When Unlikely Outcomes Occur: The Role of Communication Format in Maintaining Communicator Credibility

Sarah C. Jenkins[a], Adam J. L. Harris[a] and R. M. Lark[b]

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

[b] British Geological Survey (BGS), Environmental Science Centre, Keyworth, Nottingham NG12 5GG.

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): 8052 (excluding abstract and figure captions)

Abstract

The public expects science to reduce or eliminate uncertainty (Kinzig & Starrett, 2003), yet scientific forecasts are probabilistic (at best) and it is simply not possible to make predictions with certainty. Whilst an 'unlikely' outcome is not expected to occur, an 'unlikely' outcome will still occur one in five times (based on a translation of 20%, e.g. Theil, 2002), according to a frequentist perspective. When an 'unlikely' outcome does occur, the prediction may be deemed 'erroneous', reflecting a misunderstanding of the nature of uncertainty. Such misunderstandings could have ramifications for the subsequent (perceived) credibility of the communicator who made such a prediction. We examine whether the effect of 'erroneous' predictions on perceived credibility differs according to the communication format used. Specifically, we consider verbal, numerical (point and range [wide / narrow]) and mixed format probability expressions. We consistently find that subsequent perceptions are least affected by the 'erroneous' prediction when it is expressed numerically, regardless of whether it is a point or range estimate. Our findings suggest numbers should be used in consequential risk communications regarding 'unlikely' events, wherever possible.

Imagine your house is located on a floodplain, 10 km away from a river. Recently there has been considerable rainfall. Consequently, you are worried about rising river levels and the possibility of flooding. A geologist announces that "it is *unlikely* the floodwaters will extend 10 km from the river." Four days later, the riverbanks burst and floodwaters *do* extend 10 km, flooding the ground floor of your house. How would you feel? Was the prediction incorrect? Would you trust the geologist again?

In the current paper, we investigate scenarios in which events forecast to be 'unlikely' nevertheless occur. We term these forecasts 'erroneous', despite the fact scientific forecasts are probabilistic (at best) and it is simply not possible to predict with certainty the probability of, for example, floodwater extending a certain distance. An 'unlikely' (e.g. 20% likelihood; Theil, 2002) event is not impossible – one in five times it will occur, based on a frequentist interpretation of probability. People's misunderstandings of the nature of uncertainty (e.g. thinking a prediction 'erroneous'), and/or the expressions used to convey uncertainty, might negatively influence perceptions of a communicator's credibility. This paper addresses such misunderstandings in relation to three communication formats and their variants: verbal probability expressions (VPEs; e.g. 'unlikely'), numerical expressions (e.g. point – '20% likelihood', range – '10–30% likelihood'), and mixed expressions (e.g. 'unlikely [20% likelihood]')[1]. Our decision to focus on 'unlikely' and its associated numerical probabilities was made primarily because events with highly consequential outcomes, such as geological hazards, are usually unlikely (c.f. Harris & Corner, 2011; Weber & Hilton, 1990). Individuals often discount such events and thus fail to prepare for them (McClure, Henrich, Johnston, & Doyle, 2016).

---

[1] In line with standard dictionary definitions, we use 'likelihood' as a synonym for 'probability' in the present paper, though note that, mathematically, each has a unique and specific definition.

Credibility

Most individuals do not have in-depth knowledge about, nor experience of, hazards and new technologies (Siegrist, Gutscher, & Earle, 2005) (the focus of many risk communications), so they are often reliant on mediated information from an external source (Sjöberg, 2000). Credibility is one of a communicator's most precious assets (Covello & Allen, 1988); with trust and expertise identified as two key components (Hovland, Janis, & Kelley, 1953; Mowen, Wiener, & Joag, 1987; Wiener & Mowen, 1986). Trust relates to the reliable and unbiased nature of the communicator (Dieckmann, Mauro, & Slovic, 2010), with expertise relating to the knowledgeability of the communicator (Guilamo-Ramos, Jaccard, Dittus, & Bouris, 2006).

Whether a source is perceived as credible or not will influence how a risk is perceived and thus how an individual behaves upon receiving a risk communication (Wachinger, Renn, Begg, & Kuhlicke, 2013), for instance taking action to avoid, mitigate or adapt to the risks emphasised in the communication. Perceiving a source as untrustworthy could lead an individual to ignore such risks. Credibility's influence is far-reaching, though has largely been neglected in investigations of effective risk and uncertainty communication.

Credibility and Communication Format

People expect experts to provide precise information (Shanteau, 1992). Precision is often used as a cue for high expertise, with preference given to an advisor who gives precise estimates (Jerez-Fernandez, Angulo, & Oppenheimer, 2014). Given this, it is reasonable to assume a communicator who uses a (precise) numerical point estimate of uncertainty (e.g. 15%) will be perceived as more credible than one who uses a (less precise) verbal or range[2] expression. However, there is evidence to suggest the way expressions are perceived depends on the characteristics of the event in question. The congruence principle (Budescu & Wallsten, 1995)

---

[2] Throughout the paper, 'range format' refers to probability ranges (e.g. 10-30% likelihood), rather than general range estimates (e.g. 3 – 5 km).

states that the precision of the communication should be consistent with how precise one can be in describing the (un)certainty of the outcome. Using a point estimate to describe the chance of a natural hazard (which is, by nature, highly uncertain) might be perceived as overly precise. The recognition that such precision is unwarranted for such an event may reduce a communicator's perceived credibility. Use of a less precise format may consequently be perceived as more appropriate in such a situation.

Range Formats.

Research on how range formats are perceived is mixed, with Longman, Turner, King and McCaffery (2012) observing reduced understanding with range, as opposed to point, estimates. Han et al. (2011) also observed greater worry with risk estimates presented as ranges. Other research has found point and range estimates (e.g. 'as low as 1% and as high as 5%') to be perceived as equally credible, trustworthy, accurate and personally relevant (Lipkus, Klein, & Rimer, 2001), with no difference in perceived risk / harm, nor severity (Sladakovic, Jansen, Hersch, Turner & McCaffery, 2015).

Another body of research, however, suggests *benefits* of using ranges in risk communications. It has been suggested that being open about levels of uncertainty may lead to enhanced credibility (Chess, Hance, & Sandman, 1988). In this vein, range estimates have been found to be not only more useful (Dieckmann et al., 2010), but also more honest (Johnson & Slovic, 1995) than point estimates, perhaps because they openly acknowledge uncertainty at the outset (Joslyn & LeClerc, 2012). The usefulness of a range may, however, depend on its width. Previous research has observed a preference for narrow range forecasts (Du, Budescu, Shelly, & Omer, 2011) over wider ranges, with the former perceived as more competent and trustworthy (Jørgensen, 2016). Indeed, people trust precise forecasts more, despite the fact such forecasts have a smaller chance of capturing the actual outcome than less precise forecasts (Løhre & Teigen, 2017).

The aforementioned research has so far solely focused on how communicators will be perceived with reference to their predictions, rather than in combination with actual outcomes. When the actual outcome is opposite to what was predicted (an 'erroneous' prediction), a communicator who used a range format was perceived as more credible, and blamed less than one who used a point (percentage) format to describe the risk of a terrorist attack (Dieckmann, et al., 2010).

Verbal Formats.

It could be suggested that VPEs may be perceived similarly to range formats, given they openly acknowledge uncertainty and have consistently been shown to refer to a range of probabilities (e.g. Budescu & Wallsten, 1995). However, research generally highlights the *disadvantages* of using VPEs in risk communications. Gurmankin, Baron and Armstrong (2004) investigated the effect of verbal and numerical statements of risk (percentage / fraction) on trust and comfort in a physician in a hypothetical medical communication. Participants were more trusting of, and more comfortable with, numerical versions of the information, though this effect decreased with lowering levels of numeracy. Such findings coincide with the observation that individuals prefer to receive information in a numerical format, though prefer to communicate information using VPEs – the so-called Communication Mode Preference paradox (Erev & Cohen, 1990).

Even behavioural differences have been observed between verbal and numerical communication formats. Peters, Hart, Tusler and Fraenkel (2014) compared the use of numeric and non-numeric formats (including a VPE condition) to describe the likelihood of medication side effects (probabilities $\leq 14\%$). Participants were more willing to take the medication when side effects were presented in a numeric format, suggested to result from the heightened risk perceptions of those in the non-numeric condition. In contrast, Doyle, McClure, Paton and Johnston (2014) used range expressions featuring higher probabilities (e.g. 45-55%, 73-83%) and found that more people recommended evacuating when the risk of a volcanic eruption was

described using ranges compared to VPEs. They attributed this to the ambiguity of VPEs – numerical terms were perceived as more certain and thus required more immediate action.

Further evidence of the disparity between numerical formats and VPEs derives from research adopting the 'which-outcome' approach, used to examine understanding of VPEs (e.g. Juanchich, Teigen, & Gourdon, 2013; Løhre & Teigen, 2014; Teigen, Juanchich, & Filkuková, 2014; Teigen, Juanchich, & Riege, 2013). In this approach, participants are shown a histogram of potential outcomes and asked to complete a probability statement (e.g., "It is *unlikely* that the floodwater will extend __ km") with a value considered appropriate. Participants tend to complete the 'unlikely' sentence with a value that exceeded any represented in the histogram, equivalent to a 0% likelihood of occurrence – the 'extremity effect' – an effect which was not observed with numerical probabilities (Juanchich et al., 2013; Jenkins, Harris & Lark, in press). The 'extremity effect' could plausibly be attributed to directionality (Teigen, 1988; Teigen & Brun, 1995, 1999), which drives expectations about outcomes (Teigen & Brun, 2003). Phrases which have negative directionality (e.g. 'unlikely') focus one's attention on the non-occurrence of the event, whereas those with positive directionality (e.g. 'likely') focus on the occurrence of the event. Therefore, a participant in a W-O task who reads the sentence featuring 'unlikely' will focus on the event not occurring, potentially leading them to complete the sentence with an outcome which has never previously occurred.

The 'extremity effect' poses a problem for communicators of risk and uncertainty who use communications featuring VPEs. If 'unlikely' is used to convey an intended meaning of 20%, but is instead interpreted as equivalent to 0%, such a mismatch could adversely affect confidence in subsequent communications (Breznitz, 1984). Furthermore, whilst mixed format expressions have been proposed as a solution to the problem of miscommunications (e.g. Budescu, Broomell, & Por, 2009), simply adding a numerical expression after the VPE (e.g. 'unlikely [20%]') may not be enough to wholly prevent such a mismatch occurring. Indeed,

the 'extremity effect' is observed in verbal-numerical (V-N) expressions (Juanchich & Sirota, 2017; Jenkins et al., in press), indicating such expressions will be susceptible to the effects of directionality. Altering the order of the mixed format to create a numerical-verbal (N-V) expression (e.g. '20% likelihood [unlikely]'), however, is enough to nearly eliminate the 'extremity effect' (Jenkins et al., in press). Therefore, it is not only reasonable to expect that numerical expressions will be more robust to 'erroneous' predictions than VPEs, but also reasonable to believe numerical-verbal (N-V) expressions will be superior to V-N expressions as well.

Overview of Proposed Research

The present paper advances our understanding of the effects of communication format on credibility by investigating the effect of 'erroneous' predictions. Experiment 1 incorporated mixed expressions (V-N, and the previously unconsidered N-V format) and compared them to a numerical point estimate and VPE. Building on key findings from Experiment 1, Experiments 2 and 3 investigated the influence of the precision of the expression, by including additional range (narrow / wide) formats, in conjunction with events differing in perceived predictability. All three experiments explored the potential influence of numeracy. Ascertaining the effect of the aforementioned factors is vital to building and maintaining the trust of the public, especially given the 'crisis of trust' science has suffered (House of Lords, 2000). Our main hypotheses were as follows:

Hypothesis 1.

VPEs will suffer the greatest loss of credibility after an 'erroneous' prediction (compared to numerical expressions). Similarly, V-N expressions will be less robust to 'erroneous' predictions than numerical-verbal (N-V) expressions.

Hypothesis 2.

Ranges will be more robust against 'erroneous' predictions than numerical point expressions, as per the congruence principle (Budescu & Wallsten, 1995).

Hypothesis 3.

As per the congruence principle, outcome characteristics (such as how predictable the event is perceived to be) will interact with range width. A narrow range will be perceived more positively than the wide range for the more predictable event. In contrast, for the less predictable event, the wide range will be perceived more positively than the narrow range.

Experiment 1

Method

Participants.

300 Native English speakers were recruited from Amazon Mechanical Turk. Nine 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 291 participants (146 male), aged 19 – 80 (*Mdn* = 33). Participants were paid $0.60. Ethical approval for all studies in this paper was granted from the Departmental Ethics Chair for Speech, Hearing and Phonetic Sciences (University College London).

Design.

Communication format had four levels, manipulated between participants. Participants were randomly allocated to either a verbal – "unlikely"; numerical – "20% likelihood"; V-N – "unlikely (20% likelihood)" or N-V – "20% likelihood (unlikely)" condition.

Participants answered questions regarding trust and expertise. Expertise was operationalised as: "How knowledgeable does the expert seem?" rated from 1 – 'Not at all knowledgeable' to 5 – 'Extremely knowledgeable'. Trust was operationalised as: "How much

do you trust that the expert is giving you complete and unbiased information?" (Dieckmann, Slovic, & Peters, 2009), rated from 1 – 'Not at all' to 5 – 'A great deal'. For greater clarity of results (i.e. avoiding repetition of similar analyses), these measures were averaged to form a credibility score. Credibility was found to be highly reliable, both pre- and post-outcome (α = .74, α = .86, respectively). As we were interested in how robust credibility perceptions were to 'erroneous' predictions, credibility difference scores ([post-outcome credibility rating] – [pre-outcome credibility rating]) were our main focus for analyses.

As previous research has found a disparity between explicit ratings of trust and more implicit behavioural measures of trust for various information sources (O'Neill, 2002; Twyman, Harvey, & Harries, 2008), we used 'decision to evacuate' (Doyle et al., 2014) as a measure of the latter. This was rated from 1 – 'Definitely should evacuate today' to 5 – 'Definitely should not evacuate today'. Once more, we were interested in the change in such decisions after an 'erroneous' prediction. Consequently, difference scores ([post-outcome evacuation rating – pre-outcome evacuation rating]) were calculated and served as the primary dependent measure of implicit credibility. We did not make directional predictions for this behavioural measure since the directional effect a reduction in implicit credibility will have is difficult to predict a priori. *Any* difference, however, would be suggestive of a change in implicit perceptions of trust in the communicator. Participants also indicated why they made their decision.

After providing their second credibility ratings, participants were also asked "How correct was the geologist's prediction?" rated from 1 – 'Not at all correct' to 5 – 'Completely correct' (Teigen, 1988; Teigen & Brun, 2003).

Numeracy was measured using Lipkus, Samsa, & Rimer's (2001) numeracy scale with two additional questions from the Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, &

Garcia-Retamero, 2012). The latter were included to increase variability in scores, given we have previously observed MTurk samples to be highly numerate.

Materials and Procedure.

After consenting to participate, participants indicated their age and gender before completing an attention check. Participants then read a brief introduction. On the next screen, participants read a vignette about an ongoing volcanic eruption (see Supplementary Materials 1 for full vignette), in which lava flows were expected. A volcanologist communicated the probability of the lava flows travelling a certain distance:

> "Mount Ablon has a history of explosive eruptions that have produced lava flows. An eruption is currently underway and lava flows are expected. Volcanologists from Ablon Geological Centre are communicating information about the volcano. A volcanologist has suggested that, given the volcano's recent history, there is a **20% likelihood (unlikely)** that the lava flow will extend 3.5km from the point of eruption" (N-V condition, emphasis in original).

Participants then provided initial ratings of expertise and trust in the expert's prediction. On the subsequent screen, participants were informed that the capital city was at risk from the volcanic eruption (given its location 3km from the volcano) and asked to rate whether to evacuate the city today or not, as well as why they made this decision. A mass evacuation was described as being 'very expensive and extremely disruptive to residents'.

On the following screen, participants were informed that the unlikely outcome did in fact occur – the lava flow extended beyond 3.5km. They were asked to provide further trust and expertise ratings, as well as rating how correct the volcanologist's prediction was in light

of the outcome. On the next screen, participants were informed that it was now 2019 (two years later), and were presented with the same prediction from the volcanologist. They were then asked the two evacuation questions as before.

Finally participants answered the numeracy questions and upon completing the study, they were given a code to claim their reward, thanked and debriefed (see Figure 1 for a flow chart of the procedure).

Results

As mentioned above, our focus was the change in ratings and decisions after an 'erroneous' prediction, so our analyses focus on difference scores. Analyses of all pre-outcome ratings are available in Supplementary Materials 2 (across all three experiments, there were no robust effects of format on credibility).

Answers for each numeracy question were coded as 1 if correct and 0 if incorrect, such that numeracy scores could range from 0 to 10. The distribution of numeracy scores is shown in

Table 1. Given the highly skewed distribution of responses, participants with scores of eight or under were categorised as low numeracy, and those with nine or above categorised as high numeracy. Each dependent measure was entered into a 4 (communication format) $\times$ 2 (numeracy) ANOVA.

Credibility Ratings.

As can be seen in Figure 2, all communication formats suffered from a loss of perceived credibility post-outcome, but to a different degree, $F (3, 282) = 7.61$, $p < .001$, $\eta_p^2 = .08$. The numerical format was the most robust to the 'erroneous' prediction, and the verbal format suffered from the greatest reduction in credibility. A post-hoc Gabriel's procedure [3]

---

[3] Gabriel's procedure is recommended for use with unequal group sizes (Field, 2013). Critical value includes adjustment for multiple comparisons.

demonstrated that the verbal format was significantly different to the numerical format ($p <$ .001), with the numerical format significantly different to the V-N format ($p < .001$). There was a non-significant trend for the communicator to suffer a greater reduction in perceived credibility when rated by the high numeracy group, $F(1, 282) = 3.40$, $p = .07$, $\eta_p^2 = .01$. The effect of communication format was not qualified by an interaction with numeracy, $F(3, 282) = 0.85$, $p = .47$. $\eta_p^2 = .01$[4].

Correctness Rating.

Correctness ratings followed the pattern of differences in the credibility ratings, with the numerical format seen as 'least incorrect' and the verbal format seen as 'most incorrect' after the outcome had occurred (see

Figure 3), $F(3, 282) = 26.40$, $p < .001$, $\eta_p^2 = .22$. A post-hoc Gabriel's procedure demonstrated a significant difference between the verbal and both numerical ($p < .001$) and N-V ($p < .001$) formats. The numerical format was also significantly different to the V-N format ($p < .001$). There was no effect of numeracy, $F(1, 282) = 1.69$, $p = .20$, $\eta_p^2 = .01$ nor an interaction between format and numeracy, $F(3, 282) = 0.41$, $p = .75$, $\eta_p^2 = .004$.[5]

Decisions to Evacuate.

---

[4] Across all experiments, results for interactions remain non-significant if numeracy is entered as a continuous variable.

[5] Results for credibility and correctness ratings were replicated in a Prolific Academic sample (n= 300).

There was little change in how certain participants were about evacuating (pre-outcome, M= 2.67. SD= 1.19, post-outcome M= 2.06, SD= 1.28; see Table 2), with no significant effect of communication format, numeracy, nor an interaction between the two factors (all ps > .23).

Discussion

All formats suffered a reduction in credibility after the 'erroneous' prediction. Consistent with Hypothesis 1, the numerical (verbal) format was perceived as least (most) incorrect. The poor performance of the verbal format could plausibly be explained with regard to directionality (Teigen, 1988; Teigen & Brun, 1995, 1999). We find clear evidence of people's sensitivity to the order of mixed format expressions (see Figure 2 and 3), in line with previous findings (Jenkins et al., in press).

We found no effect of communication format on decisions to evacuate, in contrast to Doyle et al. (2014). A large number of responses to the question of why people made their evacuation decision mentioned themes such as 'evacuating just in case' and 'better to be safe than sorry'. There was little cost to the participant to adopt such an approach, which could explain the relatively high proportion of people choosing to evacuate immediately. We modified this behavioural measure in subsequent experiments.

<center>Experiment 2</center>

In Experiment 1, numerical expressions outperformed both verbal and mixed format expressions. Numerical point estimates have, however, previously been criticised for portraying a false level of certainty (Schwartz, Woloshin, & Welch, 1999). Experiment 2 was therefore designed to extend the type of uncertainty estimates investigated by including a range format ('10 – 30% likelihood'). We used two scenarios which, from pilot testing, were perceived very differently by participants, both in terms of a) how predictable the event was and b) how precise the communicator could be in describing the likelihood of the event.

Method

Participants.

Two hundred and fifty participants were recruited from Amazon Mechanical Turk, with seven cases removed due to failing the attention check, leaving a final sample of 243 (114 male) participants aged between $20 - 77$ (*Mdn* = 33). Participants were paid $0.40 for this 10 minute study.

Design.

Format (verbal – "unlikely"; numerical point – "20% likelihood"; numerical range – "10 – 30% likelihood") and scenario (forest fire [rated the least predictable / precise in the pilot study] and flood [rated the most predictable / precise]) were manipulated between-participants in a 3 × 2 design. Participants were randomly assigned to conditions.

Credibility measures were the same as in Experiment 1. For the behavioural measures, participants were first given a 'willingness to pay' measure in the context of home insurance (see Figure 4)[6], and then also completed an adapted version of the tolerable risk scale[7] (Haynes, Barclay, & Pidgeon, 2008).

Materials and Procedure.

Participants read a vignette about a current forest fire or flood in which a geologist communicated the risk of fire or floodwater travelling a certain distance (see Supplementary Materials 4). The procedure was the same as Experiment 1 (see Figure 1), though included the different behavioural measures.

Results

Credibility Ratings.

---

[6] Values for this task were selected on the basis of results from pilot testing.

[7] Further details and analysis of this scale is presented in Supplementary Materials 3. No significant effects or interactions were observed.

Credibility was found to be highly reliable, both pre- and post-outcome (α = .81, α = .86,

respectively). As can be seen in Figure 5, all communication formats suffered from a loss of

perceived credibility post-outcome, but to varying degrees, F (2, 231) = 11.58, p < .001, η_p^2=

.09. The verbal format suffered from the greatest reduction in credibility and a post-hoc

Gabriel's procedure demonstrated that the verbal format was significantly different to the

numerical point (p < .001) and range (p < .001) formats, though there was no significant

difference between the latter two. Credibility difference scores were not affected by numeracy,

F (1, 231) = 0.04, p = .85, η_p^2 < .001, or scenario, F (1, 231) = 0.53, p = .47, η_p^2= .002. No two

or three way interactions were significant (all ps > .09).

Correctness Ratings.

Correctness ratings followed the pattern of the credibility ratings (see Figure 6), with the

numerical point and range formats perceived as 'most correct' and the verbal format seen as

'least correct' after the outcome had occurred, F (2, 231) = 33.97, p < .001, η_p^2= .23. A post-

hoc Gabriel's procedure demonstrated that the verbal format was significantly less correct than

the point (p < .001) and range (p < .001) formats, though the latter two were similarly perceived.

There was no significant effect of scenario, nor numeracy, nor any two or three way interactions

(all ps > .16).

Insurance Decisions.

Participants who provided inconsistent responses (e.g., indicating they would not pay £750 for

insurance, but would pay £800) were removed from analysis, even if they were only

inconsistent at one time point. All participants were willing to pay more for insurance after the

outcome occurred (see Table 2), but this was not significantly affected by communication

format, F (2, 218) = 1.12, p = .33 η_p^2= .01, scenario, F (1, 218) = 2.38, p = .13 η_p^2= .01, or

numeracy, F (1, 218) = 0.74, p = .39, η_p^2= .003. No two or three way interactions were

significant (all ps > .09).

Discussion

Whilst we observed an effect of format on both ratings of credibility and correctness, there were no differences between ratings for the numerical point and range formats, in contrast to Hypothesis 2. The effect of format is driven solely by the verbal format, replicating findings from Experiment 1. No effects of format, scenario or numeracy were observed for either of the behavioural measures.

Our results add to the mixed findings of research exploring the effect of using numerical ranges on perceptions of the communicator. Our results coincide with those of Lipkus et al. (2001) who found both range and point estimates were perceived similarly in terms of credibility and accuracy. It is possible the divergent findings in the literature may have arisen from differences in the size of the range presented, a factor considered in Experiment 3.

Experiment 3

Experiment 3 was designed to examine whether the precision of the range would affect perceptions of credibility, and follows recommendations to further explore how people respond to varying range sizes (Longman et al., 2012). It also provided a further opportunity to test the congruence principle and check the generalisability of key results in Experiments 1 and 2. We continue to use the range featured in Experiment 2, hereafter refered to as the wide range format, and include a new, narrower, range format.

We also included measures of worry / concern and likelihood ratings in order to draw further comparisons to prior research (Han et al., 2011; Lipkus et al., 2001; Sladakovic et al., 2015). We expected a prediction made using an uncertain expression (wide numerical range) would be perceived as more worrisome and less likely than a prediction using a point estimate.

Method

Participants.

Two hundred and fifty-five Native English adult speakers were recruited from Prolific Academic (www.prolific.ac), with two cases removed for failing the attention check, and seven for duplicate IP addresses, leaving a final sample of 246 (124 male) participants aged between 18 – 69 (*Mdn* = 31). Participants were paid £0.85 for this 10 minute study.

Design, Materials and Procedure.

Communication format (numerical point – "20% likelihood"; wide range – "10 – 30% likelihood"; narrow range – "15 – 25% likelihood") and scenario (forest fire and flood) were manipulated between-subjects. Participants were randomly allocated to one of the six conditions.

The dependent variables were as in Experiment 2, with two changes. The tolerable risk scale was removed and two additional measures were added – worry / concern ratings and likelihood ratings. Participants answered the following: "How worried / concerned would you be given the geologist's prediction that …" rated from 1 – 'not at all worried' to 5 – 'extremely worried' and "How likely do you think it is that the [e.g.] forest fire will extend 80km from its origin, given the geologist's prediction that there is a 20% likelihood of this occurring?" rated from 1 – 'not at all certain' – 7 'completely certain'.

Results

Credibility Ratings.

Credibility was found to be highly reliable, both pre- and post-outcome (α = .82, α = .89, respectively). The narrow range format seemed to suffer a smaller loss of credibility compared to the other numerical expressions (see Figure 7). However, a 3 × 2 × 2 mixed ANOVA revealed there were no significant differences between communication formats, F (2, 234) = 1.85, p = .16, η_p^2 = .02. There was also no effect of scenario, F (1, 234) = .1.57, p = .21, η_p^2 = .01. The communicator suffered a greater reduction in perceived credibility when rated by the

low numeracy group, F (1, 234) = 10.96, $p < .01$, $\eta_p^2 = .05$, though there was a significant scenario × numeracy interaction, F (1, 234) = 4.97, $p < .05$, $\eta_p^2 = .02$. Figure 8 demonstrates this interaction: whilst the credibility ratings of highly numerate participants were similarly affected in the flood and fire scenarios, F (1, 234) = 0.55, $p = .46$, $\eta_p^2 = .002$, ratings of those lower in numeracy were less affected by the 'erroneous' prediction in the flood scenario, F (1, 234) = 5.34, $p < .05$, $\eta_p^2 = .02$. For these participants, larger reductions in credibility were observed in the least predictable (fire) scenario (M= -1.07, SE= 0.15) compared to the most predictable (flood) scenario (M= -.62, SE= 0.13).

Correctness Ratings.

A 3 × 2 × 2 ANOVA revealed no significant effect of communication format on correctness ratings, F (2, 234) = 1.03, $p = .36$, $\eta_p^2 = .01$, nor of scenario, F (1, 234) = 0.79, $p = .38$, $\eta_p^2 = .003$. The significant effect of numeracy can be seen in Figure 9 – the low numeracy group rated the communicator as less correct (M= 2.82, SE= 0.13) compared to the high numeracy group (M= 3.22, SE= 0.12), F (1, 234) = 4.92, $p < 0.05$, $\eta_p^2 = .02$, though there were no significant two or three way interactions (all ps > 0.14).

Worry / Concern Ratings.

A 3 × 2 × 2 mixed ANOVA revealed no significant effects of communication format, scenario or numeracy (all ps > .42) nor any significant two or three way interactions (all ps > .20), (M= 3.31, SE= 0.06).

Likelihood Ratings.

A 3 × 2 × 2 mixed ANOVA revealed no significant effects of communication format or scenario. There was a non-significant trend for more numerate participants to perceive the event as less likely (M= 3.17, SE= 0.11) compared to those with lower numeracy levels (M= 3.47, SE= 0.12), F (1, 234) = 3.56, $p = .06$, $\eta_p^2 = .02$.

Insurance Decisions.

Those who provided inconsistent responses were removed as in Experiment 2. Participants were willing to pay more for insurance after the outcome, though a $3 \times 2 \times 2$ ANOVA revealed no significant effect of communication format, $F(2, 223) = 0.83$, $p = .44$, $\eta_p^2 = .01$, scenario, $F(1, 223) = 0.62$, $p = .43$, $\eta_p^2 = .003$, or numeracy, $F(1, 223) = 0.49$, $p = .48$, $\eta_p^2 = .002$. There was a significant interaction between communication format and numeracy, $F(1, 223) = 4.38$, $p < .05$, $\eta_p^2 = .04$ (see Table 2). Both the high and low numeracy groups were willing to pay similar amounts for insurance in the wide range, $F(1, 223) = 1.30$, $p = .26$, $\eta_p^2 = .01$, and narrow range conditions, $F(1, 223) = 0.12$, $p = .73$, $\eta_p^2 = .001$. However, in the numerical point condition, those in the high numeracy group were prepared to pay significantly more than those in the low numeracy group, $F(1, 223) = 8.06$, $p = .01$, $\eta_p^2 = .04$. No other interactions were significant (all ps > .17).

Discussion

The current experiment finds no evidence to suggest precision of the range affects perceptions of credibility and correctness, contrary to Hypothesis 3. No effect of precision of format was observed for worry / concern or risk perceptions, in contrast to previous research in the health domain (Han et al., 2011; Longman, et al., 2012). The only instance in which the precision of the format affected responses was on the behavioural measure, though this was as part of an (unpredicted) interaction.

This experiment is the first in the present paper to observe a moderating effect of numeracy on perceptions. Given our sample was relatively numerate and thus little separated those in the high and low numeracy groups, we caution against drawing steadfast conclusions regarding these findings until they have been replicated in a more diverse sample.

General Discussion

We presented three experiments examining how perceptions of the communicator differ across communication formats (verbal, numerical point, numerical range [narrow, wide] and mixed expressions) in response to an 'erroneous' prediction. These experiments also considered the potential influence of numeracy. All experiments yielded consistent results; the first two replicated and extended previous findings, with the verbal (numerical point) format most (least) vulnerable to an 'erroneous' prediction. Although mixed formats (e.g. 'unlikely [20%]') have been posited as a solution to the problems of mis-communications arising from use of VPEs (e.g. Budescu et al., 2009), Experiment 1 provided little evidence of their added benefit, either in terms of credibility, correctness or behavioural decisions. Experiments 2 and 3 yielded no evidence to suggest differences between numerical point and numerical ranges' vulnerability to 'erroneous' predictions.

Robustness of the Numerical Formats

Probability expressions convey a double message (Teigen & Brun, 2003), telling us that an event may occur but also that it may not. Both messages will not, however, be similarly attended to – the way the message is framed can influence which message is more prominent, as well as the message's perceived accuracy in light of an outcome (Yeung, 2014). Framing can occur in VPEs through two different mechanisms (Teigen & Brun, 2003): directly – explicit mention of one of the complementary outcomes, or indirectly – through directionality (Teigen, 1988; Teigen & Brun, 1995, 1999). The negatively directional term 'unlikely' was used in the current experiments, which focused attention on the event *not* happening. We propose that the verbal format was perceived most negatively (supporting Hypothesis 1) because, when the event occurred, it was counter to the original focus on it not happening. In contrast, the numerical formats lacked this original focus. Thus, when the event occurred, we suggest it was less unexpected. Such an explanation is in line with the findings of Teigen and Brun (2003). Following an outcome's occurrence, predictions made with negatively directional VPEs were

rated as more 'wrong', and generated more surprise than their positively directional counterparts.

We recognise that the current results could have arisen from other differences in the way individuals reason about verbal and numerical expressions. It has previously been suggested that VPEs elicit an intuitive way of thinking, in comparison to the more deliberative, analytical type of reasoning evoked by numerical expressions (Windschitl & Wells, 1996). If this is the case, the latter type of reasoning might be behind the advantages observed for numerical probabilities. Future research could aim to compare the two accounts, by comparing negatively and positively directional (low probability) VPEs in the context of an 'erroneous' outcome. On the basis of the directionality account, we would expect a smaller decrease in credibility ratings when a positively directional VPE such as 'a chance' was used to describe the event, yielding similar ratings to a numerical expression such as '20% likelihood'.

Point Versus Range Formats

We observed no evidence for differences in the way point and range expressions were perceived, a null effect which persevered even when the precision of the format and predictability of the scenario was manipulated (contrary to Hypotheses 2 and 3). This could arise from the way numerical estimates are processed. It has previously been suggested that range estimates are processed in a similar way to point estimates. When one is presented with a range estimate, it is not possible to make an imprecise decision or take imprecise action (Karelitz & Budescu, 2004), so it is more useful to focus on a single point. If range estimates are processed like point estimates, it stands to reason that they might be perceived similarly too, potentially explaining current results. However, considering the mixed findings of previous research, we also consider alternative explanations below.

A communicator who uses precise estimates is likely to be perceived positively, given precision is associated with knowledge and expertise (Jerez-Fernandez et al., 2014; Shanteau,

1992). However, the congruence principle (Budescu & Wallsten, 1995) states that the value of using a precise estimate will depend on the characteristics of the event in question. In the current studies, using a point estimate to describe geological hazards may have been viewed as artificially precise, a factor known to reduce comprehension and believability of the risk estimate (Witteman, Zikmund-Fisher, Waters, Gavaruzzi, & Fagerlin, 2011). Whilst a range estimate might be deemed as more appropriate for reflecting the uncertainty associated with a geological hazard, use of such formats have led to questions regarding a communicator's competence and ability to estimate risk and uncertainty (Dieckmann et al., 2010; Johnson & Slovic, 1995). Therefore it is possible that the advantages and disadvantages of each format served to cancel each other out in the current experiments.

Implications for Communicating Uncertainty

It is clear that how, and the degree to which, one should communicate uncertainty is still an open question. A communicator presenting uncertainty information must make a series of tradeoffs, to ensure the communication is made with confidence, but with precision (Moore, Tenney, & Haran, 2016), and is accurate, but also remains informative (Yaniv & Foster, 1995). Despite concerns that focusing on uncertainty may discourage action, distract people, or be perceived as untrustworthy (Fischoff, 2011; Frewer et al., 2003), we find no ill-effects of being open about uncertainty in terms of perceptions or behaviour, at least when it is numerically expressed (as a range versus point estimate). We therefore continue to support calls for uncertainty to be presented in order to enable people to make fully informed decisions (Fischoff & Davies, 2014; Politi, Han, & Col, 2007).

Although the work so far has focused on one low probability VPE, 'unlikely', we suggest our findings would extend to other negatively directional VPEs such as 'improbable'. We do however recognise that the extent to which our results can be generalised to positively directional VPEs is limited, especially as Smithson, Budescu, Broomell and Por, (2012) found

a larger range of interpretations were given for negatively as opposed to postively directional VPEs. Indeed, no differences between formats was observed when a 'likely' event did not occur (Jenkins, Harris, & Lark, 2017).

Conclusion

Research has predominantly studied the effects of communication format on understanding, but has largely neglected its influence on credibility and behavioural decisions. The present paper is one of the first to compare the effects of verbal, numerical [point / range] and mixed formats on these factors, in the context of an 'erroneous' prediction. Recognition of these effects is key to designing effective risk communications that build and maintain the public's trust. Our findings show that numerical formats are consistently perceived as more credible and 'less incorrect' following an 'erroneous' prediction, though no differences are observed between point and range estimates. We thus recommend numbers should be used in scientific risk communications referring to 'unlikely' events, wherever possible.

References

Breznitz, S. (1984). *The Psychology of False Alarms*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Budescu, D. V, Broomell, S. B., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, *20*(3), 299–308.

Budescu, D. V, & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation*, *32*(2), 275–318.

Chess, C., Hance, B. J., & Sandman, P. M. (1988). *Improving dialogue with communities: a short guide for government risk communication*. Division of Science and Research, New Jersey Department of Environmental Protection.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: the Berlin numeracy test. *Judgment and Decision Making*, *7*(1), 25–47.

Covello, V. T., & Allen, F. W. (1988). *Seven Cardinal Rules of Risk Communication*. Washington, D.C: US Environmental Protection Agency.

Dieckmann, N., Mauro, R., & Slovic, P. (2010). The effects of presenting imprecise probabilities in intelligence forecasts. *Risk Analysis*, *30*(6), 987–1001.

Dieckmann, N., Slovic, P., & Peters, E. M. (2009). The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. *Risk Analysis*, *29*(10), 1473–1488.

Doyle, E. E. H., McClure, J., Paton, D., & Johnston, D. M. (2014). Uncertainty and decision making: Volcanic crisis scenarios. *International Journal of Disaster Risk Reduction*, *10*, 75–101.

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.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes*, *45*(1), 1–18.

Field, A. P. (2013). *Discovering statistics using SPSS: and sex and drugs and rock "n" roll (4th Edition)*. London: Sage.

Fischoff, B. (2011). Applying the science of communication to the communication of science. *Climatic Change*, *108*(4), 701–705.

Fischoff, B., & Davies, A. L. (2014). Communicating scientific uncertainty. *Communication Research*, *111*(3), 429–456.

Frewer, L. J., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research*, *6*(1), 75–85.

Guilamo-Ramos, A. V., Jaccard, J., Dittus, P., & Bouris, A. M. (2006). Parental expertise, trustworthiness, and accessibility: parent- adolescent communication and adolescent risk behavior. *Journal of Marriage and Family*, *68*(5), 1229–1246.

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.

Han, P. K. J., Klein, W. M. P., Lehman, T. C., Killam, B., Massett, H., & Freedman, A. N. (2011). Communication of uncertainty regarding individualized cancer risk estimates: effects and influential factors. *Medical Decision Making*, *31*(2), 354–366.

Harris, A. J. L., & Corner, A. (2011). Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(6), 1571–1578.

Haynes, K., Barclay, J., & Pidgeon, N. F. (2008). Whose reality counts? Factors affecting the perception of volcanic risk. *Journal of Volcanology and Geothermal Research*, *172*(3–4), 259–272.

House of Lords. (2000). *Science and Technology - Third Report*. Retrieved from http://www.publications.parliament.uk/pa/ld199900/ldselect/ldsctech/38/3801.htm

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and Persuasion; Psychological Studies of Opinion Change*. New Haven, CT: Yale University Press.

Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (in press). Understanding "unlikely (20% likelihood)" or "20% likelihood (unlikely)" outcomes: The robustness of the extremity effect. *Journal of Behavioral Decision Making*.

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.

Jerez-Fernandez, A., Angulo, A. N., & Oppenheimer, D. M. (2014). Show me the numbers: Precision as a cue to others' confidence. *Psychological Science*, *25*(2), 633–635.

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *15*(4), 485–494.

Jørgensen, M. (2016). The use of precision of software development effort estimates to communicate uncertainty. In D. Winkler, S. Biffl, & J. Bergsmann (Eds.), *International Conference on Software Quality* (pp. 156–168). Cham: Springer International Publishing.

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology. Applied*, *18*(1), 126–140.

Juanchich, M., & Sirota, M. (2017). How much will the sea level rise ? Outcome selection and subjective probability in climate change predictions. *Journal of Experimental Psychology: Applied*, *23*(4), 386–402.

Juanchich, M., Teigen, K. H., & Gourdon, A. (2013). Top scores are possible, bottom scores are certain (and middle scores are not worth mentioning): A pragmatic view of verbal probabilities. *Judgment and Decision Making*, *8*(3), 345–364.

Karelitz, T. M., & Budescu, D. V. (2004). You say "probable" and I say "likely": improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology. Applied*, *10*(1), 25–41.

Kinzig, A., & Starrett, D. (2003). Coping with uncertainty: a call for a new science-policy forum. *AMBIO*, *32*(5), 330–335.

Lipkus, I. M., Klein, W. M. P., & Rimer, B. K. (2001). Communicating breast cancer risks to women using different formats. *Cancer Epidemiology and Prevention Biomarkers*, *10*(8), 895–898.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *21*(1), 37–44.

Løhre, E., & Teigen, K. H. (2014). How fast can you (possibly) do it, or how long will it (certainly) take? Communicating uncertain estimates of performance time. *Acta Psychologica*, *148*, 63–73.

Løhre, E., & Teigen, K. H. (2017). Probabilities associated with precise and vague forecasts. *Journal of Behavioral Decision Making*, doi: 10.1002/bdm.2021.

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.

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.

McClure, J., Henrich, L., Johnston, D. M., & Doyle, E. E. H. (2016). Are two earthquakes better than one? How earthquakes in two different regions affect risk judgments and

preparation in three locations. *International Journal of Disaster Risk Reduction*, *16*, 192–199.

Moore, D. A., Tenney, E. R., & Haran, U. (2016). Overprecision in judgment. In G Wu & G Keren (Eds.), *Handbook of Judgment and Decision Making* (pp. 182–212). New York: Wiley.

Mowen, J. C., Wiener, J. L., & Joag, S. (1987). An information integration analysis of how trust and expertise combine to influence source credibility and persuasion. *Advances in Consumer Research*, *14*(1), 564.

O'Neill, O. (2002). *A Question of Trust*. Cambridge: Cambridge University Press.

Peters, E. M., Hart, P. S., Tusler, M., & Fraenkel, L. (2014). Numbers matter to informed patient choices: a randomized design across age and numeracy levels. *Medical Decision Making*, *34*(4), 430–442.

Politi, M. C., Han, P. K. J., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695.

Schwartz, L. M., Woloshin, S., & Welch, H. G. (1999). Risk communication in clinical practice: putting cancer in context. *Journal of the National Cancer Institute. Monographs*, (25), 124–133.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior And Human Decision Processes*, *53*, 252–266.

Siegrist, M., Gutscher, H., & Earle, T. C. (2005). Perception of risk: the influence of general trust, and general confidence. *Journal of Risk Research*, *8*(2), 145–156.

Sjöberg, L. (2000). Factors in risk perception. *Risk Analysis*, *20*(1), 1–11.

Sladakovic, J., Jansen, J., Hersch, J., Turner, R., & McCaffery, K. J. (2015). The differential effects of presenting uncertainty around benefits and harms on treatment decision making. *Patient Education and Counseling*, *99*(6), 1–7.

Smithson, M., Budescu, D. V, Broomell, S. B., & Por, H. H. (2012). Never say "not": Impact of negative wording in probability phrases on imprecise probability judgments. *International Journal of Approximate Reasoning*, *53*(8), 1262–1270.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*(1), 27–38.

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.

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.

Teigen, K. H., & Brun, W. (2003). Verbal probabilities: A question of frame? *Journal of Behavioral Decision Making*, *16*(1), 53–72.

Teigen, K. H., Juanchich, M., & Filkuková, P. (2014). Verbal probabilities: An alternative approach. *The Quarterly Journal of Experimental Psychology*, *67*(1), 124–146.

Teigen, K. H., Juanchich, M., & Riege, A. H. (2013). Improbable outcomes: Infrequent or extraordinary? *Cognition*, *127*(1), 119–139.

Theil, M. (2002). The role of translations of verbal into numerical probability expressions in risk management: a meta-analysis. *Journal of Risk Research*, *5*(2), 177–186.

Twyman, M., Harvey, N., & Harries, C. (2008). Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. *Judgment and Decision Making*, *3*(1), 111–120.

Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The risk perception paradox-implications for governance and communication of natural hazards. *Risk Analysis*, *33*(6), 1049–1065.

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 781–789.

Wiener, J. L., & Mowen, J. C. (1986). Source credibility: On the independent effects of trust and expertise. *Advances in Consumer Research*, *13*, 306–310.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology*, *2*(4), 343–364.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgement under uncertainty: An accuracy-informative trade-off. *Journal of Experimental Psychology: General*, *124*(4), 424–432.

Yeung, S. (2014). Framing effect in evaluation of others' predictions. *Judgment and Decision Making*, *9*(5), 445–464.
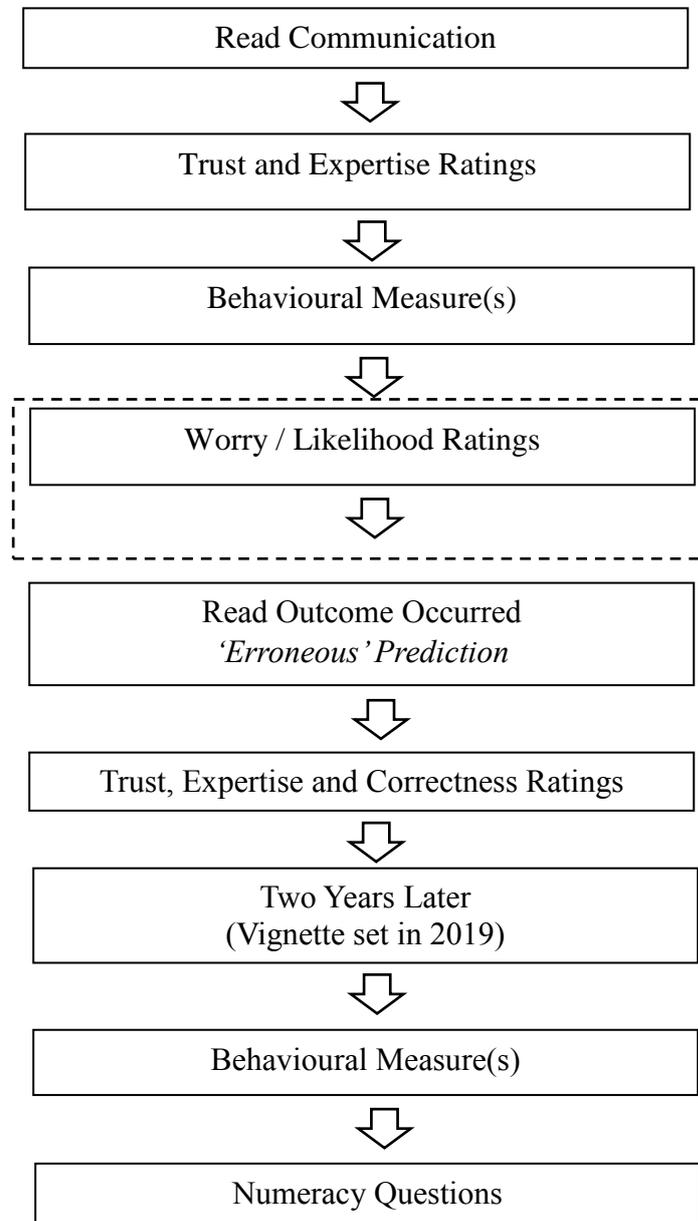
Figure 1. Procedure for experiments. Dashed box represents additional measures included in Experiment 3.
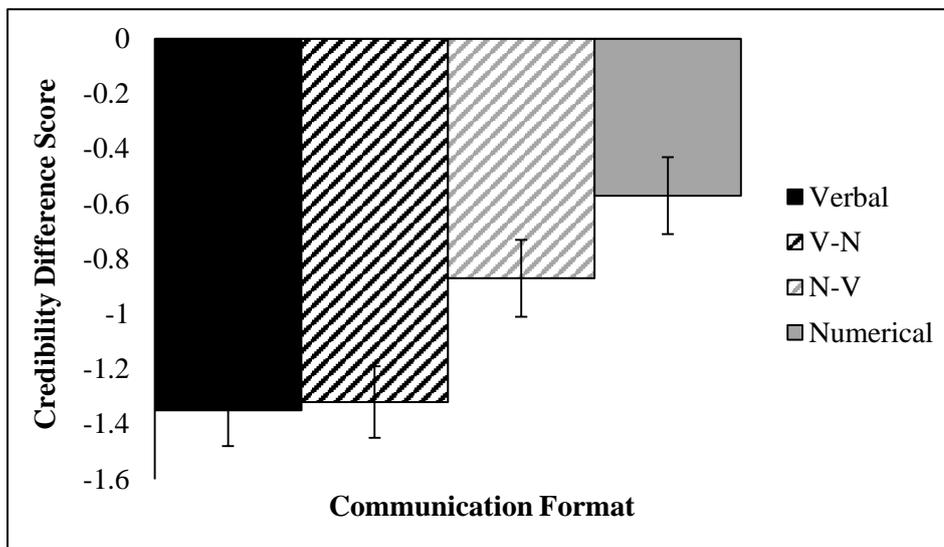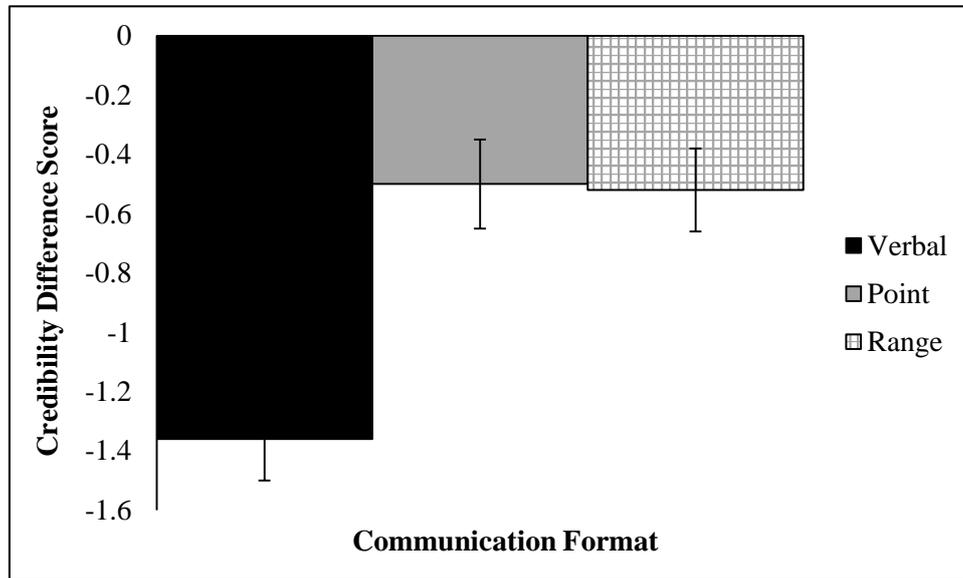
Figure 2. Credibility difference scores by communication format – Experiment 1 (Error bars represent ±1 Standard Error [SE]).
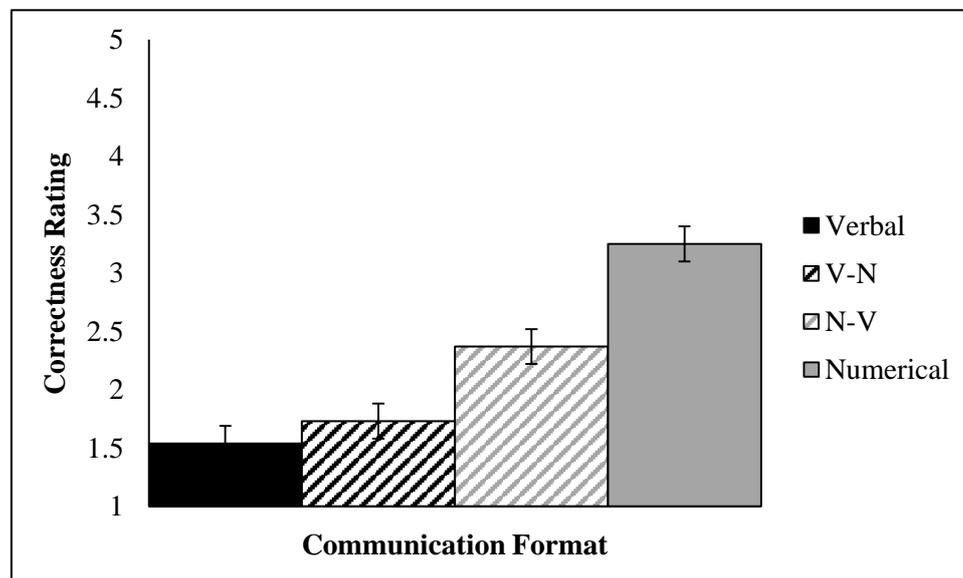
Figure 3. Correctness ratings after 'erroneous' prediction) – Experiment 1 (Error bars represent ±1 SE).

You own a house in Redmill which you have been living in for 10 years. Your house lies within 80km of the forest. The contents of your home are estimated to be worth £5000. We are interested in how much you would be willing to pay to insure these contents against any fire damage.

Based on the geologist's prediction that there is a **20% likelihood** that the forest fire will extend 80km from its origin, please indicate how much you would be willing to pay to insure your home and possessions based on the quotes listed below.

For each amount in the scale below, please indicate whether or not you would pay for the insurance if that was the price.

I would buy the insurance: I would not buy the insurance:

£600	☐	☐
£650	☐	☐
£700	☐	☐
£750	☐	☐
£800	☐	☐
£850	☐	☐
£900	☐	☐
£950	☐	☐
£1000	☐	☐
£1050	☐	☐
£1100	☐	☐
£1150	☐	☐
£1200	☐	☐

Figure 4. Willingness to Pay – Home Insurance Measure – Forest Fire Vignette – Numerical Point Condition.

Figure 5. Credibility difference scores by communication format – Experiment 2 (Error bars represent ⁺₋1 SE).
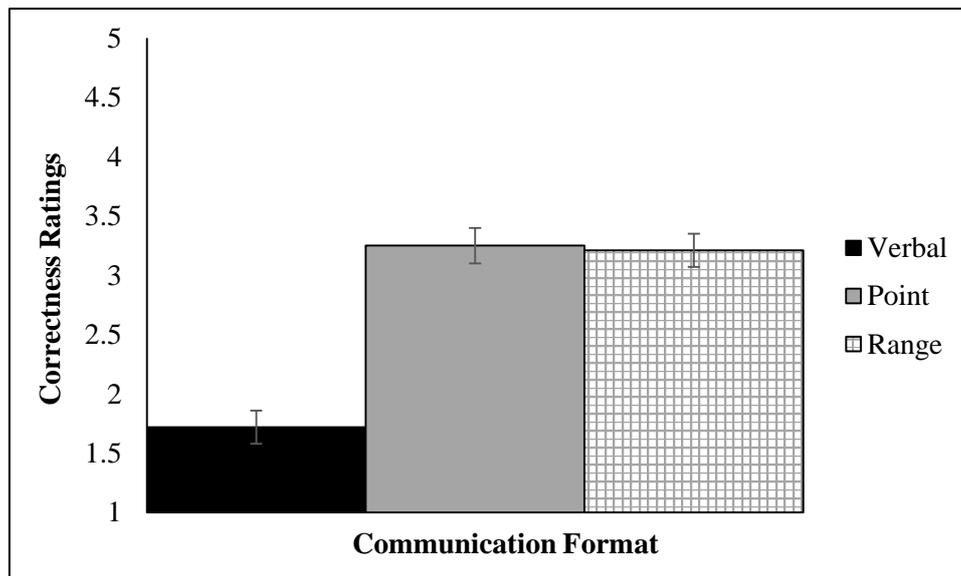
Figure 6. Correctness ratings after 'erroneous' prediction – Experiment 2 (Error bars represent ⁺₋1 SE).
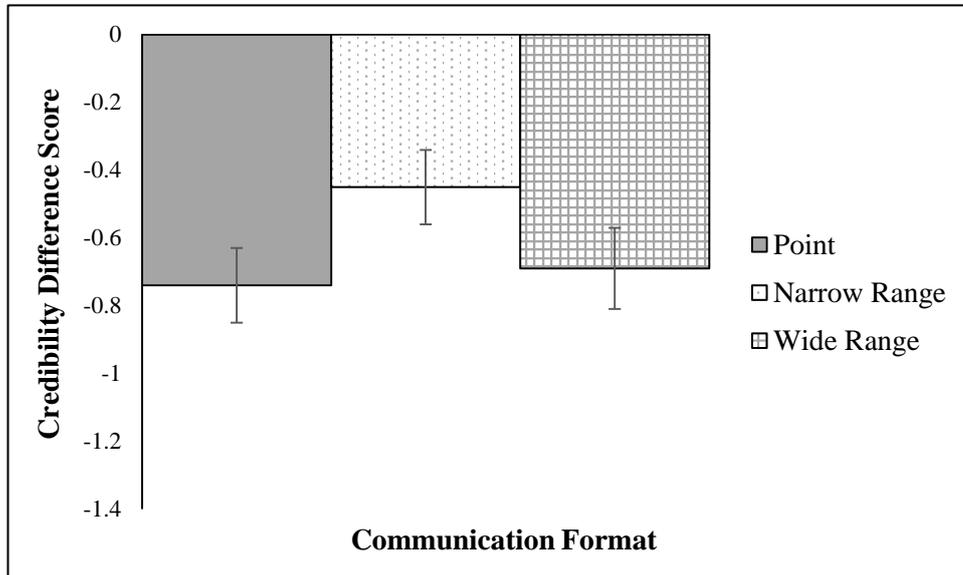
Figure 7. Credibility difference scores by communication format − Experiment 3 (Error bars represent ±1 SE).
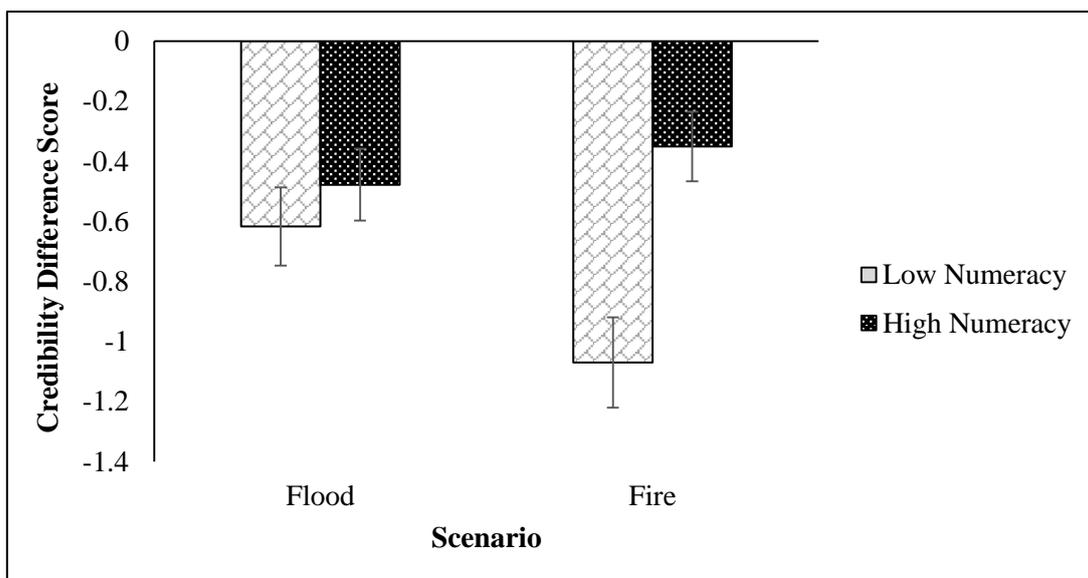
Figure 8. Credibility difference scores by communication format – Experiment 3 (Error bars represent ⁺⁻1 SE) – the scenario × numeracy interaction.
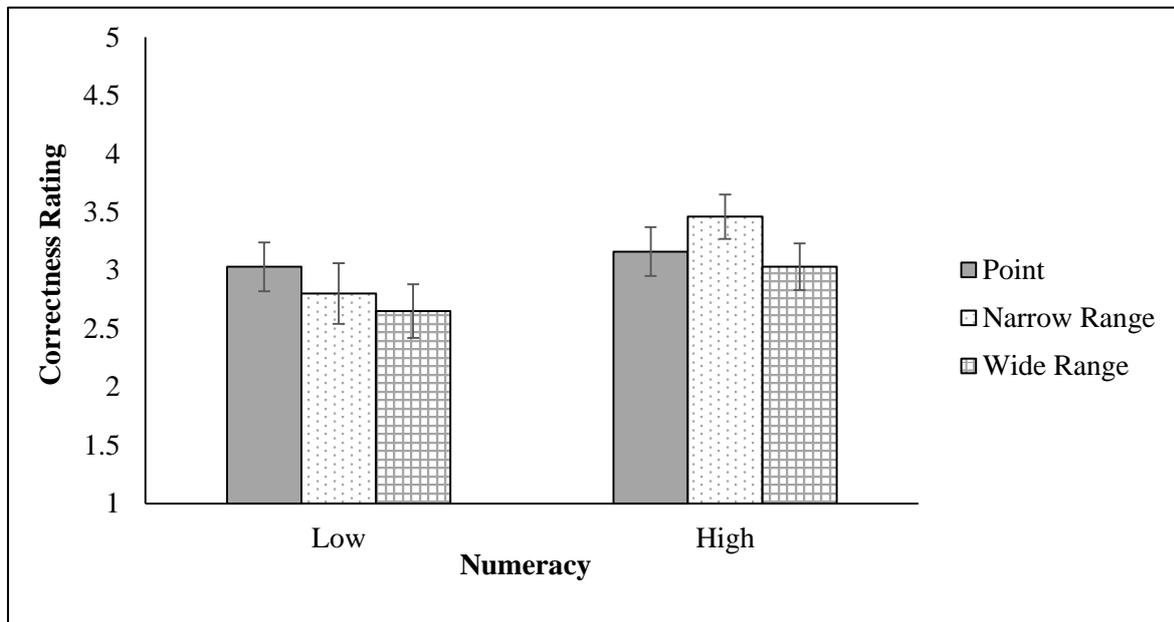
Figure 9. Correctness ratings after 'erroneous' prediction by numeracy (high/low) – Experiment 3 (Error bars represent ⁺⁻1 SE).

Table 1
Distribution (%) of Numeracy Scores by Experiment

	Numeracy Classification											
	Low									**High**		
Numeracy Score	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean (SD)**
Expt 1	0.3	1.4	1.4	0.0	1.7	3.4	5.2	13.7	30.6	25.1	17.2	7.98 (1.81)
Expt 2	0.4	0.8	2.1	1.6	1.2	5.3	6.6	18.1	26.7	20.6	16.5	7.72 (1.93)
Expt 3	0.0	0.0	0.4	0.0	1.6	2.8	4.1	11.8	23.6	33.3	24.4	8.41 (1.42)

Table 2

Behavioural Measures – Mean Scores Across All Experiments by Numeracy

Experiment Behavioural Measure Numeracy Level	Communication Format - Mean Score (SE)					
	Verbal	Numerical Point	V-N	N-V	Numerical Wide Range	Numerical Narrow Range
1. *n*	75	70	74	72	N/A	N/A
1. Evacuation Difference Score*						
Low Numeracy	-0.68 (0.22)	-0.64 (0.23)	-0.48 (0.22)	-0.56 (0.20)	N/A	N/A
High Numeracy	-0.67 (0.24)	-0.91 (0.24)	-0.88 (0.25)	-0.04 (0.29)	N/A	N/A
2. *n*	78	82	N/A	N/A	83	N/A
2. Insurance Difference Score (£)						
Low Numeracy	120.11 (24.86)	159.02 (22.49)	N/A	N/A	108.02 (22.91)	N/A
High Numeracy	188.39 (28.42)	129.17 (32.06)	N/A	N/A	125.56 (27.45)	N/A
3. *n*	N/A	84	N/A	N/A	83	79
3. Insurance Difference Score (£)						
Low Numeracy	N/A	84.82 (20.05)	N/A	N/A	122.62 (23.22)	105.75 (24.83)
High Numeracy	N/A	166.47 (20.61)	N/A	N/A	88.25 (19.25)	94.88 (18.34)

*Lower scores indicate participants were more certain about evacuating instantly