	63.87	16.56	69	65
	Not entirely definite	50	67.82	16.54	70	70
	Little doubt	45	76.09	22.74	83	82
Positive	Probable	49	74.33	15.76	75	70
	Good chance	42	76.02	16.82	80	71
	Entirely possible	44	77.68	15.02	80	71

Table S2. Low Probability Expressions

Directionality	VPE	<i>n</i>	Mean	SD	Median	Range
Negative	Improbable	45	34.04	27.65	25	87
	Doubtful	45	35.84	24.90	25	85
	Uncertain	45	51.56	24.38	50	95
Positive	Small chance	48	39.15	24.19	30.5	90
	Possible	42	63.20	18.24	62.5	70
	A chance	45	63.47	25.22	69	99

2. Pre-Outcome Credibility Ratings – Unplanned Analyses

Although the congruency of the VPE is not known pre-outcome, we have maintained this categorisation and labelling for ease of comparison with the results presented in the main text.

Experiment 1.

A 3 (communication format – congruent, incongruent, numerical) \times 2 (probability) ANOVA revealed no significant effect of probability, $F(1, 430) = 0.09$, $p = .76$, and a main effect of communication format, $F(2, 430) = 7.41$, $p = .001$, $\eta_p^2 = .03$ on pre-outcome credibility ratings (see Figure S1). There was a significant interaction between the two factors, $F(2, 430) = 11.72$, $p < .001$, $\eta_p^2 = .05$. Simple effects analyses showed a significant effect of communication format for the high probability expressions, $F(2, 430) = 19.15$, $p < .001$, $\eta_p^2 = .08$, but not for the low probability expressions, $F(2, 430) = 0.29$, $p = .75$, $\eta_p^2 = .001$.

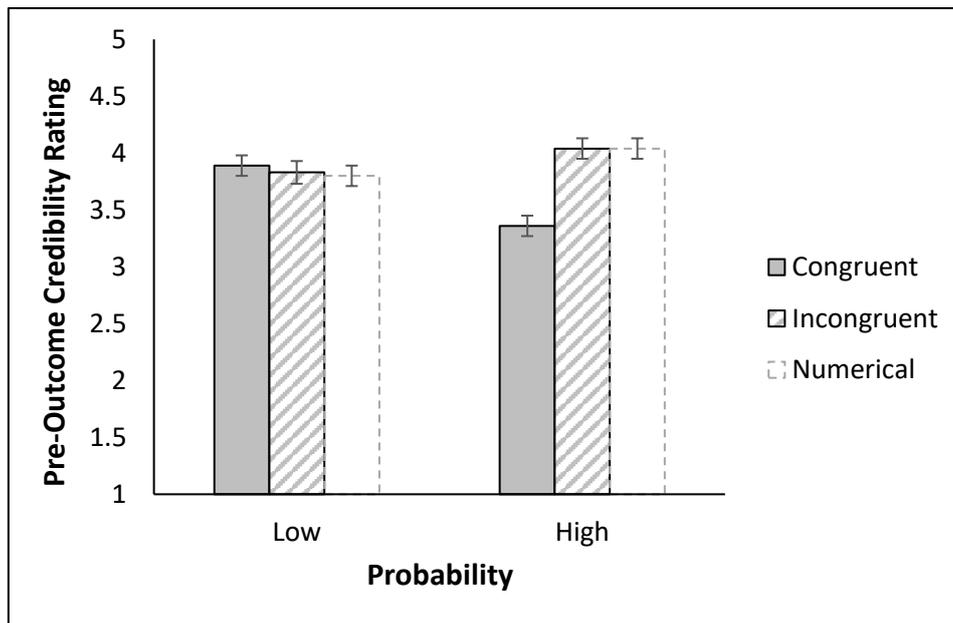


Figure S 1. Mean pre-outcome credibility ratings by probability – Experiment 1 (Error bars represent ± 1 SE).

Experiment 2.

A 3 (communication format – congruent, incongruent, numerical) × 2 (probability) ANOVA revealed no significant effect of probability, $F(1, 411) = 1.77, p = .18$, and a main effect of communication format, $F(2, 411) = 3.74, p < .001, \eta_p^2 = .02$ on pre-outcome credibility ratings (see Figure S 2). There was a significant interaction between communication format and probability, $F(2, 411) = 7.40, p = .001, \eta_p^2 = .04$. Simple effects analyses showed a significant effect of communication format for the high probability expressions, $F(2, 411) = 9.29, p < .001, \eta_p^2 = .04$, but not for the low probability expressions, $F(2, 411) = 1.91, p = .15$.

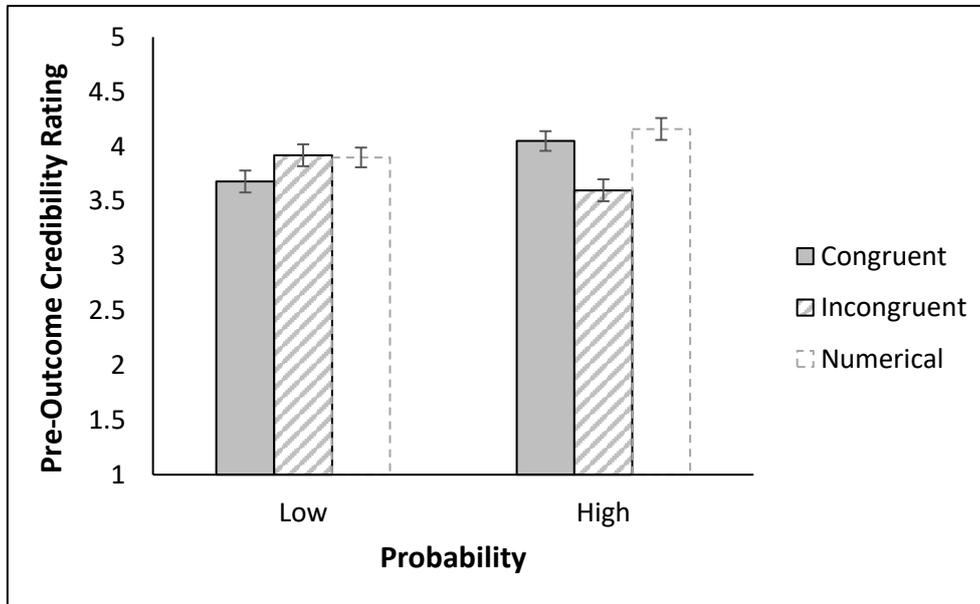


Figure S 2. Mean pre-outcome credibility ratings by probability – Experiment 2. (Error bars represent ± 1 SE).

3. Comparing VPEs and Numerical Expressions – Unplanned Analyses

In the following analyses, we categorise communication format as congruent, incongruent and numerical for ease of comparison with the results presented in the main text.

Experiment 1.

Post-outcome credibility ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 430) = 7.13, p < .001, \eta_p^2 = .03$ and a significant effect of probability, $F(1, 430) = 21.38, p < .001, \eta_p^2 = .05$ on post-outcome credibility ratings. There was a significant interaction between communication format and probability, $F(2, 430) = 6.87, p = .001, \eta_p^2 = .03$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 430) = 3.53, p = .03, \eta_p^2 = .02$, and the high probability domains, $F(2, 430) = 10.56, p < .001, \eta_p^2 = .05$.

Change in credibility ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 430) = 20.78, p < .001, \eta_p^2 = .09$ and a significant effect of probability, $F(1, 430) = 21.65, p < .001, \eta_p^2 = .05$ on credibility difference scores. There was a significant interaction between communication format and probability, $F(2, 430) = 22.21, p < .001, \eta_p^2 = .09$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 430) = 3.57, p = .03, \eta_p^2 = .02$, and the high probability domains, $F(2, 430) = 39.94, p < .001, \eta_p^2 = .16$.

Correctness ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 430) = 52.52, p < .001, \eta_p^2 = .20$ and a significant effect of probability, $F(1, 430) = 7.24, p = .007, \eta_p^2 = .02$ on correctness ratings. There was a significant interaction between communication format and probability, $F(2, 430) = 25.83, p < .001, \eta_p^2 = .11$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 430) = 21.85, p < .001, \eta_p^2 = .09$, and the high probability domains, $F(2, 430) = 57.72, p < .001, \eta_p^2 = .21$.

Surprise ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 430) = 4.15, p = .021, \eta_p^2 = .02$ and a significant effect of probability, $F(1, 430) = 5.67, p = .02, \eta_p^2 = .01$ on surprise ratings. There was a significant interaction between communication format and probability, $F(2, 430) = 9.94, p < .001, \eta_p^2 = .04$. Simple effects test showed there was a significant effect of communication format in the high probability domain, $F(2, 430) = 12.26, p < .001, \eta_p^2 = .05$, but not the low probability domain, $F(2, 430) = 1.72, p = .18$.

Experiment 2.

Post-outcome credibility ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 411) = 24.61, p < .001, \eta_p^2 = .11$, but no significant effect of probability, $F(1, 411) = 0.36, p = .55$, on post-outcome credibility ratings. There was a significant interaction between communication format and probability, $F(2, 411) = 23.45, p < .001, \eta_p^2 = .10$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 411) = 15.16, p < .001, \eta_p^2 = .07$, and the high probability domains, $F(2, 411) = 34.01, p < .001, \eta_p^2 = .14$, though the effect was larger in the latter.

Change in credibility ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 411) = 18.01, p < .001, \eta_p^2 = .08$, but no significant effect of probability, $F(1, 411) = 2.60, p = .11$, on credibility difference scores. There was a significant interaction between communication format and probability, $F(2, 411) = 12.01, p < .001, \eta_p^2 = .06$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 411) = 21.28, p < .001, \eta_p^2 = .09$, and the high probability domains, $F(2, 411) = 9.30, p < .001, \eta_p^2 = .04$.

Correctness ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed a main effect of communication format, $F(2, 411) = 63.70, p < .001, \eta_p^2 = .24$, but no significant effect of probability, $F(1, 411) = 0.46, p = .50$, on correctness ratings. There was a significant interaction between communication format and probability, $F(2, 411) = 34.66, p < .001, \eta_p^2 = .14$. Simple effects test showed there was a significant effect of communication format in both the low probability, $F(2, 411) = 51.25, p < .001, \eta_p^2 = .20$, and the high probability domains, $F(2, 411) = 48.71, p < .001, \eta_p^2 = .19$, though it was larger in the former.

Surprise ratings.

A 3 (communication format) \times 2 (probability) ANOVA revealed no significant effect of communication format, $F(2, 411) = 1.83, p = .16$, nor probability, $F(1, 411) = .015, p = .90$, on surprise ratings. There was a significant interaction between communication format and probability, $F(2, 411) = 5.06, p = .007, \eta_p^2 = .02$. Simple effects test showed there was a significant effect of communication format in the high probability domain, $F(2, 411) = 6.69, p = .001, \eta_p^2 = .03$, but not the low probability domain, $F(2, 411) = 0.42, p = .66$.

4. Weighing up Directionality Congruence and Probabilistic Expectedness

A 2 (directionality congruence) \times 2 (probabilistic expectedness) ANOVA revealed a significant effect of directionality congruence, $F(1, 849) = 70.70, p < .001, \eta_p^2 = .08$, with a greater reduction in credibility when the communicator used a directionally incongruent format ($M = -.69, SE = 0.05$) versus congruent format ($M = -.07, SE = 0.05$). There was a main effect of probabilistic congruence, $F(1, 849) = 69.98, p < .001, \eta_p^2 = .08$, with a greater reduction in credibility when the outcome was probabilistically unexpected ($M = -.69, SE = 0.05$) versus expected ($M = -.07, SE = 0.05$). There was no significant interaction between directionality congruence and probabilistic expectedness, $F(1, 849) = 0.09, p = .77$.

5. Comparing Across Experiments (Unplanned Analyses)

Table S3. Calculation of Expected Difference Scores (Underestimation)

Probability	Expression	Expected difference score	Mean credibility difference score		Overall 'expected difference score'
			Exp 1	Exp 2	
Low	Doubtful	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	- 1.18	0.62	0.44
	Small chance	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	- 0.91	- 0.12	- 0.20
	10–30% likelihood	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	- 0.71	- 0.45	- 0.48
High	Not entirely definite	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	0.41	- 0.57	- 0.47
	Good chance	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	- 0.80	0.06	- 0.03
	70–90% likelihood	$(0.1 \times \text{Exp 1}) + (0.9 \times \text{Exp 2})$	- 1.00	0.07	- 0.04

Note. Exp 1 and Exp 2 represent the mean credibility difference scores from each cell of Experiment 1 (probabilistically unexpected outcome) and Experiment 2 (probabilistically expected outcome) respectively.

Table S4. Calculation of Expected Difference Scores (Overestimation)

Probability	Expression	Expected difference score	Mean credibility difference score		Overall 'expected difference score'
			Exp 1	Exp 2	
Low	Doubtful	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	- 1.18	0.62	0.08
	Small chance	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	- 0.91	- 0.12	- 0.35
	10–30% likelihood	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	- 0.71	- 0.45	- 0.53
High	Not entirely definite	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	0.41	- 0.57	- 0.28
	Good chance	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	- 0.80	0.06	- 0.20
	70–90% likelihood	$(0.3 \times \text{Exp 1}) + (0.7 \times \text{Exp 2})$	- 1.00	0.07	- 0.25

Note. Exp 1 and Exp 2 represent the mean credibility difference scores from each cell of Experiment 1 (probabilistically unexpected outcome) and Experiment 2 (probabilistically expected outcome) respectively.

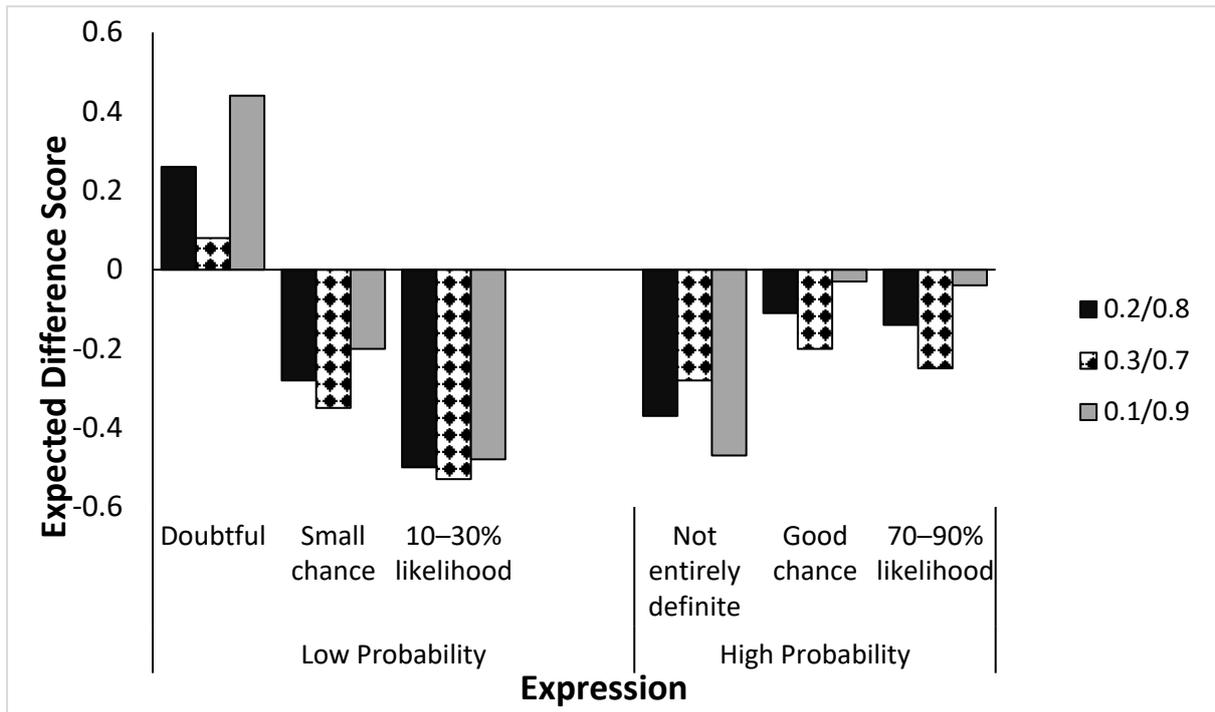


Figure S3. Expected difference scores across based on the alternative values for the communicator's calibration.