Communicating Uncertainty: The Role of Communication Format in Maximising Understanding and Maintaining Credibility

Sarah C Jenkins

Department of Experimental Psychology

University College London (UCL)

Thesis submitted for the degree of

Doctor of Philosophy

September 2019

Declaration

I, Sarah C Jenkins, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Sarah C Jenkins, 10th September 2019.

Acknowledgements

This thesis would not have been possible without the generous funding from the UCL Impact Award and the British Geological Survey (BGS).

Thank you to all of those at the BGS who contributed to the research contained within this thesis. Many people helped with the discussion of ideas and provided us with feedback on scenario design, for which we are very grateful. A special thank you goes to Murray and Hazel – it has been a pleasure working with you both. I look forward to the possibility of continuing our work with the BGS in the future.

This thesis (and the entire PhD) would not have been possible without the support and guidance of my supervisor, Adam. Working with you has been everything I ever hoped a PhD would be: inspiring, rewarding and fun. I am immensely grateful for your generosity in giving both your time and advice. Your meticulous eye for detail has made me a better scientist, so thank you. I think we've definitely earned that badge now.

To David Shanks, Dave Lagnado and their respective lab groups. Thank you for all of the advice, feedback and discussions both in and out of lab meetings. To all of those who have resided in Room 201 (and along the corridor) past and present. Thank you for the constructive work chats and the equally enjoyable (but less constructive) social and political discussions. Your company has enriched my time at UCL. A particular thank you to Lara for your continual support and the countless times we have put the world to rights.

To my friends, near and far, who have kept me grounded throughout the years. I am so grateful to all of you for offering me a listening ear, kind words and much laughter along the way. A special thank you to VJ who, in the beginning, was literally by my side, and to this day remains there metaphorically. To B, my unofficial cheerleader. Thank you for keeping me well stocked in cheese and celebratory glasses of wine, and for everything else in between.

Lastly, a thank you to my family. Throughout my education, you have always been supportive in the moments when it mattered the most – here's to ending up in a world where no-one needs to sit in a cupboard. Papa, thank you for remaining enthusiastic about my work no matter how many times you ended up reading it – your advice has always been invaluable and is appreciated more than you know. Mum, thank you for continuing to look out for me; I am grateful that I have you on my side (especially when it comes to motivational playlists). Finally, a thank you to my grandparents, whose support enabled me to start my journey at UCL back in 2012. Granny, thank you for your enthusiasm towards all of my endeavours over the years. You have always been a source of much wisdom and reason. It is now clear where I inherited my determination (and stubbornness) from! I am so glad I was able to share my journey with you.

Abstract

This thesis investigates the effect of communication format on the understanding of uncertainty communications and considers the implications of these findings for a communicator's perceived credibility. The research compares five formats: verbal probability expressions (VPEs; e.g., 'unlikely'); numerical expressions – point (e.g., '20% likelihood') and range estimates (e.g., '10–30% likelihood'); and mixed expressions in two orders (verbal-numerical, e.g., 'unlikely [20% likelihood]' and numerical-verbal format, e.g., '20% likelihood [unlikely]').

Using the 'which-outcome' methodology, we observe that when participants are asked to estimate the probability of the outcome of a natural hazard that is described as 'unlikely', the majority indicate outcomes with a value *exceeding* the maximum value shown, equivalent to a 0% probability. Extending this work to numerical and mixed formats, we find that 0% interpretations are also given to communications using a verbal-numerical format (Chapter 2).

If 'unlikely' is interpreted as referring to events which will never occur, there could be implications for a communicator's perceived credibility should an 'unlikely' event actually occur. In the low probability domain, we find a communicator who uses a verbal format in their prediction is perceived as less credible and less correct than one who uses a numerical format. However, in the high probability domain (where a 'likely' event does not occur) such an effect of format is not consistently observed (Chapter 3).

We suggest 'directionality–outcome congruence' can explain these findings. For example, the negatively directional term 'unlikely' led to harsher ratings because the outcome was counter to the original focus of the prediction (i.e., on its non-occurrence). Comparing communications featuring positively and negatively directional VPEs, we find that communicators are perceived as less credible and less correct given directionality–outcome

incongruence (Chapter 4). Our findings demonstrate the influence of pragmatics on (a) the understanding of uncertainty communications and (b) perceived communicator credibility.

Impact Statement

This thesis presents the results of an inter-disciplinary collaboration with the British Geological Survey (BGS), who were keen to improve the effectiveness of their risk and uncertainty communications. As a result of this collaboration, we situated the majority of our empirical work within the context of natural hazards. We worked closely with the BGS to develop scenarios which were reflective of the type of real-world, consequential communications one might make. Our results have been disseminated during meetings with the BGS as well as via their 'Lunchtime Lecture' series.

This thesis has also made a number of important contributions to the wider literature on effective risk and uncertainty communication. Whilst there is a wealth of research which has examined the effect of communication format on understanding, relatively little has considered format's effect on perceived credibility. The current thesis investigated these two factors together and (to our knowledge) was the first to consider the possible interplay between understanding and perceived credibility. In addition to building on previous research which has examined how verbal, numerical and mixed (verbal-numerical) formats are understood, this thesis was the first to consider a numerical-verbal format.

Notably, our results relating to the order effects seen for mixed formats (Chapters 2 and 3) have influenced public policy. Our recommendation to use a numerical-verbal format was adopted by the European Food Safety Authority in their guidelines for the communication of uncertainty (Hart et al., 2019). This recommendation was also incorporated into the UK Statutory Nature Conversation Bodies' recommendations for communicating uncertainty (Papadopoulou, McEntaggart, & Etienne, 2018).

In addition, the research in this thesis has already been disseminated to the wider academic community, via a number of presentations at conferences across a variety of disciplines, including geology and psychology. The research has also been disseminated via a number of peer-reviewed publications (see p.20).

Finally, findings from this thesis formed the basis of a successful application to UCL's 'Global Engagement Fund' to fund a future research project. In collaboration with the BGS and the Icelandic Meteorological Office (IMO), we plan to investigate if the 'extremity effect' observed in Chapter 2 extends to the IMO's volcanic ash maps.

Table of Contents

List of Figures

List of Tables

Publications

The empirical chapters are primarily based on the following articles:

Chapter 2

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.

Jenkins, S., Harris, A. J. L., & Lark, R. M. (2016). 'Unlikely' outcomes might never occur, but what about 'unlikely (20% chance)' outcomes? In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 390-395). Austin, TX: Cognitive Science Society.

Chapter 3

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.

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. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 582-587). Austin, TX: Cognitive Science Society.

Chapter 4

Jenkins, S. C. & Harris, A. J. L. (2019). Maintaining credibility when communicating uncertainty: the role of directionality in ‘erroneous’ predictions. *Invited revision, Thinking and Reasoning.*

Note to Reader

Each of the chapters within the thesis have been written such that they are standalone, and thus can be read in isolation. Some overlap within the introduction sections is therefore an unavoidable result of this endeavour, but will provide the reader with a reminder of where the research is situated and makes a contribution to the literature.

Chapter 1. Introduction

The British Geological Survey (BGS) is the UK's leading provider of objective geoscientific data, information and knowledge. One of their main responsibilities is to provide information and advice about all aspects of geoscience (including natural hazards) to the UK government, industry, academia and the public (British Geological Survey, 2019). In doing so, however, the BGS face the challenge of translating geoscience data and expert knowledge into communications which are comprehensible to a non-expert audience. This thesis aimed to investigate the BGS' question of how to improve the effectiveness of their communications concerning natural hazards. Specifically, this thesis considers effective risk and uncertainty communication as requiring two components. Firstly, for the communication to be taken on board by the recipient, the communicator must be perceived as a credible source of information. Secondly, for the information to be used effectively, the recipient must interpret the information as the communicator intended.

Problematically, in a natural hazards context, the information communicated is typically very uncertain. Complicating matters further are the high stakes attached to such communications, as illustrated by the events of the 2009 L'Aquila earthquake, which killed 308 people, injured a further 1,500 and left another 67,000 homeless. Following the earthquake, six scientists and one government official were put on trial, and found guilty of manslaughter for misleading the public with 'incomplete, imprecise and contradictory information' (Alexander, 2014). Whilst their sentences were later overturned, the case highlighted the great responsibility scientists are charged with when translating findings from Earth Science into information that the public can understand, and use, effectively. This is just one example of the wider problem faced by communicators across many domains, who are left with the question of how they can best communicate uncertainty, and maintain credibility in the face of uncertainty.

Multiple communication formats exist to express uncertainty, though the most common are verbal probability expressions (e.g., 'it is *unlikely*'), numerical probability expressions (e.g., '*20% likelihood*)[1] and mixed expressions, which are a combination of the two (e.g., 'it is *unlikely [20% likelihood]*'). This thesis investigates how these communication formats are understood, before considering their downstream effects on the manifestation and maintenance of credibility. The current chapter begins by providing a brief background to the concepts of risk and uncertainty and moves on to consider how they present challenges for communicators, specifically in the context of natural hazards. Here, effective risk and uncertainty communication is defined as a dual-component process, requiring both understanding and (perceived) communicator credibility. This chapter subsequently reviews literature which has investigated how verbal, numerical and mixed formats are understood. The final part of this chapter considers credibility and its effects, with a particular focus on credibility's two main components: trustworthiness and expertise.

Risk and Uncertainty

Risk.

Many scholars define risk as uncertainty and/or uncertainty as risk (Samson, Reneke, & Wiecek, 2009). However, within this thesis, the concepts are seen as different, but typically related, as detailed below. The word 'risk' entered the English language in the mid-1600s (Breakwell, 2014). It has been described as a 'multidimensional phenomenon', and thus it is unsurprising that it lacks a commonly agreed upon definition, both within the sciences, and within public understanding (Renn, 1992a, 1998). As a result of the multiple conceptualisations of the term, there is clear potential for problems in risk communication, for instance the provision of inadequate, or too much information (Slovic & Weber, 2002). In-depth reviews of existing definitions and their strengths and weaknesses have been

[1] In line with standard, dictionary definitions, 'likelihood' is used as a synonym for 'probability' throughout the thesis, though it is noted that, mathematically, each has a unique and specific definition.

carried out elsewhere (see Aven, 2012; Aven & Renn, 2009; Aven, Renn, & Rosa, 2011) and thus only an outline is presented here.

Knight (1921) was one of the first to draw a clear distinction between risk (as 'measurable uncertainty') and uncertainty as 'unmeasurable', a distinction which has made a significant contribution to both psychology and economics. This is despite the fact that popular and practical usage at the time typically associated risk with loss (Hubbard, 2015). Moreover, Knight's (1921) definition assumes that an objective probability distribution can be obtained. However, in most real-life situations (and especially in the context of natural hazards), such objective distributions are not available, leading to calls to move away from this definition (Aven, 2010).

Despite the variation in definitions, nearly all conceptualisations of risk make a distinction between reality and possibility, with risk relating to possibility, such as a state of affairs or outcomes (Renn, 1998). Indeed, uncertainty over the future occurrence or non-occurrence of an outcome is at the heart of the notion of risk (Politi, Han, & Col, 2007). Whilst traditionally the concept of risk was neutral, in that it took into account both positive and negative outcomes (i.e., the probability of gains or losses), more contemporary definitions of risk outside of economics have focused on the element of 'danger', and thus are less likely to include positive outcomes (Douglas, 1994). Given the nature of natural hazards, we thus move away from economic definitions of risk towards more contemporary definitions of risk, which typically focus on the possibility of undesirable or adverse effects, whether these are caused by natural or human activities (Breakwell, 2014; Renn, 1998). Indeed, definitions of risk adopted by multiple national and international organisations (e.g., the World Health Organisation [WHO], United Nations [UN], National Research Council [NRC]), are all similar in defining risk in relation to the probability of *harm* (Scheer et al., 2014). Such a focus on the negative outcomes of risk is reflected in typical dictionary definitions of risk, which relate risk to danger, chance of loss or harm (Hubbard, 2015), and are also reflected in common usage (i.e., risk as a hazard, risk as potential adversity or threat, Slovic & Weber, 2002). Whilst the majority of the literature has focused on communications featuring risks with negative outcomes, some communications within this thesis also feature neutral outcomes. As a result, our use of risk is more in line with Aven and Renn's (2009) definition, where "risk refers to uncertainty about and severity of the events and consequences (or outcomes) of an activity with respect to something that humans value" (p.6). It should be noted however, that the thesis focuses on one aspect of risk – the uncertainty of the event (specifically the language used to describe it), rather than systematically investigating the event's severity or consequences.

Objective versus subjective risk.

Typically, formal measures of risk, such as technical risk assessments assume a level of 'objective risk', focusing on specific concepts which can be measured and objectively quantified (Slovic, 1987, 1999; Verbeke, Frewer, Scholderer, & De Brabander, 2007). However, the public's intuitive risk perceptions are inherently subjective (Slovic, 1999), and thus do not always reflect the 'objective risk' resulting from technical risk assessments or mortality figures (Sjöberg, 2000). The subjective (and context sensitive) nature of risk perceptions is illustrated in research employing the psychometric paradigm. In this paradigm, participants are asked not only how they perceive a risk, but also to characterise the risk's 'personality', by rating each risk on a series of qualities such as voluntariness of exposure, familiarity, controllability and level of knowledge (Slovic, 1992). Whilst the precise nature of these factors is outside the scope of this review, psychometric research has consistently shown that these factors affect both the perception and acceptance of risk (e.g., Fischhoff, Slovic, Lichtenstein, Read, & Combs, 1978; Slovic, 1987, 1993, 2010; Starr, 1969).

The aforementioned research has led to the suggestion that risk is not something that exists 'out there', which can be measured, but is actually a concept invented to help humans cope with dangers and uncertainties of life (Slovic, 1999). Although these dangers are real, ‘objective risk’ does not exist – even probabilistic estimates are based on models which are subjective. Logically then, risk perceptions do not exist objectively either, but are constructed, and are influenced by factors other than the magnitude of risk, including people’s values. The value dimension of risk is a key part of risk perceptions, which relate to: “People's beliefs, attitudes, judgements and feelings, as well as the wider cultural and social dispositions they adopt towards threats to things that we value” (Pidgeon, 1998, p.5).

However, the value dimension has often been neglected in research, with much of previous research measuring subjective probability judgements and interpreting these as reflecting risk perceptions. In fact, the two are different concepts and thus should be distinguished (Betsch, Haase, Renkewitz, & Schmid, 2015). Subjective probability judgements refer *only* to the probability of an event, usually measured using a percentage (0–100%), or on a verbal rating scale (e.g., ‘Not at all likely’ to ‘Completely likely’). In contrast, perceived risk includes both the probability of the event and an additional value dimension, referring to how significant or severe an event is perceived to be (Betsch et al., 2015), a dimension reflected in Aven and Renn's (2009) definition of risk.

As demonstrated by research using the psychometric paradigm, risk perceptions are far more multi-dimensional, and context sensitive than other, more formal measures of risk (Bostrom, 1997). The fact that risk perceptions are actively constructed (Slovic, 1999) indicates that they could be vulnerable to the way risk information is presented, specifically the communication format. Moreover, risk perceptions are said to affect the likelihood of one responding to warnings, and engaging in preparedness behaviours to mitigate the risk (Wachinger, Renn, Begg, & Kuhlicke, 2013). In light of this, it is thus vital to take contextual factors into account in order to develop effective risk communications, or else risk the information being discounted (Frewer, 2000; Frewer, Howard, Hedderley, & Shepherd, 1997).

Risk communication.

The practice of risk communication dates back to 3200 BC (Plough & Krimsky, 1987), but the study of risk communication is far newer. The term was first used in 1984 in response to risk perception research by Slovic and colleagues, which highlighted the discord between experts and the public's views on risk (Leiss, 1996). Subsequent research interest in risk communication developed from a desire to improve the dialogue between these two groups, with the aim of achieving consensus on, and acceptance of, how risks should be managed. Risk communication can be used to refer to a wide range of activities; in its broadest form, it refers to "any public or private communication that informs individuals about the existence, nature, form, severity or acceptability of risks" (Plough & Krimsky, 1987, p. 6).

Risk communication was initially conceived of as a one-way process, with information flowing from experts to non-experts. This was based on the belief that technical, scientific risk assessments should be the basis for the assessment and management of risks (Leiss, 1996). Research using the psychometric paradigm showed that the public's risk perceptions are not just based on 'objective risk', but are sensitive to characteristics of the risk, such as its familiarity, severity and voluntariness (e.g., Slovic, 1987). As a result, risk communications originally followed a 'knowledge deficit model' (Simis, Madden, Cacciatore, & Yeo, 2016), based on the belief that the public were deficient in their knowledge about risks and 'irrational'. The assumption was that if the public were given the correct technical risk information, they would perceive the information similarly and reach the same conclusions as experts. Yet such a view fails to take into account the wide range of sociocultural and emotional factors which feed into the public's responses to risk (Joffe, 2003). It was thus unsurprising that risk communications which aimed to 'rectify the knowledge gap' were unsuccessful, with the public remaining sceptical of experts (Frewer, 2004).

The second phase of risk communication stemmed from the recognition that risk communications are an act of persuasion (Leiss, 1996) and thus serve two purposes: providing information and influencing (National Research Council, 1989). Based on findings from marketing research, an effective communication strategy should consider the characteristics of the audience, and ensure that the source is trusted and credible. According to Leiss (1996), the third stage of risk communication evolved from the observation that a lack of trust is often common when dealing with risk, and the fostering of trust does not happen instantly. Whilst the importance of trust and credibility as components of effective risk communication has long since been recognised (Covello & Allen, 1988; Kasperson, Golding, & Tuler, 1992; R. G. Peters, Covello, & McCallum, 1997), these components have been the subject of relatively little research attention, with the majority of research focusing on the 'understanding' component. The importance of effective communications cannot be understated; successful risk management relies on successful risk communication (Renn, 2005), meaning there are ramifications for the acceptance and adherence to risk management efforts.

Over time, risk communication has moved away from assuming the public is a passive audience (Rowe & Frewer, 2000) whose concerns must be diffused, or convinced that something is tolerable or not. Rather than being a top down dialogue, risk communication is now seen as a constructive, dynamic, flexible and interactive process (Löfstedt, 2003; Ng & Hamby, 1997). This process involves not just the exchange of information about the nature of risk, but also the expression of concerns, opinions and reactions to risk management efforts (National Research Council, 1989).[2] The aim of risk communication should be to foster mutual understanding (Gutteling & Kuttschreuter, 2002), resulting in an informed public who are able to make their own risk judgements and improved decisions (Árvai, 2014; Covello & Allen, 1988; Fischhoff, 1995; Gutteling & Kuttschreuter, 2002; Renn, 2005).

[2] See Lundgren and McMakin (2018) for specific approaches to risk communication.

This thesis posits that for the risk communication process to be effective, two key components are required. Firstly, the effectiveness of the process relies on the recipient understanding the information as the communicator intended, and secondly, it relies on the recipient perceiving the communicator as a credible source of information, so that the information can be used in subsequent decision-making. This thesis brings the two components together, judging the effectiveness of communications against the criteria of credibility and understanding.

Despite the progress in defining the process of risk communication, and consensus over its aims, communicators still face multiple challenges. Risk communication has been deemed 'an inherently hard task' because of the unintuitive and intangible nature of probability (Zikmund-Fisher, 2012). People split probabilities essentially into two (e.g., I will or won't get the disease) (Lippman-Hand, Fraser, & Opitz, 1979) and subsequently experience only a single outcome – it either happens or does not (Zikmund-Fisher, 2012). In addition, risk communications are typically made to a broad audience, who may differ in terms of their experience with the risk (Wachinger et al., 2013), existing level of knowledge, and their literacy/numeracy levels (Lundgren & McMakin, 2018). In light of these differences, it has been argued that it is not possible to tailor risk communications to suit everyone, and thus a 'one size fits all' solution should not be accepted, as such a solution could have a negative impact on audience trust (Löfstedt, 2003).

So far, we have viewed the communication of risk as encompassing both information about the uncertainty and severity of an event, as well as its potential consequences (in accordance with Aven and Renn, 2009). However, the communications featured in this thesis primarily convey information about the uncertainty of the natural hazard, rather than information about the hazard's severity or consequences. In the following, we thus shift our attention to the concept of uncertainty, highlighting the potential challenges uncertainty poses for communicators before considering the most common communication formats used to convey uncertainty.

Uncertainty.

Probability is typically treated as synonymous with uncertainty (Smithson, 2012), when in fact uncertainty is often used to refer to a multiplicity of concepts, leading to its labelling as 'capacious' (Morgan & Henrion, 1990). Given the broad nature of the term, uncertainty is typically subdivided into categories, which helps with the identification, assessment and communication of uncertainty (Doyle, Johnston, Smith, & Paton, 2019; van Asselt & Rotmans, 2002). There are, however, many differences in how uncertainty has been categorised in the literature (Brugnach, Dewulf, Pahl-Wostl, & Taillieu, 2008; Han, Klein, & Arora, 2011; Kahneman & Tversky, 1982; Knight, 1921; Patt & Dessai, 2005; van Asselt & Rotmans, 2002; W. Walker et al., 2003; Wynne, 1992). One of the most common categorisations of uncertainty is based on where the uncertainty derives from. A brief overview of the numerous sources of uncertainty described in the literature is presented below.

Uncertainty can result from incomplete or missing knowledge, skill or information ('epistemic uncertainty'), or can be random, resulting from chance, natural variability or stochastic processes ('aleatory uncertainty') (Fox & Ulkümen, 2017; Hacking, 2006). This distinction can be related to Kahneman and Tversky's (1982) classification, in which uncertainty can be attributed to the state of one's knowledge ('internal uncertainty') or the external world ('external uncertainty'). It is claimed that decision-makers can intuitively differentiate between the two types of uncertainty (Fox & Ülkümen, 2011), with systematic behavioural and psychological consequences resulting depending on which type of uncertainty is salient to the decision-maker (e.g., Fox & Ülkümen, 2011; Heath & Tversky, 1991; Tannenbaum, Fox, & Ülkümen, 2017).[3] Similar to the distinction between epistemic and aleatory uncertainty, Van Asselt and Rotmans (2002) distinguish between uncertainty which is due to limited knowledge, and uncertainty which is due to variability. They extend these categories by considering additional causes of limited knowledge, such as inexactness and conflicting evidence, as well as additional causes of variability, such as inherent randomness and behavioural variability. W. Walker et al. (2003) extended this classification even further, by distinguishing between location, level and nature of uncertainty.

Despite the many sources of uncertainty that exist, literature on uncertainty communication (including that which has investigated format) has largely focused on the communication of aleatory uncertainty, referring to future events which we cannot know about for certain (e.g., economic forecasts and climate change models, van der Bles et al., 2019). Van der Bles et al. (2019) thus set out a new framework (see Figure 1), which aims to provide guidance on how to communicate epistemic uncertainties honestly, without losing trust and credibility. This framework highlights individual factors which could affect the communication of uncertainty. Whilst natural hazards do not only involve epistemic uncertainty (Doyle et al., 2019), the framework provides a useful basis for situating the contribution of the current thesis within the literature. Specifically, this thesis focuses on factors relating to the form of the communication (specifically format), and its effects on the recipient, including a consideration of cognitions, emotions, trust and behaviour.

[3] See also: Juanchich, Gourdon-Kanhukamwe and Sirota (2017); Løhre and Teigen (2015); Teigen and Løhre (2017); Ülkümen, Fox and Malle (2016) for research on how these sources of uncertainty are reflected in language use.

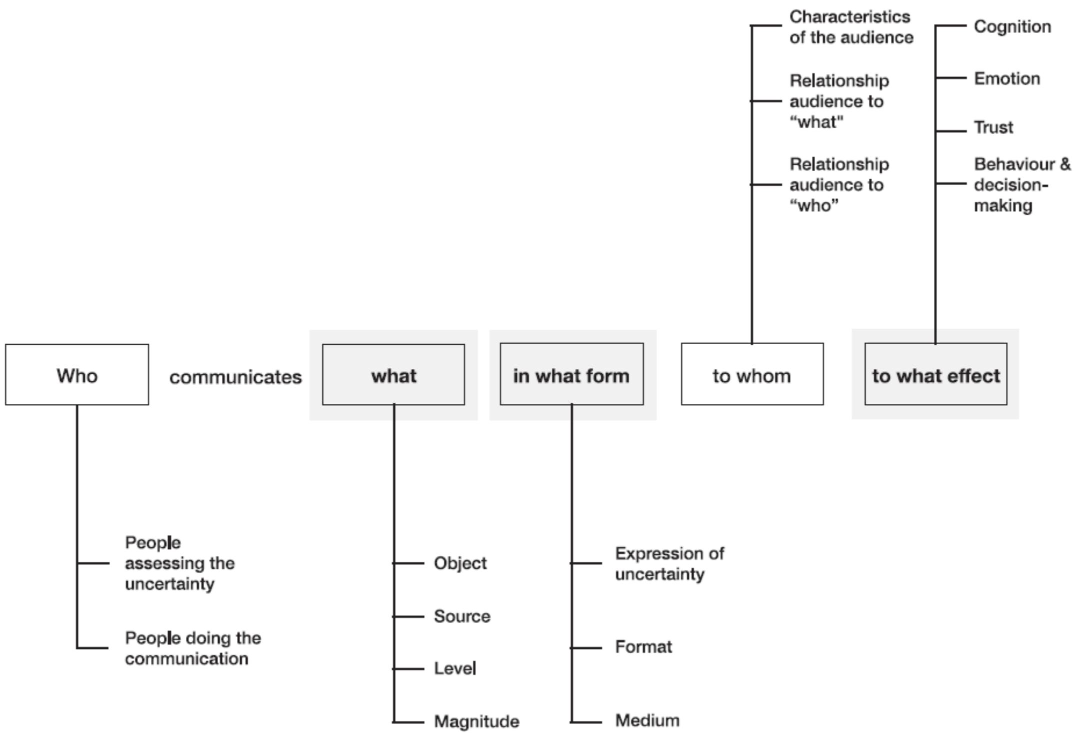


Figure 1. A framework for communicating epistemic uncertainty (van der Bles et al., 2019).

The problem of communicating uncertainty.

Uncertainty presents a whole host of challenges to experts, not just as a concept, but also in terms of how and whether one should communicate it, as well as the public's reactions to it. The public expect science to be able to reduce or eliminate uncertainty, despite the fact that it is impossible to eradicate entirely (Kinzig & Starrett, 2003). Often uncertainty is seen as something which can be eliminated or reduced by additional research, monitoring, or time, yet, even if one gains more knowledge, there is no guarantee that uncertainty will be reduced (van Asselt & Vos, 2006). Indeed, even if more scientific work reduces existing uncertainties, it also often uncovers more uncertainty, which makes the management of uncertainty difficult (National Research Council, 2005).

The public have often been assumed to be ignorant, and unable to handle scientific information, given the frequency of misunderstandings associated with such information (Davies, 2008). Indeed, experts have traditionally held the view that the public are unable to conceptualise uncertainty and believe that their own understanding differs to the public's (Frewer, Hunt, et al., 2003). As a result, many experts have expressed concerns about communicating uncertainty, for fear it will be misperceived. Indeed, research has found that the provision of uncertainty may be interpreted as a sign of the communicator's lack of knowledge, incompetence, evasiveness, equivocation or inability to make their minds up (Fessenden-Raden, Fitchen, & Heath, 1987; Fischhoff, 1995; Frewer, Hunt, et al., 2003; Johnson & Slovic, 1995; Lewandowsky, Ballard, & Pancost, 2015). Fischhoff (1995) also highlights that such interpretations could lead to the presumption that experts are confused, or that one expert's estimate is the same as another's. Such perceptions could prove deleterious to the communication, and result in increased risk perceptions, worry and concern (Frewer, 2004; Frewer, Hunt, et al., 2003; Johnson & Slovic, 1995, 1998; MacGregor, Slovic, & Morgan, 1994; Slovic, 1987; Viscusi, Magat, & Huber, 1991), increased confusion (Frewer, 2004), lower decision satisfaction (Politi, Clark, Ombao, Dizon, & Elwyn, 2011) and reduced confidence in the communicator (Ogden et al., 2002). In light of these findings, it is unsurprising that experts fear communicating uncertainty will lead to a loss of trust and credibility (Frewer, Hunt, et al., 2003), a view echoed by those working within a natural hazards context (Newhall & Punongbayan, 1996).

Moreover, much research has identified problems with the way the public respond to uncertainty. Whilst for some, uncertainty may lead one to adopt a cautious attitude, for fear of the worst, others may use uncertainty as an excuse to downplay, or even ignore the seriousness of a threat (Kuhn, 2000). As a result, uncertainty is often cited as a reason for inaction (Fischhoff, 1995; Johnson & Slovic, 1998; Kinzig & Starrett, 2003) or to justify a delay in taking mitigative action (Doyle, McClure, Paton, & Johnston, 2014), reasons cited by both politicians and the public (Lewandowsky et al., 2015). The choice to defer action in the face of uncertainty is particularly problematic in the context of natural hazards. For instance, if the time/decision window is small, uncertainty may distract people from the decision at hand (Fischhoff, 1995), when in fact prompt responses are what is required in such situations.

In light of the above, it could be suggested that a better strategy for communicators would be to underestimate, minimise or even ignore uncertainty altogether. However, doing so could seriously affect perception of risk and subsequent decision-making (National Research Council, 1989), as demonstrated in the NASA Columbia and Challenger disasters, where valuable non-quantifiable information was overlooked (Feldman, 2004). In addition, not acknowledging uncertainty may actually lead (paradoxically) to an increase in perceived uncertainty and ambiguity (Han et al., 2009; Portnoy, Han, Ferrer, Klein, & Clauser, 2013). People have intuitions about the expected level of uncertainty and they may become distrustful of information which ignores uncertainty (Han et al., 2009; Joslyn & Grounds, 2015; Joslyn & Savelli, 2010). Failures to communicate uncertainty have been associated with loss of trust, as was the case during the BSE (bovine spongiform encephalopathy) crisis (Miles & Frewer, 2003). Clearly then, ignoring uncertainty is not a solution for communicators, though exactly what is remains an open question.

In fact, research suggests that the effects of communicating uncertainty may not be as clear cut as has been made out in the literature presented so far. For instance, Bord and O'Connor (1992) found that concern levels about living near a hazardous waste site were unaffected when uncertainty about the risk was presented. In addition, the expression of uncertainty in earthquake risk estimates did not decrease judgements of an expert's perceived competence (Nakayachi, Johnson, & Koketsu, 2018). Some have even argued that the disclosure of uncertainty may be viewed as a positive attribute (Renn & Levine, 1991), as it reflects a level of honesty (Johnson & Slovic, 1995), and as a result could lead to increased perceptions of trust and credibility (Chess, Hance, & Sandman, 1988; Frewer et al., 2002). Indeed, benefits such as increased perceptions of trustworthiness (Jensen, 2008), honesty and openness (Nakayachi et al., 2018), as well as increased acceptance rates (and decreased rejection rates) of GM technology (Frewer, Howard, & Shepherd, 1998) have been observed where uncertainty has been communicated transparently.

The fact that the communication of uncertainty has not always been linked with negative effects suggests the issue is far more nuanced than has previously been thought. The effects of communicating uncertainty may depend on a range of factors, including both those related and unrelated to uncertainty. Whether disclosing uncertainty has negative effects or not may depend on the source of uncertainty. For instance, lower trust and acceptability ratings have been observed when the uncertainty arose as a result of conflict between experts or from the withholding of information, versus when it was due to a lack of data or the need to do more scientific work (Frewer et al., 2002; Markon & Lemyre, 2013). Other factors, including the type of hazard and its characteristics may also have a bearing on the effects of uncertainty. Increased perceptions of openness and honesty occurred more so when communications featured lower probabilities (20% versus 70%) and shorter time periods (10 years versus 30 years, Nakayachi et al., 2018). The disclosure of uncertainty on risk perceptions may also depend on the recipient's prior attitudes (Kuhn, 2000). Environmental risk perceptions increased when uncertainty information was provided (versus when it was not), but decreased for those who had high levels of initial environmental concern (Kuhn, 2000). Kuhn (2000) speculated that for the latter group, the acknowledgement of uncertainty may have led to increased credibility perceptions, which subsequently lowered risk perceptions. The disparity between findings highlighted previously may therefore have resulted from specific characteristics of the uncertainty, the hazard or the communication. In light of these findings, it seems reasonable to suggest that the format of the communication could influence perceptions of the communicator.

Despite the mixed evidence, it is clear that the communication of uncertainty is a balance between being transparent, but also useful – the public do not need to know all of the uncertainty that science creates (Fischhoff, 2013), as this could prove counterproductive for decision-makers and the public (Kasperson, 2014). Given that the public often look towards experts in complex situations, for instance those characterised by uncertainty, ensuring experts are able to build and maintain credibility is vital.

Natural Hazards

According to the 2019 Global Risks Report (World Economic Forum, 2019), natural hazards are in the top five risks that the world faces, both in terms of likelihood and impact. However, these two factors are often asymmetric, with many natural hazards typically low probability events, which are highly consequential (e.g., Knoblauch, Stauffacher, & Trutnevyte, 2018). This presents a challenge for communicators, as it is well documented that people struggle to accurately understand low probabilities. As a result, people may discount such events and fail to engage in preparatory behaviours that could mitigate the effects of the hazard (McClure, Henrich, Johnston, & Doyle, 2016).

Communicating information about natural hazards is a complex task, largely because they are associated with 'multiple layers of uncertainty', relating to their exact timing, location, intensity and frequency (Sword-Daniels et al., 2018). Some of this uncertainty is aleatory uncertainty, relating to the variability of physical processes, whereas some is epistemic uncertainty, resulting from a lack of knowledge (Doyle et al., 2019; Patt & Dessai, 2005). In addition, scientists may be unsure about their knowledge and data, or disagree with others, as a result of factors including 'incomplete information', 'inadequate understanding' and 'undifferentiated alternatives' (Lipshitz & Strauss, 1997) and there may be conflicting advice given from other organisations (e.g., Newhall & Punongbayan, 1996; Peterson & Tilling, 1993). All of this uncertainty means that it is impossible to predict precisely what will happen in the future, so any estimates provided are probabilistic at best. This will in turn have implications for the decisions emergency officials make (which may not always be in line with scientific advice, Doyle et al., 2014).

Further complicating matters are the differences between natural hazards, across characteristics such as predictability, scope, onset delay and lethality (Sapir & Lechat, 1986). For instance, earthquakes are less predictable than floods. Hazards such as earthquakes are not preventable, and can be predicted with great difficulty, which often leads people to adopt a fatalistic attitude in believing that such events are out of their control. Such attitudes can affect both risk perceptions and willingness to engage in preparedness behaviours (McClure, Walkey, & Allen, 1999). This is problematic for risk communications which aim to convince people that there are mitigative actions which can be taken, such as securing furniture and making evacuation plans (Eiser et al., 2012).

In addition, even within one hazard type, there may be aspects which are not predictable. For instance, whilst scientific understanding of some volcanoes is relatively well developed, there are still many intrinsic uncertainties which relate to many issues of public concern, such as the duration of an eruption, and its day to day impacts (Campbell, 2011). Volcanic eruptions involve a number of different hazards, which may not all have the same levels of risk, for instance high risk of ash fall, but a low risk of pyroclastic flows (Haynes, Barclay, & Pidgeon, 2007). Clearly then, the effectiveness of communications may differ according to the natural hazard, and more specifically, the particular risk being communicated.

Complicating matters further are the far-reaching consequences of communication which 'goes wrong'. This is perhaps best illustrated by the events of the 2009 L'Aquila earthquake, in which a 6.3 magnitude earthquake hit the Italian city of L'Aquila, killing 308 people, injuring 1,500 and leaving 67,000 homeless. Following this, six scientists and one government official were put on trial and found guilty of manslaughter for misleading the public with 'incomplete, imprecise and contradictory information' (Alexander, 2014). They were sentenced to six years in prison, and whilst their sentences were overturned in 2014, their treatment sparked outrage amongst the scientific community, who believed that ‘science was on trial’. Although this belief turned out to be somewhat misguided (Alexander, 2014), the case highlights the challenge for scientists, who are responsible for translating findings from Earth Science (which are characterised by uncertainty) into information that can be used by the public.

Clearly, the criminal proceedings following the L’Aquila earthquake are an extreme example of what could happen to communicators following a failure in communication (Benessia & De Marchi, 2017). However, there can still be serious repercussions for communicators in terms of credibility, if the public do not recognise the high level of uncertainty associated with natural hazards, or recognise it but expect scientists to reduce the uncertainty (Campbell, 2011). This has been observed in studies of the volcanic crisis in Montserrat (e.g., Haynes, Barclay, & Pidgeon, 2008a, 2008b). Haynes, Barclay and Pidgeon (2008b) found that people often inflate the predictive abilities of volcanic observatories, both in terms of their accuracy and timing, which can lead to dismay if there is no volcanic activity after an evacuation. Even if uncertainties are conveyed to the scientists’ best ability, an alert which subsequently did not manifest may be judged as false, and the communication interpreted as a failure (Campbell, 2011). Such instances are colloquially known as ‘crying wolf’, or more formally, ‘false alarms’ (e.g., Breznitz, 1984), and have a series of psychological consequences. Some studies have observed concern over ‘crying wolf’ amongst decision-makers, who fear a loss in trust and thus a reduction in the public’s subsequent compliance to future warnings (e.g., L. R. Barnes, Gruntfest, Hayden, Schultz, & Benight, 2007; Sharma & Patt, 2012; Woo, 2008). Clearly then, effective communication of uncertain hazards is a delicate balance between precaution and warnings, and most of all requires trust in the team responsible for managing and communicating the risk (Haynes et al., 2008a).

Given that trust in authorities and experts is one of the most significant contributors to risk perceptions of natural hazards (Wachinger et al., 2013), it is unsurprising that recommendations place a great deal of importance on scientific advice being transparent

(Guzzetti, 2016), and coming from a credible source (Doyle, McClure, et al., 2014). Indeed, recommendations from the International Association of Volcanology and Chemistry of the Earth's Interior (IAVCEI) emphasise that scientific knowledge about hazards, risks and their related uncertainties should be disseminated to the public (IAVCEI Task Group on Crisis Protocols, 2016). However, precisely how to communicate these uncertainties remains an open research question (see also, Doyle et al., 2019).

Communicating Uncertainty – The Role of Communication Format

Over the last three decades, there has been a burgeoning of research focusing on the issue of communicating uncertainty, both inside and outside of natural hazard contexts (e.g., Doyle et al., 2019). Probability is the best known, and most commonly used way to quantify uncertainty (Morgan & Henrion, 1990), with uncertainty often represented by providing a list of possible outcomes, along with an expression of the probability of the outcome. A vast amount of research has focused on the multiple formats that probabilistic information can be expressed in, for instance: using verbal probability expressions (VPEs; e.g., 'unlikely', 'doubtful', 'good chance', 'probable'); numerical expressions, such as percentages (e.g., '20% likelihood'), frequencies (e.g., '1 in 100'); mixed expressions (e.g., 'unlikely [20% likelihood]') and visual formats, such as graphs (line, histogram, pie charts, icon arrays), risk ladders or risk scales.

Much research has focused on how format influences understanding, but it has typically viewed the communication of uncertainty as an 'information processing' task, which assumes that the communicator's probabilistic beliefs are well-calibrated and clearly presented, and that the recipient adopts these beliefs as such (Fox & Irwin, 1998). Such a view fails to take into account the wider context of the communication, such as the motives and constraints of the communicator, the goals of the listener, and the source of the information. This thesis therefore considers not just the information processing aspect of uncertainty communication, but also part of the wider social context, in terms of the recipient's evaluation of the communicator.

Whilst a large body of research has examined how communication format affects understanding, relatively little has examined its effect on perceived credibility, and only a tiny proportion has investigated both factors together. Yet such an investigation is vital given the two way nature of communications – the communication will influence perceptions of the source, and perceptions of the source will influence how the communication is interpreted (Eiser, Miles, & Frewer, 2002). Even if a communication is understood as the communicator intended, if the source is not perceived as credible, the communication could be discounted. Moreover, the perceived credibility of a source will influence how a risk is perceived (Trumbo & McComas, 2003), which could influence how an individual responds to the risk, for instance taking action to avoid, mitigate or adapt to the risks emphasised in the communication (Wachinger et al., 2013). Therefore, it is of paramount importance that both factors are considered when investigating how to improve uncertainty communications. Thus, this thesis investigates three of the most commonly used formats: verbal probability expressions, numerical percentages and mixed expressions, and systematically considers their effect on both understanding and perceived credibility.

Verbal probability expressions.

Probability is typically conceived of as a numerical construct, yet Zimmer (1983) suggested that people are more used to processing the probability of uncertain events according to verbal rules of language, rather than the mathematics of probability. Given the frequent unavailability of precise numerical estimates, probabilistic information is often presented using VPEs (Zimmer, 1983). VPEs openly acknowledge uncertainty and refer to a range of probabilities (e.g., Budescu & Wallsten, 1995). They are natural and easy for individuals to use (Budescu & Wallsten, 1987; Wallsten, Budescu, Zwick, & Kemp, 1993), as well as being thought to be easier to understand than numerical expressions (Erev & Cohen, 1990), especially given no numerical ability is required for comprehension. In addition, they can convey additional information, such as the communicator's knowledge or opinions (Budescu & Wallsten, 1995). VPEs also allow individuals to express the vagueness and imprecision of their judgement (Budescu, Weinberg, & Wallsten, 1988; Hamm, 1991; Wallsten, Budescu, Rappaport, Zwick, & Forsyth, 1986). It is likely for these reasons that individuals have a general preference to *communicate* risks to others using VPEs,[4] despite preferring to *receive* such information in a numerical format (Wallsten et al., 1993)– the so-called 'communication mode preference paradox' (Erev & Cohen, 1990).

Whilst in some contexts the imprecision of such terms is an advantage, imprecision is often cited as a fundamental flaw of VPEs. VPEs are fuzzy concepts (Teigen, Juanchich, & Riege, 2013) and inherently vague (Wallsten, 1990; Wallsten, Budescu, et al., 1986), which makes them difficult to standardise. Moreover, the vagueness of VPEs differs across phrases, for instance 'absolutely certain' refers to a narrower range than phrases with meanings of around 75% (e.g., 'quite likely') (Hamm, 1991). Here, participants tended to prefer phrases with broader meanings, and also seemed aware that the meaning of such phrases had a greater level of interpersonal variability (Hamm, 1991). The problem of variability is compounded given the large number of expressions which can be used to convey uncertainty. Budescu et al. (1988) asked participants to generate phrases to convey the chance of a dart hitting a certain target, with 20 participants using a total of 111 different expressions. Similarly large numbers of VPEs have been observed in the lexicons of intelligence analysts (Dhami, 2018).

As noted previously, over the last five decades there has been an abundance of research investigating the variability in interpretations of VPEs, and this has used a variety of

[4] Though some research indicates that the preference to use VPEs may depend on the type of uncertainty being communicated, and the level of vagueness (Juanchich & Sirota, 2019). VPEs were preferred when conveying epistemic (arising from lack of knowledge) and dispositional (relies on causal assessment) uncertainties. Numbers were preferred when conveying distributional uncertainties (stemming from the world and the frequency of similar outcomes seen previously).

paradigms. A brief outline of these paradigms is given below, before presenting a more in-depth review of some of the contextual and cultural factors known to influence interpretations. The most common paradigm has been the ‘how likely’ translation approach, in which participants are asked to translate a VPE into a corresponding numerical probability, either as a single estimate (e.g., Weber & Hilton, 1990), or a minimum and maximum (e.g., Mosteller & Youtz, 1990). For example:

> "After your annual medical check-up, your doctor tells you that there is a *slight chance* that you will develop an ulcer during the next year. What do you think is your doctor's estimate of the probability of your developing an ulcer during the next year?” (Weber & Hilton, 1990, p.784, emphasis added)

Use of these methods has consistently shown that VPEs are associated with large amounts of variation in how they are interpreted (e.g., Beyth-Marom, 1982; Brun & Teigen, 1988; Bryant & Norman, 1980; Budescu, Broomell, & Por, 2009; Budescu, Por, & Broomell, 2012; Budescu, Por, Broomell, & Smithson, 2014; Clarke, Ruffin, Hill, & Beamen, 1992; Doupnik & Richter, 2004; Hamm, 1991; Harris & Corner, 2011; Lichtenstein & Newman, 1967; Mazur & Hickam, 1991; Mosteller & Youtz, 1990; Patt & Schrag, 2003; Reagan, Mosteller, & Youtz, 1989; Theil, 2002; Wallsten, Fillenbaum, & Cox, 1986; Weber & Hilton, 1990). This research has provided valuable insights into the contextual factors which influence interpretations, as detailed later in this section. However, translation tasks have been criticised for being artificial, and not representative of the cognitive processes people use when assessing VPEs (Renooij & Witteman, 1999). It is perhaps for these reasons that other paradigms have been employed.

An alternative approach, which elicits participant’s fuller understanding of VPEs (Ho, Budescu, Dhami, & Mandel, 2016), is the ‘membership function’ paradigm, in which participants are asked to rate the appropriateness of a list of VPEs across the probability continuum (Budescu, Karelitz, & Wallsten, 2003; Budescu & Wallsten, 1990; Dhami &

Wallsten, 2005; Fillenbaum, Wallsten, Cohen, & Cox, 1991; Karelitz & Budescu, 2004; Rapoport, Wallsten, & Cox, 1987; Wallsten, Budescu, et al., 1986). For instance, participants are asked to indicate how well a probability of 0.8 describes 'likely', with a membership value of zero indicating that the probability is not a substitute for the VPE (at all), and a membership value of one indicating that the probability is a complete substitute for the VPE. Use of such methods has been shown to reliably measure how people understand and use VPEs, and has also observed considerable interpersonal differences in interpretations (Budescu & Wallsten, 1990; Dhami & Wallsten, 2005; Rapoport et al., 1987; Wallsten, Budescu, et al., 1986).

A less common approach is the ranking paradigm, which, as the name implies, involves participants ranking VPEs in a list (Budescu & Wallsten, 1985; Dhami & Wallsten, 2005; Karelitz & Budescu, 2004; Reyna, 1981). Results from ranking studies give cause for a less pessimistic view of VPEs; when participants made similarity judgements and then ranked VPEs, there was a high degree of consistency between ratings (Reyna, 1981). Indeed, whilst precise numerical understandings are not consistent, people do seem to be able to differentiate and rank VPEs relative to each other, ranking them in a similar/stable order (Renooij & Witteman, 1999). Thus, effective translation may be possible if relative rankings are ascertained (Dhami & Wallsten, 2005; Karelitz & Budescu, 2004).

More recent research has utilised the novel 'which-outcome' paradigm (Juanchich, Teigen, & Gourdon, 2013; Løhre & Teigen, 2014; Teigen & Filkuková, 2013; Teigen, Juanchich, & Filkuková, 2014; Teigen et al., 2013).[5] Here, participants are typically shown a graph (usually a histogram, but sometimes a line graph) depicting a distribution of outcomes and asked to complete probability statements (e.g., "It is *unlikely* that a battery will last __ hours") with a value they consider appropriate. This research has found that when VPEs such

[5] See also Juanchich and Sirota (2017); Teigen, Andersen, Alnes and Hesselberg (2019) and Teigen, Filkuková and Hohle (2018) for further research using the 'which-outcome' paradigm published since the current research (see Chapter 2) was carried out.

as ‘unlikely’ and ‘improbable’, or expressions such as ‘can’ or ‘may’ were used to describe outcomes which could be ordered on a unipolar dimension (e.g., battery life), participants interpreted the term as referring to outcomes from the high end of the distribution. Most often participants completed the sentence with a lifetime that *exceeded* the maximum time any sampled battery had lasted – the ‘extremity effect’, equating to a translation of 0%. However, when an expression such as ‘will’ was used to describe the outcome, participants complete the sentence with either the lowest, or the most probable outcome. The precise mechanisms underlying the ‘extremity effect’ have yet to be elucidated.

Whilst the outlined paradigms investigate understanding of VPEs in different ways, their results are similar in that they generally observe considerable variability in the way the VPEs are used and interpreted. For instance, high levels of inter-individual variability have been consistently observed. That is, different individuals use different expressions to describe the same level of uncertainty, or use the same expression to describe different levels of uncertainty (e.g., Beyth-Marom, 1982; Bryant & Norman, 1980; Budescu & Wallsten, 1985; Clarke, Ruffin, Hill, & Beamen, 1992). Such variability has also been evidenced in real-world contexts. For instance, Morgan (1998) asked members on the Executive Committee of the Environmental Protection Agency to give numerical translations of VPEs featured in the EPA’s cancer guidelines, and found four to five orders of magnitude difference between experts. The existence of inter-individual differences in interpretations of VPEs is clearly demonstrated in the above, but a further body of evidence indicates the variability could be further amplified by a range of contextual and cultural factors, as detailed below.

Influence of contextual factors.

Despite the fact that the same individuals seem to interpret the same VPEs consistently over time (e.g., Beyth-Marom, 1982; Bryant & Norman, 1980; Budescu & Wallsten, 1985; Rapoport et al., 1987), there is much research indicating that this will not be the case when the same VPEs are presented across different contexts. A number of contextual effects arising from methodological design can influence the interpretations of VPEs. Firstly, the context within which the VPE is presented matters, for instance, whether participants are asked to assign a number to a VPE when it features in a sentence, versus on its own (Brun & Teigen, 1988). Secondly and relatedly, the response scale has also been shown to influence subjective probability judgements (Haase, Renkewitz, & Betsch, 2013). Whilst these are important to consider in experimental design, perhaps more interestingly, and of more relevance to the current research, are the wider contextual factors which can influence interpretations.

Interpretations of VPEs have been shown to be sensitive to the base-rate of the event being described (Wallsten, Fillenbaum, et al., 1986; Weber & Hilton, 1990). For example, different numeric interpretations are given upon hearing there is a *slight chance* of rain in London versus in Madrid. Wallsten, Fillenbaum and Cox (1986) found these effects differed according to probability level, with high and neutral expressions more strongly influenced by base-rate than low probability phrases. This was suggested to have arisen as a result of the greater vagueness of the high probability and neutral expressions. Indeed, high probability phrases have been described as more flexible in definition (Pepper & Prytulak, 1974).

Relatedly, the valence of an event, specifically its severity, has been found to influence interpretations of VPEs. Physicians gave different interpretations of a 'a low probability of infection' versus 'a low probability of death' – though it was unclear whether this was attributable to the severity or base rate effects (Merz, Druzdzel, & Mazur, 1991). Studies which have separated the two factors have found both an effect of severity of the outcome, and, as seen above, the base-rate of the event described by the VPE. VPEs referring to more severe outcomes are perceived as more likely than those referring to neutral events (Harris & Corner, 2011; Villejoubert, Almond, & Alison, 2009; Weber & Hilton, 1990). For instance, when an 'unlikely' sea level rise would cause an island to disappear, higher numerical translations were observed than when this island was protected by high cliffs (Harris & Corner, 2011). The effect of severity does, however, seem to disappear when participants are provided with numerical base-rates (Fischer & Jungermann, 1996).

Despite their international use, interpretations of VPEs are also vulnerable to large amounts of cultural variability. When presented with VPEs from the International Accounting Standards (IAS), German and English interpretations varied significantly, with English VPEs given broader interpretations than their German equivalent (Doupnik & Richter, 2003). Differences in interpretations of IAS VPEs were also found between English and Russian participants (Du, Stevens, Ahern, & Shigaev, 2015). Similar cultural differences have been found for VPEs used by other international organisations such as the Intergovernmental Panel on Climate Change (IPCC) (Harris, Corner, Xu, & Du, 2013). In comparing English and Chinese interpretations, Harris et al. (2013) found Chinese interpretations were regressive (i.e., closer to the middle of the scale) and showed less differentiation between VPEs, compared to UK participants. Correspondence of VPEs to IPCC guidelines was greater for UK participants compared to Chinese participants (Harris et al., 2013). Cultural differences in likelihood ratings of VPEs have been found even when the terms are cognates in English and Spanish (Cohn, 2009).

As previously highlighted, much research investigating VPEs has ignored the social, informational, motivational and discourse context that uncertainty communications take place in (Fox & Irwin, 1998). Fox and Irwin (1998) argue that this wider context can provide additional cues which influence what the speaker communicates, as well as what is understood by recipients. In addition to making inferences based on the format of the expression, recipients will also make inferences about the context when the communicator formed the judgement. These include their motivations, level of accountability and goals, which will influence how credible and accurate the communicator is perceived to be. Fox and Irwin (1998) suggest that these perceptions in turn influence how much weight is given to the information, and the recipient's willingness to update their beliefs on the basis of the information communicated.

Functions of VPEs.

As well as communicating uncertainty, VPEs may be used to fulfil a range of pragmatic functions. For example, VPEs can be used to convey a belief that an event will (not) happen via directionality. VPEs can also be used for face-management purposes – either on the part of the speaker (e.g., to avoid blame) or the hearer (e.g., in order to be polite), described in more detail below. Given that use of a VPE can be motivated by any of these functions, greater ambiguity regarding the VPE's meaning may be a consequence of the recipient not knowing the intended reason for its use and attempting to ascertain the intended reason for its use (Holtgraves & Perdew, 2016).

Directionality refers to the notion that VPEs do not simply convey probability information, but also additional, pragmatic information, by focusing attention on either the occurrence (positive directionality, e.g., 'a chance') or non-occurrence of the event (negative directionality, e.g., 'doubtful') (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 2003). Importantly, the focusing of attention occurs even when the VPEs are interpreted as conveying the same numerical probability. In a sentence completion task (e.g., 'It is *unlikely* that Clinton will become a good president because…'), those 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., 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; Teigen & Keren, 2003).

Directionality has also been shown to lead to differences in behavioural outcomes (Honda & Yamagishi, 2006, 2009; Teigen & Brun, 1999). For instance, if a doctor describes the likelihood of side effects from a new medication as ‘doubtful’, patients are more inclined to take the medication than when there is ‘some possibility’ of side effects (Teigen & Brun, 1999). The fact that these effects still occur even if the VPEs are perceived as conveying similar numerical probabilities further supports the notion that VPEs do not just function as probability communicators, but also fulfil other, pragmatic, argumentative functions.

Face-management functions can occur on both the part of the hearer, and the speaker (Bonnefon & Villejoubert, 2006; Holtgraves & Perdew, 2016; Juanchich & Sirota, 2013; Juanchich, Sirota, & Butler, 2012; Sirota & Juanchich, 2015). Hearer face-management refers to the notion that a speaker will use a VPE to avoid giving harsh news and upsetting the hearer (e.g., ‘it is possible you will develop deafness’) (Bonnefon & Villejoubert, 2006). A speaker face-management strategy refers to the idea that a speaker may use a VPE to ‘hedge one’s bets’, in order to avoid blame in case the prediction is wrong (Juanchich et al., 2012). Speaker face-management has been found to be the most common interpretation for both predictions about positive and negative outcomes (Juanchich et al., 2012).

Both types of face-management functions have been shown to influence interpretations of the VPE – when VPEs were perceived as (hearer) face-management devices, they were interpreted as referring to higher probabilities than when they were interpreted as communicating likelihood (Bonnefon & Villejoubert, 2006; Juanchich et al., 2012). Relatedly, face-management functions also influence estimates provided by a speaker. When conveying information about a negative outcome, a speaker instructed to perform either of the face-management functions gave a lower probability estimate than one who was instructed to be informative and communicate likelihood (Juanchich & Sirota, 2013; Sirota & Juanchich, 2015). Considering both speakers and hearers in two studies, Holtgraves and Perdew (2016) found that participants chose and produced expressions conveying greater uncertainty when the event was more severe (and thus more facethreatening). The greater the uncertainty communicated, the lower the probability estimate perceived by the hearer.

Further support for the use of VPEs as face-management devices comes from Piercey (2009), who studied probability assessments in an accounting context. Piercey (2009) suggested that the vagueness of VPEs allows an individual to retain the ability to re-define the expression in light of an outcome – what he terms 'elastic redefinition'. This redefinition can be done in three ways – in terms of the expression's probability level, vagueness (increased/decreased) and skewness (emphasis on lower/higher probabilities). Participants were firstly asked to provide a verbal probability expression of the likelihood that a company was overstating the value of an asset. Participants were asked to define the membership function of this expression either immediately before or after learning that the value had been overstated. Responses given after the outcome information was presented were more vague than responses given before, indicating that participants were trying to save face, and still appear correct.

In light of results demonstrating the vagueness of VPEs, it is perhaps inevitable that such variability in interpretations has been found to have an impact on decision-making (e.g., E. M. Peters, Hart, Tusler, & Fraenkel, 2014; Young & Oppenheimer, 2009). Such variability can have far-reaching consequences, as illustrated by this example set in a natural hazards context. When a (hypothetical) volcanic eruption forecast was communicated using a VPE, there were fewer evacuations than when the same information was presented using numerically equivalent terms, a difference that was greater for scientists (versus non-scientists) (Doyle, McClure, et al., 2014). Use of a VPE was seen as more ambiguous, and the difference between scientists and non-scientists was attributed to the former's belief that numbers are more certain.

Yet despite the vast variability associated with VPEs, they are commonly used in a range of domains, including accounting (Doupnik & Richter, 2003), intelligence analysis (Ho et al., 2016) and climate change (Mastrandrea et al., 2010). Their use is based on the underlying assumption that the speaker and the listener interpret the terms similarly (an 'illusion of communication', Budescu & Wallsten, 1985), yet the above research demonstrates this is rarely the case, greatly increasing the chances of mis-communications. The seriousness of mis-communications in a natural hazards context where lives are at stake cannot be understated, as illustrated by the events of the 2009 L'Aquila earthquake (Sellnow, Iverson, & Sellnow, 2017). Given variability (and thus the potential for mis-communications) has been shown to be lower for numerical rather than verbal formats (Budescu et al., 1988), the following section moves on to consider the pros and cons of numerical formats.

Numerical expressions.

Probabilities (e.g., 20%) and frequencies (e.g., 1 in 5) are two of the most common ways of presenting risk information numerically (Reimer, Jones, & Skubisz, 2014). Despite being mathematically equivalent, the specific numerical format influences both how people perceive, interpret and use the information (E. M. Peters et al., 2009; E. M. Peters, Hart, & Fraenkel, 2011; E. M. Peters et al., 2006; Skubisz, 2010; Slovic, Monahan, & MacGregor, 2000). This thesis focuses on *probabilities*, in which the likelihood of an event or an outcome occurring is expressed from zero (it will definitely not occur) to 1 (it will definitely occur), with this number often translated into a percentage (Timmermans & Oudhoff, 2011). Thus, this thesis specifically focuses on percentages,[6] and begins by initially considering the use of point formats (e.g., 20%), before extending findings to range formats (e.g., 10–30%). The term 'numerical expression' is used to refer to both types of percentage formats.

Numerical expressions have many advantages over VPEs; first and foremost, they are more precise and allow a communicator to be exact and specific. They have a clear, fixed-

[6] These percentage formats are used to reflect subjective probabilities, rather than frequentist probabilities. This is because natural hazard communications typically refer to unique events (as opposed to repetitions of the same event), for which there is little previous knowledge (e.g., Marzocchi, Newhall, & Woo, 2012).

rank order (e.g., 90% is more than 70%) and can easily be translated to other formats (e.g., 20% is equivalent to 1 in 5). They are also argued to be more reliable, given they are less easily redefined than words (Piercey, 2009). Their advantages also seem to translate into benefits in ‘real world’ domains, such as medicine. When healthcare professionals were asked to make treatment decisions which were described using verbal and numerical probability expressions, there was a greater level of agreement between doctors for decisions made when the information was presented using numerical terms (Timmermans, 1994).

Point estimates seem advantageous, in that there are fewer possible outcomes to consider, which makes them ‘psychologically easier’ and less cognitively demanding than ranges (Joslyn, Nemec, & Savelli, 2013). Although point estimates are numerically precise, it is no guarantee that they will be understood as intended. A classic illustration of misunderstandings of point estimates is associated with the US National Weather Service ‘probability of precipitation’ forecasts, which have been issued since 1965. These forecasts are a point percentage estimate as to the probability of rain for a certain area, in a given time frame (e.g., ‘there is a X% probability of rain tomorrow’). However, much research has demonstrated that people misunderstand what these forecasts mean (Gigerenzer, Hertwig, Van Den Broek, Fasolo, & Katsikopoulos, 2005; Joslyn, Nadav-Greenberg, & Nichols, 2009; Juanchich & Sirota, 2016; Morss, Demuth, & Lazo, 2008; Peachey, Schultz, Morss, Roebber, & Wood, 2013). Common misunderstandings include the belief that it will rain for X% of the day, or that rain will cover X% of the area. These misunderstandings are argued to result from the lack of reference class (i.e., what the probability refers to). This problem has led to recommendations to improve understanding of weather forecasts by specifying the reference class (Gigerenzer et al., 2005; Juanchich & Sirota, 2016). Similar misunderstandings have been observed for patients in medical contexts (Brun & Teigen, 1988), a finding which is all the more striking given the majority of patients indicated a preference for receiving information in numerical terms, and believed they were easier to understand (versus VPEs).

Some of the highlighted misunderstandings may result from the fact that numerical formats are ambiguous in how they should be intuitively interpreted (Bilgin & Brenner, 2013; Flugstad & Windschitl, 2003; Moxey & Sanford, 2000; Teigen & Brun, 2000; Windschitl, Martin, & Flugstad, 2002; Windschitl & Weber, 1999). Hearing that the probability of an event is 20% does not indicate whether it is conveying a possibility or doubt; numerical expressions carry few affective or intuitive implications, which makes them difficult to evaluate (Flugstad & Windschitl, 2003). Similar to VPEs, numerical expressions are also susceptible to context effects (Flugstad & Windschitl, 2003; Verplanken, 1997), illustrating their 'intuitive flexibility' (Windschitl, Smith, Scherer, & Suls, 2017).

Framing, specifically in this case, 'attribute framing', refers to instances in which only one option (in this thesis, an event) is described along one dimension – either in terms of its occurrence, or non-occurrence.[7] Although the two frames are mathematically equivalent (given the probabilities total 100%), they differ in the amount of information conveyed, which has been shown to differentially influence judgements and decisions. For instance, when mince is described as '75% lean', it is perceived as less greasy, and better quality than when it is described as '25% fat' (Levin & Gaeth, 1988). People showed similarly positive feelings towards '75% fat-free' products, which were perceived as more healthy (Sanford, Fay, Stewart, & Moxey, 2002). The communicator chooses how to frame the information, and selection may depend on the purpose they are trying to achieve. A successful communication can be said to be one in which the recipient not only assesses the information, but also understands the communicator's reasons for framing the information in the way that they did, either for informational and/or argumentative purposes (Teigen,

[7] We recognise that framing is not solely limited to numerical expressions, but make the point here given that the attribute framing literature has predominantly focused on such effects in relation to numerical information.

2015). It is plausible that assessments of these reasons could feed into judgements of a communicator's credibility.

An 'information leakage' account has been put forward to explain attribute framing effects (McKenzie & Nelson, 2003; Sher & McKenzie, 2006, 2008). This account suggests that although the frames are mathematically equivalent, they are not informationally equivalent. The chosen frame 'leaks' information beyond the frame's explicit content, for instance signalling a communicator's reference point, attitude or implicit recommendation. In one example, McKenzie and Nelson (2003) found that the choice of frame leaked information about the communicator's reference point, which recipients were able to accurately infer. For instance, recipients were more likely to infer that a cup which was described as 'half empty' was originally full versus when it was described as 'half full'. The 'information leakage' explanation suggests that attribute framing effects occur as a result of the different inferences recipients draw when presented with different frames. Given these findings, it is reasonable to predict that when a communicator makes a probabilistic forecast, their choice of frame will influence the inferences the recipient makes, namely relating to the communicator's expectations of the outcome. If the outcome is contrary to these expectations, it could have ramifications for the communicator's perceived credibility.

Relatedly, the expressions used to present numerical ranges also leak information. For instance, the range '10–30%' can also be described using single bound expressions, such as 'more than 10%'/'above 10%' (lower bounds) or 'less than 30%'/'below 30%' (upper bounds), with lower bound expressions used more frequently (Halberg & Teigen, 2009). Whilst one might presume these to be neutral expressions of magnitude, and thus similar ways of framing the same estimate, they act as reference points (Halberg, Teigen, & Fostervold, 2009), conveying additional pragmatic information which suggests how the magnitude should be interpreted or evaluated (Teigen, 2008; Teigen, Halberg, & Fostervold, 2007). For instance, the expressions may be used by the speaker to convey attitudes or expectations, or to encourage a certain course of action. The expressions are directional, in that they guide one's attention to either the occurrence (lower), or non-occurrence (upper) of an event (Hohle & Teigen, 2018). They also indicate a communicator's belief about whether the trend will increase or decrease (Hohle & Teigen, 2018); generally, lower (upper) bounds are perceived as indicating large, increasing (small, decreasing) quantities (Teigen, 2008; Teigen et al., 2007).

People are sensitive to the objective base rate of events, in that perceived certainty of an event was higher when the event had a high rather than low base rate (Windschitl & Weber, 1999). For instance, a 20% chance of contracting a disease related to malaria was perceived as more likely on a trip to India (versus Hawaii), as indicated on a subjective graphic line scale, ranging from 'definitely wouldn't get the disease' to 'definitely would get the disease'. The extent to which perceived base rate affects the interpretation of numerical probabilities may depend on the size of the base rate. Perceived base rates were found to affect the interpretation of numerical probabilities more so for lower rather than higher probabilities (Bilgin & Brenner, 2013). Building on work by Teigen and Brun (1995), Bilgin and Brenner (2013) attributed this to the directional ambiguity of the former, in that they can be interpreted either positively or negatively. Thus, even using numerical expressions does not guarantee that interpretations will be stable across contexts.

Even information which is not directly relevant to the forecast at hand can influence interpretations of numerical probabilities. When given information about prevalence rates for a disease for both a target group and a context group, prevalence information for the context group was used as a comparison point. The target group were perceived as more (less) likely to get the disease when the prevalence rate for the context group was lower (higher) than that for the target group (Windschitl et al., 2002). After being given a numerical prediction about the success of surgery, reasons (positive or negative) given for the doctor's probability estimate influenced levels of optimism and worry (Flugstad & Windschitl, 2003).

Despite being (numerically) precise, these findings provide a further example of how even numerical expressions can be flexibly interpreted.

Aside from the flexibility of interpretations, numerical expressions lack sensitivity in conveying gut-level reactions and intuitions (Lipkus, 2007), and are inappropriate when it comes to conveying risks smaller than 1% (Timmermans & Oudhoff, 2011). Furthermore, communicators are often reluctant to use them for fear of portraying a false level of certainty (Schwartz, Woloshin, & Welch, 1999) or precision (Budescu & Wallsten, 1985). If a communicator is perceived as too precise (i.e., more than is warranted by their actual level of knowledge), then the credibility of both the information, and the reliability of the communicator may be brought into question (Zhang & Schwarz, 2012). Such concerns were echoed by intelligence analysts, who were reluctant to use numerical expressions for fear of giving a false sense of precision, and the possibility of such expressions being used as a means of evaluation (i.e., how right or wrong the estimate was; A. Barnes, 2016). They were also concerned about generating numerical probabilities in cases of unmeasurable phenomena.

The issue of ability (and indeed suitability) of using numerical expressions is echoed by those charged with communicating predictions about natural hazards. There is not always adequate data (either in terms of quality or quantity) to produce a numerical probability forecast for a volcanic eruption (Fearnley, Winson, Pallister, & Tilling, 2018). Scientists at the Montserrat Volcano Observatory considered use of quantitative expressions as complicating their risk communications, as "likelihoods and associated uncertainties were neither well-explained nor understood" (Haynes et al., 2008b, p. 263). It is therefore unsurprising that calls have been made to only use numerical expressions in relation to 'well-defined outcomes' (Risbey & Kandlikar, 2007).

The concern over whether an audience interprets the numerical expression as intended is amplified when one considers the general numeracy level of the public. Only 22% of working age adults in England have numeracy skills equivalent to a GCSE C Grade

(Department for Business, Innovation & Skills, 2011). It therefore seems that concerns over numerical expressions holding little meaning for most members of the public may be warranted. Many difficulties in the understanding of risk are suggested to result from a lack of numeracy (Brust-Renck, Royer, & Reyna, 2013). For instance, lower levels of numeracy are associated with increased risk and likelihood perceptions (Dieckmann, Slovic, & Peters, 2009), as well as increased vulnerability to extraneous factors, including communication format (e.g., frequencies versus percentages) and framing effects (E. M. Peters et al., 2006). It seems therefore prudent to take into account the numeracy level of the audience when developing communications, as it will have implications for the effectiveness of such communications (Lipkus, Samsa, & Rimer, 2001; E. M. Peters, 2012; E. M. Peters et al., 2006; Reyna, Nelson, Han, & Dieckmann, 2010).

Despite the problems associated with using numerical expressions, they have long been seen as a crucial part of risk communication strategies, evidenced in recommendations that suggest: "all we have to do is tell them the numbers" (Fischhoff, 1995, p. 138). Yet the literature presented above demonstrates communication is clearly not that simple – numbers can, and are, flexibly interpreted, which makes effective risk and uncertainty communication a challenge. It does not seem to be a straightforward choice between using VPEs or numerical formats (O'Brien, 1989), and thus using a combination of both formats may be more appropriate.

Mixed expressions.

In a bid to reduce the mis-communications associated with use of VPEs, one commonly proposed solution is a mixed format, combining both a VPE and its numerical translation, for example, 'It is unlikely (less than 33%)' – a verbal-numerical (V-N) format. Use of such a mixed format was first suggested in the early eighties (Beyth-Marom, 1982), but the call has been echoed in more recent years (Budescu et al., 2009, 2012, 2014; Harris, Por, & Broomell, 2017; Patt & Dessai, 2005; Renooij & Witteman, 1999; C. L. M. Witteman & Renooij, 2003; C. L. M.

Witteman, Renooij, & Koele, 2007). Such a format has the advantage of catering to everyone's preference (Budescu et al., 2012), irrespective of numerical ability, and means people can focus on the representation that they find the easiest to use (Dieckmann, Peters, Gregory, & Tusler, 2012; Gregory et al., 2012; C. L. M. Witteman et al., 2007). Indeed, a combined format was rated as the easiest to understand versus five other presentation formats (including VPEs, and numerical point expressions; Brewer et al., 2009). The format has also been said to facilitate the process of probability assessment (C. L. M. Witteman & Renooij, 2003).

Several organisations have sought to standardise the use of VPEs by developing guidelines which feature numerical translations of VPEs. The IPCC were one of the first organisations to create such guidelines, with their latest classification table featuring seven VPEs and their associated numerical values (Mastrandrea et al., 2010, see Table 1). Similar efforts to standardise uncertainty lexicons were introduced in defence in the mid-2000s, in the US by the National Intelligence Council (NIC) and in the UK by Defence Intelligence (DI) (as cited in Ho, Budescu, Dhami, & Mandel, 2016). Other adopters of an uncertainty lexicon include the World Meteorological Organisation (Gill et al., 2008) in their "Forecast Likelihood Scale", and the European Food Safety Authority (EFSA) (Hart et al., 2019).

Table 1. IPCC Likelihood Scale (Mastrandrea et al., 2010)

Term	Likelihood of outcome (% probability)
Exceptionally unlikely	0–1
Very unlikely	0–10
Unlikely	0–33
About as likely as not	33–66
Likely	66–100
Very likely	90–100
Virtually certain	99–100

The development of an uncertainty lexicon can be used as a basis for creating a mixed format, which has been shown to improve levels of understanding. When the numerical probability ranges were provided alongside the VPEs, consistency with IPCC guidelines increased (Budescu et al., 2009, 2012). This V-N format was also found to increase levels of differentiation between VPEs, as well as increasing consistency in interpretations and reducing out of range responses (Budescu et al., 2012). Importantly, the increased accuracy of interpretations for more numerate participants did not come at the expense of the performance of the less numerate participants (Budescu et al., 2012). This increase in both differentiation and correspondence between people's interpretations and the IPCC guidelines for the V-N format was further replicated in 24 countries and across 17 languages (Budescu et al., 2014). In light of these findings, Budescu and colleagues have advocated a mixed format approach.

The aforementioned benefits of mixed formats could result from two mechanisms: the numerical expression can act as an anchor, which guides subsequent numerical estimates, or the numerical expression could increase the likelihood that the expression will be perceived as a likelihood-communication device (Juanchich et al., 2012). Overall, the advantages of using mixed formats over solely using verbal or numerical expressions is clear. A mixed format caters to people's preferences, but also allows fine-tuning of the range values to indicate more precision if required (Ho et al., 2016).

Whilst the mixed, V-N format is clearly beneficial in reducing variability in understandings of VPEs, it is not a perfect solution. The benefits gained may depend on how the mixed format is presented, with improvements only appearing when the numerical guidelines were directly presented within text (versus appearing via a new window, or via a tooltip; Wintle, Fraser, Wills, Nicholson, & Fidler, 2019). In addition, estimates were better differentiated when the numerical bounds were explicitly stated (e.g., '0–33%') versus when they were not (e.g., '< 33%'; Harris et al., 2017).

Further problems include the considerable variation in how organisations define the same VPEs. Table 2 shows how a variety of organisations translate 'unlikely' and 'likely' (the two VPEs the thesis predominantly focuses on). Alarmingly, the variation in numerical guidelines for the same VPE occurs even within similar contexts (e.g., intelligence analysis). Furthermore, the IPCC state that the translation table guidelines are *not exhaustive*, as technically 'likely' could be used for anything between 66%–100%, despite the fact that there are other terms for some of these probabilities (Budescu et al., 2012). As a result, the IPCC guidelines have been criticised for being confusing (Wintle et al., 2019), with greater benefits shown for guidelines which have mutually exclusive categories.

Table 2. Organisational Guidelines for 'Unlikely' and 'Likely'

VPE	Translation	Organisation
Unlikely	0–33%	IPCC
	10–29%	WMO
	10–33%	EFSA
	20–45%	NIC
	15–20%	DI
Likely	66–100%	IPCC
	70–89%	WMO
	66–90%	EFSA
	55–80%	NIC
	55–70%	DI

Note. IPCC = Intergovernmental Panel on Climate Change; WMO = World Meteorological Organisation; EFSA = European Food Safety Authority; NIC = National Intelligence Council; DI = Defence Intelligence.

Dhami (2018) presents an in-depth discussion of the problems associated with such standardised guidelines, focusing on those used by intelligence analysts. In the NIC and DI guidelines, the categories overlap, which makes choosing a specific phrase difficult when communicating uncertainty at the top or bottom of the category. Moreover, guidelines which are not evidence-based (i.e., do not reflect what we know about how people use these phrases) cannot capture whether individuals are familiar and comfortable with the language featured, or if the expressions are considered to be interchangeable. Dhami (2018) also highlights the lack of clarity over whether individuals (a) use phrases in the guidelines and (b) use them as intended (i.e., expressions presented alongside the numerical uncertainty value), finding that analysts did not use some of the mid-point expressions prescribed by the NIC at all. More promisingly, when evidence-based methods, such as the 'membership function' approach are used as a basis for formulating numerical equivalents for VPEs, improved understanding and reduced variability is observed, an effect demonstrated in both climate science and intelligence analysis domains (Ho et al., 2016; Wintle et al., 2019).

Of particular significance to this thesis is the fact that the aforementioned research has only ever considered mixed formats in one presentational order – V-N, rather than also including numerical-verbal (N-V) formats. Given Grice's (1975) conversational maxims, specifically the 'maxim of manner', which stipulates that one should be as orderly as possible, it is reasonable to assume that individuals could still be predominantly focusing on the VPE, which is presented first. As evidenced previously, such a focus could lead to variations in interpretations. This thesis takes research on mixed formats one step further by considering the order with which the translations are presented.

Interim Summary

As is apparent from the literature reviewed above, there is no optimal communication format which can drastically improve the quality of communications (Budescu et al., 2012; Politi et al., 2007). Given that poor communications could have as big an impact on people than the risk being described, it is said that it should be no more acceptable to release untested communications than untested drugs (Fischhoff, 2019). Importantly, this testing should not only include a consideration of understanding, but should also consider the effects on perceptions of a communicator (Wiedemann, Schtz, & Thalmann, 2008). So far, we have considered research which has investigated the former. We now move on to consider research which focuses on how a communicator is perceived, specifically in relation to credibility. Credibility is one of a communicator's most precious assets (Covello & Allen, 1988) – even if a communication is understood as intended, it is of little use if the source is not perceived as credible (Berry, 2004). It is therefore unsurprising that credibility has been put forward as a fundamental part of effective risk communication practices (Reynolds, 2011). However, systematic research examining the influence of credibility in this context is lacking (Frewer et al., 1997).

Credibility and Trust

Credibility and trust are commonly treated as synonyms, but the two concepts have been differentiated (Fogg & Tseng, 1999; Lucassen & Schraagen, 2011). Trust has been said to encompass the willingness to depend on the credibility of the information or source (Lucassen & Schraagen, 2011) – trust is essentially a response based, in part, on a perception of credibility (Trettin & Musham, 2000). Therefore, at its simplest, trust can be seen as 'dependability' and credibility as 'believability' (Fogg & Tseng, 1999). More detailed definitions relate credibility to "the state of being believable, trustworthy and reliable" (Trettin & Musham, 2000, p. 412). Trust and credibility are clearly linked, and we thus present an overview of both below. Credibility forms the main focus of this thesis, which we see as a pre-cursor to trust.

Trust.

When faced with risk and uncertainty, one way of coping with insufficient knowledge is to rely on trust, which reduces the complexity and uncertainty people are faced with (Earle & Cvetkovich, 1995; Luhmann, 1979; Savadori et al., 2004; Siegrist & Cvetkovich, 2000; Slovic, Fischhoff, & Lichtenstein, 1982). Trust means individuals can interact and behave without full knowledge (Kasperson et al., 1992). As such, it is generally assumed that trust and knowledge are negatively correlated – the less knowledge one has about a hazard, the more likely they are to trust someone more knowledgeable. Reversely, the more knowledge one has, the less reliant they are on others (Siegrist & Cvetkovich, 2000).

Relatedly, trust is particularly important when the risks involved are not within the realm of everyday experience (Earle & Cvetkovich, 1995). Natural hazards are complex and uncertain; people do not always have specialist knowledge or experience of them (for instance, hazards with 100 year return periods), which means they are particularly reliant on expert sources (Paton, 2008). These characteristics mean trust is of paramount importance to both experts and the public in such contexts. For instance, the public must trust that experts will provide information about the risk's severity and urgency. In turn, experts must trust that the public will (a) take actions to mitigate risk and (b) react to events in a predictable and socially acceptable way (Eiser et al., 2012).

Trust is not something which is attained permanently, but rather requires maintenance and reinforcement (Kasperson et al., 1992). The fragile nature of trust is a common adage in everyday life, and much has also been made of its fragility within the risk communication literature (Slovic, 1993, 1999). The asymmetry principle refers to the fact that trust takes a long time to build, but can be destroyed in an instant, with actions that erode trust more impactful than actions which build trust (Slovic, 1993).[8] A lack of trust is often invoked as a reason for the disparity between expert and the public's risk perceptions (Slovic, 1993, 1999). It is perhaps for these reasons that fears of a global 'crisis of trust' have been expressed by many in recent decades (Cvetkovich & Löfstedt, 1999; Edelman, 2018). Specific reference has been made to the crisis within both science (House of Lords, 2000), and risk regulation (van Kleef et al., 2007). Despite such fears, in actuality, levels of trust have stayed relatively stable across recent decades in the United Kingdom (Richards & Heath, 2015).

It is commonly assumed that levels of trust and risk perceptions are related, with trust argued to explain up to half of the public's perceptions of risk (Löfstedt, 2005; Slovic, 1993).[9] If trust is lacking, then individuals may have higher risk perceptions (Slovic, 1993). The link between levels of trust and levels of perceived risk has been evidenced in many contexts, including risks associated with natural hazards (Wachinger et al., 2013), artificial

[8] There is some dispute as to the existence of trust asymmetry, see Earle et al. (2012).

[9] Though Sjöberg (2001) argues that trust and risk perceptions may only be weakly associated and thus increased trust may not always lead to reduced risk perceptions.

food additives (Bearth, Cousin, & Siegrist, 2014) and new technologies (Siegrist, Gutscher, & Earle, 2005).

Relatedly, trust has been found to positively influence perceived benefits (Siegrist, 2000; Siegrist, Cvetkovich, & Roth, 2000). Perceived risks and benefits have been found to affect acceptance of gene technology, and thus trust has been said to have an indirect effect on acceptance (Siegrist, 1999, 2000). This influence of trust on acceptance has also been shown for a number of other food technologies (Eiser et al., 2002). Furthermore, trust has been shown as an important determinant of responses to a risk/benefit communication (Frewer, Howard, Hedderley, & Shepherd, 1999).

As well as having consequences for risk and benefit perceptions, trust also has consequences for the intentions and actions people take on the basis of the information communicated, with the most effects seen when there is low familiarity and little information available (Siegrist & Cvetkovich, 2000). Increased trust has been associated with an increased likelihood of taking precautionary action (Losee & Joslyn, 2018), a finding demonstrated in communities affected by natural hazards. When people trusted civic institutions, they were more likely to take the protective measures featured in risk communications (Paton, 2008). Conversely, low levels of trust can lead to inaction, as was the case in Hurricane Katrina, where many people were slow to evacuate as a result of not trusting the authorities' warnings, which were perceived as 'scaremongering' (Eisenman, Cordasco, Asch, Golden, & Glik, 2007). Given these findings, it is unsurprising that trust is a key aspiration for risk communicators, especially in the realm of natural hazards (Doyle, McClure, et al., 2014).

Many of the problems associated with risk communication are said to be due to the lack of trust the public have in experts (Margolis, 1996). It is unsurprising, thus, that trust has been called the 'currency of transaction' within risk communication (World Health Organization, 2014), and forms one of its major objectives. Yet despite the importance of trust, we are only just developing an understanding of the 'mechanics of trust' (Bostrom & Löfstedt, 2003). This has led to calls for future research to investigate how communications can be used to build trust (Infanti et al., 2013).

Credibility.

Self (2009) refers to credibility as a "familiar and yet strangely illusive concept" (p.449). Its illusory nature likely results from the fact that it is a 'perceived quality' which "doesn't reside in an object, a person, or a piece of information" (Fogg & Tseng, 1999, p. 80). However, characteristics of an object, person or information may form the basis of credibility assessments (Rieh, 2009). Throughout the literature, credibility has been used as a broad term, but various permutations also exist, including source credibility and message credibility (Rieh & Danielson, 2007). Whilst there are some similarities between the concepts, message credibility refers to the characteristics of the message (e.g., structure, content and delivery), which can influence perceptions of believability (either of the source, or the message; Metzger, Flanagin, Eyal, Lemus, & Mccann, 2003). The focus in this thesis, however, is on source credibility, which refers the perceptions of the person or institution who is communicating the information (Frewer, Scholderer, & Bredahl, 2003), rather than reflecting something specific about the content of the information itself.

Source credibility is commonly conceived of as relating to the believability of the communicator (Fogg & Tseng, 1999; Hilligoss & Rieh, 2008; Metzger & Flanagin, 2013; O'Keefe, 2016; Wathen & Burkell, 2002). In addition to believability, there are dozens of other concepts suggested to contribute to source credibility (Self, 2009), with research uncovering a multitude of dimensions which are proposed to underlie credibility, including expertise, competence, trustworthiness, dynamism, objectivity, reliability and authoritativeness (for reviews see Eisend, 2006; Pornpitakpan, 2004b).

Two of the most commonly identified dimensions initially derived from source credibility theory (Hovland, Janis, & Kelley, 1953; Hovland & Weiss, 1951) are trustworthiness and expertise, which have been observed in subsequent research (Fogg & Tseng, 1999; Freeman & Spyridakis, 2009; Hovland et al., 1953; Lucassen & Schraagen, 2011; Mowen, Wiener, & Joag, 1987; Wiener & Mowen, 1986; Yoo & Gretzel, 2008), including that which has specifically focused on risk communication (Jungermann, Pfister, & Fischer, 1996; R. G. Peters et al., 1997).[10] Indeed, many of the factors which affect credibility judgements can be said to be components of trustworthiness and expertise (Freeman & Spyridakis, 2009), defined below. We thus adopt such a dual component view of credibility in this thesis.

Problematically, the concepts of credibility, trustworthiness and expertise have been used in many different ways by authors from different disciplines, often without clearly defining the terms. For clarity, in the following literature review we use trustworthiness and expertise in accordance with how the terms have been typically defined within argumentation literature (e.g., Walton, 1997). We see trustworthiness as relating to the integrity and reliability of the communicator, specifically whether they are unbiased, honest and conscientious (Dieckmann et al., 2009; Walton, 1997). We see expertise as relating to the communicator's knowledge and the general characteristic of 'knowledgeability' (Guilamo-Ramos, Jaccard, Dittus, & Bouris, 2006).

The study of credibility, and recognition of its importance to risk communications, stemmed from persuasion and attitude change research. Studies have consistently found that a highly credible source leads to greater persuasion, and greater attitude change than a less credible source (e.g., Heesacker, Petty, & Cacioppo, 1983; Hovland et al., 1953; Hovland & Weiss, 1951; Jain & Posavac, 2001; McGinnies & Ward, 1980; Wiener & Mowen, 1986; Wilson & Sherrell, 1993; for reviews see Ohanian, 1990; Pornpitakpan, 2004b). High levels of source credibility have been found to influence information adoption (Watts & Zhang, 2008) and intentions to seek further information/advice from the source (Joiner, Leveson, &

[10] Although some studies have identified three-factor solutions of source credibility (e.g., Eisend, 2006; Metzger et al., 2003; Ohanian, 1990, 1991), the featured components either specifically name trustworthiness and expertise, or clearly map to them.

Langfield-Smith, 2002). Of particular relevance to the current thesis are findings from Trumbo and McComas (2003), who investigated the role of credibility in responses to risk communications about regional cancer rates. They found that greater levels of perceived credibility directly led to lower risk perceptions.

Traditional models of risk communication put forward perceptions of expertise and trustworthiness as central to the success of a risk communication (Trettin & Musham, 2000). Expertise and trustworthiness dimensions are linked (Guilamo-Ramos et al., 2006; Reichelt, Sievert, & Jacob, 2014). For instance, if a source is perceived as possessing expertise, then people are more likely to judge the source as trustworthy (Hilligoss & Rieh, 2008). In addition, when assessing expertise, individuals are sensitive to its 'epistemic features' (i.e., the level of objectivity and competence, and thus whether it is genuine, dubious or partial expertise) and adjust their reasoning in relation to this (Cummings, 2014). The two dimensions are, however, conceptually distinct (Harris, Hahn, Madsen, & Hsu, 2015; O'Hara, Netemeyer, & Burton, 1991) and have been shown to be discriminant (Reichelt et al., 2014). People are able to make distinctions between the perceived expertise and trustworthiness of sources (Marquart, O'Keefe, & Gunther, 1995), and they seem to have different determinants.

Trustworthiness.

Whilst trust is an attribute of a relationship between partners, trustworthiness is an attribute of an individual partner (Barney & Hansen, 1994). Many of the definitions given to trustworthiness in the literature refer to the same core constructs, such as the honest, benevolent or non-deceptive nature of the source (Dieckmann et al., 2009; Fogg & Tseng, 1999; Guilamo-Ramos et al., 2006; Hilligoss & Rieh, 2008; McGinnies & Ward, 1980; Walton, 1997). As previously outlined, these constructs are central to our conceptualisation of trustworthiness within this thesis. Such constructs have also been highlighted as being of central importance to risk communicators – in Covello and Allen's (1988) cardinal rules for risk communication, rule three centres upon the importance of being 'honest, frank and open'.

Given that trustworthiness is a component of credibility, it is unsurprising that increased levels of trustworthiness have a similar effect to credibility. Specifically, increased trustworthiness has been linked to increased levels of persuasion and attitude change (e.g., Hovland & Weiss, 1951), though some studies have found trustworthiness is insufficient to cause attitude change on its own, additionally requiring expertise (Kelman & Hovland, 1953).

Expertise.

When individuals lack sufficient knowledge to make a judgement or decision, they often look to experts to provide further information to enable them to act. In natural hazard contexts, people are especially reliant on experts, as a result of both the complexity of such hazards, and the relative lack of experience that people have with them (Paton, 2008).

An objective definition of precisely what expertise is remains elusive (Shanteau, 1992). Dictionary definitions typically refer to an expert as one who is skilful, well informed or a reliable source of knowledge, technique or skill (Ericsson, 2018). Expertise can be defined absolutely (e.g., ten years of training) or relatively (e.g., 'experts' versus 'novices'), and is generally used either to refer to top performance (e.g., Olympic athletes) or knowledge (e.g., doctors; Thon & Jucks, 2017). For the purposes of this thesis, expertise is used in reference to the latter – knowledge, and the general characteristic of 'knowledgeability' (Guilamo-Ramos et al., 2006), rather than in terms of making a distinction between 'experts' and 'novices'.

Expertise has commonly been defined in terms of competence (Hovland et al., 1953; McGinnies & Ward, 1980), which involves not just knowledge, but also recognising the limits to one's knowledge (Sjöberg, 2001). Judgements of expertise have also been found to be based on the perception of a source as well informed and intelligent, and capable of making valid assertions (McCracken, 1989). The validity of an assertion is linked to qualities of the information, such as its completeness and accuracy, which Freeman and Spyridakis (2009) classify as components of expertise. Indeed, the ability of the source to provide correct information is central to some definitions of expertise (Rieh, 2009; Treise, Walsh-Childers, Weigold, & Friedman, 2003). This aspect of expertise is of key importance to this thesis, given its setting in a natural hazards context. As indicated previously, the uncertain nature of such hazards means that it is not always possible to make predictions, let alone guarantee their accuracy. In Chapters 3 and 4, we examine the change in perceptions of credibility (and within that, trustworthiness and expertise) when an initial prediction of an event subsequently turns out to be 'erroneous', given a probabilistically unexpected outcome (i.e., an 'unlikely' outcome occurs, or a 'likely' event does not occur).

One source of support for the role of expertise in credibility judgements comes from Peters et al. (1997).[11] In a study of environmental risk communication, perceptions of knowledge and expertise were found to be two determinants of perceived credibility (alongside openness and honesty, and care and concern – the former which relates to our trustworthiness component). Perceptions of knowledge and expertise most strongly explained the largest proportion of variance when it came to explaining perceived credibility of NGOs (versus industry, government and society).

Research exploring the effects of expertise originated in studies of persuasion and attitude change,[12] with source expertise found to influence attitude change (Chaiken, 1980; Petty, Cacioppo, & Goldman, 1981). If a source is perceived as high in expertise, more positive attitudes are found towards the source (Braunsberger, 1996). Given these findings, it is unsurprising that individuals are more likely to be persuaded by a source who is perceived as possessing expertise (McGuire, 1985; Mowen et al., 1987; Sternthal, Phillips, & Dholakia, 1978; Wiener & Mowen, 1986).

[11] Though see also Harris et al., (2015); Shafto, Eaves, Navarro and Perfors (2012) for further support.
[12] Note that much of this research sees expertise as not only relating to knowledgeability but also as relating to experience and/or capability, as these are typically related.

Some studies have observed greater effects of expertise on persuasion than trustworthiness. For instance, in a car buying context, Wiener and Mowen (1986) found that expertise measures accounted for more than twice the size of variance accounted for by trustworthiness. Echoing findings from Mowen and colleagues, Wilson and Sherrell's (1993) meta-analysis also observed a stronger influence of expertise on persuasion (versus other source characteristics). Akin to findings in the persuasion literature, expertise was found to be more influential than trustworthiness and attractiveness on intention to purchase (Pornpitakpan, 2004a). In Mowen et al. (1987), whilst higher expertise of the source was linked with greater persuasion, no such effect was found for trustworthiness.

The strength of the effect of expertise over trustworthiness has also been found given a three-component solution of credibility (made up of expertise, attractiveness and trustworthiness). Ohanian (1991) found that perceived expertise was the only dimension significantly associated with purchase intentions. Relatedly, expertise has consistently been found to be a key predictor of credibility (e.g., Slater & Rouner, 1996). In a meta-analysis of 20 studies, which explored the effect of expertise or trustworthiness on perceived credibility of online health information, expertise was found to account for a greater proportion of variance in credibility judgements than trustworthiness (Yang & Beatty, 2016). This was suggested to result from the use of 'authority heuristics', which are more available for expertise than trustworthiness. For instance, it may be easier to instantly tell if someone is knowledgeable or not (as there are more cues available, e.g., qualifications, years of experience) whereas trustworthiness usually requires time and interpersonal interaction to develop. The effect of expertise may be context dependent, in that some contexts require more specialised knowledge than others, for instance with new and technological hazards (Savadori et al., 2004).

Importantly, the presence of an expert source is not enough to ensure effective uncertainty communication. Whilst ensuring the information conveyed is accurate is undoubtedly important, it is unlikely information will be believed and/or acted upon if the source is perceived as possessing expertise, but untrustworthy (Berry, 2004; Frewer et al., 1999). The challenge for communicators is to therefore be able to demonstrate both attributes, though this is particularly problematic within a natural hazards context. Credibility has been shown to be vulnerable to the quality of information, such as its accuracy and whether it has previously been proved wrong (Frewer, Howard, Hedderley, & Shepherd, 1996; Lucassen & Schraagen, 2011). Yet given the many uncertainties associated with natural hazards (Sword-Daniels et al., 2018), complete accuracy is not possible all of the time. Furthermore, if the presence of uncertainty is perceived as an indicator that the communicator is incompetent or lacking the necessary knowledge or expertise (Johnson & Slovic, 1995, 1996), it is clear to see how this could prove deleterious to perceptions of a communicator's credibility. It is therefore important to pay more attention to the role of credibility in risk and uncertainty communications (Löfstedt, 2003) – a key motivation for the current research. As indicated previously, credibility is conceptualised within this thesis as trustworthiness (i.e., relating to the integrity and reliability of the communicator, Dieckmann et al., 2009; Walton, 1997) and expertise (i.e., relating to the communicator's knowledge, Guilamo-Ramos et al., 2006).

Chapter Summary

This chapter has argued that effective communication of risk and uncertainty information relies on two components. The communicator must be perceived as credible source of information and the information must be understood as the communicator intended. Taking three of the most common communication formats for presenting uncertainty information (verbal, numerical and mixed), this chapter has considered existing research which has investigated how these formats are understood. In the remaining section, this chapter has considered the concept of credibility and its two components: trustworthiness and expertise, as well as their effects. The chapter concludes by highlighting the challenge of demonstrating credibility within a natural hazards context.

Thesis Overview

Chapter 2 extends the research on how mixed formats are understood firstly by using the relatively new 'which-outcome' paradigm, and secondly by considering a previously untested N-V format. The replication of the 'extremity effect', and its extension to a V-N format lends further support to the notion that VPEs fulfil a range of pragmatic functions, such as conveying a communicator's expectations about what will happen. Chapter 2 concludes by considering the implications of these findings for a communicator's credibility, the premise of which forms the focus of Chapter 3. Here, the effect of communication format on a communicator's perceived credibility is investigated, firstly on the basis of an initial prediction, and then following a probabilistically unexpected outcome (i.e., an 'unlikely' event occurs, or a 'likely' event does not occur). Differences in the effect of communication format are observed in the high and low probability domain, which could be as a result of probability, or as a result of the directionality of the VPEs featured. Chapter 4 therefore aims to disentangle the effects of directionality and probability, in order to reach an explanation of what is driving the effect of communication format on perceived credibility, and finds support for a 'directionality–outcome congruence' account. The findings and conclusions are brought together in Chapter 5, where a pragmatic account of communication format is proposed. Implications of our findings for the communication of uncertainty within a natural hazards context are considered, before considering ideas for future research.

Chapter 2. The Robustness of the 'Extremity Effect'

Chapter Overview

As already highlighted in Chapter 1, there has been much research investigating how verbal probability expressions (VPEs) are understood, which has used a variety of paradigms. Despite the differences in methodology, the research is consistent in its overall conclusions – there is much (inter-individual) variability in the way that VPEs are used and interpreted. Most recently, research using the 'which-outcome' paradigm has demonstrated that use of VPEs can lead to an 'extremity effect' – that is, when asked to indicate an outcome that is 'unlikely', participants typically indicate extreme outcomes, often with a value exceeding the maximum value shown, equivalent to a 0% probability (e.g., Teigen, Juanchich, & Riege, 2013). It has been noted that a mixed format can help reduce misunderstandings associated with use of VPEs (e.g., Budescu et al., 2009, 2012, 2014). This chapter therefore addresses the question of whether the 'extremity effect' still occurs given use of a mixed, verbal-numerical format (e.g., 'unlikely [20% likelihood]'), as well as a previously unconsidered numerical-verbal format (e.g., '20% likelihood [unlikely]'). We further examine how robust the effect is in the context of consequential outcomes and over non-normal distributions. We also investigate whether participants are aware of the inconsistency in their responses from a traditional 'how likely' task, and a 'which-outcome' task. We replicate and extend previous findings, with the 'extremity effect' observed in both verbal and verbal-numerical formats. These findings suggest that there could be important, downstream effects for the perceived credibility of a communicator, if a communicator uses 'unlikely' to mean 20% and the audience expects 'unlikely' to refer to outcomes which will never happen.

Introduction

Often, people may lack in-depth knowledge and experience of natural hazards, which means they are reliant on mediated information, typically from an expert source (Paton, 2008; Sjöberg, 2000). Experts are thus presented with the challenge of how to effectively communicate information about uncertainty. They must ensure not just that the information is understood correctly, but also that the information is perceived as coming from a credible and reliable source. The extent to which a communicator achieves one or both aims will have implications for how (and even whether) the recipient responds following the communication.

Challengingly for scientists working within Earth Science, information relating to natural hazards is typically characterised by large amounts of uncertainty (Doyle et al., 2019; Sword-Daniels et al., 2018), meaning forecasts are typically probabilistic (at best). It is not possible to predict with certainty whether a destructive earthquake will occur in a certain place within the next month, for example. A prediction that such an event is 'unlikely' does not imply that the event will not occur. Given that an estimate of 'unlikely' might be used to describe the probability of events with a 20% likelihood of occurrence (e.g., Theil, 2002), approximately 20% of the time such events will occur. Problematically, it is well established that such uncertainty is not always well understood, nor well received by the public (e.g., Frewer et al., 2003; Johnson & Slovic, 1995). The majority of the public desire certainty and simply want to know whether something will happen or not, or is safe or not (Johnson, 2003; Johnson & Slovic, 1998). As a result, there is clear potential for uncertainty to lead to reduced levels of trust in scientists who make such predictions, and, in the extreme case of L'Aquila, legal liability (Benessia & De Marchi, 2017).

Communication format.

Budescu and Wallsten (1995) proposed that the choice of format for communicating probability information should be governed by the congruence principle: the precision of the communication should be consistent with the degree of certainty that can reasonably be expected for estimates about the event described. In the domain of natural hazards, estimations of events, such as the probability of a large earthquake, might not be precisely quantifiable. In such instances, a specific numerical expression of the probability of this event might be perceived as overly precise. Using a VPE would seem to better represent the uncertainty and underlying imprecision associated with the probability estimate, as they refer to a range of values (e.g., Budescu & Wallsten, 1995). VPEs are also thought to be easier to understand and more natural for individuals to produce (Budescu & Wallsten, 1987; Erev & Cohen, 1990). There is, however, considerable variability in people's usage and interpretations of VPEs (e.g., Budescu & Wallsten, 1985). In addition to 'natural' inter-individual variability (Beyth-Marom, 1982), interpretations of VPEs are susceptible to contextual and cultural influences (Bonnefon & Villejoubert, 2006; Fischer & Jungermann, 1996; Harris & Corner, 2011; Harris et al., 2013; Juanchich et al., 2012; Teigen & Brun, 1999, 2003; Wallsten, Fillenbaum, et al., 1986; Weber & Hilton, 1990). Despite this variability, VPEs continue to be commonly used in a range of domains, including accounting (Doupnik & Richter, 2003), intelligence analysis (Ho et al., 2016) and climate change (Mastrandrea et al., 2010).

Studies investigating interpretations of VPEs have typically used the 'how likely' translation approach, whereby people are asked to translate VPEs to corresponding numerical probabilities. However, more recently, Teigen and colleagues have demonstrated that a 'which-outcome' (W-O) approach to understanding people's interpretations of VPEs paints rather a different picture (Juanchich et al., 2013; Løhre & Teigen, 2014; Teigen & Filkuková, 2013; Teigen et al., 2014, 2013).[13] In this approach, participants are shown a graph (typically a histogram) depicting a distribution of outcomes and asked to complete probability statements (e.g., "It is *unlikely* that a battery will last __ hours") with a value considered appropriate (see Figure 2 for an example histogram). This approach has highlighted the potential for a large qualitative disparity between the probability statement's meaning intended by the communicator and that which is expected/understood by the

[13] See also Juanchich and Sirota (2017); Teigen, Andersen, Alnes and Hesselberg (2019) and Teigen, Filkuková and Hohle (2018) for further research using the 'which-outcome' paradigm published since the current research was carried out.

recipient of the information. Specifically, this research has demonstrated that when the term 'unlikely' was used to describe outcomes such as battery life, or weight loss, participants interpreted the term as referring to outcomes from the high end of the distribution (Teigen et al., 2013). Most often, participants completed the sentence with a value that *exceeded* the maximum shown in the graph – hereafter the 'extremity effect'. This was despite a mean translation of 'unlikely' as 40% in a pre-test.

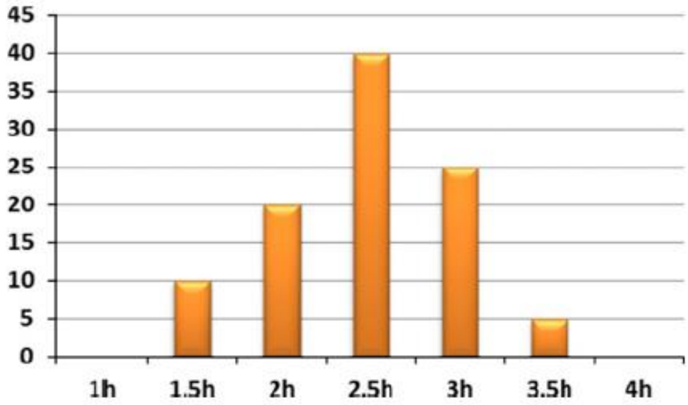


Figure 2. Battery life histogram featured in Teigen et al., (2013).

Findings from the W-O methodology are of importance for our understanding of how people understand risk communications, especially given many real-world, natural hazards concern continuous outcomes, such as the extent of coastal erosion or lava flows. Ultimately, the number which a participant assigns to a VPE (e.g., in a translation task) is not of great import. The critical element is how people conceptualise these expressions and, ultimately, use them to guide their decision-making. Whilst the mechanisms behind the results from alternative methodologies (such as the W-O task) have yet to be identified, this chapter's focus is on extending the communication formats used within this paradigm, given the W-O task can provide additional information over and above typical 'how likely' translation tasks. This information should be heeded for a full understanding of people's conceptualisations of VPEs, which is especially relevant when making recommendations for effective communication formats. It is therefore worrying that the 'extremity effect' is considerably out of line with the prescribed usage of the term ‘unlikely’ in communications of uncertainty. For instance, the UK’s Defence Intelligence developed a six category translation table in which ‘unlikely’ was translated as 15–20% (as cited in Ho, Budescu, Dhami, & Mandel, 2016; see Table 2, Chapter 1 for further examples).

On the one hand, the *use* of ‘unlikely’ to communicate outcomes with a 0% likelihood of occurrence could be argued to be exaggerating the risk, and an appropriate strategy to minimise losses (e.g., by encouraging preventative action) where those associated with an underestimate of the risk are greater than an overestimate (e.g., Batchelor & Peel, 1998; Clatworthy, Peel, & Pope, 2012; Granger, 1969; Harris, Corner, & Hahn, 2009; Lawrence & O’Connor, 2005; Weber, 1994). We propose, however, that the W-O task is informative with respect to how VPEs will be *understood,* and hence acted upon, by recipients of a risk communication. From this perspective, it is easy to foresee how the ‘extremity effect’ could prove deleterious to effective risk communication. If phrases such as ‘unlikely’ are seen as most appropriate for communicating outcomes with no chance of occurring, the mismatch between this and an intended communication of ‘20% likelihood’ could adversely affect confidence in subsequent communications. More immediately, Teigen and colleagues’ findings (Juanchich et al., 2013; Løhre & Teigen, 2014; Teigen & Filkuková, 2013; Teigen et al., 2014, 2013) suggest there could be extreme consequences for citizens who, upon hearing that the chance of a volcanic lava flow extending as far as their village is ‘unlikely’, may potentially discount the information, believing it will not happen, and thus choose not to evacuate their homes.

The possibility of catastrophic consequences, however, relies on the assumption that the ‘extremity effect’ will be obtained even when one potential outcome is of particular consequence, and thus salient. This is especially relevant given that consequential communications about natural hazards will, by definition, refer to consequential outcomes. However, previous research using the ‘how likely’ methodology suggests that such an assumption might not necessarily hold, as people's interpretations of VPEs are higher when those VPEs describe a severe outcome than a neutral outcome (e.g., Harris & Corner, 2011; Weber & Hilton, 1990). More generally, making one outcome particularly consequential in the W-O methodology will enhance its salience, a characteristic present in all real-world risk communications. It is possible that increased saliency of a location (or even multiple locations) could lead one to assume all communications are relevant to that particular location. For example, when considering the potential extent of a volcanic lava flow, the location of a school a certain distance from the volcano might consume the attention of a communicator, such that all communications are deemed to be relevant to the school, in which case the effects reported in Teigen, Juanchich and Filkuková (2014) (see also Juanchich & Sirota, 2013; Løhre & Teigen, 2014; Teigen & Filkuková, 2013; Teigen et al., 2013) may not occur.

Improving risk and uncertainty communications.

One commonly proposed solution to reduce mis-communications (from researchers employing the 'how likely' translation methodology) is the use of a mixed format approach to express uncertainty, for example, 'It is unlikely (less than 33%)' (Beyth-Marom, 1982; Budescu et al., 2009, 2012, 2014; Harris et al., 2017; Patt & Dessai, 2005; Renooij & Witteman, 1999; C. L. M. Witteman & Renooij, 2003; C. L. M. Witteman et al., 2007). Budescu and colleagues have demonstrated that such a 'verbal-numerical' format increased the correspondence between people's interpretations and the IPCC guidelines, an effect that was replicated in 24 countries (Budescu et al., 2014). A question which arises from the W-O methodology and has, of yet, received little attention (but see Juanchich & Sirota, 2017) is whether the addition of a VPE could actually *harm* the effectiveness of communications, in comparison with those that use only numerical expressions. The 'extremity effect' is one example of how this harm could occur. It is thus imperative to ascertain if the 'extremity effect' is observed with mixed formats and consequential scenarios, given the potential for negative consequences arising from such communications.

The low probability domain was of particular interest for study because natural hazards typically lead to highly consequential outcomes (which are usually unlikely e.g., Knoblauch, Stauffacher, & Trutnevyte, 2018). Additionally, negatively directional expressions such as 'unlikely' have been found to induce a large range of interpretations (Smithson, Budescu, Broomell, & Por, 2012).[14]

Overview of experiments.

This chapter therefore aims to further our understanding of the ramifications of 'extremity effect' work of Teigen and colleagues (Juanchich et al., 2013; Løhre & Teigen, 2014; Teigen & Filkuková, 2013; Teigen et al., 2014, 2013). Specifically, the present research compares five communication formats: VPEs (e.g., 'unlikely'); numerical expressions, (point, e.g., '20% likelihood' and range formats e.g., '10–30% likelihood') and mixed expressions in two orders (verbal-numerical, e.g., 'unlikely [20% likelihood]', and the previously unstudied numerical-verbal format, e.g., '20% likelihood [unlikely]'). We examine how robust the 'extremity effect' is to consequential outcomes (Experiments 1, 2, 4–6), over distributions other than the commonly used normal (bell-shaped) distribution (Experiment 3) and following another (different) probability estimation task (a translation task, Experiment 6). We also explore the potential influence of numeracy. Examining the effects of using different communication formats is instructive for designing effective future instruments for the communication of risk and uncertainty.

Experiment 1

The 'extremity effect' was initially tested in non-consequential domains such as battery life and mailing letters (Teigen et al., 2013).[15] We thus sought to use scenarios featuring

[14] See also more recent work by Wintle et al. (2019).
[15] Though see Juanchich & Sirota (2017) for research which later addressed this.

outcomes differing in consequence, in order to examine whether the 'extremity effect' would still occur even when potential outcomes were of particular consequence and thus salient. We manipulated the location of the salient site(s) in Experiment 1.

Method.

Participants.

One hundred and fifty-five participants were recruited for this online experiment via Prolific Academic (www.prolific.ac) and were paid £0.85. Eight participants were excluded (six due to duplicate IP addresses and two due to lack of consent) leaving a final sample of 147 (83 male) participants, aged 18–60 years (*Mdn*= 27). Ethical approval was granted from the Departmental Ethics Chair for Speech, Hearing and Phonetic Sciences (University College London).[16]

Design.

Communication format (verbal – 'unlikely'; numerical – '20% chance' and mixed – 'unlikely [20% chance]') was manipulated between-participants. 20% was a plausible value for 'unlikely', given the IPCC's likelihood scale (which suggests using 'unlikely' to communicate probabilities of between 0 and 33%). 20% was also the average numerical translation of 'unlikely' in Theil's (2002) meta-analysis. Numerical point estimates were used rather than range estimates in order to maximise differences between conditions. Scenario (volcano; flood; earthquake; landslide) and saliency (no salient site; close site; far site; multiple sites) were manipulated within-participants. Scenario and saliency were randomised using the Latin Square Confounded method (Kirk, 1969), such that each participant only saw each scenario and each saliency once, but the combinations of these differed systematically across participants.

[16] As was the case for all experiments within this thesis.

Materials.

The introductory text informed participants that they would see reliable projections of a model designed to predict future geological events and be asked to make a series of judgements about these. All participants read four vignettes (developed in conjunction with geologists at the British Geological Survey [BGS] to ensure they reflected plausible real-world situations) describing outcomes of how far lava flows, floodwater, earthquake tremors and debris flows would extend (see Figure 3 for an example and Appendix A for all of the vignettes). Each vignette was illustrated by a histogram which showed the frequency of model-outcomes in each of ten interval bins. The shapes of the distributions were similar and approximately normal across the scenarios, though the distributions in the volcano and floods scenarios had a slightly negative skew. The *y*-axes deliberately featured no values, as we were interested in people's understanding of risk and uncertainty communications, rather than their responses to what they might otherwise have perceived as a mathematical problem. Participants were required to type a numerical response which corresponded to the outcome that was being described.

Saliency was manipulated through the inclusion of sites of particular scientific interest, to which the impact of the natural hazard might extend. These sites were either home to rare plants or critically-endangered animal species (e.g., the last habitat of 'white-spotted Antis' in Figure 3). There were four saliency conditions: no site of interest; one close site (located in the second bin of the histogram); one far site (last bin of the histogram) or multiple sites (second bin, modal bin and last bin, see Figure 3).

Procedure.

The experiment was run using Qualtrics.[17] Participants were first informed about the nature of the experiment and told they could withdraw at any time during the experiment. After consenting to participate, they were asked to indicate their age and gender, before reading

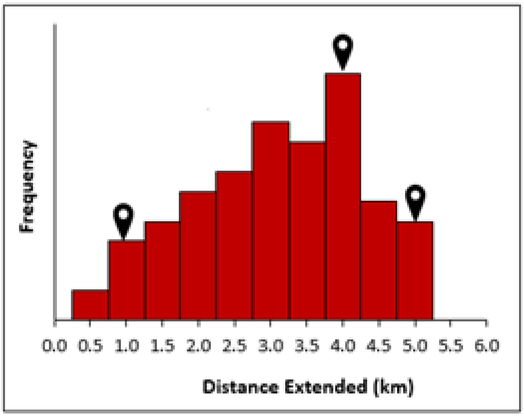
[17] All of the experiments in the thesis were run using Qualtrics (www.qualtrics.com).

the introductory text. The next four screens contained the four vignettes and judgement tasks. Upon completion, participants were given a code to claim their payment, thanked and debriefed.

Reminder: The number of times the model has produced each outcome is a reliable indication of how likely that particular outcome is.

Mount Ablon has a history of explosive eruptions forming lava flows. An eruption has been predicted; the figure below shows the model's predictions of the distance extended by lava flows for this eruption, given the volcano's situation and recent scientific observations.

Due to the highly fertile soil and rich vegetation, multiple sites of special scientific interest home to the critically endangered 'white-spotted Antis', exist in the area surrounding the volcano. Sites A, B and C lie 1km, 4km and 5km respectively away from the volcano (shown below). If lava flows reach any of these sites, the last surviving populations of 'white-spotted Antis' in the wild (at the site) would be lost.



Complete the sentence below with a number that seems appropriate in this context.

In the event of an eruption, it is unlikely (20% chance) that the lava flow will extend to a distance of __ km.

Figure 3. Experiment 1: Example vignette (volcano scenario, multiple salient sites, mixed format).

Results.

Effect of saliency.

Outcomes on the *x*-axes were standardised across scenarios by 'binning' responses in all scenarios in relation to the salient points, as if they were in the multiple site condition (1 =

below minimum, 2 = minimum, 3 = low saliency point, 4 = between low and mid saliency points, 5 = middle saliency point, 6 = between mid and high saliency points, 7 = high saliency point [maximum] and 8 = above maximum; see Appendix B). Given the expected saliency × communication format interaction, an ANOVA was not appropriate because of the Latin Square Confounded Design (Kirk, 1969) and therefore three Kruskal-Wallis tests were performed. These showed that responses were not significantly affected by saliency in either the verbal, χ^2 (3) = 0.53, p = .92, numerical, χ^2 (3) = 4.00, p = .26 or mixed format conditions, χ^2 (3) = 3.05, p = .39. Whilst we randomised scenario, we also checked for an effect of scenario using three Kruskal-Wallis tests, which showed responses were not affected by scenario in either the verbal, χ^2 (3) = 1.56, p = .67, numerical, χ^2 (3) = 5.67, p = .13 or mixed format conditions, χ^2 (3) = 1.33, p = .72.

Effect of communication format.

Given the non-significant effect of saliency, we code responses in relation to the bars shown in the histogram (ignoring the salient sites) in all subsequent analyses. For instance, the first bin contains all responses reflecting outcomes to the left of the first histogram bar, the second bin containing all responses reflecting outcomes included in the first histogram bar, and so on (see Figure 3). With ten bars in each histogram and the additional minimum/maximum bins at either end, there were 12 bins in total.

Typical outcomes for 'unlikely' were chosen from the higher end of the distribution, from maximum and above maximum values (bins 11 and 12) – high amplitude outcomes. In contrast, typical outcomes for '20% chance' tended to correspond to lower values, primarily chosen from the intermediate outcomes.[18] Results for the mixed format fell between those observed with the verbal and numerical formats; outcomes tended to be chosen from the intermediate outcomes, but this did not preclude a sizeable proportion (45.1%) choosing

[18] Effects of communication format were unchanged if responses were binned into five categories (below minimum, minimum, intermediate, maximum and above maximum), as in Teigen et al. (2013).

high amplitude outcomes. The contrasting patterns of responses are clearly evidenced in Figure 4.

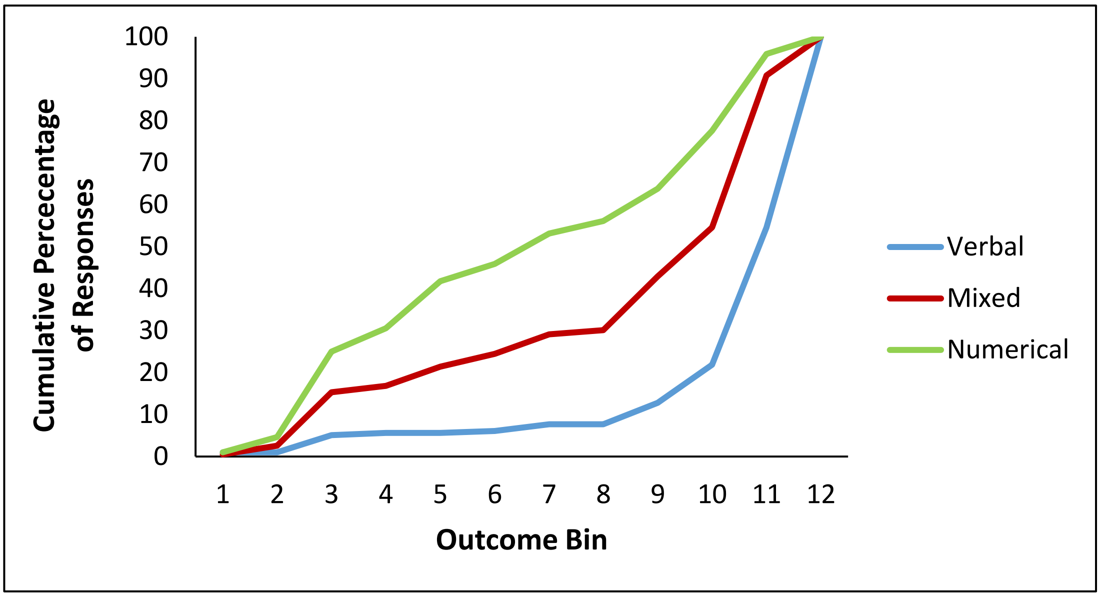


Figure 4. Experiment 1: Cumulative distribution of responses by communication format.

To enable comparison with Teigen et al. (2013), responses were coded according to whether they indicated high amplitude outcomes (bins 11 or 12) or not. The proportion of responses indicating high amplitude outcomes was highest in the verbal condition (replicating the 'extremity effect'), followed by the mixed format condition. The numerical condition had the lowest proportion of responses indicating high amplitude outcomes (see Figure 5).[19]

Further differences between the numerical and mixed formats are observed when one considers that the '20%' in the numerical and mixed communication formats enables the calculation of an objectively correct answer to the statement "there is a 20% chance that the *x* will extend to a distance of __" for the four scenarios. Using data the histograms were created with, the correct answer was defined cumulatively – representing the forecasted outcome where 20% of forecasted outcomes were that distance or higher. This calculation

[19] The same format differences are observed even if only 'above maximum' responses are included in analyses. This is also the case for Experiments 2–6.

of the correct answer implies an 'at least' reading of the sentence that is to be completed by participants: 'there is a 20% chance the *x* will extend *at least* ___ km', rather than '… *exactly* ___ km'.[20] Whilst some might intuit an 'exact' reading of the sentence, searching for an outcome which occurred on *exactly* 20% of occasions, mathematically this is a less appropriate interpretation. The probability of a continuous random variable (such as the distance of a lava flow) taking a single specified value is (strictly) zero. An 'exact' reading could be justified were 5km (for example) given to mean a range from (for example) 4.75 to 5.24km. We maintain, though, that the 'at least' reading is more appropriate, both for the reasons above and because if a lava-flow does extend 5 km from the volcano, sites at any distance up to 5km along the path of the flow are all affected. We conjecture that these reasons serve as 'background implicature' (Breheny, 2008), which leads one to make an 'at least' interpretation. The correct answer fell in the 7th bin in three of the four scenarios (for the earthquake scenario it fell in the 8th bin). For these three scenarios, in the numerical condition responses were fairly evenly distributed above and below the correct response (46.9% and 49% respectively), whilst responses in the mixed format condition were more skewed, with 72.8% of responses above the correct response and only 25.2% below.[21] This pattern was also observed in the earthquake scenario. A consideration of responses in the verbal condition was not appropriate here, given the lack of an objective correct response.

[20] There are various linguistic accounts of how numeric quantifiers are interpreted. One neo-Gricean account suggests that numeric quantifiers a unilateral, 'at least' interpretations (Levinson, 2000, see also Musolino, 2004). Breheny (2008) reviews arguments against a unilateral account of numerical quantifiers, and suggests a bilateral account, where unilateral interpretations may occur as a result of pragmatic influences.

[21] The reader might question why a more quantitative comparison of the accuracy of different formats is not included. We see 'above maximum' responses as qualitative errors which do not reflect the primary intended meaning of a communication of 'unlikely'. Given that the correct answer lies towards the upper end of the scale, such errors were not too far from the correct answer. Despite, we argue, being qualitatively inappropriate, these responses (most plentiful in the V-N condition) would be coded as more accurate than some solely quantitative errors in the opposite direction. Further quantitative analysis of accuracy could thus be misleading for purely statistical reasons.

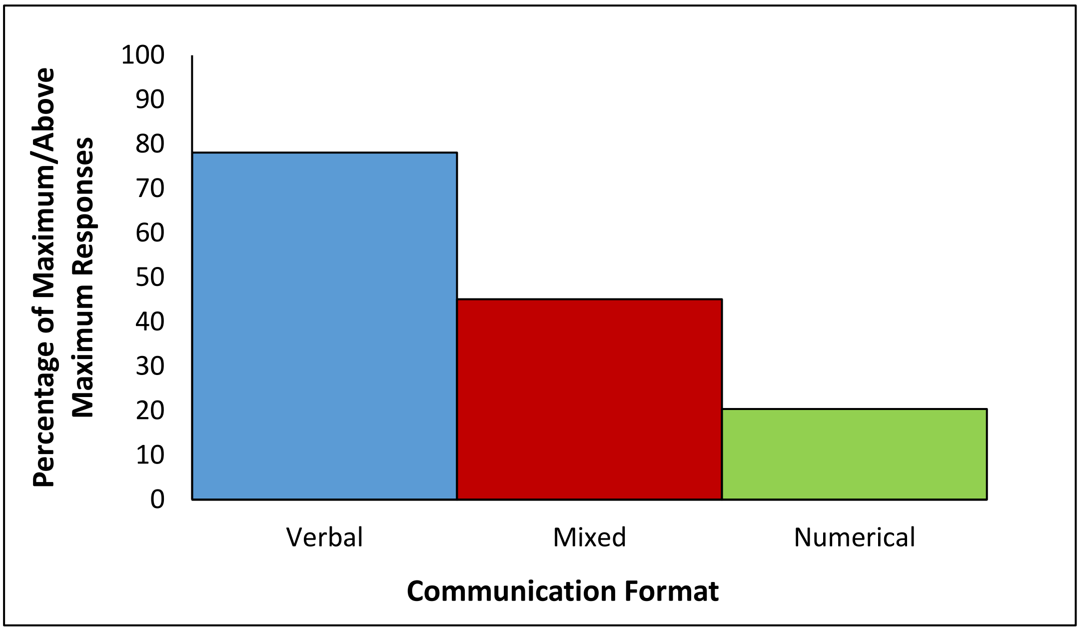


Figure 5. Experiment 1: Percentages of maximum and above maximum responses by communication format.

Experiment 2

Although the numerical format was shown to be effective at reducing the 'extremity effect' (proportion of maximum/above maximum responses), it is conceivable that the effectiveness of communication formats could vary across individuals. Less numerate people tend to rely on non-numerical information and have been shown to be more vulnerable to extraneous factors, such as framing, and the format in which the information is presented (E. M. Peters et al., 2006; Reyna et al., 2010). Furthermore, lower numeracy is associated with increased risk and likelihood perceptions (Dieckmann et al., 2009), and its influence has been shown to prevail even in consequential situations (Lipkus, Peters, Kimmick, Liotcheva, & Marcom, 2010). The influence of numeracy is of particular relevance when considering the effectiveness of using a mixed format, given such formats allow one to focus on the expression that is most easy to use (Dieckmann et al., 2012; Gregory et al., 2012; C. L. M. Witteman et al., 2007). For instance, more numerate people might focus on the numerical expression, whereas less numerate might choose to focus on the VPE.

We therefore repeated Experiment 1 in a controlled laboratory session, with the addition of a numeracy measure.

Method.

Participants.

Eighty-three participants were recruited for this online experiment via a University College London (UCL) first-year undergraduate ‘Introduction to Psychological Experimentation’ class and completed the experiment for course credit. Two participants were excluded due to skipping the consent question, leaving a final sample of 81 (15 male) participants, aged 18–20 years (*Mdn* = 19).

Design, procedure and materials.

Design, procedure and materials were as in Experiment 1, though included the addition of the Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), a series of four questions designed to test numeracy, which was the final task in this experiment. Participants were then thanked and debriefed.

Results.

Effect of saliency.

Three Kruskal-Wallis tests were run to investigate if there was an influence of saliency in (a) verbal (b) numerical and (c) mixed format conditions. These showed that responses were not significantly affected by saliency in either the verbal, χ^2 (3) = 2.22, p = .53, numerical, χ^2 (3) = 3.53, p = .32 or mixed format conditions, χ^2 (3) = 0.55, p = .91. Again, we checked for an effect of scenario. Responses were not significantly affected by scenario in either the verbal, χ^2 (3) = 0.42, p = .94, numerical, χ^2 (3) = 1.47, p = .69 or mixed format conditions χ^2 (3) = 0.14, p = .99. In the following analyses, we therefore coded the data as in Experiment 1.

Effect of communication format.

The proportion of responses indicating high amplitude outcomes (the maximum/above maximum value present in the histogram) was highest in the verbal condition (60.2%), followed by the mixed format condition (38%). The numerical condition had the lowest proportion of responses (14.7%) indicating high amplitude outcomes. The distribution of responses followed a very similar pattern to those in Experiment 1 (see Figure 6).

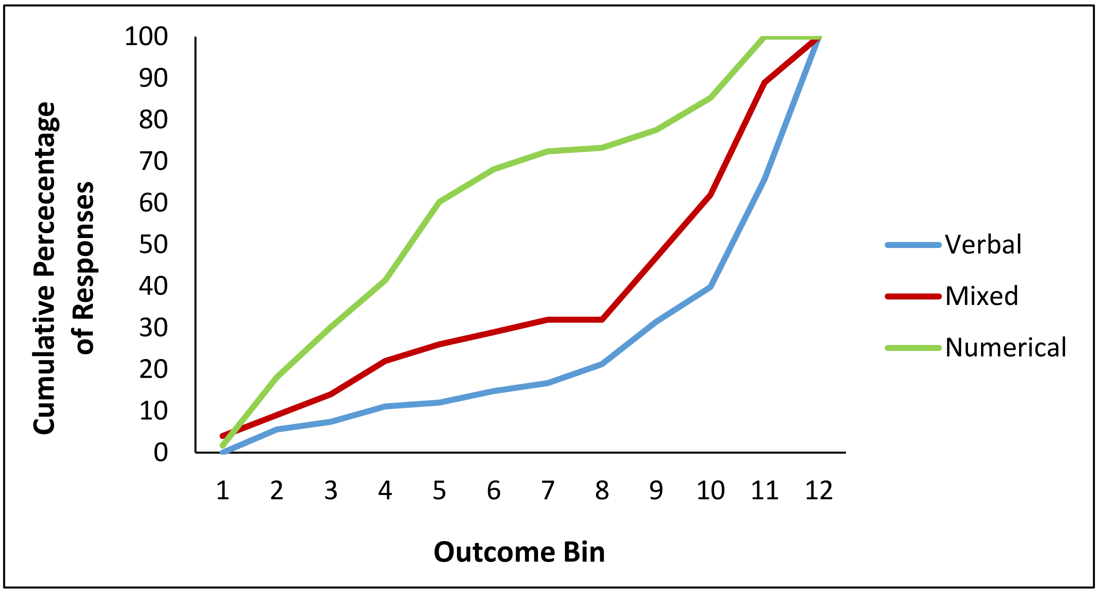


Figure 6. Experiment 2: Cumulative distribution of responses by communication format.

As in Experiment 1, for the scenarios in which the correct answer fell in the 7th bin, responses in the mixed format condition were skewed above the correct answer (69.3%), with 30.7% below. The opposite pattern of responses was found in the numerical condition, in which 28.7% of responses were skewed above the correct answer, with 67.8% below, differing slightly to the even distribution observed in Experiment 1. A similar pattern of results was found for the earthquake scenario.

Effect of numeracy.

Answers for each question were coded as 1 if correct and 0 if incorrect, such that numeracy scores could range from 0 to 4. The distribution of numeracy scores is shown in Table 3.

Given that we observed no effects of saliency or scenario, we averaged the four responses provided by each participant. A 4 (communication format) × 2 (numeracy – high/low) ANOVA revealed no significant effect of numeracy, $F(1, 75) = 0.52$, $p = .47$, nor a significant interaction between communication format and numeracy, $F(2, 75) = 0.92$, $p = .40$. We finally note that there was a reduction in above maximum responses as numeracy increased, reducing from 42.1% of responses from participants with numeracy scores of 0 or 1, to 15.8% for those with scores of 3 or 4.

Table 3. Experiments 2 and 3: Distribution of Numeracy Scores (%) Using the Berlin Numeracy Test (Cokely et al., 2012)

	Numeracy Classification (%)				
	Low		High		
Numeracy Score	0	1	2	3	4
Exp 2	18.5	28.4	34.6	16.0	2.5
Exp 3	32.9	30.1	18.2	12.1	6.7

Note. Responses were divided into low and high numeracy (as specified) given the uneven distribution of scores.

Discussion.

We replicated Teigen et al.'s (2013) results and tested whether these would hold for numerical and mixed format expressions of probability. We found evidence that the tendency to describe outcomes at the very end, or beyond the range of, a distribution generalised to consequential scenarios. The 'extremity effect' occurred wherever the word 'unlikely' was included – the verbal and mixed format conditions. This tendency was not apparent in the numerical condition. We found no effect of including consequential outcomes, that is to say that the presence of a salient site(s) did not affect responses, so this variable was not included in the following experiments.

Whilst the results from Experiment 2 replicate our main findings from Experiment 1, it is interesting to note the difference in distribution of responses in the numerical condition relative to the correct answer. Whilst responses in the numerical condition in Experiment 1 were evenly split above and below the correct answer, Experiment 2 showed a tendency for more responses being provided below it. A potential to err towards lower estimates might arise because our numerical condition was not purely numerical – '20%' was accompanied by the positive directional term 'chance'. Teigen and colleagues have demonstrated that VPEs can be distinguished in terms of their directionality (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 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 attention on the occurrence of the event. As an expression with positive directionality (Teigen & Brun, 2003), the term 'chance' could thus have led participants in the numerical condition to provide estimates closer to the likely end of the scale (the left). In contrast, those in the mixed format condition may have seen the '20% chance' in parentheses, considered it non-essential information (J. Walker, 1823), discounted it, and focused on 'unlikely'.

The parenthesis account (J. Walker, 1823) suggests that the order of the mixed format expression might influence estimates, depending on what information is contained within the parentheses. An order effect would also be in line with Grice's (1975) conversational maxims. The 'maxim of manner' prescribes that utterances should be 'orderly', such that more important information is presented before less important information. Whilst existing research has explored whether a mixed format increases understanding (Budescu et al., 2009, 2012, 2014; C. L. M. Witteman & Renooij, 2003), it has not considered the order of the mixed format expression, using only V-N expressions (e.g., 'unlikely [20% likelihood]'), as opposed to N-V expressions (e.g., '20% likelihood [unlikely]'). We therefore include this additional communication format in subsequent experiments.

It is also possible that people who focused more on the term ‘unlikely’ in the mixed expression had lower numeracy levels and felt uncomfortable using the ‘20%’ to form their estimates, though we found no evidence to support this assertion in Experiment 2. Nonetheless, our results provide some suggestion numeracy is having some influence on responses, in that a greater proportion of less numerate participants demonstrated the ‘extremity effect’.

Experiment 3

Not all outcomes for which probabilities must be communicated follow a normal distribution. Examining the ‘extremity effect’ in relation to different distributions can tell us more about the mechanisms behind the effect, for instance whether it is driven by extremeness, or by the frequency of the outcomes. As the name suggests, the effect has been proposed to reflect an extremity preference, with a preference for values from the higher end of the distribution driven by the potential usefulness of being informed about extreme outcomes, irrespective of actual probability (Juanchich et al., 2013). Juanchich et al. (2013) observed the ‘extremity effect’ for ‘possible’ outcomes, across positive skew, negative skew and bi-modal distributions.

Experiment 3 was designed to extend the aforementioned research by examining whether the order of the mixed format expression influenced estimates over varying distributions. It also provided the opportunity to check the generalisability of the results for the numerical and mixed format (V-N) conditions. Setting this experiment in a natural hazards context was not appropriate here, given negative skew distributions are uncommon (and less plausible) for natural hazards. We therefore used an altered version of Teigen et al.’s (2013) original battery life scenario where different shaped distributions were entirely plausible. Additionally, we wanted to investigate participants’ understanding of the W-O sentence – are participants completing the sentence with a figure that represents an ‘at least’ or ‘exactly’ interpretation?

Method.

Participants.

Seven hundred and fifty-one participants were recruited for this online experiment from Amazon Mechanical Turk via the TurkPrime platform (Litman, Robinson, & Abberbock, 2017) and were paid $0.30. Fifty-four cases were removed (due to duplicate IP addresses or for failing the attention check), leaving a final sample of 697 (364 male) participants, aged between 18–84 (*Mdn* = 30).

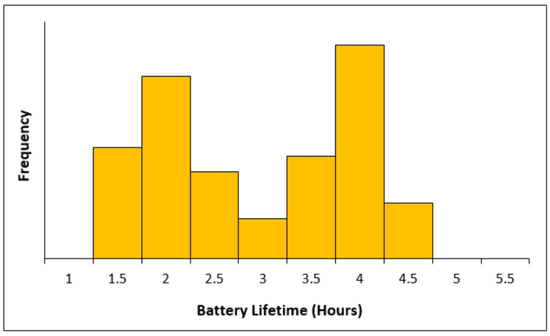
Design.

Communication format (verbal – 'unlikely'; numerical – '20% likelihood'; V-N – 'unlikely [20% likelihood]'; N-V – '20% likelihood [unlikely]') and distribution (positive skew; negative skew; bi-modal; normal) were manipulated in a 4 × 4 between-participants design. As per the previous experiments, participants were required to type a numerical response which corresponded to the outcome being described.

Materials.

The battery life scenario from Teigen et al. (2013, Experiment 3) was used for this experiment. The introductory text informed participants that they would read a short vignette and be asked to make a judgement about it. The text of the vignette was identical to that used originally, though the accompanying histograms illustrating how long the previously tested batteries lasted included seven observed durations of battery life (in comparison to the five in Teigen et al.'s [2013] experiment). This facilitated the manipulation of the distribution (see Appendix C for all histograms). A sentence completion task was presented at the bottom of the vignette (see Figure 7 for an example).

A sample of computer batteries of the brand ''Comfor'' were tested to check how long the batteries last before they need to be recharged. All computers were used by students for lecture notes and similar purposes. The figure below shows how many batteries lasted how many hours (duration is rounded to the nearest half hour).



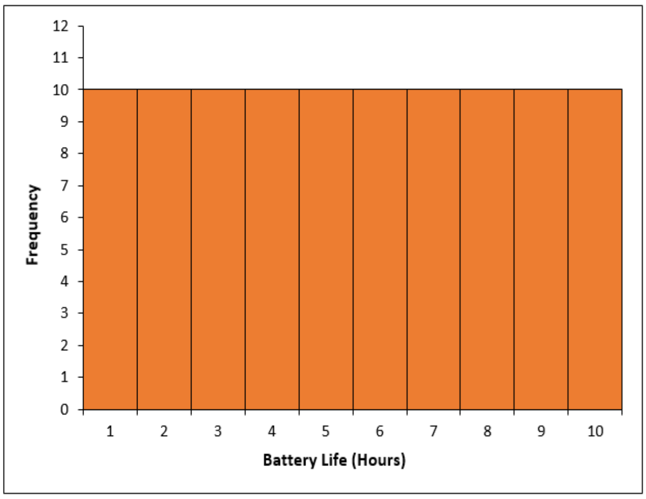
Complete the sentence below with a number that seems appropriate in this context.

It is **unlikely (20% likelihood)** that the battery in a Comfor computer will last for: ________ hours.

Figure 7. Experiment 3: Computer battery vignette, V-N format, bi-modal distribution.

After completing the battery task, participants were asked to complete the Berlin Numeracy Test (Cokely et al., 2012). Participants then saw a final histogram, this time of the lifetime of 'Powerplus' batteries, in which 10 batteries lasted one hour, 10 batteries lasted two hours, and so on up to 10 hours (see Figure 8). Participants were then asked which of the following statements was 'most' correct: "There is a 10% likelihood the battery will last for one hour" versus "There is a 100% likelihood the battery will last for one hour" and why. This task was included in order to establish how participants interpret the W-O sentence. The 10% response denotes an 'exactly' interpretation, whilst the 100% response denotes an 'at least' interpretation.

A sample of computer batteries of the brand ‘‘Powerplus’’ were tested to check how long they lasted before they need to be recharged. The figure below shows how many batteries lasted how many hours (duration is rounded to the nearest half hour).



Which of these two statements do you think is more correct? Please note there is no correct answer.

- There is a 10% likelihood the battery will last for one hour
- There is a 100% likelihood the battery will last for one hour

Why?

Figure 8. Experiment 3: Interpreting the W-O sentence – ‘correct statement’ question.

Procedure.

Participants saw one experimental task screen followed by two screens featuring the Berlin Numeracy Test, followed by the ‘correct statement’ screen. Upon completion, participants were given a code to claim their payment, thanked and debriefed.

Results.

Responses were coded into nine bins, with the first bin equivalent to a below minimum response, and the ninth equivalent to an above maximum response. The middle seven represented the seven bars in the histogram (e.g., Figure 7). The distribution of numeracy scores is shown in Table 3.

The 'extremity effect' (with regards to maximum and above maximum responses) was found in all distributions. In each distribution, there was a significant association between communication format and choosing a maximum/above maximum outcome in the histogram (positive: χ^2 (3) = 31.36, p < .001; normal: χ^2 (3) = 46.36, p < .001; negative: χ^2 (3) = 42.93, p < .001; bi-modal: χ^2 (3) = 57.69, p < .001). The verbal format was consistently associated with the largest proportion of responses indicating high amplitude outcomes and numerical format associated with the smallest proportion (see Figure 9) replicating Experiments 1 and 2. We also found a significant association between numeracy and maximum/above maximum responses, χ^2 (4) = 14.15 p = .007, with fewer maximum/above maximum responses as numeracy increased, reducing from 66.8% of responses from participants with numeracy scores of 0 or 1, to 17% for those with scores of 3 or 4.

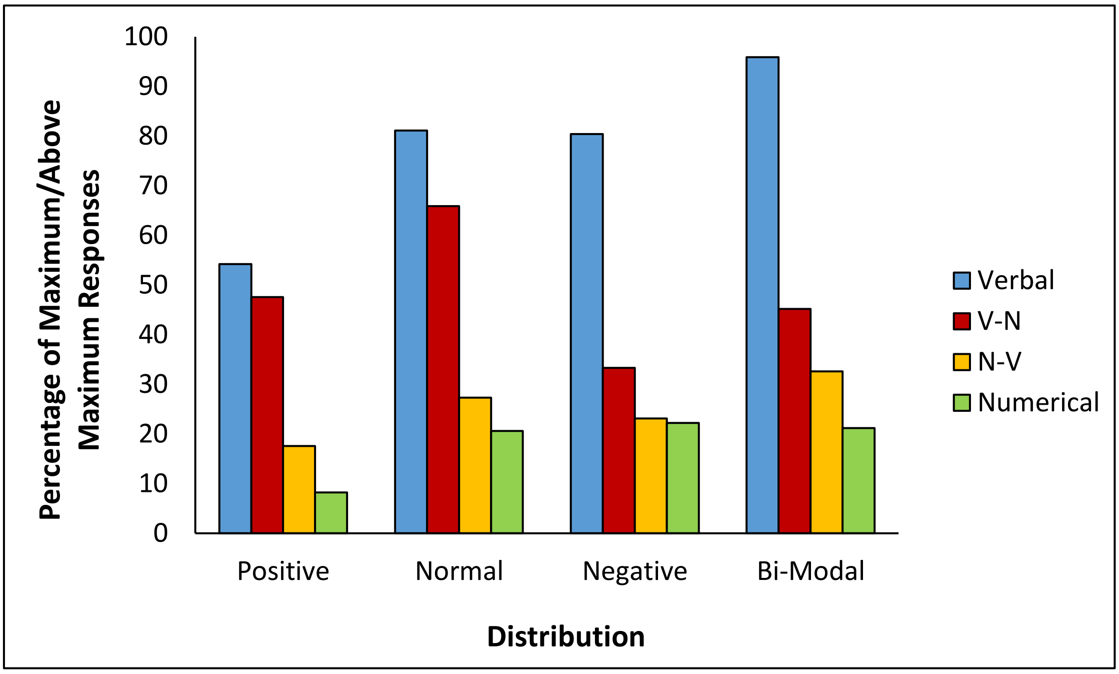


Figure 9. Experiment 3: Percentage of maximum and above maximum responses by communication format for each distribution.

Figure 10 shows the mean response by outcome bin for each condition. Consistent with the maximum/above maximum analysis, the highest responses were in the verbal format across all distributions. A 4 (distribution) × 4 (communication format) × 2 (numeracy

– high/low) between-subjects ANOVA revealed a significant effect of communication format, $F(3, 661) = 29.41$, $p < .001$, $\eta_p^2 = .12$, and a marginally significant effect of distribution, $F(3, 661) = 2.61$, $p = .05$, $\eta_p^2 = .01$, on responses. There was no significant effect of numeracy, $F(1, 661) = 0.89$, $p = .35$. Although the format × distribution interaction approached significance, $F(9, 681) = 1.85$, $p = .06$, $\eta_p^2 = .03$, the effect of communication format was similar across all distributions (see Figure 10). No other interactions approached significance (all *p*s > .51).[22]

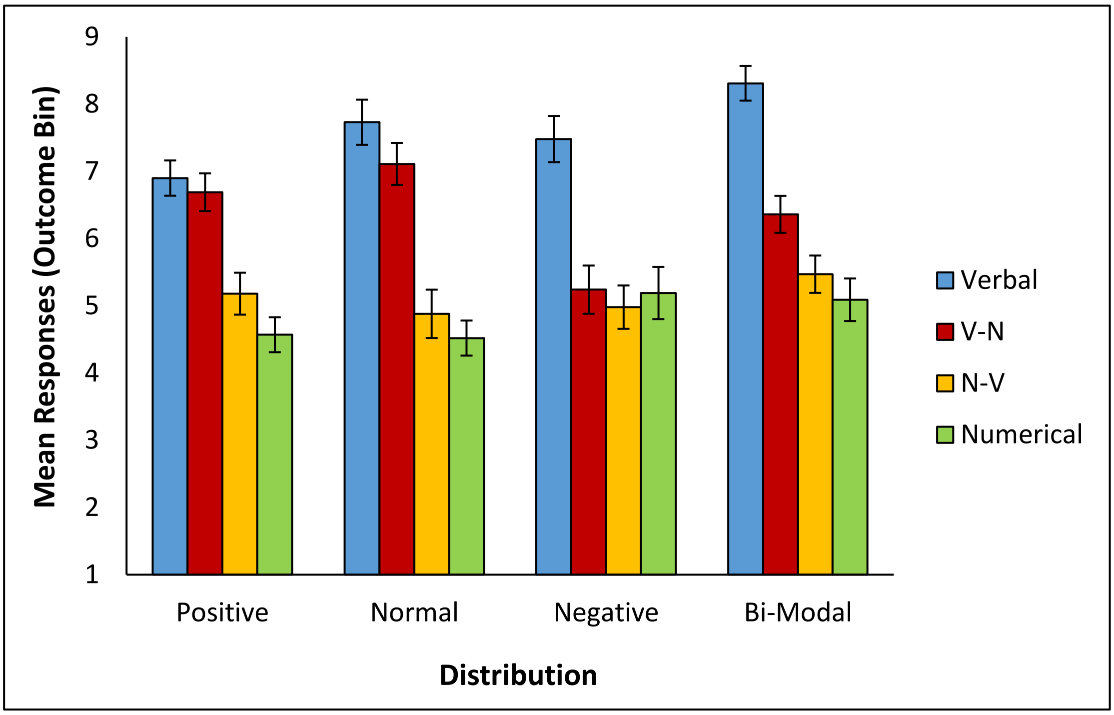


Figure 10. Experiment 3: Responses by communication format for each distribution (error bars represent ±1 Standard Error [SE]).

Following Experiments 1 and 2, we analysed the distribution of responses around the correct answer for the numerical and mixed formats in the normal distribution. The correct answer fell in bin six. As in Experiment 2, the numerical condition had a higher proportion of responses below the correct answer (61.9%) than above (20.6%). Responses in the V-N condition again replicated, with responses skewed above the correct answer (65.9%)

[22] Cumulative distribution graphs for each distribution can be found in Appendix D.

compared to 13.6% below. In the N-V condition 51.5% of responses were below the correct answer and 27.3% were above the correct answer.

Correct statement.

Nearly three-quarters of participants (73.2%) endorsed the statement 'there is a 100% likelihood the battery will last for one hour', with 26.8% endorsing the statement 'there is a 10% likelihood the battery will last for one hour' as correct. The majority of participants therefore gave the W-O sentence an 'at least' interpretation. There was a significant association between numeracy score (high/low) and statement endorsement, χ^2 (1) = 4.65, *p* = .03, with higher numeracy scores linked to the '100% likelihood' statement (an 'at least' interpretation).

Discussion.

The results of Experiment 3 support the conclusions of Experiments 1 and 2 with one additional caveat: the 'extremity effect' extends to scenarios wherever the word 'unlikely' is included *at the beginning of the probability phrase.* Participants seem to be sensitive to the order of the mixed format expression, attributing greater weight to the information preceding that which is presented in parentheses – in line with Grice's (1975) stipulation that cooperative utterances should be 'orderly'. However, it is not known whether this effect occurs as a direct result of the parentheses, or is merely one of order, a question investigated in Experiment 5.

The majority of participants endorsed the 100% likelihood statement as correct, indicating they interpreted the sentence in the way we have argued is most justifiable. It is interesting to note that more numerate participants provide fewer responses from the high end of the distribution, thus not demonstrating the qualitative inconsistency between indicating an outcome that will *never* happen, and standard uses of 'unlikely' to imply probabilities greater than zero.

Overall, the results indicate that the 'extremity effect' occurs whenever a VPE is presented first in a probability phrase. The effect seems relatively robust against contextual influences such as the saliency of certain possible outcomes, and the shape of the distribution (similar to Juanchich et al.'s, 2013, findings). Numeracy appears to have only a limited influence on responses.

Experiment 4

So far, the experiments have used point estimates in the numerical formats. However, Budescu and colleagues (Budescu et al., 2009, 2012, 2014) used numerical ranges in their mixed format condition, such as 'unlikely (less than 33%)'.[23] Range estimates could be said to be more similar to VPEs in that they refer to a range of outcomes (e.g., Budescu & Wallsten, 1995). Moreover, range estimates might be more suitable in situations where point estimates are unrealistic. It however remains to be seen whether the 'extremity effect' will replicate when range expressions are explicitly included. We returned to using consequential geological scenarios, having established that our results generalise across different shapes of distribution.

Method.

Participants.

Three hundred and twelve participants were recruited for this online experiment from Amazon Mechanical Turk via the TurkPrime platform (Litman et al., 2017) and were paid $0.20 upon completion. Twenty-three cases were removed (due to duplicate IP addresses, for failing the attention check or completing the experiment quicker than reasonably expected) leaving a final sample of 289 (114 male, 1 'other') participants, aged between 19–70 (*Mdn* = 32).

[23] Though see more recent research which has demonstrated that VPEs presented accompanied by translations with the lower and upper bound explicitly stated (e.g., 'unlikely [0–33%]') are better differentiated than when only one bound is stated (e.g., 'unlikely [<33%]') (Harris et al., 2017).

Design, procedure and materials.

Communication format (verbal – 'unlikely'; numerical range – '10–30% likelihood'; V-N – 'unlikely [10–30% likelihood]' and N-V – '10–30% likelihood [unlikely]') was manipulated between-participants. Scenario (flood) was the same for all participants. As per the previous experiments, participants were required to type a numerical response which corresponded to the outcome being described. Finally participants completed a numeracy scale (Lipkus, Samsa, et al., 2001), with two additional questions from the Berlin Numeracy Test (Cokely et al., 2012) included to increase variability in numeracy scores. Upon completion, participants were given a code to claim their payment, thanked and debriefed.

Results.

Effect of communication format.

The proportion of responses indicating high amplitude outcomes (the maximum/above maximum value present in the histogram) was highest in the verbal condition, followed by the mixed format condition. The numerical range condition had the lowest proportion of responses indicating high amplitude outcomes, χ^2 (2) = 76.35, p < .001. The distribution of responses followed a very similar pattern to those in previous experiments, see Figure 11.

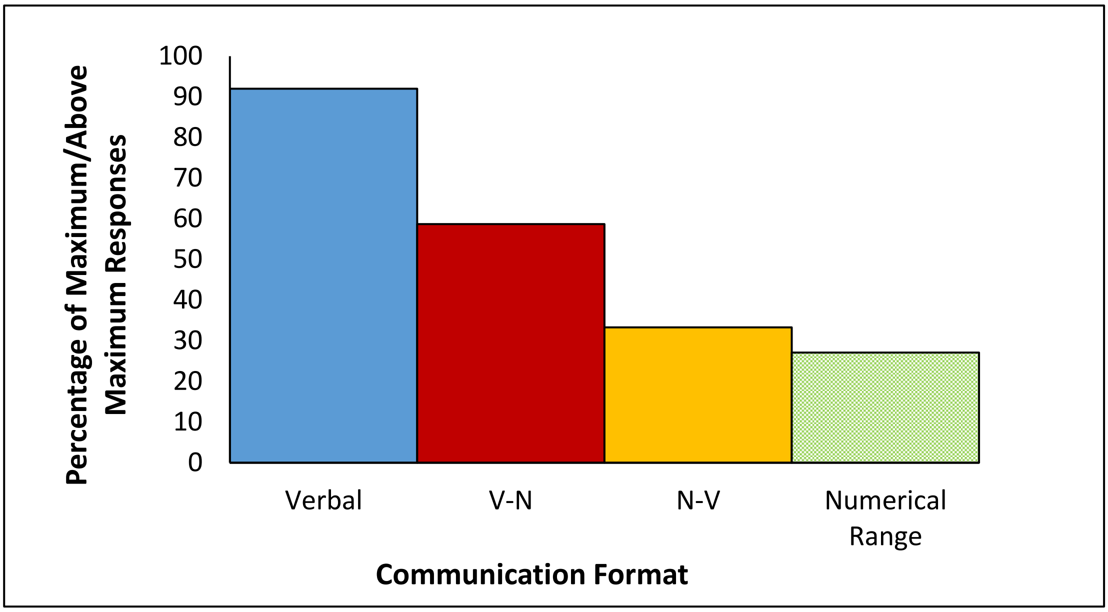


Figure 11. Experiment 4: Percentages of maximum and above maximum responses by communication format.

A one-way ANOVA showed a significant effect of communication format $F(3, 285) = 32.74$, $p < .001$, $\eta_p^2 = .26$. A post-hoc Ryan, Einot, Gabriel and Welsh Q (REGWQ) procedure revealed that responses in the verbal condition (M= 9.41, SD= 1.87) were significantly higher than responses in the other three conditions. Responses in the V-N condition (M= 7.60, SD= 2.81) were significantly higher than those in the numerical range condition (M= 5.47, SD= 3.05) and the N-V condition (M= 5.87, SD= 2.90). Responses were similar in the numerical range and N-V conditions. Replicating findings from Experiment 3, there is clear evidence for order effects, see Figure 11 and Figure 12.

Similar to the findings for the numerical point condition in Experiments 2 and 3, the numerical range condition had a higher proportion of answers below the correct answer (60%) than above (37.1%). To a lesser extent, the N-V condition followed this pattern, with fewer answers above the correct answer (39.1%) than below (52.2%). Responses in the V-N condition had a greater proportion of responses above the correct answer (68%) compared to below (29.3%).

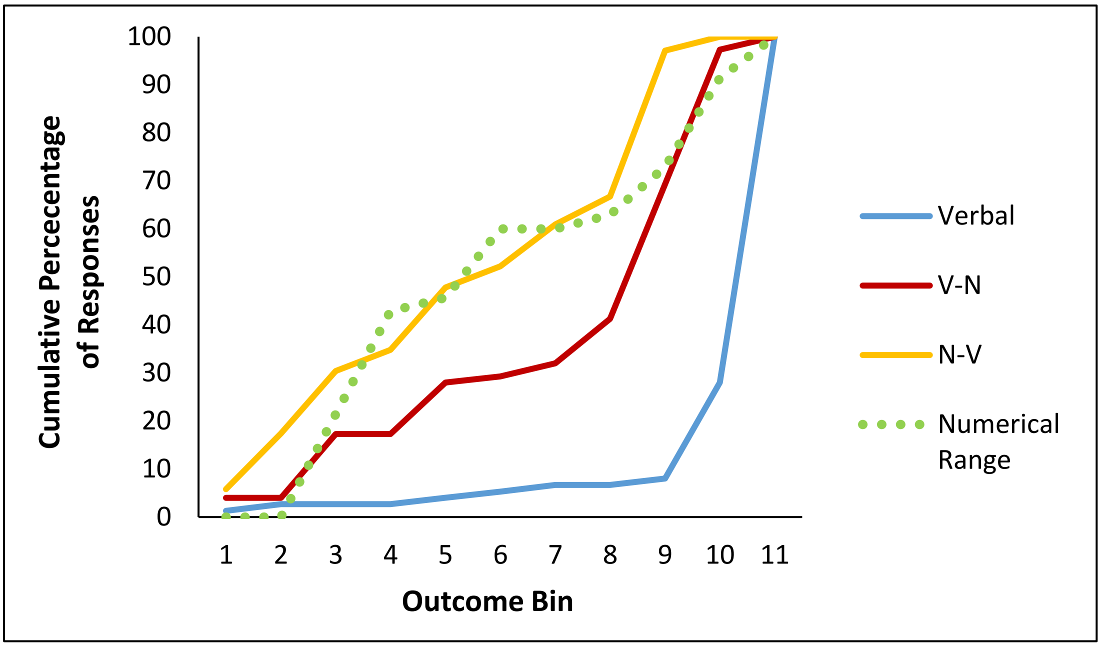


Figure 12. Experiment 4: Cumulative distribution of responses by communication format.

Effect of numeracy.

Answers for each 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 4.

Table 4. Experiments 4 and 5: Distribution of Numeracy Scores (%) using Lipkus et al. (2001) Numeracy Scale and 2 Questions from the Berlin Numeracy Test (Cokely et al., 2012)

	Numeracy Score											
	0	1	2	3	4	5	6	7	8	9	10	Mean (SD)
Exp 4	0	2.1	0.3	1.7	2.4	3.1	5.5	13.1	29.4	31.8	10.4	7.84 (1.82)
Exp 5	0.3	0.3	0.3	1.0	1.7	4.5	5.5	15.1	26.7	30.8	13.7	8.00 (1.64)

The effect of communication format held over differing levels of numeracy. A two-way ANOVA with numeracy (high/low) as a fixed factor showed there was a significant effect of communication format $F(3, 281) = 30.83$, $p < .001$, $\eta_p^2 = .25$ and no significant effect of numeracy, $F(1, 281) = 2.67$, $p = .10$, nor a significant format × numeracy interaction, $F(3, 281) = 1.56$, $p = .20$. There was no significant association between high and low numeracy and maximum/above maximum responses, χ^2 (1) = 1.19, $p = .29$.

Discussion.

As in previous experiments, we replicate the 'extremity effect' in both the verbal and V-N formats, as well as replicating the difference in responses between the two mixed formats. Whether the mixed format includes a numerical point, or range expression does not significantly affect the proportion of high amplitude responses. Rather, the more important factor seems to be the position of the numerical expression within the mixed format, though whether this relates to the parentheses, or order remains an open question.

Experiment 5

Whilst Experiments 3 and 4 revealed a difference in responses between V-N and N-V formats, it was not possible to distinguish whether this difference was driven by an effect of parentheses or an effect of order. Walker (1823) suggests that in both writing and speech, parentheses are given little emphasis, with the option of either ignoring them, or pronouncing them with a lower tone. The order account is subtly different, and is based on Grice's (1975) conversational maxim that cooperative utterances should be 'orderly', with the most important information presented first. We therefore designed a further experiment to differentiate between the two accounts.[24]

[24] Thank you to the named reviewer of the paper which this chapter is based on, Karl-Halvor Teigen, for suggesting this experiment.

Method.

Participants.

Three hundred participants were recruited for this online experiment from Amazon Mechanical Turk via the TurkPrime platform (Litman et al., 2017) and were paid $0.20. Eight cases were removed (due to duplicate IP addresses, for failing the attention check or completing the experiment quicker than reasonably expected) leaving a final sample of 292 (118 male) participants, aged between 18–73 (*Mdn* = 32).

Design, procedure and materials.

Punctuation (brackets [B] or dashes [D]) and order (V-N or V-N) were manipulated in a 2 × 2 between-participants design. As such, there were four communication formats: (V-N-B– ‘unlikely [10–30% likelihood]’; N-V-B – ‘10–30% likelihood [unlikely]’; V-N-D – ‘unlikely – 10–30% likelihood’ and N-V-D – ’10–30% likelihood – unlikely’). Scenario (flood) was the same for all participants. As per the previous experiments, participants were required to type a numerical response which corresponded to the outcome being described. Following this participants completed the 10 question numeracy scale, as in Experiment 4. Upon completion, participants were given a code to claim their payment, thanked and debriefed.

Results.

Effect of communication format.

The proportion of responses indicating high amplitude outcomes (the maximum/above maximum value present in the histogram) was highest in the V-N-B and V-N-D conditions. The N-V-B and N-V-D conditions had the lowest proportion of responses indicating high amplitude outcomes, χ^2 (3) = 13.62, p = .003. The distribution of responses reflected the order effects seen in Experiments 3 and 4, see Figure 13.

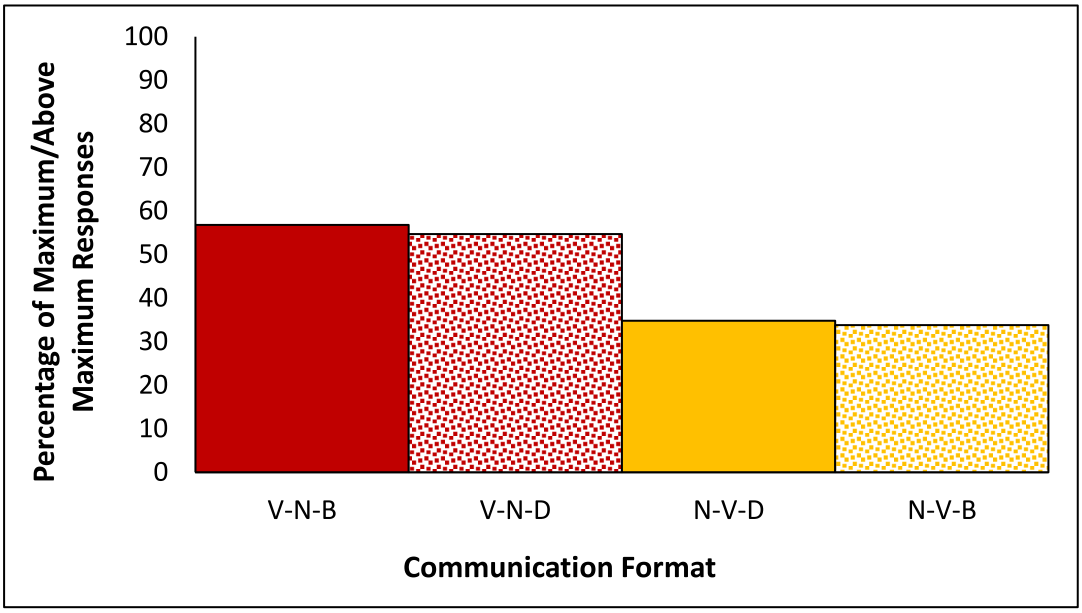


Figure 13. Experiment 5: Percentages of maximum and above maximum responses by communication format.

A two-way ANOVA showed a significant effect of order, F(1, 288) = 18.64, p < .001, η_p^2= .06 and no significant effect of punctuation, F(1, 288) = 0.40, p = .53. There was no interaction between order and punctuation, F(1, 288) = 0.15, p = .70. A post-hoc REGWQ procedure revealed that responses in the V-N-B (M= 7.32, SD= 2.99) and V-N-D (M= 7.41, SD= 2.89) were not significantly different. Responses in both V-N formats were significantly higher than responses in the N-V-B (M= 5.64, SD= 3.26) and N-V-D conditions (M= 6.00, SD= 3.12). Responses were similar in the N-V-B and N-V-D conditions. Figures 13 and 14 clearly display the effect of order.

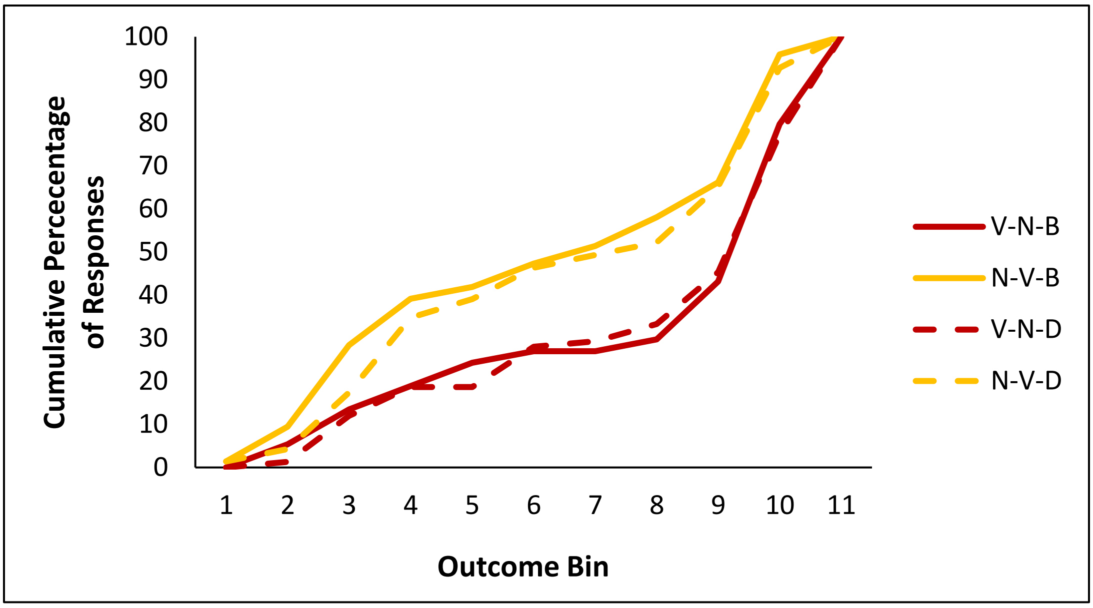


Figure 14. Experiment 5: Cumulative distribution of responses by communication format.

As observed in Experiments 3 and 4, expressions where the numerical phrase came first had a higher proportion of answers below the correct answer (N-V-B: 51.4%, N-V-D: 49.3%) than above (N-V-B: 41.9%, N-V-D: 47.8%). Responses in the V-N-B and V-N-D conditions had a greater proportion of responses above the correct answer (V-N-B: 70.3%, V-N-D: 66.7%) compared to below the correct answer (V-N-B: 27.0%, V-N-D: 29.3%).

Effect of numeracy.

Numeracy scores were coded as in Experiment 4 and can be found in Table 4. The effect of communication format held over differing levels of numeracy. A two-way ANOVA with numeracy (high/low) as a fixed factor showed there was a significant effect of communication format, $F(3, 284) = 5.83$, $p = .001$, $\eta_p^2 = .06$ and no significant effect of numeracy, $F(1, 284) = 0.03$, $p = .86$. There was no significant communication format × numeracy interaction, $F(3, 284) = .58$, $p = .63$.

There was a significant association between high and low numeracy and maximum/above maximum responses, χ^2 (1) = 5.34, $p = .02$, with those lower in numeracy giving a higher proportion of maximum/above maximum responses than those higher in numeracy.

Discussion.

Experiment 5 confirmed that it is the order of the information that drives the effect seen in mixed formats, rather than the presence of parentheses. It therefore seems that the pragmatics of communication are responsible for the 'extremity effect', rather than the effect reflecting a general feature of how people understand probabilities and frequencies. The negative directionality of the expression 'unlikely' focuses attention on the non-occurrence of the event and shifts estimates to the 'unlikely' (right) end of the scale, whereas '20% likelihood' could be seen as focusing attention on the occurrence of the event and shifting estimates to the 'likely' (left) end of the scale.

Experiment 6

The 'extremity effect' observed in W-O tasks conflicts with typical responses in translation tasks of VPEs. Even in a within-participants study, Teigen et al. (2013, Experiment 5b) found participants provided translations of around 30% after having completed an 'improbable' sentence with maximum or above maximum responses (0% probability). These results suggest that participants might not be sensitive to the inconsistency between two such responses. We extend this research by providing participants with a second W-O task after a translation task. We propose that participants who select an 'unlikely' outcome with a 0% frequency of occurrence, but yet translate 'unlikely' as 20% *should* recognise the inconsistency between their two responses and thus adjust their answers in a second W-O task.

Method.

Participants.

One hundred and fifty-one participants were recruited for this online experiment via Prolific Academic and were paid £0.50. One participant was excluded for failing the attention check, leaving a final sample of 150 participants (75 male), aged between 18–65 (*Mdn* = 31).

Design.

Communication format (verbal – 'unlikely'; numerical – '20% likelihood'; V-N – 'unlikely [20% likelihood]'; N-V – '20% likelihood [unlikely]') was manipulated between-participants. Scenario (volcano; flood; earthquake) and W-O task (before and after the translation task) were manipulated within-participants. Scenario was randomised such that each participant only saw each scenario once, but the task they saw it in (e.g., first or second W-O task/translation task) differed systematically across participants. Different scenarios for these tasks were used in order to reduce demand characteristics. No numeracy measure was included in this experiment. For the W-O tasks, participants responded as in previous experiments. For the translation task, we asked participants what they thought the probability conveyed by the expert was by indicating a minimum and maximum estimate. This meant the translation made sense for expressions featuring 20% – it would have seemed rather odd to ask for a best estimate of the numerical probability implied by 20% in any of the conditions including the precise number(!)

Materials and procedure.

The introductory text informed participants that they would read various predictions about future geological events and be asked to make a series of judgements about these. All participants read three vignettes describing outcomes of how far lava flows, floodwater and earthquake tremors would extend, as in Experiment 1.

Participants completed two W-O tasks. The shape of the distributions were similar and approximately normal across the three (volcano, flood and earthquake) scenarios (see Appendix E).

The Wayston flood plain has a history of flooding due to its flat terrain and proximity to the east side of the River Wayston. Given the river's situation and recent scientific observations, a senior hydrologist has reported that there is a **20% likelihood** that given a flood, the floodwater will extend to 7.0km.

What do you think is the senior hydrologist's estimate of the probability of the floodwater extending to 7.0km? (Please indicate a number between 0 and 100%)

Minimum Estimate:

Maximum Estimate:

Figure 15. Experiment 6: Example of translation task (numerical condition).

After completing the first W-O task, participants then moved on to a translation task, in which they were required to provide a numerical translation of the probability term used, by giving their minimum and maximum estimates (see Figure 15). They then completed a final W-O task before being given a code to claim their payment, thanked and debriefed (see Figure 16 for a flow chart of the procedure).

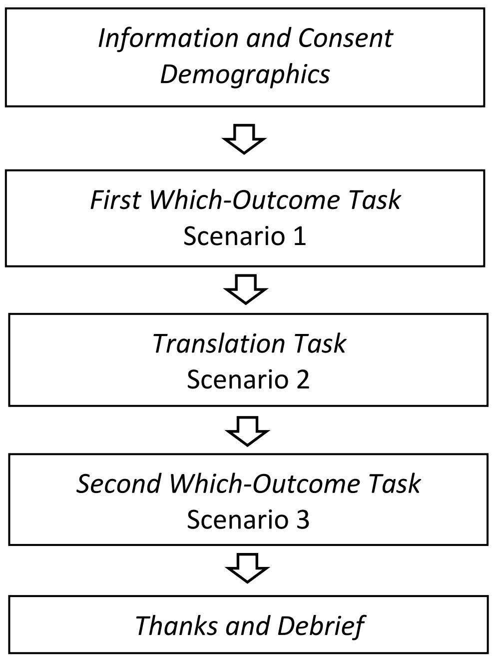


Figure 16. Experiment 6: Procedure. The scenario used as 1, 2 and 3 was randomised across participants.

Results.

The first W-O task enabled a replication of the 'extremity effect' over all four communication formats with the geological scenarios. The primary aim of the experiment was, however, to examine whether the W-O effect was robust against an interim translation task, which was expected to yield an 'inconsistent' response, equating 'unlikely' with values greater than 0%. In the analysis, the translation task results are presented first, followed by results from the first and then second W-O tasks. Finally we consider the effect of the translation task by examining the difference between the two W-O tasks.

Translation task.

The midpoint of each participant's minimum and maximum translations was taken as their 'best estimate'. There was no effect of communication format on estimates, $F(3, 146) = 0.65$, $p = .59$ – estimates in the verbal format were similar to those in the other formats (see Table 5). In the verbal format, only one participant gave an estimate of zero, with 29.7% of participants giving an estimate of 5% or under.

Table 5. Experiment 6: Mean Best Estimate for Translation Task by Communication Format

	Mean (SD)	Median	Range
Verbal	31.38 (32.83)	15	0 – 96.5
Numerical	24.47 (16.42)	20	10 – 85
V-N	25.76 (22.48)	20	10 – 99
N-V	24.47 (15.44)	20	2.4 – 80

First 'which-outcome' task.

As in previous experiments, responses on the W-O task were standardised across scenarios by 'binning' responses, in accordance with the order of bars in the histogram. With nine bars in each histogram and an additional minimum/maximum bin at each end, there were 11 bins in total. Responses were similar across scenarios. Similar to previous experiments, the 'extremity effect' replicated, with the proportion of responses indicating high amplitude outcomes largest in the verbal condition, followed by the V-N condition. The numerical condition had the lowest proportion of responses indicating high amplitude outcomes, χ^2 (3) = 58.72, p < .001 (see Figure 17).

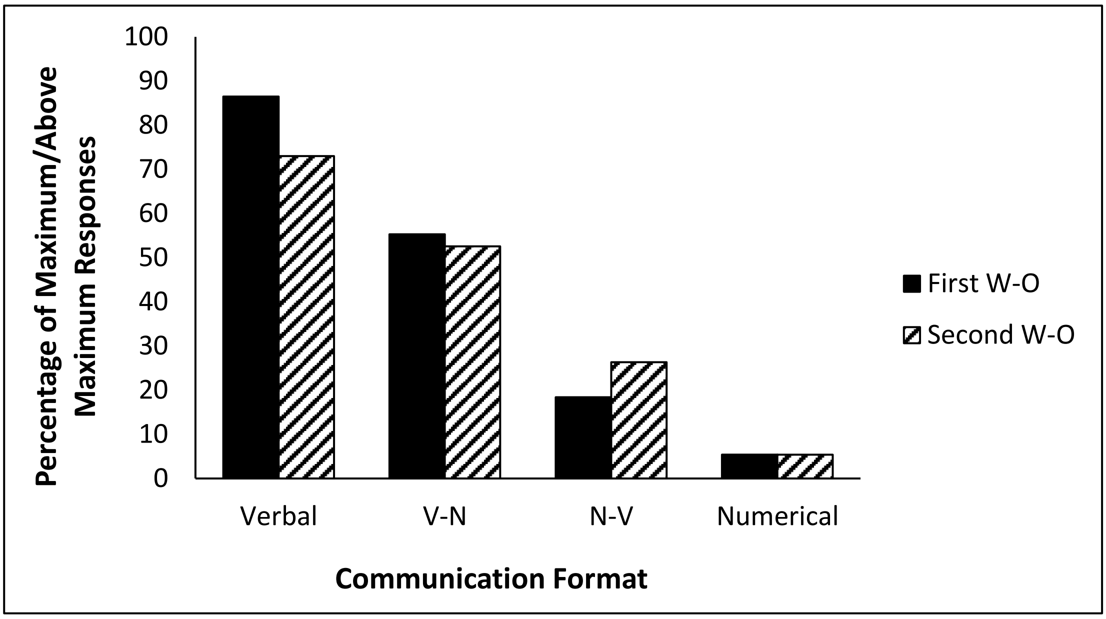


Figure 17. Experiment 6: Percentage of maximum and above maximum responses by communication format for first and second W-O tasks.

Figure 18 shows responses by bin, further illustrating the effect of communication format, specifically that the verbal format leads to high amplitude responses. A one-way ANOVA showed a significant effect of communication format $F(3, 146) = 31.83$, $p < .001$, $\eta_p^2 = .40$. A REGWQ procedure revealed that responses in the verbal condition (M= 10.00, SD= 2.07) and the V-N condition (M= 9.00, SD= 2.23) were similar. They were significantly higher than responses in the N-V (M= 6.76, SD= 3.04) and numerical (M= 4.89, SD= 2.45) conditions. Responses were significantly different in the numerical and N-V conditions.

Bin eight included the outcome reflecting a probability of 20% across all three scenarios – the correct answer. As observed in previous experiments, the numerical format had a higher proportion of answers below the correct answer (83.8%) than above (13.5%).

Responses from the V-N format again replicated, with a greater proportion of responses above the correct answer (78.9%) compared to below (13.2%). Responses in the N-V format were evenly distributed, with 47.4% below the correct answer and 50% above the correct answer.

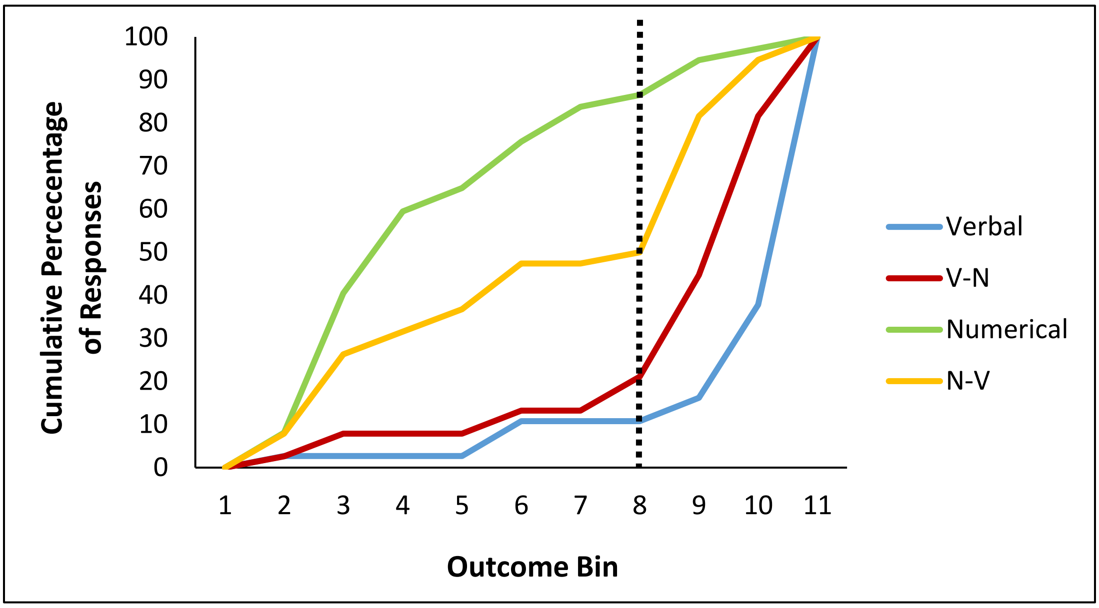


Figure 18. Experiment 6: Cumulative distribution of responses by communication format pre-translation task (dotted line represents outcome bin containing correct answer).

Second 'which-outcome' task.

The proportion of responses indicating high amplitude outcomes was again largest in the verbal condition, followed by the V-N condition. The numerical condition had the lowest proportion of responses indicating high amplitude outcomes, χ^2 (3) = 40.91, p < .001 (see Figure 17).

A one-way ANOVA showed a significant effect of communication format, F(3, 146) = 15.70, p < .001, η_p^2= .24, with the general distribution of responses for each communication format being relatively similar to the first W-O task. A REGWQ procedure revealed that responses in the verbal condition (M= 9.16, SD= 2.98) and the V-N condition (M= 8.29, SD= 2.89) were similar. They were significantly higher than responses in the numerical condition

(M= 4.89, SD= 2.60). Responses were not significantly different in the V-N and N-V conditions (M= 7.47, SD= 2.84).

The numerical format had a higher proportion of answers below the correct answer (78.4%) than above (21.6%). Responses from the V-N format again replicated, with a greater proportion of responses above the correct answer (71.1%) compared to below (23.7%). The pattern of responses in the N-V format was less uneven, with 34.2% below the correct answer and 55.3% above the correct answer.

The effect of the translation task.

Most notably, a repeated measures ANOVA revealed no significant difference between responses on the first and second W-O tasks, $F(1, 146) = 0.80$, $p = .37$. There was an inevitable (given the results above) effect of communication format, $F(3, 146) = 30.96$, $p < .001$, $\eta_p^2 = .40$. There was a trend towards an interaction between W-O task and communication format but this did not reach conventional levels of significance, $F(3, 146) = 2.36$, $p = .07$, $\eta_p^2 = .05$ (see Figure 19).

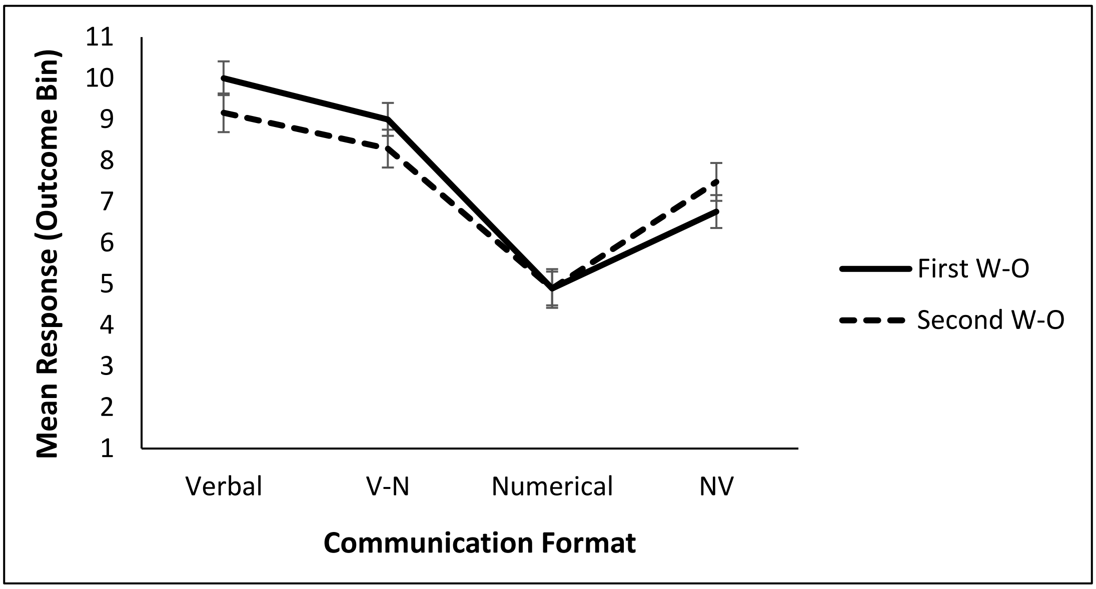


Figure 19. Experiment 6: Mean responses by communication format pre and post translation task (error bars represent $\pm$1 Standard Error).

To investigate directly whether participants continued to endorse maximum/above maximum responses to the same degree in the second W-O task as in the first (given that the W-O task × communication format interaction term approached significance), four McNemar tests were carried out. No significant differences were observed (all *p*s > .22; see Figure 17).

Discussion.

The initial W-O task replicated the 'extremity effect' observed in previous work: responses in the verbal condition were typically taken from the high end of the distribution. This condition also had the most maximum/above maximum responses. Responses in the V-N condition followed this pattern to a lesser extent. Responses in the numerical condition were typically from the intermediate values in the distribution and had very few maximum/above maximum response (as in the N-V condition). There were no significant differences between responses on the first and second W-O tasks.[25]

Results from the translation task showed there were no differences in translations between communication formats. The mean translation of unlikely was just over 30% – slightly higher than in some previous studies (e.g., Fillenbaum, Wallsten, Cohen, & Cox, 1991; Theil, 2002). Translations were still, however, in line with prescriptions of usage in the real-world (e.g., the Intergovernmental Panel on Climate Change prescribes using 'unlikely' for probabilities of less than 33%).

There was no effect of the translation task on participants' responses in the second W-O task. Participants thus seem unaware of an inconsistency between their responses to 'unlikely' on a translation and W-O task. This might be because participants see the two tasks as unrelated and/or because different processes are involved in the two tasks.

[25] The reported experiment was a replication of an undergraduate project. The original experiment (thanks to Duyen Tran, Jay See Tow, & Pauline Gordon for data collection) showed a similar pattern of findings but the order effect was not significant. An analysis across the two experiments did, however, reveal a significant order effect, with no moderation by 'Experiment'.

General Discussion

This chapter presents six experiments which used the W-O methodology to test the robustness of the 'extremity effect' against the presence of a consequential outcome, across differing distributions and communication formats (verbal, numerical and mixed [V-N & N-V]). We also examined whether participants were sensitive to the inconsistencies in their responses in W-O and translation tasks, as well as investigating the potential influence of numeracy.

All experiments yielded consistent results, replicating and extending previous findings, with preference for outcomes at the high end of a distribution (including above maximum values) present in both verbal and V-N conditions. Experiment 6 replicated the effects of communication format, in terms of preferences for the upper end of a distribution in the verbal and V-N conditions, which persisted despite an interim translation task eliciting different responses.

Verbal, numerical, verbal-numerical or numerical-verbal formats?

The differences in responses in the V-N and N-V formats observed in Experiments 3–6 we label an 'order effect', in which the first part of the expression is given the most weight. This is in line with Grice's (1975) conversational maxims, particularly the 'maxim of manner', which states one should be orderly, with the most important information first. The fact that responses in the single format conditions are not exactly the same as their mixed format counterparts implies that the secondary information is underweighted (relative to the initial information) rather than being wholly disregarded. Furthermore, the perceived usefulness (and function) of the secondary information varies between the two mixed format conditions. In the V-N condition, participants may have an initial estimate which is then anchored upwards or downwards with the following numerical expression. In contrast, in the N-V condition, the participant has a very clear number to start with.

That the 'extremity effect' continued to occur in the V-N format is of particular relevance to current literature, given recent recommendations to use a mixed format approach to express uncertainty (Budescu et al., 2009, 2012, 2014; Harris et al., 2017; Patt & Dessai, 2005; Renooij & Witteman, 1999; C. L. M. Witteman & Renooij, 2003; C. L. M. Witteman et al., 2007). The present results suggest that simply including numerical translations alongside VPEs will potentially be of only limited benefit in improving the effectiveness of risk communications. Whilst using a V-N format has been found to increase the differentiation of participants' interpretations of VPEs, as well as increasing the level of agreement with IPCC guidelines (Budescu et al., 2014), our findings highlight a potential downside of the V-N format, as well as a *potential* upside of the N-V format. Whilst the numerical [V-N] format led to a preponderance of responses below [above] the correct answer, responses in the N-V format were more evenly distributed above and below the correct answer. Of most consequence, however, was that the V-N format showed an increased endorsement of outcomes with a 0% frequency of occurrence compared to the numerical and N-V formats. This endorsement could be seen as beneficial from the perspective of the communicator, who recognises the costs of underestimating the risk, in line with the asymmetric loss function account (Harris et al., 2009; Weber, 1994), and sees overestimation of a lava flow as a way of motivating action. From the perspective of the recipient, however, such an endorsement is problematic, given that organisations such as the IPCC use 'unlikely' to mean less than 33%, rather than 0%. A failure to act (e.g., evacuate a danger area) based on a 0% interpretation could therefore have life-threatening consequences. Ultimately, the appropriate choice of communication format will depend on the purpose of the communication, and how important minimisation of the 'extremity effect' is to the communicator. At the very least, the current results suggest that refinement of recent recommendations to use a mixed (V-N) communication format is required.

Our findings clearly demonstrate that communication format influences interpretations of risk communications. The fact that the ‘extremity effect’ persists in the mixed V-N format, but not to phrases where the numerical expression comes first indicates that the effect is not related to how people understand probability and frequencies in general, but rather the pragmatics of communication.

A potential critique of the work presented thus far is its focus on one low probability VPE, ‘unlikely’. This decision was made in light of the larger range of interpretations given to negative directionality expressions, particularly ‘unlikely’ (Smithson et al., 2012; see also Wintle et al., 2019) and the relevance of using such phrases to describe highly consequential events (which are usually unlikely) such as natural hazards. However, previous investigations have shown that effects observed with ‘unlikely’ are also observed with positively directional VPEs such as ‘possible’ and ‘chance of occurring’ (Teigen et al., 2014). It therefore seems reasonable to expect similar results to the current research for other low probability verbal expressions. We do, however, recognise that it is a more open question as to whether these results will extend to high probability VPEs such as ‘quite probable’ (see Teigen, Juanchich, & Filkuková, 2014).

Effect of numeracy.

When presented with numerical information, those lower in numeracy are “left with information that is less complete and less understood” (E. M. Peters et al., 2006, p. 412) versus those higher in numeracy. Varying numeracy levels (and comfort with numerical information) have therefore been cited as a reason for presenting information in a mixed format (C. L. M. Witteman & Renooij, 2003). We originally conjectured that the high amplitude responses present in 45% of cases in the V-N condition of Experiment 1 (see Figure 5) could have been due to numeracy levels, with responses varying according to two different focuses. The less numerate participants may have felt uncomfortable using ‘20% chance’ to form their estimates and thus focused on ‘unlikely’, but the more highly numerate may have focused on the opposite, which could have explained why the ‘extremity effect’ persisted, but less frequently. However, the contrasting pattern of responses in the N-V conditions in our subsequent experiments suggest that the effect observed in this chapter is one of order, where whatever information is presented first is deemed most important.

Nevertheless, over all communication formats and all experiments featuring a numeracy measure, more numerate participants demonstrated the ‘extremity effect’ less, giving fewer maximum and above maximum responses (except in Experiment 5). This result challenges the notion that VPEs are particularly useful when communicating with less numerate audiences (Gurmankin, Baron, & Armstrong, 2004), instead indicating that they could do most harm for such audiences. We speculate this could be because they encourage a non-mathematical interpretation of a mathematical question, leaving participants susceptible to the effects of the directionality of the expression. It is however, important to recognise that our sample were from Western, educated, industrialised, rich and democratic (WEIRD) countries (Henrich, Heine, & Norenzayan, 2010). Our findings are not necessarily generalisable to those who are less numerate or from non-WEIRD countries.

Implications for credibility.

The results from the current research using the W-O paradigm continues to highlight the variability in people’s understanding of VPEs, demonstrating the ‘extremity effect’ is pervasive and robust and occurs even if the proposed solution of a mixed (V-N) format is used. At this point, the more critical reader might posit that the effect is merely an experimental artefact, and thus question the ‘real-world’ relevance of the current findings, given that risk and uncertainty communications do not present information as in the W-O paradigm. However, we believe that there could be important (and yet to be researched) consequences for the perceived credibility of the communicator. For instance, a communicator who uses ‘unlikely’ to mean 20% will quickly lose the trust of an audience if that audience expects ‘unlikely’ to refer to outcomes which never happen. Yet scientific forecasts are probabilistic (at best) and thus it is not possible to predict with certainty the probability of a landslide on a given day (for example). Even if an event is predicted to be 'unlikely' to occur, the very fact that the outcome is not certain means that it could still happen. If it does occur (a *probabilistically unexpected* outcome), the prediction might be perceived as 'erroneous', which could have implications for how a communicator is subsequently perceived. The following chapter thus features such 'erroneous' predictions, and investigates their potential repercussions for the perceived credibility of a communicator.

Chapter Summary

The present research provides an exploration of the effects of communication format on the understanding of risk and uncertainty communications. We provide further evidence of the potential for discrepancies between the way these expressions are typically used in formal risk communications (e.g., by those from the UK Defence Intelligence or the IPCC) and the way they are understood by recipients. Whilst it is generally acknowledged that there is no 'optimal' presentation format, and no single 'fix' for communications of risk and uncertainty (Budescu et al., 2012), identifying instances in which the communication format has a significant impact on audience's understanding is key to improving the efficacy of such communications. Findings from these experiments show that the 'extremity effect' extends to communications which use a V-N format, and appears robust across differing distributions, as well as in the presence of consequential outcomes. Overall, the results add to previous research demonstrating that understanding differs according to communication format, which raises the possibility that perceptions of credibility could also differ as a result of communication format – a possibility that is directly tested in the next chapter.

Chapter 3. The Role of Communication Format in Maintaining Communicator Credibility

Chapter Overview

The public expects science to reduce or eliminate uncertainty, yet scientific forecasts are probabilistic (at best) and it is simply not possible to make predictions with certainty. Whilst an ‘unlikely’ event is not expected to occur, an ‘unlikely’ event will still occur one in five times (based on a translation of 20%), according to a frequentist perspective. Yet findings from Chapter 2 indicate that ‘unlikely’ is most often interpreted as referring to events that will never occur. Thus if an ‘unlikely’ event does occur (a *probabilistically unexpected* outcome), a communicator’s prediction may be deemed ‘erroneous’, which could have ramifications for the subsequent (perceived) credibility of the communicator. This chapter examines 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 expressions, in both high and low probability domains. In the low probability domain, 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. Such an effect of communication format is not, however, consistently observed in the high probability domain, a point we return to in Chapter 4.

Introduction

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 still consider the geologist to be trustworthy?

In the current chapter, we investigate scenarios in which events forecast to be ‘unlikely’ (or ‘likely’) nevertheless (do not) occur – *probabilistically unexpected* outcomes.

We term these forecasts ‘erroneous’, despite the fact scientific forecasts are probabilistic (at best). Although the public expects science to reduce or eliminate uncertainty (Kinzig & Starrett, 2003), 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 chapter addresses such misunderstandings in relation to five communication formats and their low and high probability variants: verbal probability expressions (VPEs; e.g., ‘unlikely’, ‘likely’), numerical expressions (e.g., point – ‘20% likelihood’, ‘80% likelihood’; range – ‘10–30% likelihood’, ’70 –90% likelihood’), and mixed expressions (e.g., verbal-numerical – ‘unlikely [20% likelihood]’, ‘likely [80% likelihood]’; numerical verbal – ‘20% likelihood [unlikely]’, ‘80% likelihood [likely]’). Our decision to initially 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 et al., 2016). We subsequently consider the high probability domain (e.g., a ‘likely’ event does not occur), given fears of ‘crying wolf’ have been documented amongst decision-makers in a natural hazards context (e.g., L. R. Barnes, Gruntfest, Hayden, Schultz, & Benight, 2007; Sharma & Patt, 2012; Woo, 2008).

Credibility.

Most individuals do not have in-depth knowledge about, nor experience of, hazards and new technologies (Siegrist et al., 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 trustworthiness and expertise identified as two key components (e.g., Hovland, Janis, & Kelley, 1953; Hovland & Weiss, 1951). Trustworthiness relates to the integrity and reliability of the communicator, specifically whether they are unbiased, honest and conscientious (Dieckmann et al., 2009; Walton, 1997), with expertise relating to the knowledgeability of the communicator (Guilamo-Ramos et al., 2006).

Credibility has been cited as a fundamental part of risk communication practices (Reynolds, 2011). Credibility has been found to significantly influence risk perceptions (Frewer et al., 1997; Trumbo & McComas, 2003) and therefore has implications for how an individual behaves upon receiving a risk communication (Wachinger et al., 2013), for instance taking action to avoid, mitigate or adapt to the risks communicated. If a source is not perceived as credible, then this could lead an individual to ignore such risks. Credibility's influence is thus 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 expression.[26] 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) 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)

[26] Throughout the chapter, 'range' is used to refer to probability ranges (e.g., 10–30% likelihood), rather than outcome range estimates (e.g., 3–5 km).

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, and thus preferred in such a situation.

Range formats.

Probability ranges are widely used, largely due to the assumption that they reflect the precision of an estimate, and thus can be used to refer to a range of possible future states (Dieckmann, Peters, & Gregory, 2015). Despite their widespread use, there is mixed evidence relating to how well they are understood and how they are perceived. Some research suggests that probability ranges are no more advantageous than point estimates, with similar levels of verbatim and gist understanding observed for both formats (Moraes & da Silva Dal Pizzol, 2018). Other research has even found reduced understanding with (both narrow and wide) range formats versus point formats (Longman, Turner, King, & McCaffery, 2012). Likewise, Sladakovic, Jansen, Hersch, Turner and McCaffery (2015) observed reduced understanding of harm and benefit information when this information was presented using a range versus a point format. Increased risk perceptions and worry have been observed when information is presented in a range versus point format (Han, Klein, Lehman, et al., 2011; Han et al., 2009). Increased risk perceptions and worry may occur from the interpretation of range formats as reflecting higher (and more variable) perceived likelihoods (Dieckmann, Mauro, & Slovic, 2010) and/or from the potential focus on the upper bound of the range as a 'worst case scenario' (Han et al., 2009). Not all studies, however, have observed disadvantages to using range formats over point format, with both formats perceived as equally credible, trustworthy, accurate and personally relevant (Lipkus, Klein, & Rimer, 2001). Similarly, Han et al. (2009) found point and range formats were both susceptible to concerns over the source's credibility. Despite finding reduced understanding with range formats, Sladakovic et al. (2015) found no difference in levels of perceived risk, or severity of harm across range and point formats.

Another body of research, however, suggests *benefits* of using ranges in risk communications. It has long been suggested that being open about levels of uncertainty may lead to enhanced credibility (Chess et al., 1988). In this vein, range formats have been found to be not only more useful (Dieckmann et al., 2010), but communicators who use them are perceived as more honest (yet still competent) versus one who uses a point format (Johnson & Slovic, 1995). This may perhaps be because range formats openly acknowledge uncertainty at the outset (Joslyn & LeClerc, 2012). Research investigating outcome ranges indicates that the usefulness of a range may, however, depend on its width. Previous research has observed a preference for narrow outcome range forecasts (Du, Budescu, Shelly, & Omer, 2011) over wider outcome ranges, with the former perceived as more competent and trustworthy (Jørgensen, 2016). Indeed, an expert who makes a precise (narrow interval) forecast is perceived as more competent and trustworthy than one who makes a less precise (wider interval) forecast, despite the fact the former has a smaller chance of capturing the actual outcome than the latter (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 a probabilistically unexpected outcome occurred (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 formats. E. M. 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 ≤ 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 (for similar findings see also Young & Oppenheimer, 2009). 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).[27] In this approach, participants are shown a histogram of potential outcomes and asked to complete a probability statement (e.g., "It

[27] See also Juanchich and Sirota (2017); Teigen, Andersen, Alnes and Hesselberg (2019) and Teigen, Filkuková and Hohle (2018) for further research using the 'which-outcome' paradigm published since the current research was carried out.

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 in Chapter 2 (see also Juanchich et al., 2013).

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. Findings from Chapter 2, specifically the occurrence of the 'extremity effect' in verbal-numerical (V-N) expressions, highlight the possibility that such expressions could also be associated with a loss of confidence. 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'. 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 experiments.

The present chapter aims to advance our understanding of the effects of communication format on credibility by investigating the effect of 'erroneous' predictions, in both a low probability domain (an 'unlikely' event occurs), and a high probability domain (a 'likely' event does not occur). Experiments 7 (low probability) and 10 (high probability) incorporate mixed expressions (V-N, and the previously unconsidered N-V format) and compare them to a numerical point estimate and VPE. Given the benefits observed with use of numerical expressions, Experiments 8, 9 and 11 investigate the influence of the precision of the numerical expression, by including additional range (narrow/wide) formats, in conjunction with events differing in perceived predictability. All five experiments in this chapter additionally explore the potential influence of numeracy. Ascertaining the effect of communication format and the consequences of 'erroneous' predictions is vital to ensure that the public continue to perceive scientists as a credible source of information.

In light of findings from Chapter 2 and the literature reviewed above, the main hypotheses for the Experiments 7–9 (low probability domain) were as follows:

Hypothesis 1.

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

Hypothesis 2.

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

Hypothesis 3.

In Experiments 8 and 9, 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 7

Method.

Participants.

Three hundred Native English speakers were recruited from Amazon Mechanical Turk via the TurkPrime platform (Litman et al., 2017). 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.

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 trustworthiness and expertise. 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?" (Dieckmann et al., 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. 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 the main focus for analyses.[28]

As previous research has found a disparity between explicit ratings of trustworthiness and more implicit behavioural measures of trustworthiness for various information sources (Twyman, Harvey, & Harries, 2008), we used 'decision to evacuate'

[28] This was the case for all experiments within this chapter.

(Doyle, McClure, 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 the communicator's trustworthiness. 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 et al., 2012). The latter were included to increase variability in scores, given previous experiments in Chapter 2 found that MTurk samples tended 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 Appendix F 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 the communicator's trustworthiness and expertise. 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 trustworthiness 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 20 for a flow chart of the procedure).

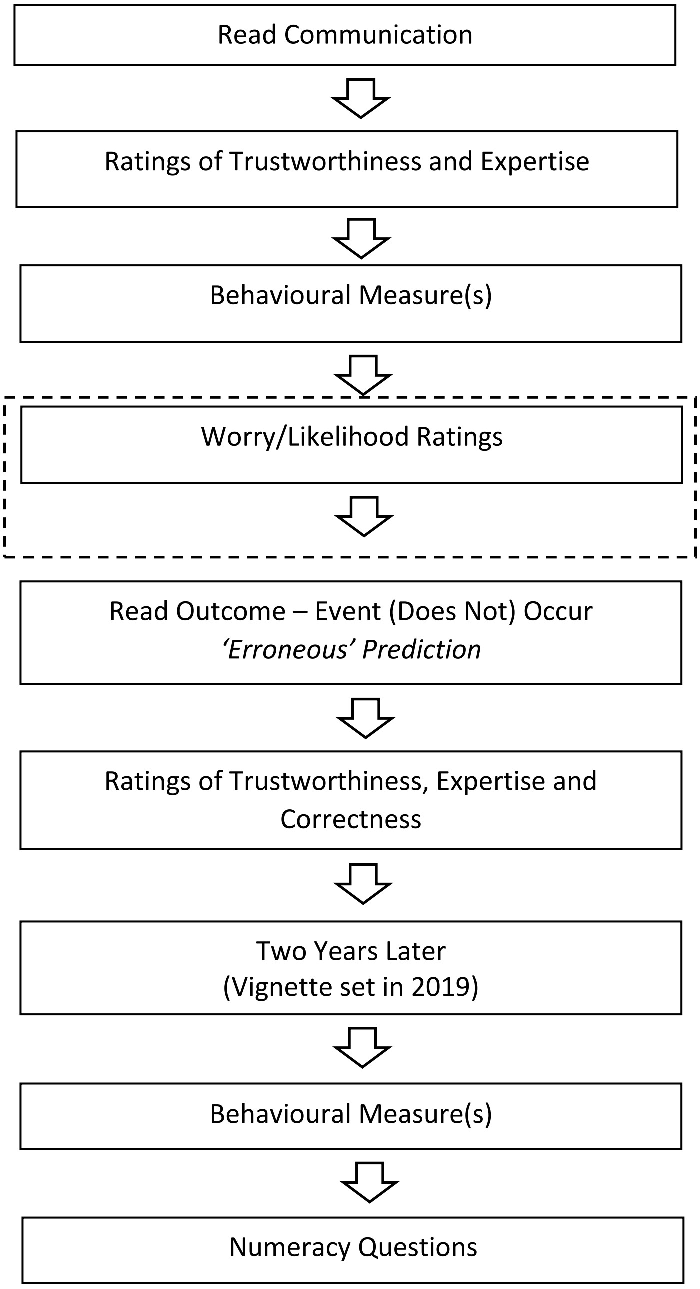


Figure 20. Procedure for Experiments 7–11. Dashed box represents additional measures included in Experiments 9 and 11.

Results.

As mentioned above, our focus was the change in ratings and decisions after an 'erroneous' prediction, so our analyses throughout this chapter focus on difference scores ([post-outcome rating] – [pre-outcome rating]). Analyses of all pre-outcome ratings are available in Appendix G (across all five experiments, there were no robust effects of communication format on pre-outcome credibility ratings).

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 6. 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.[29] Each dependent measure was entered into a 4 (communication format) × 2 (numeracy) ANOVA.

Change in credibility ratings.

Trustworthiness and expertise were positively correlated, both pre- and post-outcome (r = .60, p < .001, r = .76, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .74, α = .86, respectively). As can be seen in Figure 21, all communicators suffered from a loss of perceived credibility post-outcome, but the extent of the loss differed according to communication format, $F(3, 282) = 7.61$, $p < .001$, $\eta_p^2 = .08$. The numerical format was the most robust to the 'erroneous' prediction, with the greatest reduction in credibility observed for the verbal format. A post-hoc Gabriel's procedure 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).[30]

[29] This categorisation method was used across Experiments 7–10.

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

There was a trend for larger reductions in credibility in 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, $F(3, 282) = 0.85$, $p = .47$.[31]

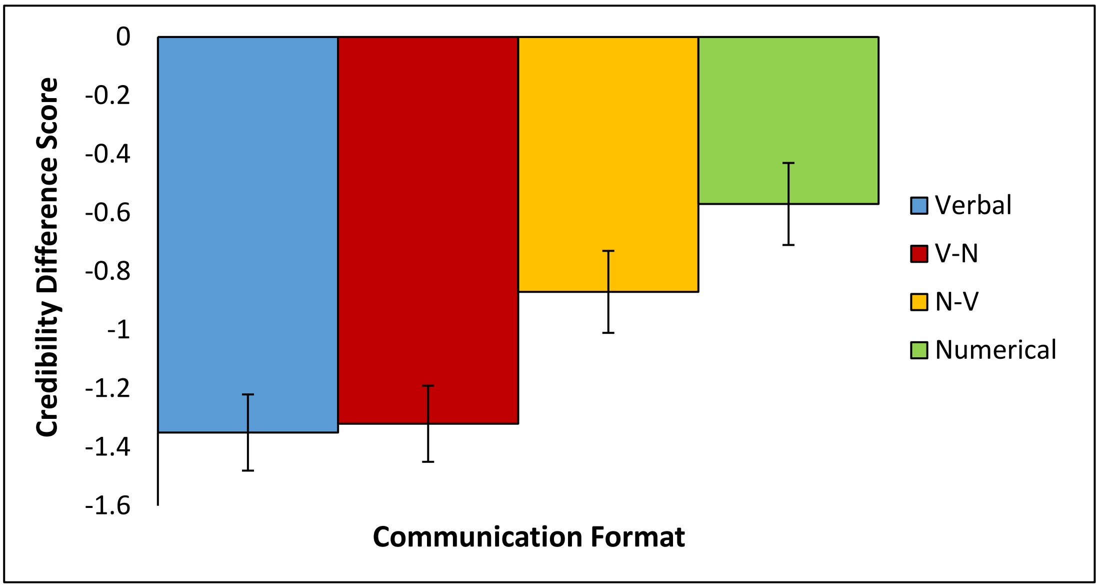


Figure 21. Experiment 7: Mean credibility difference scores by communication format (error bars represent $\pm$1 Standard Error [SE]).

[31] Across all experiments in this chapter, results for format × numeracy interactions remain non-significant if numeracy is entered as a continuous variable.

Table 6. Experiments 7–11: Distribution of Numeracy Scores (%)

	Numeracy Classification											
	Low									High		
Numeracy Score	0	1	2	3	4	5	6	7	8	9	10	Mean (SD)
Exp 7	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)
Exp 8	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)
Exp 9	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)
Exp 10	0.0	0.7	0.7	1.1	1.1	5.3	6.4	14.2	24.9	29.2	16.4	8.00 (1.70)
Exp 11*	0.0	0.8	0.4	2.4	4.9	7.3	24.9	33.9	24.5	0.8	0.0	6.55 (1.36)

*Given the highly skewed distribution of responses in Experiment 11, participants with scores of six or under were categorised as low numeracy, and those with seven or above categorised as high numeracy.

Correctness ratings.

Correctness ratings followed the pattern of differences in the credibility ratings, with the numerical format perceived as 'most correct' and the verbal format seen as 'least correct' after the outcome had occurred (see Figure 22), $F(3, 282) = 26.40$, $p < .001$, $\eta_p^2 = .22$. 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, $F(1, 282) = 1.69$, $p = .20$, nor an interaction between format and numeracy, $F(3, 282) = 0.41$, $p = .75$.[32]

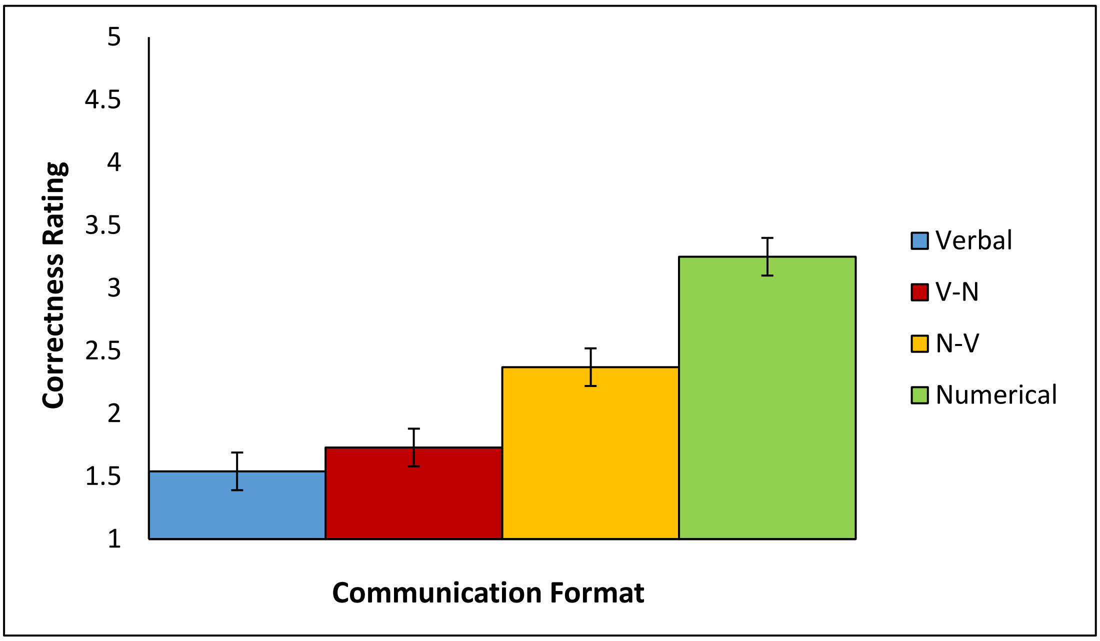


Figure 22. Experiment 7: Mean correctness ratings by communication format (error bars represent ±1 SE).

Decisions to evacuate.

There was no significant effect of communication format on evacuation difference scores, $F(3, 282) = 1.45$, $p = .23$ (see Table 7). There was no effect of numeracy, $F(1, 282) = 0.04$, $p = .84$, nor an interaction between the two, $F(3, 282) = 1.42$, $p = .24$.

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

Table 7. Experiments 7–10: Mean Behavioural Measure Scores 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
7. *n*	75	70	74	72	N/A	N/A
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
8. *n*	78	82	N/A	N/A	83	N/A
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
9. *n*	N/A	84	N/A	N/A	83	79
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

Discussion.

All formats suffered a reduction in credibility after the 'erroneous' prediction. Consistent with Hypothesis 1, the numerical (verbal) format was perceived as most (least) correct. The poor performance of the verbal format can 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 21 and Figure 22), in line with findings from Chapter 2.

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 low probability experiments.[33]

Experiment 8

In Experiment 7, the numerical format outperformed both verbal and mixed formats. Numerical point formats have, however, previously been criticised for portraying a false level of certainty (Schwartz et al., 1999). Experiment 8 was therefore designed to extend the type of probability expressions 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.

[33] Experiment 10 (featuring high probability expressions) was run before Experiments 8 and 9.

Method.

Participants.

Two hundred and fifty participants were recruited from Amazon Mechanical Turk via the TurkPrime platform (Litman et al., 2017), 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.

Design.

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

Credibility measures were the same as in Experiment 7, with the exception of trustworthiness, where an additional clause was added. Trustworthiness was operationalised as: "How much do you trust that the expert is giving you complete and unbiased information, to their best of their knowledge?" (Dieckmann et al., 2009), rated from 1 (*Not at all*) to 5 (*A great deal*). For the behavioural measures, participants were first given a 'willingness to pay' measure in the context of home insurance (see Figure 23),[34] and then also completed an adapted version of the tolerable risk scale (Haynes et al., 2008b).[35] Participants also completed the mixed numeracy scale (Cokely et al., 2012; Lipkus, Samsa, et al., 2001), as in Experiment 7.

[34] Values for this task were selected on the basis of results from pilot testing.
[35] Further details and analysis of this scale is presented in Appendix H. No significant effects or interactions were observed.

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 23. Experiment 8: Willingness to pay home insurance measure, forest fire vignette, numerical point condition.

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 a current forest fire or flood in which a geologist communicated the risk of fire or floodwater travelling a certain distance (see Appendix I for both vignettes). For instance:

> "A site in South Redmill has a history of forest fires due to a lack of rainfall in the area. A forest fire is currently blazing. Geologists from Redmill Geological Centre are communicating information about the forest fire. A geologist has suggested that, given the recent weather, it is **unlikely** that the forest fire will extend 80km from its origin" (Verbal condition, emphasis in original).

The procedure was the same as Experiment 7 (see Figure 20), though included the different behavioural measures.

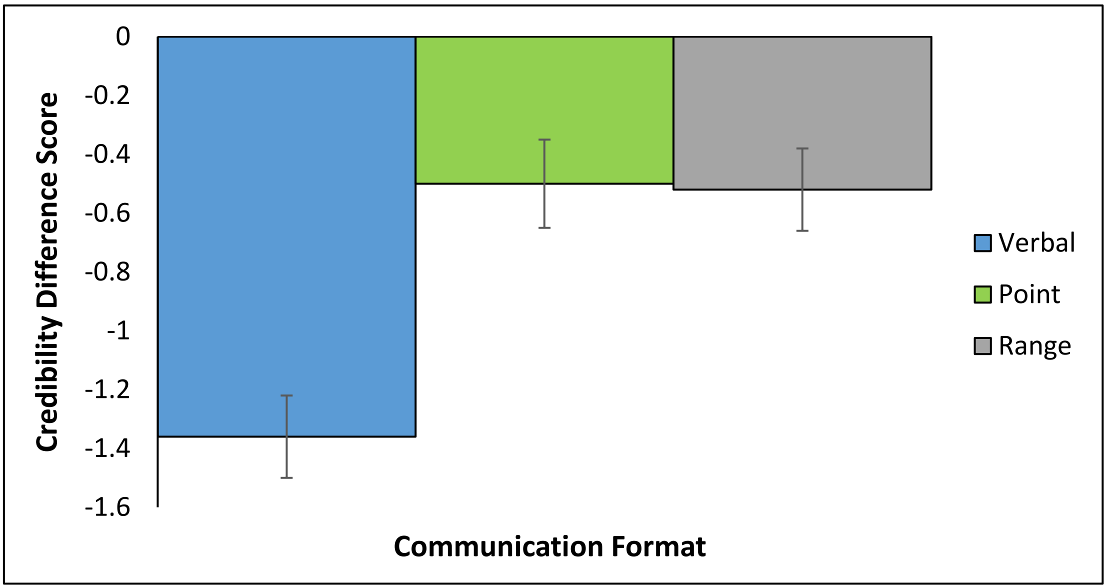
Results.

The distribution of numeracy scores is shown in

Table 6. Each dependent measure was entered into a 3 (communication format) × 2 (scenario) × 2 (numeracy) ANOVA.

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .68, p < .001, r = .75, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .81, α = .86, respectively). As can be seen in Figure 24, all communicators suffered from a loss of perceived credibility post-outcome, but the extent of the loss differed according to communication format, $F(2, 231) = 11.58$, $p < .001$, $\eta_p^2 = .09$. The greatest reduction in credibility was observed for the verbal format 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 (p = .10). There was no significant effect of scenario, $F(1, 231) = 0.53$, p = .47, or numeracy, $F(1, 231) = 0.04$, p = .85. No two- or three-way interactions



were significant (all ps > .09).

Figure 24. Experiment 8: Mean credibility difference scores by communication format (error bars represent ±1 SE).

Correctness ratings.

Correctness ratings followed the pattern of the credibility ratings (see Figure 25), with the numerical point and range formats perceived as 'most correct' and the verbal format seen as 'least correct' after the outcome had occurred, $F(2, 231) = 33.97$, $p < .001$, $\eta_p^2 = .23$. A post-hoc Gabriel's procedure demonstrated that the verbal format was significantly different to the point ($p < .001$) and range ($p < .001$) formats, though the latter two were similarly perceived ($p = 1.00$). There was no significant effect of scenario, $F(1, 231) = 1.96$, $p = .16$, nor numeracy $F(1, 231) = 1.76$, $p = .19$, nor any two- or three-way interactions (all *p*s > .38).

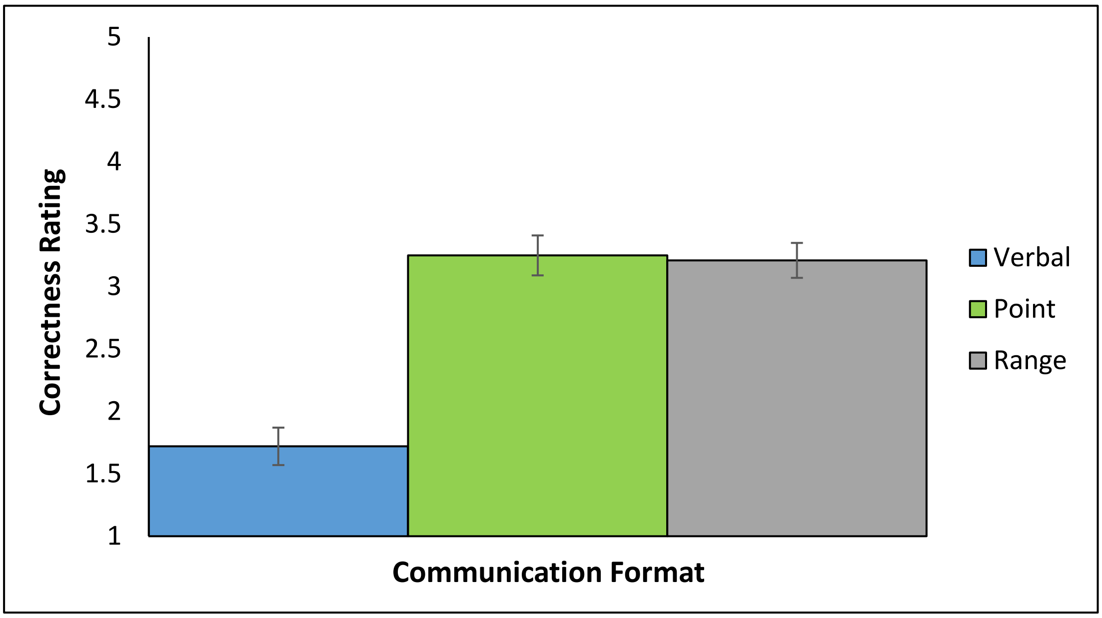


Figure 25. Experiment 8: Mean correctness ratings by communication format (error bars represent $\pm$1 SE).

Insurance decision.

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 7), but this was not significantly affected by communication format, $F(2, 218) = 1.12$, $p = .33$, scenario, $F(1, 218) = 2.38$, $p = .13$, or numeracy, $F(1, 218) = 0.74$, $p = .39$. No two- or three-way interactions were significant (all *p*s > .09).

Discussion.

Whilst we observed an effect of communication 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 communication format is driven solely by the verbal format, replicating findings from Experiment 7. 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, Klein 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 9.

Experiment 9

Experiment 9 was designed to examine whether the precision of the range would affect perceptions of credibility, following 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 7 and 8. We continue to use the range featured in Experiment 8, hereafter referred 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, Klein 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, 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 experiment.

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.

The dependent variables were as in Experiment 8, with two changes. The tolerable risk scale was removed and two additional (pre-outcome) 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 [e.g.,] there is a 20% likelihood that the forest fire will extend 80km from its origin" 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*) to 7 (*Completely certain*).

Results.

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .70, p < .001, r = .80, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .82, α = .89, respectively). There seemed a smaller loss of credibility for the narrow range format compared to the other numerical formats (see Figure 26). However, a 3 × 2 × 2 mixed ANOVA revealed there were no significant differences between communication formats, $F(2, 234) = 1.85$, $p = .16$. There was also no effect of scenario, $F(1, 234) = 1.57$, $p = .21$. The communicator suffered a greater reduction in perceived credibility when rated by the low numeracy group, $F(1, 234) = 10.96$, $p = .001$, $\eta_p^2 = .05$ (see

Table 6 for distribution of numeracy scores), though there was a significant scenario × numeracy interaction, $F(1, 234) = 4.97$, $p = .03$, $\eta_p^2 = .02$. Figure 27 demonstrates this interaction: whilst the credibility ratings of highly numerate participants were similarly affected in the flood and fire scenarios, $F(1, 234) = 0.55$, $p = .46$, ratings of those lower in numeracy were less affected by the 'erroneous' prediction in the flood scenario, $F(1, 234) = 5.34$, $p = .02$, $\eta_p^2 = .02$. 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). No other two- or three-way interactions were significant (all ps > .56).

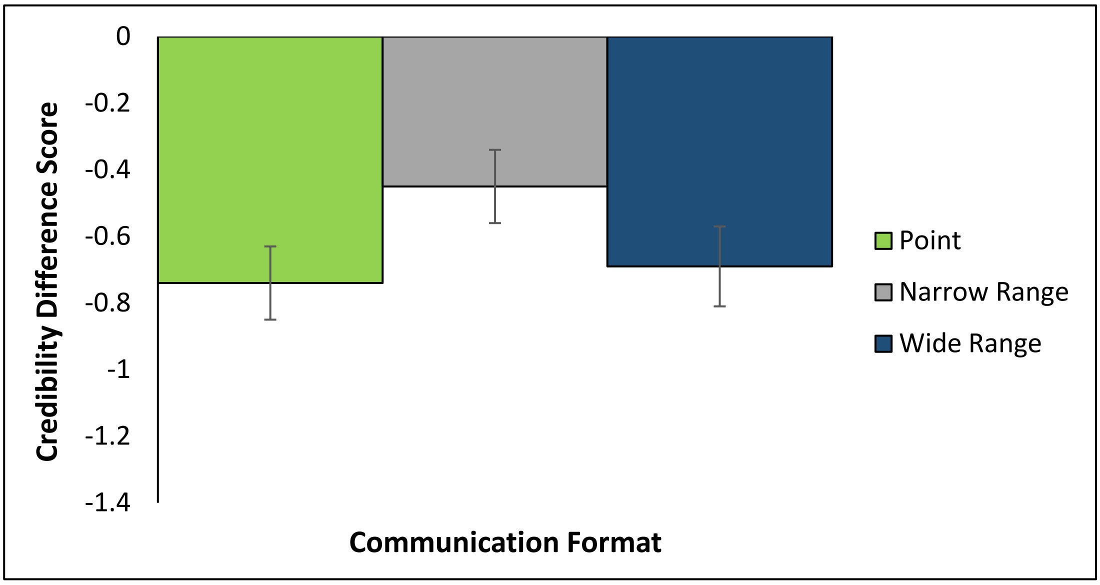


Figure 26. Experiment 9: Mean credibility difference scores by communication format (error bars represent ±1 SE).

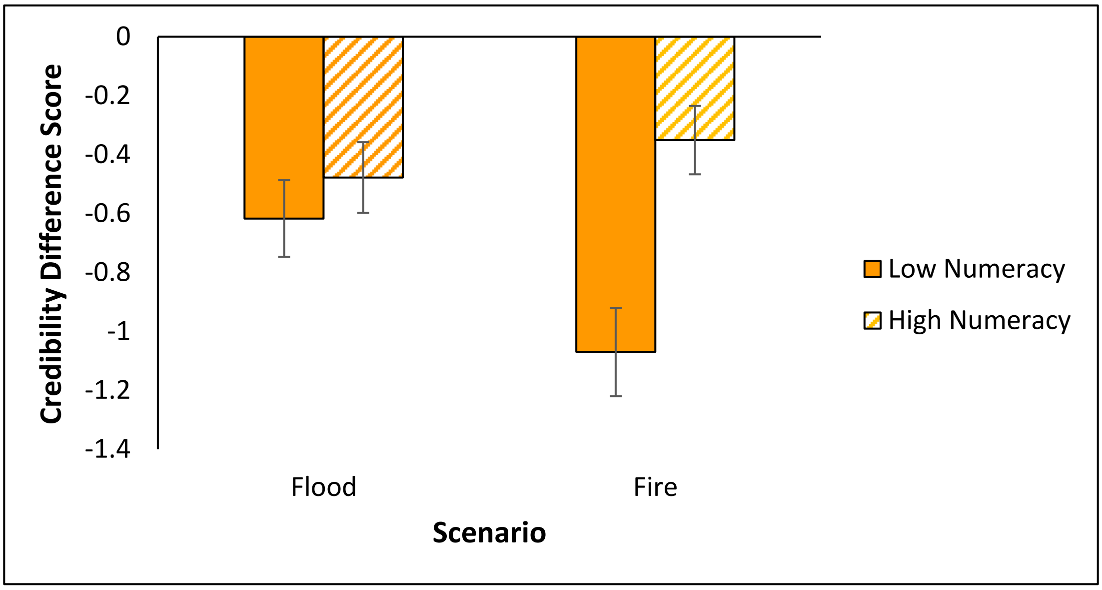


Figure 27. Experiment 9: Mean credibility difference scores – the scenario × numeracy interaction (error bars represent ±1 SE).

Correctness ratings.

A 3 × 2 × 2 ANOVA revealed no significant effect of communication format, *F(*2, 234) = 1.03, *p* = .36, or scenario, *F(*1, 234) = 0.79, *p* = .38, on correctness ratings. The significant effect of numeracy can be seen in Figure 28 – 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* = .03, η_p^2= .02. There were no significant two- or three- way interactions (all *p*s > .14).

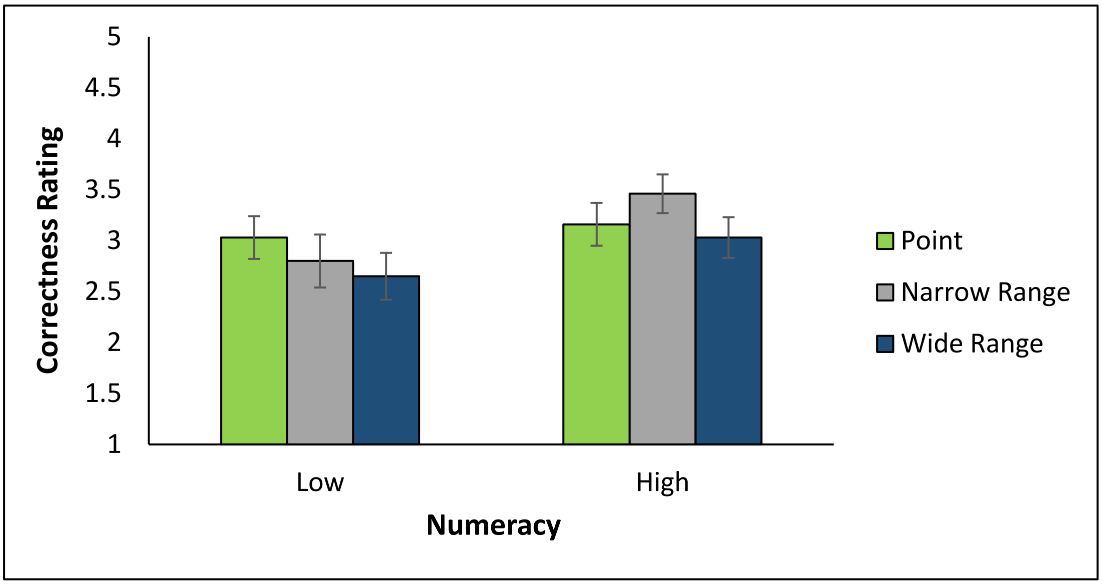


Figure 28. Experiment 9: Mean correctness ratings by communication format and numeracy (high/low) (error bars represent ±1 SE).

Insurance decision.

Those who provided inconsistent responses were removed as in Experiment 8. Participants were willing to pay more for insurance after the outcome, though a 3 × 2 × 2 mixed ANOVA revealed no significant effect of communication format, $F(2, 223) = 0.83$, $p = .44$, scenario, $F(1, 223) = 0.62$, $p = .43$, or numeracy, $F(1, 223) = 0.49$, $p = .48$. There was a significant interaction between communication format and numeracy, $F(1, 223) = 4.38$, $p = .01$, $\eta_p^2 = .04$ (see Table 7). Both the high and low numeracy groups were willing to pay similar amounts for insurance in the wide range, $F(1, 223) = 1.30$, $p = .26$, and narrow range conditions, $F(1, 223) = 0.12$, $p = .73$. 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 *p*s > .17).

Worry/concern ratings.

A 3 × 2 × 2 mixed ANOVA revealed no significant effect of communication format, $F(2, 234) = 0.83$, $p = .44$, scenario, $F(1, 234) = 0.59$, $p = .44$, or numeracy, $F(1, 234) = 65$, $p = .42$, on worry ratings. No significant two- or three-way interactions were observed (all *p*s > .20).

Likelihood ratings.

A 3 × 2 × 2 mixed ANOVA revealed no significant effects of communication format, $F(2, 234) = 0.06$, $p = .95$ or scenario, $F(1, 234) = 0.70$, $p = .40$. There was a 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), $F(1, 234) = 3.56$, $p = .06$, $\eta_p^2 = .02$.

Discussion.

This experiment found 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 likelihood 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 current chapter to have observed a moderating effect of numeracy on perceptions. Given that we have not observed an effect of numeracy previously, and in light of the fact that our sample was relatively numerate, with little separating 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.

So far, the experiments have only considered 'erroneous' predictions in the low probability domain. Experiments 10 and 11 sought to investigate if the effect of communication format would occur within a high probability domain (a 'likely' event does not occur).

Experiment 10

Extending the work to the high probability domain is important for several reasons. First and foremost, communications about high probability events may simply be more common. In the Intergovernmental Panel on Climate Change's (IPCC) Fifth Report, the word 'likely' was mentioned 10-20 times more often than its opposite counterpart 'unlikely' (Fløttum & Dahl, 2011, as cited in Teigen, 2014). Moreover, it is far more useful for the public to know what events might occur, rather than what might not occur. This notion ties in with Grice's (1975) 'maxim of quantity', which states that communications should be as informative as is required for the purpose of the exchange. Secondly, investigating whether credibility is affected following the non-occurrence of a high probability event is particularly relevant in a natural hazards context, given fears of 'crying wolf' have been documented amongst decision-makers (e.g., L. R. Barnes, Gruntfest, Hayden, Schultz, & Benight, 2007; Sharma & Patt, 2012; Woo, 2008).

Furthermore, results of previous research highlight the possibility that perceptions of a communicator might be different for the two probability domains. Bagchi and Ince (2016) found that a forecaster who made a higher forecast (70%) was perceived as more confident, trustworthy and accurate than one who made a lower forecast (30%), despite participants never being informed of the actual outcome. Similarly, Teigen and Brun (2003) observed a trend in which the higher the probability a VPE was perceived to reflect, the more correct it were considered.

Relatedly, Keren and Teigen (2001b) observed a preference for probabilities close to the extreme, as well as a preference for high over low probabilities,[36] with higher probabilities perceived as more valuable (Keren & Teigen, 2001b). This preference for higher probabilities was most commonly a result of the belief that they were more certain and aided decision-making. They also found that participants preferred an overconfident forecaster over a well-calibrated one – a forecaster who provided a higher probability estimate (90%) was preferred than a forecaster who provided a lower probability estimate (75%), even when the latter forecaster was perfectly calibrated.

Furthermore, findings from Keren and Teigen (2001a) indicate that numerical probabilities are perceived as indicating not only information about likelihood, but also characteristics of the predicted outcome, such as its magnitude and timing, which could have implications for the way the prediction is evaluated. Specifically, high probabilities are associated with strong, immediate outcomes, whereas low probabilities are associated with weak, delayed outcomes. It is possible, therefore, that in the context of an 'erroneous' prediction in the high probability domain, a communicator may lose a greater amount of credibility because the outcome was expected to be both strong and immediate, which was more of a contrast when the outcome did not occur.

[36] Though their use of the term 'high' and 'low' probability stemmed from comparisons of numbers within a probability domain (e.g., '70%' v '90%'; '30%' v '10%') rather than across probability domains (e.g., '70%' v '30%').

Extending the 'erroneous' prediction paradigm to the high probability domain was primarily done in order to investigate if the effect of communication format seen in the low probability domain held.

Method.

Participants.

Two hundred and ninety-nine Native English speakers were recruited from Amazon MTurk via the TurkPrime platform (Litman et al., 2017). Seventeen cases were removed for failing the attention check, leaving a final sample of 281 participants (138 male) aged between 18–74 (*Mdn* = 32). Participants received $0.60.

Design, materials and procedure.

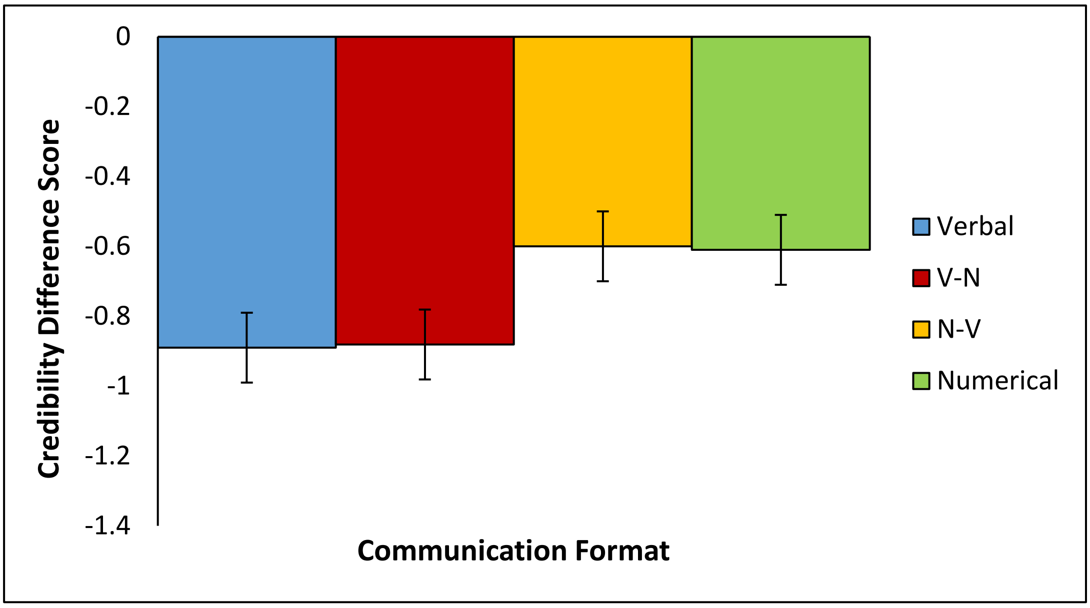
Design, materials and procedure were as in Experiment 7, though this time communication format was set in the high probability domain: verbal – 'likely'; numerical – '80% likelihood'; V-N – 'likely (80% likelihood)' and N-V – '80% likelihood (likely)'. Post-outcome, the likely event did *not* occur.

Results.

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .60, $p < .001$, r = .74, $p < .001$, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .74, α = .85, respectively). All communicators suffered from a loss of perceived credibility following the 'erroneous' prediction. There was a trend for a communicator who used a verbal or V-N format to suffer more of a loss in credibility than one who used a numerical point, or N-V format, $F(3, 273) = 2.53$, $p = .06$, $\eta_p^2 = 0.03$ (see Figure 29). A post-hoc Gabriel test revealed there were no significant differences between communication formats (all ps > .20). There was no significant effect of numeracy, $F(1, 273) = 0.62$, $p = .43$ (see

Table 6 for distribution of numeracy scores). There was a trend towards an interaction between communication format and numeracy, $F(3, 273) = 2.53$, $p = .06$, $\eta_p^2 = .03$. Simple effects analyses showed a significant effect of communication format for low numeracy



group, $F(3, 273) = 3.16$, $p < .001$, $\eta_p^2 = .08$, but not for the high numeracy group, $F(3, 273) = 1.88$, $p = .13$.

Figure 29. Experiment 10: Mean credibility difference scores by communication format (error bars represent $\pm$1 SE).

Correctness ratings.

As was found in the low probability experiments, there was a significant effect of communication format on correctness ratings, $F(3, 273) = 4.90$, $p = .002$, $\eta_p^2 = 0.05$. A post-hoc Gabriel's procedure demonstrated a significant difference between the verbal and both numerical ($p < .001$) and N-V ($p < .001$) formats. Replicating results in the low probability domain, the numerical format was seen as 'most correct' and the verbal format seen as 'least correct' (see Figure 30). There was no significant effect of numeracy, $F(1, 273) = 2.70$, $p = .10$, nor was there an interaction between communication format and numeracy, $F(3, 273) = 0.20$, $p = .90$.

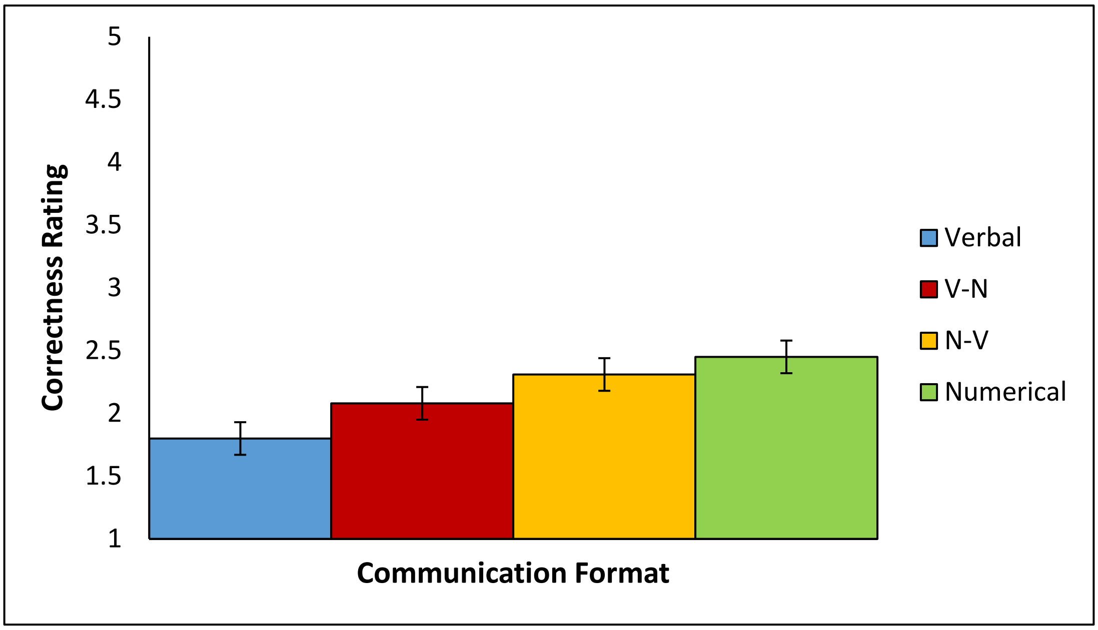


Figure 30. Experiment 10: Mean correctness ratings by communication format (error bars represent ±1 SE).

Decisions to evacuate.

There was no significant effect of communication format on evacuation difference scores, $F(3, 273) = 0.22$, $p = .89$. There was no significant effect of numeracy, $F(1, 273) = 2.11$, $p = .15$, nor was there an interaction between communication format and numeracy, $F(3, 273) = 0.54$, $p = .66$.

Discussion.

This experiment provides partial support for the robustness of the numerical format following an 'erroneous' prediction. There was a trend for a communicator who used a numerical format to lose less credibility than one who used a verbal format, an effect which was more clearly demonstrated in correctness ratings. The fact that no evidence for an effect of communication format was observed for evacuation ratings is unsurprising, given responses in the open question indicated a desire to evacuate to be on the safe side, similar to responses in Experiment 7. The behavioural decision task was thus changed in the following experiment.

Experiment 11

Whilst Experiments 8 and 9 did not observe a difference between numerical range and point formats, it is possible that ranges could be perceived differently when used to convey higher probabilities. Thus as in Experiment 9, we sought to manipulate the precision of the range, using a narrow – ('75–85% likelihood') and wide ('65–95% likelihood') range format, exploring both perceptions of credibility, and behavioural decisions, in the form of the previously used insurance task. The experiment also provided an opportunity to test whether the difference between verbal and numerical point formats seen in Experiment 10 replicated.

Whilst some studies have found that range formats are perceived as more useful and more honest (Dieckmann et al., 2010; Johnson & Slovic, 1995), these have been set in the low probability domain. It is possible that in a high probability context, use of a range format could be more likely to be perceived as indicative of incompetence, ignorance and/or a lack of confidence, as a range conflicts with the assumption that high probabilities are more certain (Keren & Teigen, 2001b) and reflect higher levels of confidence (Bagchi & Ince, 2016). Furthermore, following the non-occurrence of an outcome, one might think that range formats should have included probabilities closer to, or under, 50%. Thus, we predicted that a communicator who used a range format would be perceived as less credible than one who used a point format.

Method

Participants.

Two hundred and fifty Native English speakers were recruited via Prolific Academic and paid £0.85 for participation. Four cases were removed for failing the attention check, with one for duplicate IP address, leaving a total of 245 participants (156 female, 2 selecting other), aged between 18–72 (*Mdn* = 33).

Design, materials and procedure.

Communication format was manipulated between-participants and had four levels: verbal – ‘likely’; numerical point – ‘80% likelihood’; narrow range – ‘75–85% likelihood’ and wide range – ‘65–95% likelihood’. Scenario had two levels, manipulated between-participants. The predictability of the event was manipulated (based on pilot work), with forest fire and flood scenarios, as in Experiments 8 and 9.

The measures completed were as in Experiment 9 – participants gave ratings of trustworthiness, expertise, worry/concern, and likelihood, though worry/concern was rated on a scale ranging from 1 (*Not at all worried/concerned*) to 7 (*Extremely worried/concerned*). Participants also completed the numeracy questions.

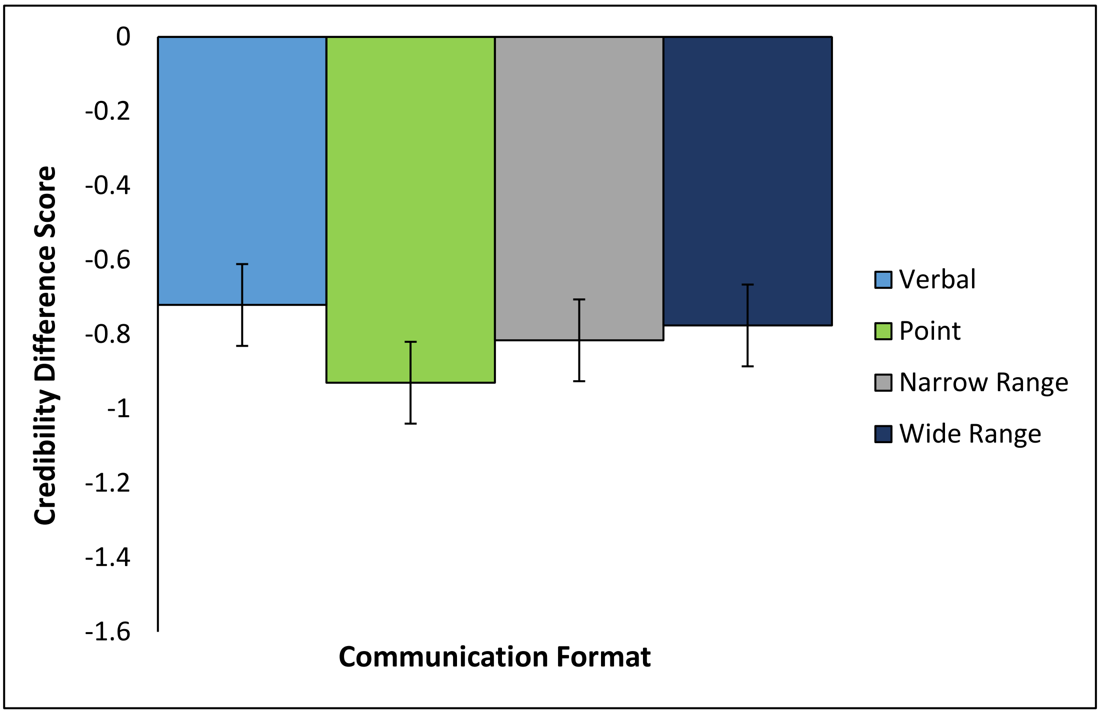
Results.

The distribution of numeracy scores can be seen in

Table 6. Given the slightly lower numeracy scores (relative to Experiments 7–10), numeracy was categorized differently, with scores of six or under categorised as low numeracy, and scores of seven or above categorised as high numeracy. Each dependent variable was entered into a 4 (communication format) × 2 (scenario) × 2 (numeracy) ANOVA.

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .65, p < .001, r = .71, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .79, α = .83, respectively). All communicators suffered from a



loss of perceived credibility following the event's non-occurrence, but the level of reduction did not significantly differ according to communication format, $F(3, 229) = 0.63$, p = .60 (see Figure 31). There was no significant effect of numeracy, $F(1, 229) = 1.08$, p = .30, nor scenario, $F(1, 229) = 0.02$, p = .88. No two- or three-way interactions were significant (all ps > .10).

Figure 31. Experiment 11: Mean credibility difference scores by communication format (error bars represent ±1 SE).

Correctness ratings.

Correctness ratings were not affected by communication format, $F(3, 229) = 0.95$, $p = .42$ (see Table 8), nor were ratings affect by scenario, $F(1, 229) = 1.27$, $p = .26$, or numeracy, $F(1, 229) = 2.73$, $p = .10$. No two- or three-way interactions were significant (all *p*s > .54).

Table 8. Experiment 11: Mean Scores for Each Dependent Variable by Communication Format

Dependent Variable	Communication Format – Mean Score (SE)			
	Verbal	Point	Narrow Range	Wide Range
Credibility Difference Score	- 0.72 (0.11)	- 0.93 (0.12)	- 0.82 (0.11)	- 0.78 (0.11)
Correctness	1.92 (0.14)	2.16 (0.16)	2.25 (0.14)	2.10 (0.14)
Insurance Difference Score (£)	- 23.97 (19.11)	- 22.75 (20.94)	- 30.55 (19.00)	- 36.86 (19.20)
Worry/Concern	5.84 (0.14)	5.71 (0.15)	6.07 (0.14)	6.05 (0.14)
Likelihood	5.15 (0.13)	5.44 (0.14)	5.38 (0.12)	5.35 (0.13)

Insurance decision.

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. Participants were less willing to pay for insurance after the event did not occur, but this was not significantly affected by communication format, $F(3, 208) = 0.26$ $p = .85$ (see Table 8), scenario, $F(1, 208) = 1.39$, $p = .24$, or numeracy, $F(1, 208) = 0.25$, $p = .62$. No two- or three-way interactions were significant (all *p*s > .30).

Worry/concern ratings.

There was no significant effect of communication format, $F(3, 229) = 1.47$, $p = .22$ (see Table 8), scenario, $F(1, 229) = 2.28$, $p = .13$, or numeracy, $F(1, 229) = 1.25$, $p = .27$, on worry/concern ratings. There was an unpredicted significant three-way interaction between communication format, scenario and numeracy, $F(3, 229) = 3.09$, $p = .03$, $\eta_p^2 = .04$, which was not clearly interpretable (see Figure 32). No other interactions were significant, (all *p*s > .13).

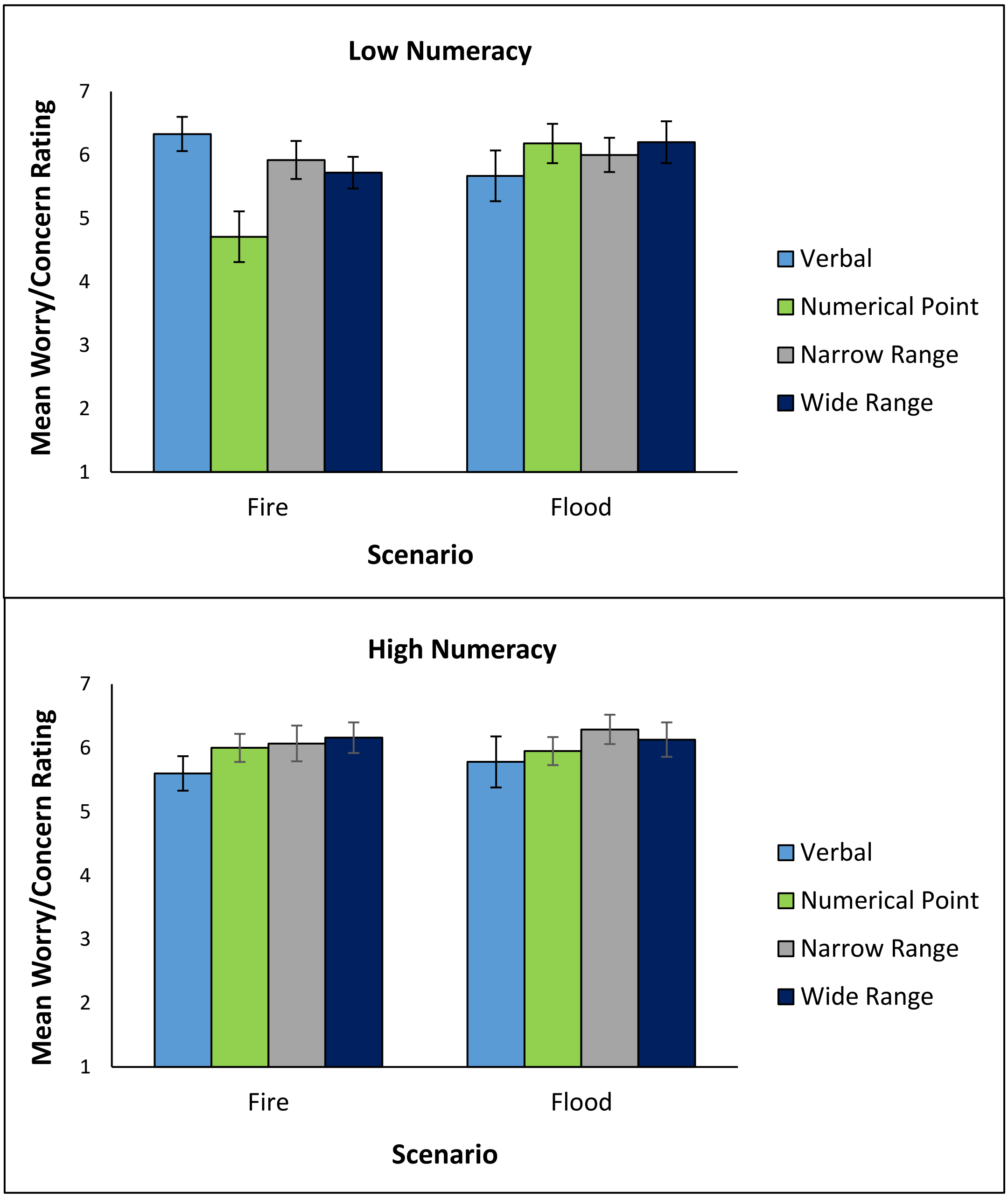


Figure 32. Experiment 11: Mean worry/concern rating – the communication format × scenario × numeracy interaction (error bars represent ±1 SE).

Likelihood ratings.

There was no significant effect of communication format, $F(3, 229) = 0.94$, $p = .42$, scenario, $F(1, 229) = 2.85$, $p = .09$, or numeracy, $F(1, 229) = 0.04$, $p = .83$, on likelihood ratings. No two- or three-way interactions were significant (all *ps* > .09).

Discussion.

Similar to results from Experiments 8 and 9, there does not appear to be difference in the way that numerical point and range formats are perceived, with the precision of the range having little impact. Of particular note is the failure to replicate findings from Experiment 10, where an effect of communication format was seen for correctness ratings and (less strongly) for credibility difference scores. The effect of communication format on perceptions following an ‘erroneous’ prediction therefore seems to be different across high and low probability domains. A consideration of possible explanations for this difference follows below.

General Discussion

In this chapter we present five 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, across high and low probability domains. These experiments also considered the potential influence of numeracy. Experiments set in the low probability domain (7–9) yielded consistent results, with the verbal (numerical point) format most (least) vulnerable to an ‘erroneous’ prediction. Such a difference between verbal and numerical formats was not consistently observed in the high probability domain (Experiments 10 and 11). 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), Experiments 7 and 10 provided little evidence of their added benefit, either in terms of credibility, correctness or behavioural decisions. Experiments 8, 9 and 11 yielded no evidence to suggest differences in the way numerical point and range formats are perceived, nor differences in their vulnerability to ‘erroneous’ predictions.

Directionality.

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 – through explicit mention of one of the complementary outcomes, or indirectly – through directionality (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 2003). The negatively directional term 'unlikely' was used in the current low probability 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 – there was 'directionality–outcome incongruence'. In contrast, the numerical formats lacked this original (negative) focus. In fact, some evidence even indicates that numerical expressions are biased towards positive interpretations (Teigen & Brun, 2000). 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 suggest that a 'directionality–outcome congruence' account can also explain the difference in the effect of communication format between the high and low probability domains. In the low probability domain, the verbal and numerical formats differed in directionality. However, in the high probability domain, 'likely' and '80% likelihood' could be said to both be positively directional. As a result, they were both incongruent with the outcome's non-occurrence and thus lead to similar reductions in credibility.

We acknowledge that the current results (at least for the low probability domain) could have arisen from other differences in the way that 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. Research in Chapter 4 aims to test the 'directionality–outcome congruence' account, by comparing predictions featuring negatively and positively directional VPEs across both high and low probability domains. Although the low probability experiments so far have focused on one low probability VPE, 'unlikely', on the basis of a 'directionality–outcome congruence' account, we suggest our findings would also extend to other negatively directional VPEs such as 'improbable'. In contrast, we would expect a smaller decrease in credibility ratings when positively directional VPEs such as 'a chance' were used, 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 (Experiments 8, 9 and 11), 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 (H. O. 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 trade-offs, 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 (Fischhoff, 2011; Frewer, Hunt, 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 (in a range versus point format). 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 et al., 2007).

Chapter Summary

Previous research has predominantly studied the effects of communication format on understanding, largely neglecting format's influence on credibility and behavioural decisions. Research in this chapter 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 communication strategies. Findings from these experiments show that numerical formats are consistently perceived as more credible and less incorrect following an 'erroneous' prediction, at least in the low probability domain. Across both high and low probability domains, no differences are observed between perceptions of numerical point and range formats. That the effect of communication format following an 'erroneous' prediction differs across the two probability domains could be due to the difference in the directionality of the expressions – an explanation tested in Chapter 4.

Chapter 4. The Role of Directionality

Chapter Overview

Risk communicators often need to communicate information about the likelihood of uncertain events. On occasion, an event with 10% likelihood will occur, or one with 90% likelihood will not – a probabilistically unexpected outcome. Manipulating communication format, Chapter 3 showed that communicators lose more credibility, and are perceived as less correct, when an 'unlikely' event occurs than when a '10–30% likelihood' event occurs. The expression 'unlikely' has negative directionality, which directs attention towards the non-occurrence of an event. Thus, when the 'unlikely' event occurs, the prediction and outcome are directionally incongruent. The inconsistent effects of communication format in the high probability domain may have arisen because both 'likely' and '80% likelihood' are positively directional, and thus were both directionally-incongruent when the outcome did not occur. We therefore examine whether this 'directionality–outcome congruence' is the key characteristic underlying the perception of predictions as 'erroneous'. We compare positively and negatively directional communications across high and low probability events, in the context of both a probabilistically unexpected outcome (Experiment 12) and a probabilistically expected outcome (Experiment 13). We find communicators are generally perceived less favourably (in terms of credibility and correctness ratings) given directionality–outcome incongruence, with such incongruence also linked with higher levels of surprise.

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 has therefore been identified as one of a communicator’s most precious assets (Covello & Allen, 1988). In the current chapter, we consider how the communication format used to describe the (un)certainty of an event influences the maintenance of credibility in uncertainty 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’) in order to reduce misunderstandings (e.g., Budescu, Broomell, & Por, 2009). Relatively little research has examined the effect of format 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). The majority of this research has focused on whether credibility ratings of a communicator are affected by the format of the communication (e.g., VPE or numerical estimate). More recently, researchers have extended this to investigate how credibility ratings of communicators change once focal event outcomes are known (Dieckmann, Mauro, & Slovic, 2010; see also Chapter 3). The latter context 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).

Results from Chapter 3 showed 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'. We posit that the differences in the effect of format, and the disparity between findings for 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').[37] 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.

We suggest that the results in Chapter 3, 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

[37] See also Moxey and Sanford (2000) for a review of similar work in relation to quantifiers.

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 also 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 '80% 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.[38]

Overview of experiments.

The present chapter aims to gain an understanding of the mechanisms behind the effects of communication format on credibility by testing the 'directionality–outcome congruence' account. We firstly test the account in the context of a probabilistically unexpected outcome (an 'erroneous' prediction – Experiment 12) and then in the context of a probabilistically expected outcome (Experiment 13). Having previously only focused on the VPEs 'likely' and

[38] 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 Experiment 10. This is not problematic for our account, as it has previously been noted that numerical expressions are more directionally ambiguous than VPEs (Budescu et al., 2003; Teigen & Brun, 1995).

‘unlikely’, the present research provided a chance to examine whether findings from Chapter 3 would replicate with different VPEs, which share the same directionality and approximate numerical probability translation. 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 12

Our predictions first focus 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.

Hypothesis 2_A.

As was found in Experiments 7 – 9, 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 ≥ .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 Experiments 7 and 8). 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 et al., 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.

Design.

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

A series of pilot studies featuring 18 VPEs was carried out in order to identify four VPEs for use in Experiments 12 and 13 (a positive and negative 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 Appendix J). 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), as used in Experiments 8, 9 and 11.

The dependent variables used were as in previous experiments in Chapter 3: 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 unexpected outcomes, credibility difference scores ([post-outcome credibility rating] – [pre-outcome credibility rating]) were the main focus for analyses.[39] 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).

[39] Credibility difference scores continued to be the focus in Experiment 13.

Procedure and materials.

First, participants were asked to respond to a number of demographic questions, including age and gender, before completing the attention check. Participants then read a 'flood' vignette (see Appendix I) 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 communicator's trustworthiness and expertise. 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 ratings of trustworthiness and expertise, as well as rating how correct the geologist's prediction was in light of the outcome and how surprised they were (see Figure 33 for a flow chart of the procedure).

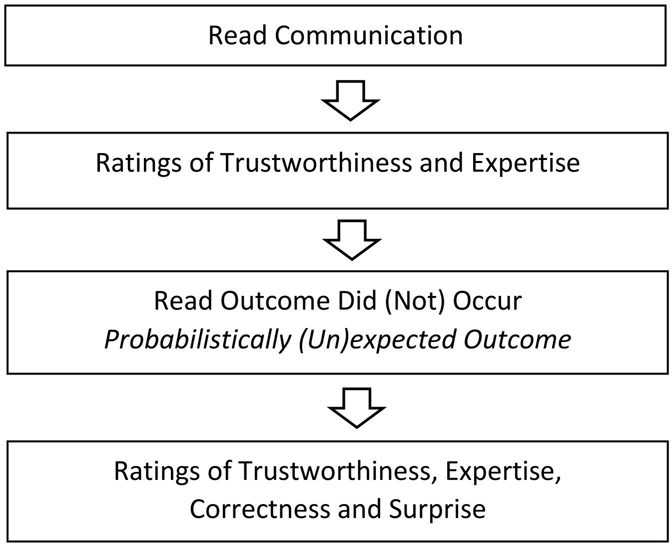


Figure 33. Procedure for Experiments 12 and 13.

Results.

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.

Analysis of the pre-outcome credibility ratings is available in Appendix K. There was an interaction between probability and communication format, with a significant effect of communication format for the high, but not the low, probability expressions, with ‘not entirely definite’ perceived as less credible than ‘70–90% likelihood’, and ‘good chance’. The following analyses were planned, as specified in the pre-registration.

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .75, p < .001, r = .79, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .85, α = .88, respectively). A 2 (congruence) × 2 (probability) ANOVA 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 34). 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, $F(1, 285) = 49.56$, $p < .001$, $\eta_p^2 = .15$, not for the low probability expressions: $F(1, 285) = 2.26$, $p = .13$.

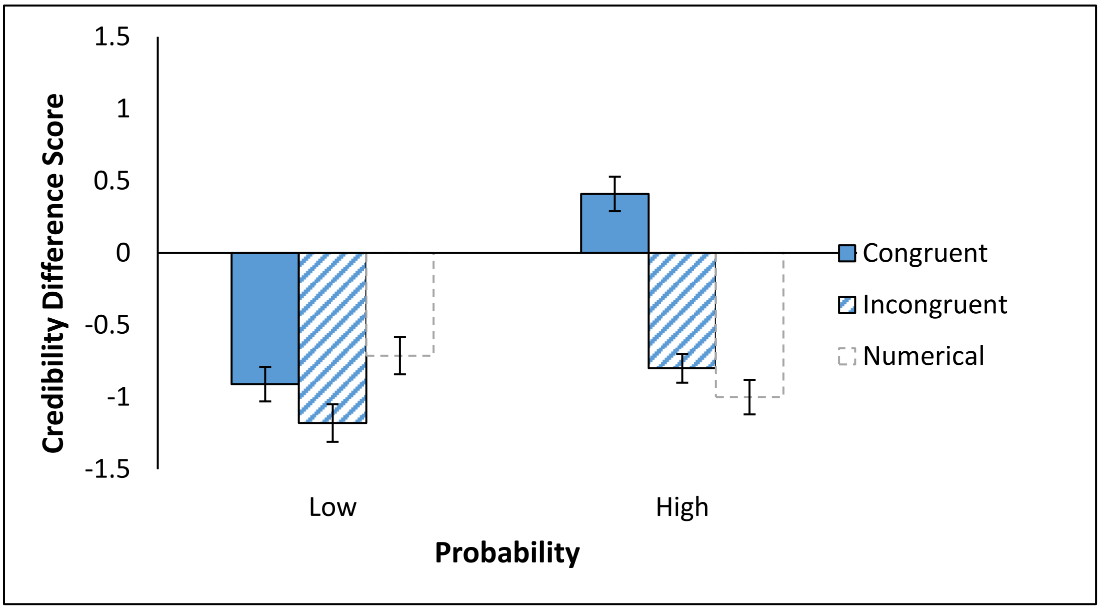


Figure 34. Experiment 12: Mean credibility difference scores: the congruence × probability interaction (error bars represent ±1 Standard Error [SE]).

Correctness ratings.

A 2 (congruence) × 2 (probability) ANOVA revealed the predicted main effect of congruence, $F(1, 285) = 117.53$, $p < .001$, $\eta_p^2 = .29$, and a main effect of probability, $F(1, 285) = 39.20$, $p < .001$, $\eta_p^2 = .12$. Additionally, there was a significant interaction, $F(1, 285) = 9.75$, $p = .002$, $\eta_p^2 = .03$ (see Figure 35). 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, $F(1, 285) = 29.13$, $p < .001$, $\eta_p^2 = .09$.

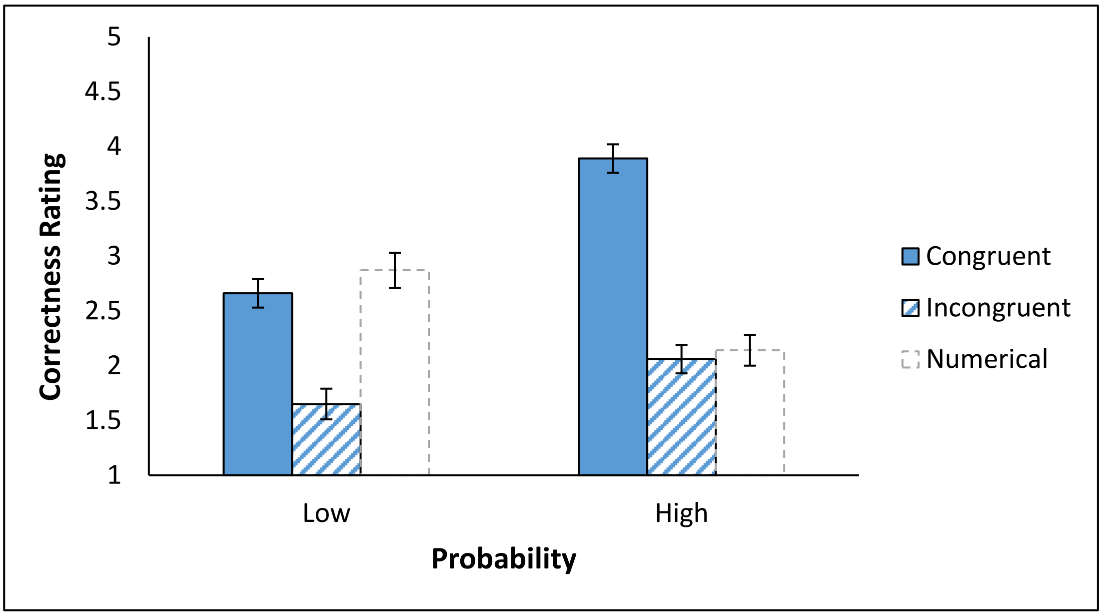


Figure 35. Experiment 12: Mean correctness ratings: the congruence × probability interaction (error bars represent ±1 SE).

Surprise ratings.

A 2 (congruence) × 2 (probability) ANOVA revealed the predicted main effect of congruence, $F(1, 285) = 4.36$, $p = .04$, $\eta_p^2 = .02$, and a main effect of probability, $F(1, 285) = 19.43$, $p < .001$, $\eta_p^2 = .06$ (see Figure 36). The interaction between congruence and probability was not significant, $F(1, 285) = 1.07$, $p = .30$.

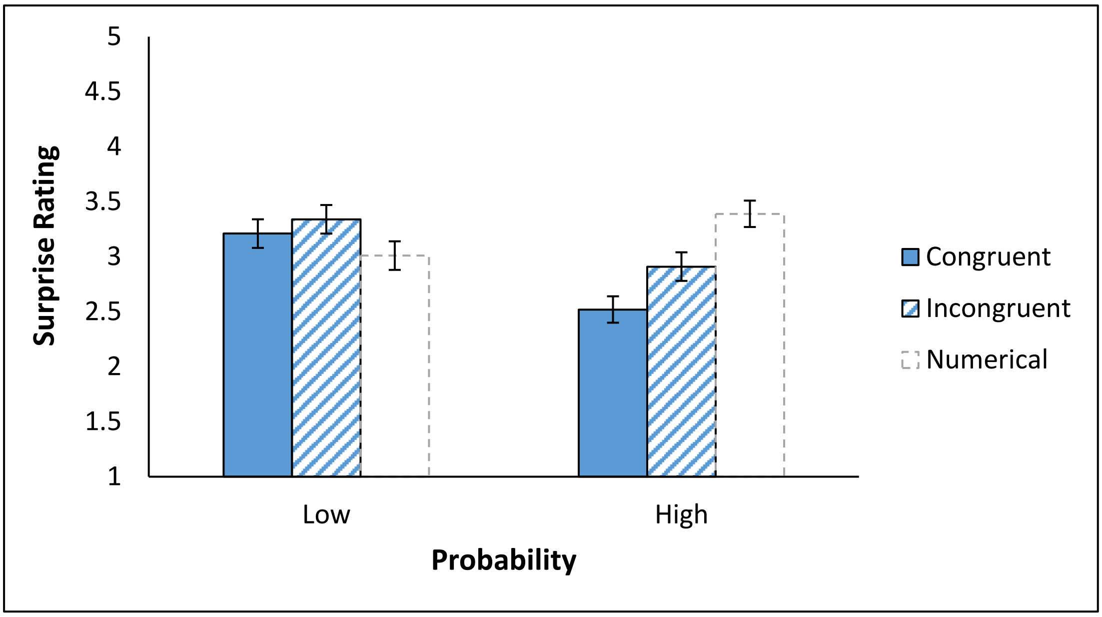


Figure 36. Experiment 12: Mean surprise ratings (error bars represent ±1 SE).

Comparisons to previous results.

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 Table 9, 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, in which a communicator who used '10–30% likelihood' was seen as more correct than one who used 'doubtful', $t(138.43) = -6.08$, $p < .001$, $d = 1.13$.[40] These

[40] 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.

results replicate findings in Chapter 3 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, $t(140.89) = 1.96$, $p = .05$, $d = -0.32$.

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 Table 9, 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 perceived as more correct than one who used 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, $t(144.45) = 4.79$, $p < .001$, $d = 0.78$.

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 communicators were perceived, $t(126.60) = 0.50$, $p = .62$. 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 Experiment 10 – although no effect for format was observed for credibility ratings, one was previously observed for correctness ratings, but this was in relation to a numerical point, rather than numerical range format.

Table 9. Experiment 12: Mean Ratings Across Low and High Probability Expressions

Probability		Communication Format – Mean Rating (SE)		
		Congruent VPE	Incongruent VPE	Numerical Range
Low	Credibility Difference Score	- 0.91 (0.12)	- 1.18 (0.13)	- 0.71 (0.13)
	Correctness	2.66 (0.13)	1.65 (0.14)	2.87 (0.16)
	Surprise	3.21 (0.13)	3.34 (0.13)	3.01 (0.13)
High	Credibility Difference Score	0.41 (0.14)	- 0.80 (0.10)	- 1.00 (0.12)
	Correctness	3.89 (0.13)	2.06 (0.14)	2.14 (0.14)
	Surprise	2.52 (0.12)	2.91 (0.13)	3.39 (0.12)

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.

Discussion.

In Experiment 12, 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. This 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 12 (as well as those previously reported in Chapter 3), 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 their very definition, these will occur more often than probabilistically unexpected outcomes. In Experiment 13, 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).

Experiment 13

Following a pragmatic function account of VPEs, a communicator may opt to use a specifically directional VPE in order to convey their belief about the (non)occurrence of an event. In light of this account, and given 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 used a congruent VPE to 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.[41] 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 a communicator's credibility 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 12, 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

[41] 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 equivalent increase may not be seen given a probabilistically expected outcome.

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, procedure and materials.

Probability (low; high) and communication format (negatively directional VPE [$n_{LowProbability}$=70, $n_{HighProbability}$= 70]; positively directional VPE [$n_{LowProbability}$= 63, $n_{HighProbability}$= 77]; numerical range [$n_{LowProbability}$ = 76, $n_{HighProbability}$= 61]) were manipulated in a between-participants design. The procedure and materials were the same as in Experiment 12, with the exception of outcome. In this experiment, outcomes were 'probabilistically expected'. In the low probability domain, the event did not occur, and in the high probability domain, the event occurred.

Results.

As in Experiment 12, 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.

Analysis of the pre-outcome credibility ratings is available in Appendix K. Overall, there was a significant effect of communication format, with communicators who used a numerical expression or positively directional VPE perceived as most credible. The communicator who used a negatively directional VPE was perceived as least credible. The following analyses were planned, as specified in the pre-registration.

Correctness ratings.

A 2 (congruence) × 2 (probability) ANOVA revealed the predicted main effect of congruence, $F(1, 276) = 133.78$, $p < .001$, η_p^2= .33, and a main effect of probability, $F(1, 276) = 31.21$, $p <$

.001, η_p^2= .10 (see Figure 37). The interaction between congruence and probability was not significant, *F(*1, 276) = 0.73, *p* = .40.

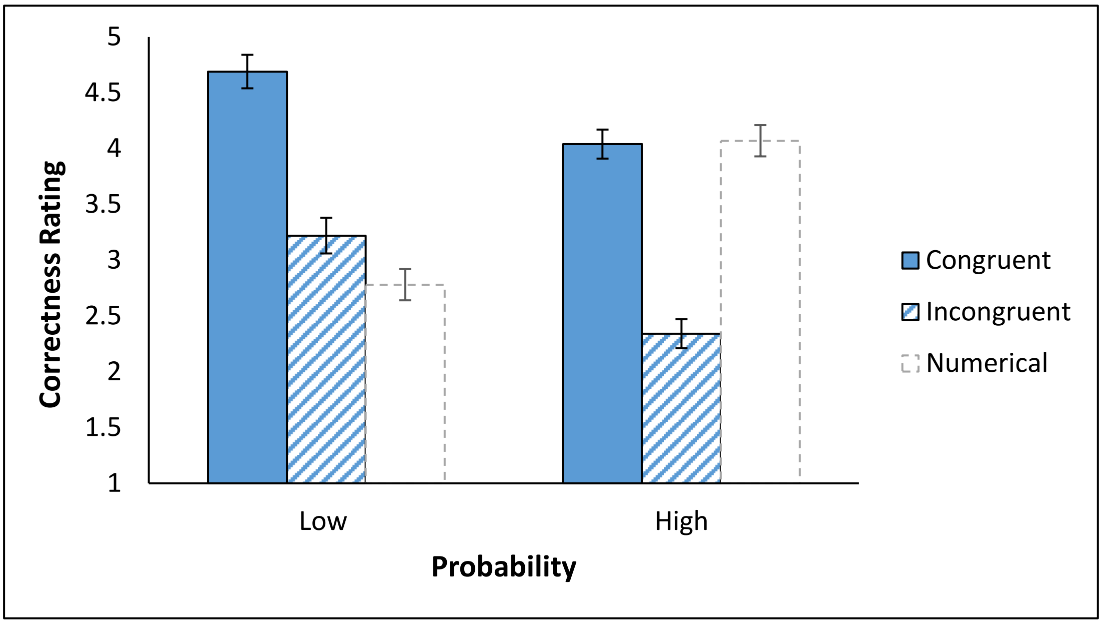


Figure 37. Experiment 13: Mean correctness ratings (error bars represent ±1 SE).

Change in credibility ratings.

Trustworthiness and expertise were highly correlated, both pre- and post-outcome (r = .74, p < .001, r = .83, p < .001, respectively). Credibility was found to be highly reliable, both pre- and post-outcome (α = .85, α = .90, respectively). A 2 (congruence) × 2 (probability) ANOVA revealed a predicted main effect of congruence, *F(*1, 276) = 32.26, p < .001, η_p^2= .11, and a main effect of probability, *F(*1, 276) = 17.32, p < .001, η_p^2= .06 (see Figure 38). The interaction between congruence and probability was not significant, *F(*1, 276) = 0.18, p = .67.

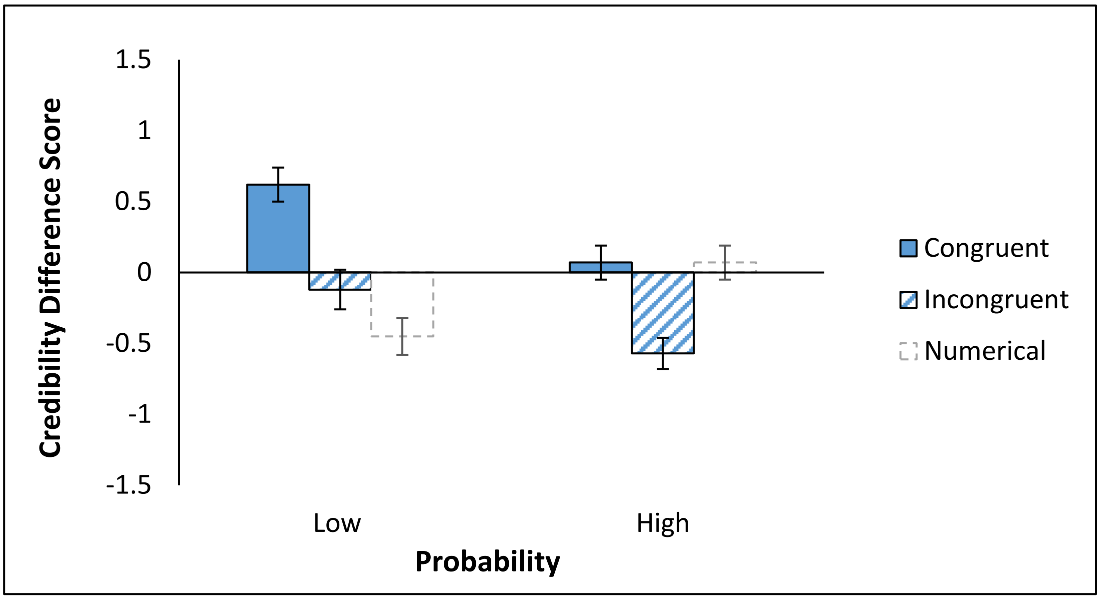


Figure 38. Experiment 13: Mean credibility difference scores (error bars represent ±1 SE).

Surprise ratings.

A 2 (congruence) × 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 39).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, $F(1, 276) = 10.85$, $p = .001$, $\eta_p^2 = .04$.

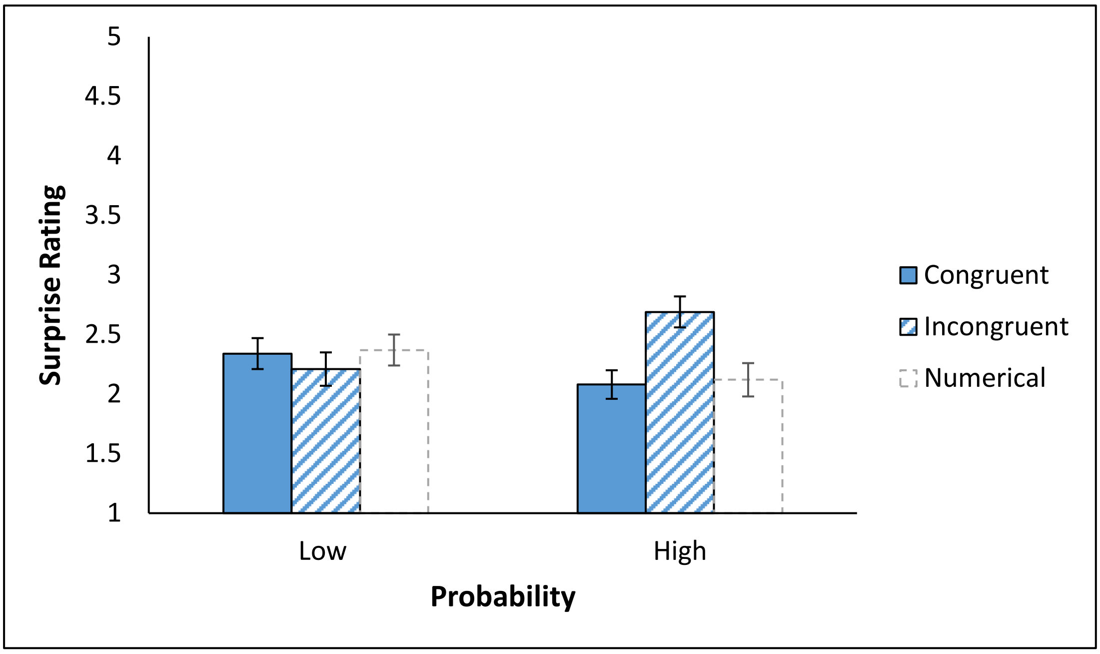


Figure 39. Experiment 13: Mean surprise ratings (error bars represent ±1 SE).

Comparing verbal and numerical formats.

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 Table 10, 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 the communicator who used 'doubtful' was perceived as more correct than the one who used '10–30% likelihood', $t(122.30) = 10.03$, $p < .001$, $d = -1.63$.[42] There was no difference in level of surprise when a communicator described the event using the congruent VPE compared to a range expression, $t(144) = -.137$, $p = .89$.

[42] 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 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 a communicator who used ‘small chance’ to be perceived as more correct than one who used ‘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, $t(117.21) = -3.95$, $p < .001$, $d = 0.67$. These differences were also reflected in the correctness ratings, in which a communicator who used ‘70–90% likelihood’ was seen as more correct than one who used ‘not entirely definite’, $t(128.21) = -8.65$, $p < .001$, $d = 1.50$. Similarly, lower surprise ratings were observed when a communicator described the event using a numerical range compared to the incongruent VPE, $t(129) = -3.07$, $p = .003$, $d = 0.55$.

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.

Table 10. Experiment 13: Mean Ratings Across Low and High Probability Expressions

Probability		Communication Format – Mean Rating (SE)		
		Congruent VPE	Incongruent VPE	Range
Low	Credibility Difference Score	0.62 (0.12)	- 0.12 (0.14)	- 0.45 (0.13)
	Correctness	4.69 (0.15)	3.22 (0.16)	2.78 (0.14)
	Surprise	2.34 (0.14)	2.21 (0.14)	2.37 (0.13)
High	Credibility Difference Score	0.07 (0.10)	- 0.57 (0.11)	0.07 (0.12)
	Correctness	4.04 (0.13)	2.34 (0.13)	4.07 (0.14)
	Surprise	2.08 (0.12)	2.69 (0.13)	2.12 (0.14)

Comparing across experiments.

So far, the analyses from both Experiments 12 and 13 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 12 or 13. 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 credibility difference scores for each potential outcome ('probabilistically unexpected' – E1; 'probabilistically expected' – E2'), 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 11, 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%). Given the assumptions involved, and the reliance on aggregate level data, the expected difference scores will only be compared descriptively in an indicative analysis.

Table 11. Calculation of Expected Difference Scores

Probability	Expression	Expected difference score	E1	E2	Overall 'expected difference score'
Low	Doubtful	(0.2 × E1) + (0.8 × E2)	- 1.18	0.62	0.26
	Small chance	(0.2 × E1) + (0.8 × E2)	- 0.91	- 0.12	- 0.28
	10–30% likelihood	(0.2 × E1) + (0.8 × E2)	- 0.71	- 0.45	- 0.50
High	Not entirely definite	(0.2 × E1) + (0.8 × E2)	0.41	- 0.57	- 0.37
	Good chance	(0.2 × E1) + (0.8 × E2)	- 0.80	0.06	- 0.11
	70–90% likelihood	(0.2 × E1) + (0.8 × E2)	- 1.00	0.07	- 0.14

Note. E1 and E2 represent the mean credibility difference scores from each cell of Experiment 12 (probabilistically unexpected outcome) and Experiment 13 (probabilistically expected outcome) respectively.

Based on the results in Table 11, 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.

Discussion.

In Experiment 13, we found that communicators who used a congruent VPE were perceived as more correct and experienced an increase in credibility ratings than communicators who used an incongruent VPE, following a probabilistically expected outcome (i.e., when a 'doubtful' event did not occur, or when an event with a 'good chance' occurred). Lower levels of surprise at the outcome were observed for congruent expressions, but only in the high probability domain. Overall, these results provide further support for our 'directionality–outcome congruence' account of the effect of communication format on the change in credibility.

General Discussion

Across two experiments, we observed consistent evidence for the role of 'directionality–outcome congruence' in driving evaluations of a communicator, following either a probabilistically expected outcome, or a probabilistically unexpected outcome. In addition, Experiment 12 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 of 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 in order 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. 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 changes 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. In a similar vein, we suggest the directionality of the VPE drove expectations about the event, resulting in more negative credibility ratings when there was an incongruent outcome.

Despite the similar numerical translations given to congruent and incongruent VPEs (see Appendix J), the different downstream effects 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 evidence demonstrating that individuals are sensitive to the pragmatic subtleties associated with use of VPEs.

Directional ambiguity.

The fact that no effect of congruence was observed for credibility difference scores (Experiment 12) or surprise ratings (Experiment 13) 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 clearly perceived as positively directional, affirmative 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.

The ambiguous directionality of 'small chance' is clearly a limitation of the current experiments. However, the phrase's inclusion was somewhat unavoidable given our 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.[43] 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 given a probabilistically expected outcome, as 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

[43] We do however, acknowledge, that in Experiment 12 surprise ratings were significantly different in the high probability domain. We also note that for credibility difference scores and correctness ratings in Experiment 13, the effect of communication format was close to (but did not reach) conventional levels of significance.

& 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 use of VPEs, they have also been shown to occur given use of numerical probabilities (Sirota & Juanchich, 2012, see also Yeung, 2014). Although the addition of numbers to create a mixed format expression (e.g., 'unlikely [20%]') has been proposed as a solution to the problem of mis-communications associated with use of VPEs (e.g., Budescu, Broomell, & Por, 2009), in light of the above, communicators should be aware that numerical expressions are not completely devoid of pragmatic implicatures.

Further considerations.

The results of Chapter 3, specifically the different effect 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'. 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 from the current chapter 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 et al., 2009; Weber, 1994), for the natural hazards we have considered throughout the thesis (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. 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 13, 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 mis-understandings associated with use of VPEs has long since been 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 use of a VPE, and recognise the possible implications their use could have on perceived credibility. Although the results from Chapters 2 and 3 indicated that numerical formats held greater advantages over verbal and mixed formats, the results in the current chapter highlight that numerical formats are still vulnerable to pragmatic influences.

Chapter 5. General Discussion

Summary

This thesis systematically investigated the effect of communication format on understanding, before considering the implications of these findings for a communicator's credibility. This thesis focused on one aspect of risk – the uncertainty of the event, specifically the probabilistic language used to describe this uncertainty. Five different communication formats were investigated: verbal probability expressions (VPEs) (e.g., 'unlikely'); numerical expressions (point, e.g., '20% likelihood' and range, e.g., '10–30% likelihood') and mixed expressions in two orders (verbal-numerical, e.g., 'unlikely [20% likelihood]' and numerical-verbal format, e.g., '20% likelihood [unlikely]'). Firstly, this thesis furthered existing 'which-outcome' research (e.g., Teigen, Juanchich, & Riege, 2013), extending it to numerical and mixed formats, including the previously unstudied N-V format. Secondly, this thesis investigated the effect of communication format on perceived communicator credibility – an important, yet relatively understudied component of effective risk and uncertainty communication. Specifically, we investigated the effect of communication format on how credible a communicator was perceived to be, firstly given an initial prediction, and secondly in light of an outcome. This thesis brought these two (previously unlinked) strands of research together, situating them in the context of consequential, natural hazard communications.

In Chapter 2, we extended previous 'which-outcome' research in a number of ways. We investigated five communication formats, including the previously unstudied numerical-verbal format. We extended the paradigm such that it featured consequential, 'real world' outcomes and included distributions other than the commonly used normal (bell-shaped) distribution. We replicated previous findings, with the majority of participants equating an 'unlikely' outcome to one that had never been shown to occur (equivalent to a 0% likelihood) – the 'extremity effect'. The 'extremity effect' extends to communications which use a verbal-numerical (V-N, ‘unlikely [20% likelihood]’) format and was robust across differing distributions. Furthermore, the ‘extremity effect’ persisted even when participants completed an interim ‘how likely’ task, in which ‘unlikely’ was translated as indicating probabilities much greater than 0% (around 30%, see Experiment 6).

If individuals expect an ‘unlikely’ event to never occur, there could be consequences for a communicator’s perceived credibility if the event subsequently occurs. In Chapter 3, we investigated whether the effects on a communicator’s perceived credibility differed according to the communication format, following a probabilistically unexpected outcome (i.e., an ‘unlikely’ event occurs; a ‘likely’ event does not occur). We found no effect of communication format on initial perceptions of the communicator, though this was not the case following the probabilistically unexpected outcome. In the low probability domain, communicators who used ‘unlikely’ in their prediction suffered a greater reduction in credibility and were perceived as less correct than a communicator who described the event as having a ‘10–30% likelihood’, following the event’s occurrence. This effect of format was not consistently observed in the high probability context, where communicators who used ‘likely’ in their prediction suffered a similar reduction in credibility and were perceived as similarly incorrect versus a communicator who described the event as having a ‘70–90% likelihood’, following the event’s non-occurrence. We posited that the different effects of communication format across probability domains may have resulted from the opposing directionality of the expressions used in the prediction.

We suggested a ‘directionality–outcome congruence’ account to explain the results of Chapter 3, which drew on previous research on directionality and the contrast theory of surprise (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 2003; Teigen & Keren, 2003). ‘Unlikely’ is negatively directional, focusing one’s attention on the non-occurrence of the event, resulting in directionality–outcome incongruence when the event occurred in our low probability scenarios. In contrast, numerical expressions have been said to be positively directional, which meant that the expression was directionally congruent in the low probability domain. However, in the high probability domain, 'likely' and '80% likelihood' are both positively directional and thus both were directionally-incongruent following the event's non-occurrence. In Chapter 4, we tested the 'directionality–outcome congruence' account of these findings, featuring both probabilistically expected and unexpected outcomes. We found that communicators who used expressions which were directionally incongruent with the outcome were generally perceived as less credible and less correct than those who used directionally congruent expressions, with such incongruence also linked with higher levels of surprise.

Overall, our results indicate that there are clear differences between communication formats, in terms of the way that they are understood, as well as in terms of how a communicator is perceived when using them in predictions. That differences between communication formats still occur, despite the fact that they refer to similar numerical probabilities indicates that more than just numerical probability information is being conveyed. Whilst this has long since been recognised to be the case for VPEs, we argue that the influence of pragmatics has largely been neglected when considering numerical and mixed formats. In the following, we consider each communication format and highlight how pragmatic considerations can explain results of the empirical work presented in this thesis. Drawing on insights from the risk and decision-making literature, we move on to consider the implications of this work for risk and uncertainty communication practices, before highlighting outstanding questions which could form the focus of future research.

Towards a Pragmatic Understanding of Communication Format

Verbal probability expressions.

As highlighted in the introduction, there is a wealth of previous research which has shown that VPEs fulfil a variety of pragmatic functions as well as simply communicating likelihood information. These include face-management functions, relating to both the speaker (e.g., trying to avoid blame), and the hearer (e.g., softening bad news) (Bonnefon & Villejoubert, 2006; Juanchich & Sirota, 2013; Juanchich et al., 2012). In the context of the current research, the most important function of VPEs is their role in focusing attention on either the occurrence or non-occurrence of an outcome, via directionality (Teigen, 1988; Teigen & Brun, 1995, 1999, 2000, 2003). Directionality thus enables a communicator to convey a particular perspective (Moxey & Sanford, 2000).

The results from the empirical work in this thesis complement previous directionality research, which has shown that use of a specifically directional VPE influences judgements, predictions and behavioural decisions (Honda & Yamagishi, 2006, 2009; Teigen, 1988; Teigen & Brun, 1999, 2003; Teigen & Keren, 2003). Our work also adds to previous research on directionality by investigating its influence in the English language and demonstrates that directionality's influence extends to communications in a (previously unexplored) natural hazards context. This is in line with previous work which has found that the effects of directionality were invariant across contexts (Budescu et al., 2003). Moreover, our results show that directionality has downstream effects on the perceived credibility of a communicator.

Specifically, we have shown that communicators are evaluated against the extent to which their prediction is directionally congruent with an outcome. We suggest that participants recognised that the communicator's choice of VPE was not simply arbitrary but rather reflected a belief in what the communicator thought was the most likely outcome. This is in line with pragmatic theories of language – Grice's (1975) conversational maxims, specifically the 'maxim of relation', states that utterances should be relevant, a notion echoed in relevance theory (e.g., Sperber & Wilson, 1996). One of the central principles of relevance theory is 'the communicative principle of relevance', which suggests that every communication conveys a "presumption of its own optimal relevance" (Sperber & Wilson, 1996, p. 260). That is, the communicator is assumed to have used the utterance which is (a) worth the recipient's efforts to process it and (b) most compatible with the communicator's abilities and preferences. If the communicator had wanted to communicate something different, then it is assumed they would have chosen a different, more relevant utterance. It is therefore assumed that the choice of communication format was made for a specific reason.

The fact that VPEs have been shown to fulfil a range of pragmatic functions over and above merely communicating likelihood could explain why the 'extremity effect' occurred with verbal, but not numerical formats (Chapter 2). Essentially, we suggest that the two communication formats were perceived as fulfilling two different functions. In the verbal condition, we suggest that participants interpreted the use of 'unlikely' as involving pragmatic considerations, rather than simply communicating likelihood. On this basis, a 0% interpretation is not necessarily non-sensical – if a communicator refers to an 'unlikely' outcome which has a 0% likelihood of occurrence, one possible inference is that the communicator does not believe the event will *ever* occur. In this instance, therefore, the communicator is essentially conveying information that the event is not worth consideration, rather than merely providing a synonym for a numerical probability. In contrast, in the numerical condition, we suggest the use of '20% likelihood' was most likely to be perceived as a 'likelihood communicator', as the expression directly maps to a specific probability.[44] Given such a function, a 0% interpretation would not make sense.

Initial (unpublished) work which has investigated the role of perceived function supports this explanation of the 'extremity effect' (Jenkins & Harris, 2019). As indicated in Chapters 3 and 4, we suggested that the cost to a communicator's credibility associated with use of 'unlikely' or 'doubtful' (in the context of a probabilistically unexpected outcome)

[44] It is important to note that we do not see function as dichotomous (i.e., an expression is either a likelihood communicator or not at all), but rather more relative. Crucially though, we suggest that VPEs are less likely to be perceived as likelihood communicators than the equivalent numerical expression.

occurred as a result of the framing of the prediction, specifically the directionality of the VPE. Using a framing manipulation, we compared a single frame prediction (e.g., ‘it is unlikely to occur….’) to a dual-frame prediction (e.g., ‘It is *unlikely* the event *will* occur [it is *likely* the event *will not* occur]’). We suggested that the presence of the additional frame made a non-likelihood interpretation of the VPE less likely, as the presence of the additional frame means the prediction ‘sums to one’, a characteristic more commonly associated with numerical expressions (Welkenhuysen, Evers-Kiebooms, & D’Ydewalle, 2001). In one experiment, participants were asked to complete a ‘which-outcome’ task (as in Experiment 4, e.g., ‘it is *unlikely* the floodwaters will extend 7 km’). Following this task, participants were subsequently presented with a communication about a volcanic eruption and the chance of lava flows (e.g., ‘it is *unlikely* the lava flows will travel 3.5km from the point of eruption’). Participants were asked “Why do you think the volcanologist chose to use the expression unlikely in this sentence?” (as in Bonnefon & Villejoubert, 2006; Juanchich et al., 2012) and answered by selecting one or more reasons from a list of nine (presented in a randomised order), adapted from those in Juanchich et al. (2012). Participants were more likely to select the ‘likelihood communication’ function in the dual-frame condition than the single frame condition. Importantly, perceived function affected responses in the ‘which-outcome’ task, with those who endorsed ‘likelihood communication’ as a function significantly less likely to show the ‘extremity effect’ than those who selected alternative (pragmatic) functions.

Overall, our findings highlight an additional (previously unseen) downside of using VPEs in predictions about consequential events, in terms of a communicator’s credibility. Our results add further support to the previously suggested notion that VPEs do not communicate information neutrally, but rather convey a particular perspective. Clearly, VPEs fulfil additional, pragmatic functions, over and above simply communicating likelihood. Failure to recognise that VPEs convey more than just numerical information could have unintended (potentially catastrophic) consequences if they are used in real-world, natural hazard communications. We thus echo previous calls to avoid the use of VPEs in such communications.

Mixed expressions.

Whilst mixed expressions have commonly been cited as a way of reducing mis-communications associated with VPEs (Beyth-Marom, 1982; Budescu, Broomell, & Por, 2009; Budescu, Por, & Broomell, 2012; Budescu, Por, Broomell, & Smithson, 2014; Harris, Por, & Broomell, 2017; Patt & Dessai, 2005; Renooij & Witteman, 1999; Witteman & Renooij, 2003; Witteman, Renooij, & Koele, 2007), there has previously been relatively little research which has systematically compared the performance of mixed formats to both verbal and numerical formats. Moreover, research had never specifically investigated a numerical-verbal (N-V) format. Similar to previous findings (such as those using a 'how likely' translation paradigm, e.g., Budescu et al., 2009), we observed improved understanding with the V-N format versus the verbal format.

In Chapter 2, whilst the V-N format led to lower levels of the 'extremity effect' versus the verbal format, the effect was still present for a sizable proportion of participants (around 50%) who saw the V-N format. In contrast, the 'extremity effect' was not observed with the N-V format, nor the numerical format. A difference in findings between the two mixed formats was also found in Chapter 3. A communicator who used a V-N format in their prediction suffered a greater reduction in credibility and were perceived as less correct than a communicator who used a N-V format, following a probabilistically unexpected outcome. These results demonstrate that the order of the mixed format is integral to how the expression is understood and perceived.

The order effects observed are in accordance with Grice's (1975) conversational maxims, particularly one of his 'maxims of manner', which prescribes that utterances should be orderly, with the most important information presented first. The fact that responses in the single format (i.e., verbal or numerical) are different to the mixed format indicates that individuals do not completely disregard the information which comes second, but rather underweight it (versus the initial information). We suggest that when an individual is presented with a mixed format, they infer some level of intentionality or purpose on the part of the communicator, relating to which information was presented first. In addition, the poorer performance of the dual-format (versus the single format) may have also arisen from the perception that the secondary format obscured the meaning of the communication, thus violating the additional 'maxims of manner' in that one should avoid "obscurity of expression…ambiguity… [and] unnecessary prolixity" (Grice, 1975, p. 46).

Calls to use a mixed format seem to rest on the assumption that the addition of numerical expressions can 'cure' the problem of mis-communications associated with VPEs, yet results from Chapter 2 indicate that the addition of a numerical expression to the end of a VPE is not enough to fully 'correct' interpretations. Moreover, findings from Chapter 3 highlight another potential cost associated with use of a V-N format, namely a significant reduction in credibility following a probabilistically unexpected outcome. Although the N-V format fared better on both counts, it is nevertheless worth noting that use of an N-V format was no more advantageous than the numerical format, and was often less advantageous, in terms of credibility. Overall, our findings bring into question how much V-N formats are an appropriate solution to the problem of mis-communications. We therefore suggest that if a communicator is to use a mixed format, the numerical expression should be presented first.

Numerical expressions.

The use of numbers naturally encompasses a level of implicit precision not seen in the aforementioned communication formats. Consequently, numerical expressions may be perceived as objective, certain and trustworthy (Doyle, McClure, et al., 2014; Gurmankin et al., 2004; Schwartz et al., 1999). These beliefs could contribute to the preference for *receiving* information in a numerical format (Erev & Cohen, 1990). Yet as the 'communication mode preference paradox' states, people would prefer to *use* a verbal format to communicate uncertainty, which may in part result from the view that VPEs are more informative, conveying pragmatic as well as likelihood information (Hilton, 2008; Moxey & Sanford, 2000; Teigen & Brun, 1995, 2003). On the basis of these characteristics, one might assume that numerical expressions carry no pragmatic meaning and are thus inherently neutral (Hilton, 2008), though previous literature and our current findings indicate that this is not case.

Previous work has detailed the pragmatics associated with use and interpretations of numerical quantities (e.g., Mandel, 2014; for a review see Moxey & Sanford, 2000). Relatedly, research investigating numerical probability expressions has shown that they are susceptible to contextual influences such as base rate, representativeness and severity (Bilgin & Brenner, 2013; Pighin, Bonnefon, & Savadori, 2011; Windschitl & Weber, 1999). These findings suggest that numerical expressions fulfil functions other than just likelihood communication, with people inferring pragmatic information from their use. Indeed, previous research has found that numerical expressions are susceptible to face-management functions (Sirota & Juanchich, 2012). Participants perceived a communicator as indicating a greater numerical probability when the communicator was perceived as being polite versus when they were plain speaking (i.e., communicating likelihood).

Numerical probability expressions convey additional pragmatic meaning as a result of their directionality. Although numbers are more directionally ambiguous than VPEs, they are not directionally neutral. Rather, they have been said to be biased towards positively directional interpretations (Teigen & Brun, 2000), a notion which is supported by our findings in Chapters 3 and 4. In particular, in Chapter 4, we observed no consistent differences in the way numerical expressions and positively directional VPEs were interpreted, across credibility and correctness ratings. The positively directional nature of the numerical expression could also explain why, in the low probability domain, a communicator who used such a format suffered a greater loss in credibility than one who used a negatively directional

VPE, following a probabilistically expected outcome. We suggest that participants made inferences about why the communicator chose to use a numerical expression, as well as why they chose to frame the prediction in terms of the event's occurrence.

In light of the above, it seems reasonable to conclude that numerical expressions are not devoid of pragmatic information, which has implications for recommended practices in risk and uncertainty communication. The long-held belief in the power of numbers to improve communications is reflected in the development of risk communication strategies (Fischhoff, 1995), with recommendations such as: "all we have to do is get the numbers right"; "all we have to do is tell them the numbers" and "all we have to do is explain what we mean by the numbers" (p.138–140). Yet such a view neglects the fact that numerical expressions can, in fact, be likened to verbal expressions in that they convey a particular perspective (albeit less strongly). Whilst the present results demonstrate that the numerical format is often more beneficial than a verbal format, communicators should be aware that numerical expressions still contain pragmatic information and are thus not completely neutral.

Summary.

In light of the above, it is clear that the role of format in communicating risk and uncertainty information is more nuanced than has previously been acknowledged. It is not simply a case of 'words versus numbers', as the two formats share some common characteristics which have implications for how communications are understood and perceived. Human language is not neutral, but inherently polarised (Hilton, 2008). Part of the meaning of probability expressions comes from their directionality (positive or negative), which enables a communicator to express an attitude relating to the subject in question (Hilton, 2008). Thus, each format can be said to generate 'conversational implicatures' (Levinson, 2000), which influence the inferences that people make. In this thesis, we have shown how these inferences affect not just how a communication is understood, but also how a communicator is evaluated. Importantly, the extent to which these inferences are triggered varies according to communication format.

Credibility

As previously indicated, much of the research on the effect of communication format has focused on the interpretation and use of these expressions, but has neglected the wider social context of communications (Fox & Irwin, 1998). No matter how well risk and uncertainty information is understood, it may not be acknowledged if the source of the information is not perceived as credible. This thesis made a unique contribution to the literature by bringing together these two parts of the communication process, investigating not just understanding, but also the recipient's evaluation of the communicator, and considering the potential relationship between them. Moreover, our work aimed to take into account the dynamic, two-way process of communication (e.g., Slater & Rouner, 1996), as the majority of previous research (investigating either communication format or credibility) has typically only considered 'one shot' communications. That is, only measuring understanding and perceptions at one time point, rather than measuring them over time (e.g., before and after message exposure/an outcome). The time element is especially pertinent given that those living in areas near natural hazards will likely hear multiple communications and experience multiple outcomes over the course of their lifetime.

Overall, we did not observe an effect of communication format on initial perceptions of a communicator (i.e., based solely on a prediction). Had our experiments only measured perceived credibility at one time point, then one might have concluded that it does not matter which format a communicator uses in their prediction, as they will be perceived as equally credible. However, over time, in light of an outcome, we found clear differences in the change in credibility ratings by communication format. Our findings are in line with the notion of 'experienced credibility', which is based on one's experience of the communicator and their interactions over time (Fogg & Tseng, 1999; Önkal, Sinan Gönül, Goodwin, Thomson, & Öz, 2017). In particular, the accuracy of a communicator's previous predictions has been found to influence how much weight people attach to a communicator's advice (Önkal et al., 2017). Our results indicate that accuracy does play a role in 'experienced credibility' judgements, with format differences in correctness ratings mirrored in the change in credibility ratings. Indeed, credibility has been said to relate to perceptions of the past performance of the communicator (Renn & Levine, 1991) and has been shown to be vulnerable to inaccuracy (Frewer et al., 1996; Lucassen & Schraagen, 2011).

If a communicator's credibility is judged according to how 'accurate' they are, the finding that a prediction may be perceived as 'erroneous' in light of a probabilistically unexpected outcome presents a problem for communicators. The highly uncertain nature of natural hazards (Sword-Daniels et al., 2018) means that communicators in this domain are especially vulnerable. Moreover, the asymmetric loss function account (Harris et al., 2009; Weber, 1994) implies that not all 'incorrect' predictions will be evaluated similarly. Whilst we did not explicitly test this account, findings from the experiments featuring probabilistically unexpected outcomes seem to suggest that people may be more forgiving with 'mistakes' which have lower costs. For example, the prediction that a likely event will occur, which subsequently does not occur not may be more forgivable than the prediction that an unlikely event will not occur, which subsequently does. The costs associated with 'erroneous' predictions are clearly something which natural hazard communicators should be aware of, but in reality makes communication more complex. Communicators must strike a delicate balance – repeated over-predictions and erring too much on the side of caution risks people ignoring communications in the future. However, too much the other way and one risks not leaving the public with enough time to take preparatory action, risking damage to both people and property.

Openness about Uncertainty

Many experts have articulated concerns over communicating uncertainty, for fear that it will be misperceived and interpreted as indicating ignorance, incompetence or evasiveness (Fessenden-Raden et al., 1987; Frewer, Hunt, et al., 2003; Johnson & Slovic, 1995, 1996; Lewandowsky et al., 2015). As a result, concerns have been raised that openness about uncertainty will lead to a loss in trust and credibility. The empirical work in this thesis speaks to this concern, given our comparison of numerical range and point formats. Contrary to previously articulated concerns, our results show no ill-effects of expressing uncertainty – we observed no differences in comprehension (i.e., occurrence of the 'extremity effect'), credibility ratings or decision-making between the two types of numerical formats. A communicator who used a range format in their prediction was perceived as equally credible as one who used a point format. Notably, following a probabilistically unexpected outcome (where one might assume an advantage of acknowledging uncertainty) a communicator who used a range format was no more protected against a loss in credibility versus one who used a point format. We therefore echo previous calls for uncertainty to be presented in order to enable people to make fully informed decisions (Fischoff & Davies, 2014; Politi et al., 2007). We do, however, recognise that we only explored the expression of uncertainty in assessments of probability ranges. Thus, the extent to which this conclusion applies to the expression of uncertainty in outcomes (e.g., a sea level rise of 5–25cm) remains an open question.

Implications

After presenting the results of our empirical work, we have often been met with the question of 'which communication format should I be using?' As previous literature has identified, there is no one optimal communication format which will 'fix' the quality of communications (Budescu et al., 2012; Politi et al., 2007). This thesis did not set out to pit each communication format against one another in order to identify a 'winner', but rather aimed to develop a greater understanding of the effects of communication format on understanding and perceived credibility. After all, an awareness of these effects is the first step to being able to mitigate against them (should one wish to).

Hilton (2008) argued that “human logical language serves the function of co-ordinating actions to manage uncertainty by focusing participants’ attention on what is important” (p.108). Indeed, we provide evidence demonstrating that verbal, numerical and mixed formats directionally focus attention to either the occurrence, or non-occurrence of an outcome, albeit to varying degrees. Importantly, the communicator’s choice of where to focus attention may in part depend on the purpose of the communication, with some formats better suited to achieving a particular purpose than others.

One of the main purposes of risk communication is to increase the audience’s understanding of the particular risk (Renn, 1992b; Walaski, 2011). It is therefore unsurprising that research investigating the effect of communication format has typically focused on comprehension, with a desirable communication format arguably seen as one which maximises understanding and aligns experts and the public’s views. In light of this, we firstly consider the implications of our findings in relation to understanding, before considering the wider implications of our research.

The variability in understanding we saw for verbal and V-N formats has implications for organisations whose official guidelines still continue to permit the use of VPEs (albeit more commonly as a part of a standardised uncertainty lexicon). These guidelines come from organisations across a range of domains including meteorology (Gill et al., 2008), climate change (Mastrandrea et al., 2010), accounting (Doupnik & Richter, 2003), intelligence analysis (Ho, Budescu, Dhami, & Mandel, 2016) and food safety (Hart et al., 2019). Our findings mean that we continue to support calls to avoid the sole use of VPEs in communications. In addition, the order effects seen for mixed formats in this thesis indicate that recommendations advocating the use of a V-N format should be refined. If the aforementioned organisations continue to use a mixed format, then we suggest the numerical information should be presented first – a recommendation adopted by the European Food Safety Authority (Hart et al., 2019).

Our findings relating to numerical expressions are particularly important given the National Research Council's (2000) recommendation to avoid the use of expressions such as '1 in 100 year floods' – terminology which has been shown to be confusing and widely misunderstood (Ludy & Kondolf, 2012). Instead, the National Research Council (2000) suggest communications should use numerical expressions similar to those featured in our research, such as 'annual exceedance probability' (AEP) terminology (e.g., '1% AEP flood') or the percentage chance during a 30 year mortgage period (e.g., '26% chance'). Communicators should be aware of the pragmatics associated with their use – particularly when recommendations result in communications featuring overly precise numerical expressions.

Whilst maximising understanding is clearly a worthy purpose of risk communication, it is not the only purpose that a communicator might want to achieve. A communicator might seek to change the recipient's attitudes or behaviour, for instance persuading people them to take a particular course of action. Alternatively, a communicator might wish to highlight the uncertainty associated with an outcome, or might use the communication as a way of seeking to demonstrate that they are a credible and reliable source of information. In light of the multi-purpose nature of risk and communication, we suggest that understanding should not be the sole criterion that communication format is judged against. Rather, when deciding how to frame a communication, communicators should consider the purpose of their communication and which format makes that best clear.

Future Research

As highlighted in the introduction, in the interests of narrowing the scope of the research, we chose to focus on one aspect of risk – the uncertainty of an event, specifically the language used to describe it. As a result, we do not know whether our findings extend to communications where other components of risk are also communicated. Further research could therefore seek to examine whether the effect of communication format interacts with event severity, for example.

More specifically, in order to be able to systematically explore the effect of communication format on understanding and perceptions of credibility (which was already a broad research question), we necessarily had to simplify the nature of the communications featured in the experiments. As a result, we were not in a position to be able to fully explore many other factors which might have also influenced the effectiveness of a communication. These factors include (but are not limited to) *individual differences*: personal experience (Barnett & Breakwell, 2001; van der Linden, 2014); tolerance of uncertainty (O'Neill et al., 2006); *characteristics of the communication*: medium/channel used (Byrne & Curtis, 2000); additional formats such as visual formats (Garcia-Retamero & Cokely, 2017); time frame (Doyle, Mcclure, Johnston, & Paton, 2014; McClure, Doyle, & Velluppillai, 2015) and the presence of conflicting information (Regan et al., 2014). In turn, these factors might have an effect not just on understanding and perceptions, but also *behavioural outcomes* including adherence to warnings (Rød, Botan, & Holen, 2012) and adoption of protective measures (Paton et al., 2010). Whilst we touched on factors relating to individual differences (numeracy) and behavioural outcomes (insurance and evacuation decisions), we did not observe consistent effects of communication format. However, in the interests of brevity, we move on to consider future research questions that are directly related to the main effects observed in this thesis.

We recognise that the majority of the empirical work presented in this thesis featured the same high and low VPEs (e.g., 'likely' and 'unlikely'), coupled with their equivalent numerical and mixed variants. We sought to address this in Chapter 4 by including two new (similarly translated) VPEs for each probability domain – 'doubtful', 'small chance', ‘not entirely definite’ and ‘good chance’. Nevertheless, given the preponderance of VPEs that people use (e.g., Budescu, Weinberg, & Wallsten, 1988), there are clearly many more VPEs which could be tested. We would, however, expect our results to replicate to equivalently directional VPEs, as per our explanation of ‘directionality–outcome congruence’.

Whilst we did not observe a difference between point and range formats, either in terms of the way they were understood, or how credible a communicator was perceived, we only used numerical expressions which were rounded to the nearest five or ten and did not include decimal points. Such rounded expressions are suggested to be easier to understand (Fagerlin & Peters, 2012; Zhang & Schwarz, 2013) and importantly, more believable than more precise estimates (H. O. Witteman et al., 2011). It is possible that if numerical expressions were to be made more precise, more pragmatic inferences would be drawn from their use (such as the communicator’s level of confidence) and thus the benefits associated with their use might reduce.

On a related note, given recommendations to use a mixed format, there is a question over how the numerical expressions should be framed. In their latest guidance document on the communication of uncertainty, the European Food Safety Authority suggest that communications should be framed “positively, as % certainty for the outcome, conclusion or range of values…that is considered most likely” (Hart et al., 2019, p. 18). However, this recommendation has not been tested. Previous research on directionality, in combination with results from this thesis, suggests that there may be unanticipated costs to adopting such an approach. In line with previous results, we found evidence to suggest that numerical expressions are positively directional. If a number is paired with the positively directional VPE ‘certain’ (e.g., ‘80% certainty that the event will occur’), the resulting expression might be perceived as even more positively directional than if it were simply ‘80% likelihood’. Using the certain frame may lead people to overestimate how certain the communicator is, especially given higher forecasts are perceived as signalling high confidence (Bagchi & Ince, 2016). This framing could prove harmful to a communicator's credibility following a probabilistically unexpected outcome, where a communicator may be judged even more harshly than if they had used the '% likelihood' frame.

The research exploring the effect of communication format on perceptions of credibility has (with the exception of Experiment 13) focused on how communicators are perceived in light of a probabilistically unexpected outcome. Given that probabilistically expected outcomes will, by definition, occur more frequently, a more detailed investigation of the effects of communication format in such a context is warranted. Whilst mixed formats were not as advantageous as a numerical expression in light of a probabilistically unexpected outcome, it is possible that they might be more beneficial following a probabilistically expected outcome. For instance, in the low probability domain, a negatively directional VPE might be used as a way of 'cancelling out' the positive directionality of the numerical expression, thus reducing pragmatic interpretations.

In this thesis, we extended the which-outcome work to numerical and mixed formats, and we have shown that the 'extremity effect' has implications for a communicator's perceived credibility. However, we only investigated written communications. As highlighted at the beginning of this section, audiences do not always receive communications in the same medium. For example, with natural hazards, information may often be communicated using maps (e.g., Kunz, Grêt-Regamey, & Hurni, 2011). It is important to investigate whether the 'extremity effect' occurs in such a format, given that the effect could have extreme consequences for those deciding whether to evacuate their homes on the basis of an 'unlikely' prediction.

Conclusion

Overall, this thesis has presented a systematic investigation into the effects of communication format on understanding and perceived communicator credibility. The results add to the large literature on the downsides of using VPEs, and highlight an additional cost to a communicator who uses them, namely, a significant loss in credibility in the face of probabilistically unexpected outcomes. Importantly, this thesis has indicated that recommendations to use mixed format should be refined, with use of a N-V format shown as more beneficial than use of a V-N format, in terms of both understanding and credibility. Notably, use of a N-V format was shown to be no more (and in some cases was less) advantageous than use of numerical format. However, communicators should be aware numerical formats are not completely devoid of pragmatic implicatures. Moving forward, communication strategies should take into account not only how communications are explicitly understood, but should also consider the pragmatic inferences which could be drawn from the choice of communication format.

References

Alexander, D. E. (2014). Communicating earthquake risk to the public: the trial of the "L'Aquila Seven." *Natural Hazards*, *72*(2), 1159–1173. https://doi.org/10.1007/s11069-014-1062-2

Árvai, J. (2014). The end of risk communication as we know it. *Journal of Risk Research*, *17*(10), 1245–1249. https://doi.org/10.1080/13669877.2014.919519

Aven, T. (2010). *Misconceptions of Risk*. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9780470686539

Aven, T. (2012). The risk concept—historical and recent development trends. *Reliability Engineering & System Safety*, *99*, 33–44. https://doi.org/10.1016/J.RESS.2011.11.006

Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, *12*(1), 1–11. https://doi.org/10.1080/13669870802488883

Aven, T., Renn, O., & Rosa, E. A. (2011). On the ontological status of the concept of risk. *Safety Science*, *49*(8), 1074–1079. https://doi.org/10.1016/j.ssci.2011.04.015

Bagchi, R., & Ince, E. C. (2016). Is a 70% forecast more accurate than a 30% forecast? How level of a forecast affects inferences about forecasts and forecasters. *Journal of Marketing Research*, *53*(1), 31–45. https://doi.org/10.1509/jmr.12.0526

Barnes, A. (2016). Making intelligence analysis more intelligent: Using numeric probabilities. *Intelligence and National Security*, *31*(3), 327–344. https://doi.org/10.1080/02684527.2014.994955

Barnes, L. R., Gruntfest, E. C., Hayden, M. H., Schultz, D. M., & Benight, C. (2007). False alarms and close calls: A conceptual model of warning accuracy. *Weather and Forecasting*, *22*(5), 1140–1147. https://doi.org/10.1175/WAF1031.1

Barnett, J., & Breakwell, G. M. (2001). Risk perception and experience: hazard personality profiles and individual differences. *Risk Analysis*, *21*(1), 171–178. https://doi.org/10.1111/0272-4332.211099

Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal*, *15*(S1), 175–190. https://doi.org/10.1002/smj.4250150912

Batchelor, R., & Peel, D. A. (1998). Rationality testing under asymmetric loss. *Economics Letters*, *61*(1), 49–54. https://doi.org/10.1016/S0165-1765(98)00157-8

Bearth, A., Cousin, M.-E., & Siegrist, M. (2014). The consumer's perception of artificial food additives: Influences on acceptance, risk and benefit perceptions. *Food Quality and Preference*, *38*, 14–23. https://doi.org/10.1016/j.foodqual.2014.05.008

Benessia, A., & De Marchi, B. (2017). When the earth shakes … and science with it. The management and communication of uncertainty in the L'Aquila earthquake. *Futures*, *91*, 35–45. https://doi.org/10.1016/j.futures.2016.11.011

Berry, D. (2004). *Risk, Communication and Health Psychology*. Maidenhead, England: Open University Press.

Betsch, C., Haase, N., Renkewitz, F., & Schmid, P. (2015). The narrative bias revisited: What drives the biasing influence of narrative information on risk perceptions? *Judgment and Decision Making*, *10*(3), 241–264.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, *1*(3), 257–269. https://doi.org/10.1002/for.3980010305

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

Bonnefon, J. F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*(9), 747–751. https://doi.org/10.1111/j.1467-9280.2006.01776.x

Bord, R. J., & O'Connor, R. E. (1992). Determinants of risk perceptions of a hazardous waste site. *Risk Analysis*, *12*(3), 411–416. https://doi.org/10.1111/j.1539-6924.1992.tb00693.x

Bostrom, A. (1997). Risk perceptions: Experts vs. lay people. *Duke Environmental Law & Policy Forum*, *8*(1), 101–113.

Bostrom, A., & Löfstedt, R. E. (2003). Communicating risk: Wireless and hardwired. *Risk Analysis*, *23*(2), 241–248. https://doi.org/10.1111/1539-6924.00304

Braunsberger, K. (1996). *The effects of source and product characteristics on persuasion*. (Doctoral dissertation, University of South Florida St. Petersburg). Retrieved from https://digital.usfsp.edu/fac_publications

Breakwell, G. M. (2000). Risk communication: factors affecting impact. *British Medical Bulletin*, *56*(1), 110–120. https://doi.org/10.1258/0007142001902824

Breakwell, G. M. (2014). Risk communication. In *The Psychology of Risk* (pp. 143–189). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139061933.007

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, *25*(2), 93–139. https://doi.org/10.1093/jos/ffm016

Brewer, Noel T, Tzeng, J. P., Lillie, S. E., Edwards, A. S., Peppercorn, J. M., & Rimer, B. K. (2009). Health literacy and cancer risk perception: Implications for genomic risk communication. *Medical Decision Making*, *29*(2), 157–166. https://doi.org/10.1177/0272989X08327111

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

British Geological Survey. (2019). About the British Geological Survey. Retrieved August 26, 2019, from https://www.bgs.ac.uk/about/home.html

Brugnach, M., Dewulf, A., Pahl-Wostl, C., & Taillieu, T. (2008). Toward a relational concept of uncertainty: about knowing too little, knowing too differently, and accepting not to know. *Ecology and Society*, *13*(2), 30. https://doi.org/10.5751/ES-02616-130230

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

Brust-Renck, P. G., Royer, C. E., & Reyna, V. F. (2013). Communicating numerical risk: Human factors that aid understanding in health care. *Review of Human Factors and Ergonomics*, *6*(8), 235–276. https://doi.org/10.1021/nn300902w

Bryant, G. D., & Norman, G. R. (1980). Expressions of probability: Words and numbers. *New England Journal of Medicine*, *302*(7), 411–411. https://doi.org/10.1056/NEJM198002143020717

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. H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change*, *113*(2), 181–200. https://doi.org/10.1007/s10584-011-0330-3

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. (1987). Subjective estimation of precise and vague uncertainties. In G. Wright & P. Ayton (Eds.), *Judgemental Forecasting* (pp. 63–82). Chichester, England: Wiley.

Budescu, D. V., & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior and Human Decision Processes*, *46*(2), 240–263. https://doi.org/10.1016/0749-5978(90)90031-4

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

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

Budescu, D. V, Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 281–294. https://doi.org/10.1037/0096-1523.14.2.281

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*. Philadelphia, PA. Retrieved from https://bit.ly/2CeBKVn

Byrne, M., & Curtis, R. (2000). Designing health communication: Testing the explanations for the impact of communication medium on effectiveness. *British Journal of Health Psychology*, *5*(2), 189–199. https://doi.org/10.1348/135910700168856

Campbell, P. (2011). Understanding the receivers and the reception of science's uncertain messages. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1956), 4891–4912. https://doi.org/10.1098/rsta.2011.0068

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

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.

Clarke, V. A., Ruffin, C. L., Hill, D. J., & Beamen, A. L. (1992). Ratings of orally presented verbal expressions of probability by a heterogeneous sample. *Journal of Applied Social Psychology*, *22*(8), 638–656. https://doi.org/10.1111/j.1559-1816.1992.tb00995.x

Clatworthy, M. A., Peel, D. A., & Pope, P. F. (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting*, *31*(8), 736–756. https://doi.org/10.1002/for.1253

Cohn, L. (2009). Quantifying risk: Verbal probability expressions in Spanish and English. *American Journal of Health Behavior*, *33*(3), 244–255. https://doi.org/10.5993/AJHB.33.3.3

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., and F. Allen. (1988). *Seven cardinal rules of risk communication*. U.S. Environmental Protection Agency. Washington, DC.

Cummings, L. (2014). The "trust" heuristic: Arguments from authority in public health. *Public Health, Health Communication*, *29*(10), 1043–1056. https://doi.org/10.1080/10410236.2013.831685

Cvetkovich, G., & Löfstedt, R. E. (1999). *Social Trust and the Management of Risk*. London: Earthscan.

Davies, S. R. (2008). Constructing communication. *Science Communication*, *29*(4), 413–434. https://doi.org/10.1177/1075547008316222

Department for Business, Innovation & Skills. (2011). *Skills for Life Survey 2011: Headline Findings*. Retrieved from UK Government website: https://www.gov.uk/government/publications/2011-skills-for-life-survey

Dhami, M. K. (2018). Towards an evidence-based approach to communicating uncertainty in intelligence analysis. *Intelligence and National Security*, *33*(2), 257–272. https://doi.org/10.1080/02684527.2017.1394252

Dhami, M. K., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: toward translating linguistic probabilities. *Memory & Cognition*, *33*(6), 1057–1068. https://doi.org/10.3758/BF03193213

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

Dieckmann, N., Peters, E. M., & Gregory, R. (2015). At home on the range? Lay interpretations of numerical uncertainty ranges. *Risk Analysis*, *35*(7), 1281–1295. https://doi.org/10.1111/risa.12358

Dieckmann, N., Peters, E. M., Gregory, R., & Tusler, M. (2012). Making sense of uncertainty: Advantages and disadvantages of providing an evaluative structure. *Journal of Risk Research*, *15*(7), 717–735. https://doi.org/10.1080/13669877.2012.666760

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. https://doi.org/10.1111/j.1539-6924.2009.01279.x

Douglas, M. (1994). *Risk and Blame: Essays in Cultural Theory*. London: Routledge.

Doupnik, T. S., & Richter, M. (2003). Interpretation of uncertainty expressions: a cross-national study. *Accounting, Organizations and Society*, *28*(1), 15–35.

Doupnik, T. S., & Richter, M. (2004). The impact of culture on the interpretation of "in context" verbal probability expressions. *Journal of International Accounting Research*, *3*(1), 1–20. https://doi.org/10.2308/jiar.2004.3.1.1

Doyle, E. E. H., Johnston, D. M., Smith, R., & Paton, D. (2019). Communicating model uncertainty for natural hazards: A qualitative systematic thematic review. *International Journal of Disaster Risk Reduction*, *33*, 449–476. https://doi.org/10.1016/j.ijdrr.2018.10.023

Doyle, E. E. H., Mcclure, J., Johnston, D. M., & Paton, D. (2014). Communicating likelihoods and probabilities in forecasts of volcanic eruptions. *Journal of Volcanology and Geothermal Research*, *272*, 1–15. https://doi.org/10.1016/j.jvolgeores.2013.12.006

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. https://doi.org/10.1016/j.ijdrr.2014.07.006

Du, N., Budescu, D. V., Shelly, M. K., & Omer, T. C. (2011). The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, *114*(2), 179–189. https://doi.org/10.1016/j.obhdp.2010.10.005

Du, N., Stevens, K., Ahern, J., & Shigaev, A. (2015). Cross-cultural differences in interpreting IAS 37 probability phrases. *International Journal of Financial Research*, *7*(1), 1–8. https://doi.org/10.5430/ijfr.v7n1p1

Earle, T. C., & Cvetkovich, G. T. (1995). *Social Trust. Towards a Cosmopolitan Society.* London: Praeger.

Earle, T. C., Siegrist, M., & Gutscher, H. (2012). Trust, Risk Perception and the TCC Model of Cooperation. In *Trust in Risk Management*. Routledge. https://doi.org/10.4324/9781849773461

Edelman. (2018). Edelman Trust Barometer. Retrieved from Edelman website: https://www.edelman.com/sites/g/files/aatuss191/files/2018-10/2018_Edelman_TrustBarometer_Executive_Summary_Jan.pdf

Eisend, M. (2006). Source credibility dimensions in marketing communication - a generalized solution. *Journal of Empirical Generalisations in Marketing Science*, *10*(2), 1–33.

Eisenman, D. P., Cordasco, K. M., Asch, S., Golden, J. F., & Glik, D. (2007). Disaster planning and risk communication with vulnerable communities: Lessons from Hurricane Katrina. *American Journal of Public Health*, *97*, S109–S115. https://doi.org/10.2105/AJPH.2005.084335

Eiser, R. J., Bostrom, A., Burton, I., Johnston, D. M., McClure, J., Paton, D., van der Pligt, J., & White, M. P. (2012). Risk interpretation and action: A conceptual framework for responses to natural hazards. *International Journal of Disaster Risk Reduction*, *1*, 5–16. https://doi.org/10.1016/j.ijdrr.2012.05.002

Eiser, R. J., Miles, S., & Frewer, L. J. (2002). Trust, perceived risk, and attitudes toward food technologies. *Journal of Applied Social Psychology*, *32*(11), 2423–2433. https://doi.org/10.1111/j.1559-1816.2002.tb01871.x

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. https://doi.org/10.1016/0749-5978(90)90002-Q

Ericsson, K. A. (2018). An introduction to the second edition of The Cambridge Handbook of Expertise and Expert Performance: Its development, organization, and content. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 3–20). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781316480748.001

Fagerlin, A., & Peters, E. M. (2012). Quantitative Information. In N T Brewer & J. S. Downs (Eds.), *Communicating Risks and Benefits: An Evidence-Based User's Guide* (pp. 53–64). Silver Spring, Maryland: Food and Drug Administration (FDA), US Department of Health and Human Services.

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

Fearnley, C. J., Winson, A. E. G., Pallister, J., & Tilling, R. (2018). Volcano crisis communication: Challenges and solutions in the 21st century. In C. J. Fearnley, D. K. Bird, K. Haynes, W. J. McGuire, & G. Jolly (Eds.), *Observing the Volcano World: Volcano Crisis Communication* (pp. 3–21). Cham: Springer International Publishing. https://doi.org/10.1007/11157_2017_28

Feldman, S. P. (2004). The culture of objectivity: Quantification, uncertainty, and the evaluation of risk at NASA. *Human Relations*, *57*(6), 691–718. https://doi.org/10.1177/0018726704044952

Fessenden-Raden, J., Fitchen, J. M., & Heath, J. S. (1987). Providing risk information in communities: factors influencing what is heard and accepted. *Science, Technology, and Human Values*, *12*(3), 94–101.

Field, A. P. (2013). *Discovering Statistics Using SPSS: And Sex and Drugs and Rock "n" Roll (4th Edition)*. London: Sage.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. (1991). Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *The American Journal of Psychology*, *104*(1), 35–60. https://doi.org/http://dx.doi.org/10.2307/1422850

Fischer, K., & Jungermann, H. (1996). Rarely occurring headaches and rarely occurring blindness: Is rarely = rarely? *Journal of Behavioral Decision Making*, *9*(3), 153–172. https://doi.org/10.1002/(SICI)1099-0771(199609)9:3<153::AID-BDM222>3.0.CO;2-W

Fischhoff, B. (1995). Risk perception and commuication unplugged: Twenty years of process. *Risk Analysis*, *15*(2), 137–145. https://doi.org/10.1111/j.1539-6924.1995.tb00308.x

Fischhoff, B. (2011). Applying the science of communication to the communication of science. *Climatic Change*, *108*(4), 701–705. https://doi.org/10.1007/s10584-011-0183-9

Fischhoff, B. (2013). The sciences of science communication. *Proceedings of the National Academy of Sciences*, *110*(Supplement 3), 14033–14039. https://doi.org/10.1073/PNAS.1213273110

Fischhoff, B. (2019). Risk perception and health behaviour. In S. Ayers, A. Baum, C. McManus, S. Newman, K. Wallston, J. Weinman, & R. West (Eds.), *Cambridge Handbook of Psychology, Health and Medicine* (pp. 187–191). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511543579.040

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, *9*(2), 127–152. https://doi.org/10.1007/BF00143739

Fischoff, B., & Davies, A. L. (2014). Communicating scientific uncertainty. *Communication Research*, *111*(3), 429–456. https://doi.org/10.1177/0093650214534967

Flugstad, A. R., & Windschitl, P. D. (2003). The influence of reasons on interpretations of probability forecasts. *Journal of Behavioral Decision Making*, *16*, 107–126. https://doi.org/10.1002/bdm.437

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 80–87). New York: ACM Press.

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

Fox, C. R., & Irwin, J. R. (1998). The role of context in the communication of uncertain beliefs. *Basic and Applied Social Psychology*, *20*(1), 37–41. https://doi.org/10.1207/s15324834basp2001

Fox, C. R., & Ulkümen, G. (2017). Comment on Løhre & Teigen (2016). “There is a 60% probability, but I am 70% certain: communicative consequences of external and internal expressions of uncertainty”. *Thinking & Reasoning*, *23*(4), 483–491. https://doi.org/10.1080/13546783.2017.1314939

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making* (pp. 21–35). Oslo: Universitetsforlaget.

Freeman, K. S., & Spyridakis, J. H. (2009). Effect of contact information on the credibility of online health information. *IEEE Transactions on Professional Communication*, *52*(2), 152–166. https://doi.org/10.1109/TPC.2009.2017992

Frewer, L. J. (2000). Risk perception and risk communication about food safety issues. *Nutrition Bulletin*, *25*(1), 31–33. https://doi.org/10.1046/j.1467-3010.2000.00015.x

Frewer, L. J. (2004). The public and effective risk communication. *Toxicology Letters*, *149*(1–3), 391–397. https://doi.org/10.1016/j.toxlet.2003.12.049

Frewer, L. J., Howard, C., Hedderley, D., & Shepherd, R. (1996). What determines trust in information about food-related risks? Underlying psychological constructs. *Risk Analysis*, *16*(4), 473–486. https://doi.org/10.1111/j.1539-6924.1996.tb01094.x

Frewer, L. J., Howard, C., Hedderley, D., & Shepherd, R. (1997). The elaboration likelihood model and communication about food risks. *Risk Analysis*, *17*(6), 759–770. https://doi.org/10.1111/j.1539-6924.1997.tb01281.x

Frewer, L. J., Howard, C., Hedderley, D., & Shepherd, R. (1999). Reactions to information about genetic engineering: impact of source characteristics, perceived personal relevance, and persuasiveness. *Public Understanding of Science*, *8*(1), 35–50. https://doi.org/10.1088/0963-6625/8/1/003

Frewer, L. J., Howard, C., & Shepherd, R. (1998). The influence of initial attitudes on responses to communication about genetic engineering in food production. *Agriculture and Human Values*, *15*(1), 15–30. https://doi.org/10.1023/A:1007465730039

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. https://doi.org/10.1080/1366987032000047815

Frewer, L. J., Miles, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2002). Public preferences for informed choice under conditions of risk uncertainty. *Public Understanding of Science*, *11*(4), 363–372. https://doi.org/10.1088/0963-6625/11/4/304

Frewer, L. J., Scholderer, J., & Bredahl, L. (2003). Communicating about the risks and benefits of genetically modified foods: The mediating role of trust. *Risk Analysis*, *23*(6), 1117–1133. https://doi.org/10.1111/j.0272-4332.2003.00385.x

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. *Human Factors*, *59*(4), 582–627. https://doi.org/10.1177/0018720817690634

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623–629. https://doi.org/10.1111/j.1539-6924.2005.00608.x

Gill, J., Rubiera, J., Martin, C., Cacic, I., Mylne, K., Dehui, C., Jiafeng, G., Xu, T., Yamaguchi, M., Foamouhoue, A.K., Poolman, E., & John, G. (2008). *Guidelines on Communicating Forecast Uncertainty*. (Technical report World Meteorological Organization/TD No. 4122). Retrieved from World Meteorological Organization website: https://www.wmo.int/pages/prog/amp/pwsp/documents/TD-1422.pdf

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society*, *20*(2), 199–207. https://doi.org/10.1057/jors.1969.52

Granger, M. (1998). Commentary: Uncertainty analysis in risk assessment. *Human and Ecological Risk Assessment*, *4*(1), 25–39. https://doi.org/10.1080/10807039.1998.11009680

Gregory, R., Dieckmann, N., Peters, E. M., Failing, L., Long, G., & Tusler, M. (2012). Deliberative disjunction: Expert and public understanding of outcome uncertainty. *Risk Analysis*, *32*(12), 2071–2083. https://doi.org/10.1111/j.1539-6924.2012.01825.x

Grice, P. H. (1975). Logic and Conversation. In *Syntax and Semantics 3: Speech Acts* (pp. 41–58). San Diego, CA: Academic Press. https://doi.org/10.1111/j.1365-2664.2006.01229.x

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. https://doi.org/10.1111/j.1741-3737.2006.00325.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

Gutteling, J. M., & Kuttschreuter, M. (2002). The role of expertise in risk communication: Laypeople's and expert's perception of the millennium bug risk in The Netherlands. *Journal of Risk Research*, *5*(1), 35–47. https://doi.org/10.1080/13669870010029639

Guzzetti, F. (2016). Forecasting natural hazards, performance of scientists, ethics, and the need for transparency. *Toxicological & Environmental Chemistry*, *98*(9), 1043–1059. https://doi.org/10.1080/02772248.2015.1030664

Haase, N., Renkewitz, F., & Betsch, C. (2013). The measurement of subjective probability: Evaluating the sensitivity and accuracy of various scales. *Risk Analysis*, *33*(10), 1812–1828. https://doi.org/10.1111/risa.12025

Hacking, I. (2006). *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511817557

Halberg, A.-M., & Teigen, K. H. (2009). Framing of imprecise quantities: When are lower interval bounds preferred to upper bounds? *Journal of Behavioral Decision Making*, *22*(5), 490–509. https://doi.org/10.1002/bdm.635

Halberg, A.-M., Teigen, K. H., & Fostervold, K. I. (2009). Maximum vs. minimum values: Preferences of speakers and listeners for upper and lower limit estimates. *Acta Psychologica*, *132*(3), 228–239. https://doi.org/10.1016/j.actpsy.2009.07.007

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, *48*(2), 193–223. https://doi.org/10.1016/0749-5978(91)90012-I

Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2011). Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*, *31*(6), 828–838. https://doi.org/10.1177/0272989X11393976

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. https://doi.org/10.1177/0272989X10371830

Han, P. K. J., Klein, W. M. P., Lehman, T. C., Massett, H., Lee, S. C., & Freedman, A. N. (2009). Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. *Medical Decision Making*, *29*(3), 391–403. https://doi.org/10.1177/0272989X08327396

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. https://doi.org/10.1037/a0024195

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

Harris, A. J. L., Corner, A., Xu, J., & Du, X. (2013). Lost in translation? Interpretations of the probability phrases used by the Intergovernmental Panel on Climate Change in China and the UK. *Climatic Change*, *121*(2), 415–425. https://doi.org/10.1007/s10584-013-0975-1

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2015). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, *40*(6), 1–38. https://doi.org/10.1111/cogs.12276

Harris, A. J. L., Por, H. H., & Broomell, S. B. (2017). Anchoring climate change communications. *Climatic Change*, *140*(3–4), 387–398. https://doi.org/10.1007/s10584-016-1859-y

Hart, A., Maxim, L., Siegrist, M., Von Goetz, N., da Cruz, C., Merten, C., … Younes, M. (2019). Guidance on communication of uncertainty in scientific assessments. *EFSA Journal*, *17*(1), 5520. https://doi.org/10.2903/j.efsa.2019.5520

Haynes, K., Barclay, J., & Pidgeon, N. F. (2007). Volcanic hazard communication using maps: An evaluation of their effectiveness. *Bulletin of Volcanology*, *70*(2), 123–138. https://doi.org/10.1007/s00445-007-0124-7

Haynes, K., Barclay, J., & Pidgeon, N. F. (2008a). The issue of trust and its influence on risk communication during a volcanic crisis. *Bulletin of Volcanology*, *70*(5), 605–621. https://doi.org/10.1007/s00445-007-0156-z

Haynes, K., Barclay, J., & Pidgeon, N. F. (2008b). Whose reality counts? Factors affecting the perception of volcanic risk. *Journal of Volcanology and Geothermal Research*, *172*(3–4), 259–272. https://doi.org/10.1016/j.jvolgeores.2007.12.012

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*(1), 5–28. https://doi.org/10.1007/BF00057884

Heesacker, M., Petty, R. E., & Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality*, *51*(4), 653–666. https://doi.org/10.1111/j.1467-6494.1983.tb00872.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing and Management*, *44*(4), 1467–1484. https://doi.org/10.1016/j.ipm.2007.10.001

Hilton, D. (2008). Emotional tone and argumentation in risk communication. *Judgment and Decision Making*, *3*(1), 100–110.

Ho, E. H., Budescu, D. V., Dhami, M. K., & Mandel, D. R. (2016). Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy*, *1*(2), 43–55. https://doi.org/10.1080/15614263.2012.754133

Hohle, S., & Teigen, K. H. (2018). More than 50% or less than 70% chance: Pragmatic implications of single bound probability estimates. *Journal of Behavioral Decision Making*, *31*(1), 138–150. https://doi.org/10.1002/bdm.2052

Holtgraves, T. M., & Perdew, A. (2016). Politeness and the communication of uncertainty. *Cognition*, *154*, 1–10. https://doi.org/10.1016/j.cognition.2016.05.005

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

House of Lords. (2000). *Science and Technology - Third Report*. Retrieved from Science and Technology Committee Publications website: https://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.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635. https://doi.org/10.1086/266350

Hubbard, D. W. (2015). An ivory tower of Babel: Fixing the confusion about risk. In *The Failure of Risk Management* (pp. 79–93). Hoboken, New Jersey: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119198536.ch5

IAVCEI Task Group on Crisis Protocols. (2016). Toward IAVCEI guidelines on the roles and responsibilities of scientists involved in volcanic hazard evaluation, risk mitigation, and crisis response. *Bulletin of Volcanology*, *78*(31), 1–3. https://doi.org/10.1007/s00445-016-1021-8

Infanti, J., Sixsmith, J., Barry, M., Núñez-Córdoba, J., Oroviogoicoechea-Ortega, C., & Guillén-Grima, F. A. (2013). *A literature review on effective risk communication for the prevention and control of communicable diseases in Europe*. Stockholm: European Centre for Disease Prevention and Control. https://doi.org/10.2900/64747

Jain, S. P., & Posavac, S. S. (2001). Prepurchase attribute verifiability, source credibility, and persuasion. *Journal of Consumer Psychology*, *11*(3), 169–180. https://doi.org/10.1207/S15327663JCP1103_03

Jenkins, S. C., & Harris, A. J. L. (2019). *When words function like numbers: Frame affects perceived function of verbal probability expressions*. Manuscript in preparation.

Jensen, J. D. (2008). Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility. *Human Communication Research*, *34*(3), 347–369. https://doi.org/10.1111/j.1468-2958.2008.00324.x

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. https://doi.org/10.1177/0956797613504301

Joffe, H. (2003). Risk: From perception to social representation. *The British Journal of Social Psychology*, *42*, 55–73. https://doi.org/10.1348/014466603763276126

Johnson, B. B. (2003). Further notes on public response to uncertainty in risks and science. *Risk Analysis*, *23*(4), 781–789. https://doi.org/10.1111/1539-6924.00355

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: Initial studies of its effects on risk perception and trust. *Risk Analysis*, *15*(4), 485–494. https://doi.org/10.1111/j.1539-6924.1995.tb00341.x

Johnson, B. B., & Slovic, P. (1996). *Lay views on uncertainty in health risk assessment: A report on phase II research*. Eugene, Oregon: Decision Research.

Johnson, B. B., & Slovic, P. (1998). Lay views on uncertainty in environmental health risk assessment. *Journal of Risk Research*, *1*(4), 261–279. https://doi.org/10.1080/136698798377042

Joiner, T. A., Leveson, L., & Langfield-Smith, K. (2002). Technical language, advice understandability, and perceptions of expertise and trustworthiness: The case of the financial planner. *Australian Journal of Management*, *27*(1), 25–43. https://doi.org/10.1177/031289620202700102

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., & Grounds, M. A. (2015). The use of uncertainty forecasts in complex decision tasks and various weather conditions. *Journal of Experimental Psychology: Applied*, *21*(4), 407–417. https://doi.org/10.1037/xap0000064

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. https://doi.org/10.1037/a0025185

Joslyn, S. L., Nadav-Greenberg, L., & Nichols, R. M. (2009). Probability of precipitation: Assessment and enhancement of end-user understanding. *Bulletin of the American Meteorological Society*, *90*(2), 185–193. https://doi.org/10.1175/2008BAMS2509.1

Joslyn, S. L., Nemec, L., & Savelli, S. (2013). The benefits and challenges of predictive interval forecasts and verification graphics for end users. *Weather, Climate, and Society*, *5*, 133–147. https://doi.org/10.1175/WCAS-D-12-00007.1

Joslyn, S. L., & Savelli, S. (2010). Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*, *17*(2), 180–195. https://doi.org/10.1002/met.190

Juanchich, M., Gourdon-Kanhukamwe, A., & Sirota, M. (2017). "I am uncertain" vs "It is uncertain". How linguistic markers of the uncertainty source affect uncertainty communication. *Judgment and Decision Making*, *12*(5), 445–465.

Juanchich, M., & Sirota, M. (2013). Do people really say it is "likely" when they believe it is only "possible"? Effect of politeness on risk communication. *Quarterly Journal of Experimental Psychology*, *66*(7), 1268–1275. https://doi.org/10.1080/17470218.2013.804582

Juanchich, M., & Sirota, M. (2016). How to improve people's interpretation of probabilities of precipitation. *Journal of Risk Research*, *19*(3), 388–404. https://doi.org/10.1080/13669877.2014.983945

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. https://doi.org/doi: 10.1037/xap0000137

Juanchich, M., & Sirota, M. (2019). Do people really prefer verbal probabilities? *Psychological Research*, Advance online publication. https://doi.org/10.1007/s00426-019-01207-0

Juanchich, M., Sirota, M., & Butler, C. L. (2012). The perceived functions of linguistic risk quantifiers and their effect on risk, negativity perception and decision making. *Organizational Behavior and Human Decision Processes*, *118*(1), 72–81. https://doi.org/10.1016/j.obhdp.2012.01.002

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.

Jungermann, H., Pfister, H.-R., & Fischer, K. (1996). Credibility, information preferences, and information interests. *Risk Analysis*, *16*(2), 251–261. https://doi.org/10.1111/j.1539-6924.1996.tb01455.x

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*(2), 143–157. https://doi.org/10.21236/ADA099503

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. https://doi.org/10.1037/1076-898X.10.1.25

Kasperson, R. E. (2014). Four questions for risk communication. *Journal of Risk Research*, *17*(10), 1233–1239. https://doi.org/10.1080/13669877.2014.900207

Kasperson, R. E., Golding, D., & Tuler, S. (1992). Social distrust as a factor in siting hazardous facilities and communicating risks. *Journal of Social Issues*, *48*(4), 161–187. https://doi.org/10.1111/j.1540-4560.1992.tb01950.x

Kelman, H. C., & Hovland, C. I. (1953). “Reinstatement” of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology*, *48*(3), 327–335. https://doi.org/10.1037/h0061861

Keren, G., & Teigen, K. H. (2001a). The probability-outcome correspondence principle: A dispositional view of the interpretation of probability statements. *Memory & Cognition*, *29*(7), 1010–1021. https://doi.org/10.3758/BF03195763

Keren, G., & Teigen, K. H. (2001b). Why is p = .90 better than p = .70? Preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin & Review*, *8*(2), 191–202.

Kinzig, A., & Starrett, D. (2003). Coping with uncertainty: a call for a new science-policy forum. *AMBIO*, *32*(5), 330–335. https://doi.org/10.1579/0044-7447-32.5.330

Kirk, R. E. (1969). *Experimental Design: Procedures for the Behavioral Sciences.* Belmont: CA: Brooks/Cole.

Knight, F. (1921). *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

Knoblauch, T. A. K., Stauffacher, M., & Trutnevyte, E. (2018). Communicating low-probability high-consequence risk, uncertainty and expert confidence: Induced seismicity of deep geothermal energy and shale gas. *Risk Analysis*, *38*(4), 694–709. https://doi.org/10.1111/risa.12872

Kuhn, K. M. (2000). Message format and audience values: Interactive effects of uncertainty information and environmental attitudes on perceived risk. *Journal of Environmental Psychology*, *20*(1), 41–51. https://doi.org/10.1006/jevp.1999.0145

Kunz, M., Grêt-Regamey, A., & Hurni, L. (2011). Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Natural Hazards*, *59*(3), 1735–1751. https://doi.org/10.1007/s11069-011-9864-y

Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting*, *21*(1), 3–14. https://doi.org/10.1016/j.ijforecast.2004.02.003

Leiss, W. (1996). Three phases in the evolution of risk communication practice. *The ANNALS of the American Academy of Political and Social Science*, *545*(1), 85–94. https://doi.org/10.1177/0002716296545001009

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

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press. Retrieved from http://cognet.mit.edu/book/presumptive-meanings

Lewandowsky, S., Ballard, T., & Pancost, R. D. (2015). Uncertainty as knowledge. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 1–11. https://doi.org/10.1098/rsta.2014.0462

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, *9*(10), 563–564. https://doi.org/10.3758/BF03327890

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggesting best practices and future recommendations. *Medical Decision Making*, *27*(5), 696–713. https://doi.org/10.1177/0272989X07307271

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., Peters, E. M., Kimmick, G., Liotcheva, V., & Marcom, P. (2010). Breast cancer patients' treatment expectations after exposure to the decision aid program Adjuvant Online: The influence of numeracy. *Medical Decision Making*, *30*(4), 464–473. https://doi.org/10.1177/0272989X09360371

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105

Lippman-Hand, A., Fraser, F. C., & Opitz, J. M. (1979). Genetic counseling — the postcounseling period: I. Parents' perceptions of uncertainty. *American Journal of Medical Genetics*, *4*(1), 51–71. https://doi.org/10.1002/ajmg.1320040108

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. https://doi.org/10.1006/obhd.1997.2679

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

Löfstedt, R. E. (2003). Risk communication: pitfalls and promises. *European Review*, *11*(3), 417–435. https://doi.org/10.1017/S106279870300036X

Löfstedt, R. E. (2005). *Risk Management in Post-Trust Societies*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230503946

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. https://doi.org/10.1016/j.actpsy.2014.01.005

Løhre, E., & Teigen, K. H. (2015). There is a 60% probability, but I am 70% certain: communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, *22*(4), 369–396. https://doi.org/10.1080/13546783.2015.1069758

Løhre, E., & Teigen, K. H. (2017). Probabilities associated with precise and vague forecasts. *Journal of Behavioral Decision Making*, *30*(5), 1014–1026. https://doi.org/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. https://doi.org/10.1016/j.pec.2012.07.010

Losee, J. E., & Joslyn, S. (2018). The need to trust: How features of the forecasted weather influence forecast trust. *International Journal of Disaster Risk Reduction*, *30*(A), 95–104. https://doi.org/10.1016/j.ijdrr.2018.02.032

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the Association for Information Science and Technology*, *62*(7), 1232–1242. https://doi.org/10.1002/asi.21545

Ludy, J., & Kondolf, G. M. (2012). Flood risk perception in lands "protected" by 100-year levees. *Natural Hazards*, *61*(2), 829–842. https://doi.org/10.1007/s11069-011-0072-6

Luhmann, N. (1979). *Trust and Power*. Chichester: John Wiley & Sons.

Lundgren, Regina E ; McMakin, A. H. (2018). Understanding risk communication. In *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks* (Sixth Edit, pp. 9–22). Hoboken, NJ, USA: John Wiley & Sons.

Lundgren, R. E., & McMakin, A. H. (2018). *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks* (Sixth Edit). Hoboken, NJ, USA: John Wiley & Sons.

MacGregor, D. G., Slovic, P., & Morgan, M. G. (1994). Perception of risks from electromagnetic fields: a psychometric evaluation of a risk communication approach. *Risk Analysis*, *14*(5), 815–828. https://doi.org/10.1111/j.1539-6924.1994.tb00293.x

Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General*, *143*(3), 1185–1198. https://doi.org/10.1037/a0034207

Margolis, H. (1996). *Dealing with Risk: Why the Public and the Experts Disagree on Environmental Issues*. Chicago: University of Chicago Press.

Markon, M.-P. L., & Lemyre, L. (2013). Public reactions to risk messages communicating different sources of uncertainty: An experimental test. *Human and Ecological Risk Assessment*, *19*(4), 1102–1126. https://doi.org/10.1080/10807039.2012.702015

Marquart, J., O'Keefe, G. J., & Gunther, A. C. (1995). Believing in biotech. *Science Communication*, *16*(4), 388–402. https://doi.org/10.1177/1075547095016004002

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

Marzocchi, W., Newhall, C., & Woo, G. (2012). The scientific management of volcanic crises. *Journal of Volcanology and Geothermal Research*, *247–248*, 181–189. https://doi.org/10.1016/j.jvolgeores.2012.08.016

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Kroegler E., Mach, K.J., Matschoss, P.R., Plattner, G.K. Yohee, G.W. & Zwiers, F. W. (2010). *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.* Retrieved from IPCC website: https://archive.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf

Mazur, D. J., & Hickam, D. H. (1991). Patients' interpretations of probability terms. *Journal of General Internal Medicine*, *6*(3), 237–240. https://doi.org/10.1007/BF02598968

McClure, J., Doyle, E. E. H., & Velluppillai, J. M. (2015). A tale of two cities: Judgments about earthquake and aftershock probabilities across time windows. *International Journal of Disaster Risk Reduction*, *14*, 15–26. https://doi.org/10.1016/j.ijdrr.2014.11.010

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. https://doi.org/10.1016/j.ijdrr.2016.03.003

McClure, J., Walkey, F., & Allen, M. (1999). When earthquake damage is seen as preventable: Attributions, locus of control and attitudes to risk. *Applied Psychology*, *48*(2), 239–256. https://doi.org/10.1080/026999499377637

McCracken, G. (1989). Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research*, *16*(3), 310. https://doi.org/10.1086/209217

McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, *6*(3), 467–472. https://doi.org/10.1177/0021955X9903500605

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (pp. 233–346). New York: Random House.

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

Merz, J. F., Druzdzel, M. J., & Mazur, D. J. (1991). Verbal expressions of probability in informed consent litigation. *Medical Decision Making*, *11*(4), 273–281. https://doi.org/10.1177/0272989X9101100405

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication Yearbook*, *27*(1), 293–335. https://doi.org/10.1207/s15567419cy2701_10

Miles, S., & Frewer, L. J. (2003). Public perception of scientific uncertainty in relation to food hazards. *Journal of Risk Research*, *6*(3), 267–283. https://doi.org/10.1080/1366987032000088883

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. https://doi.org/10.1002/9781118468333.ch6

Moraes, C. G., & da Silva Dal Pizzol, T. (2018). Effect of different formats for information on side effects regarding medicine users' understanding: A randomized controlled trial. *Patient Education and Counseling*, *101*(4), 672–678. https://doi.org/10.1016/j.pec.2017.11.015

Morgan, M. G., & Henrion, M. (1990). The nature and sources of uncertainty. In *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (pp. 47–72). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511840609.005

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Weather and Forecasting*, *23*(5), 974–991. https://doi.org/10.1175/2008WAF2007088.1

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, *5*(1), 2–34. https://doi.org/10.1214/ss/1177013604

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, Retrieved from http://www.acrwebsite.org/volumes/6770/volumes/v14/NA-14

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

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition*, *93*(1), 1–41. https://doi.org/10.1016/j.cognition.2003.10.002

Nakayachi, K., Johnson, B. B., & Koketsu, K. (2018). Effects of acknowledging uncertainty about earthquake risk estimates on San Francisco Bay area residents' beliefs, attitudes, and intentions. *Risk Analysis*, *38*(4), 666–679. https://doi.org/10.1111/risa.12883

National Research Council. (1989). *Improving Risk Communication*. Washington DC: National Academies Press. https://doi.org/10.17226/1189

National Research Council. (2000). *Risk Analysis and Uncertainty in Flood Damage Reduction Studies*. Washington, DC: National Academies Press. https://doi.org/10.17226/9971

National Research Council. (2005). *Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program*. Washington, DC: National Academies Press. https://doi.org/10.17226/11292.

Newhall, C., & Punongbayan, R. S. (1996). The narrow margin of successful volcanic-risk mitigation. In R. Scarpa & R. I. Tilling (Eds.), *Monitoring and Mitigation of Volcano Hazards* (pp. 807–838). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-80087-0_25

Ng, K. L., & Hamby, D. M. (1997). Fundamentals for establishing a risk communication program. *Health Physics*, *73*(3), 473–482. https://doi.org/10.1097/00004032-199709000-00005

O'Brien, B. J. (1989). Words or numbers? The evaluation of probability expressions in general practice. *Journal of the Royal College of General Practitioners*, *39*(320), 98–100.

O'Hara, B. S., Netemeyer, R. G., & Burton, S. (1991). An examination of the relative effects of source expertise, trustworthiness, and likeability. *Social Behaviour and Personality*, *19*(4), 305–314. https://doi.org/10.2224/sbp.1991.19.4.305

O'Keefe, D. (2016). *Persuasion: Theory and Research* (Third Edit). Thousand Oaks, CA: SAGE Publications.

O'Neill, S. C., DeMarco, T., Peshkin, B. N., Rogers, S., Rispoli, J., Brown, K., Valdimarsdottir, H., & Schwartz, M. D. (2006). Tolerance for uncertainty and perceived risk among women receiving uninformative BRCA1/2 test results. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *142C*(4), 251–259. https://doi.org/10.1002/ajmg.c.30104

Ogden, J., Fuks, K., Gardner, M., Johnson, S., McLean, M., Martin, P., & Shah, R. (2002). Doctors expressions of uncertainty and patient confidence. *Patient Education and Counseling*, *48*(2), 171–176. https://doi.org/10.1016/S0738-3991(02)00020-4

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.2307/4188769

Ohanian, R. (1991). The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *Journal of Advertising Research*, *31*(1), 46–54.

Önkal, D., Sinan Gönül, M., Goodwin, P., Thomson, M., & Öz, E. (2017). Evaluating expert advice in forecasting: Users' reactions to presumed vs. experienced credibility. *International Journal of Forecasting*, *33*(1), 280–297. https://doi.org/10.1016/j.ijforecast.2015.12.009

Papadopoulou, L., McEntaggart, K., & Etienne, J. (2018). *Communicating Scientific Uncertainty in Advice Provision to Decision-Makers: Review of Approaches and Recommendations for UK Statutory Nature Conservation Bodies*. Retrieved from Joint Nature Conservation Committee website: http://jncc.defra.gov.uk/default.aspx?page=6675.

Paton, D. (2008). Risk communication and natural hazard mitigation: how trust influences its effectiveness. *International Journal of Global Environmental Issues*, *8*(1/2), 2–16. https://doi.org/10.1504/ijgenvi.2008.017256

Paton, D., Sagala, S., Okada, N., Jang, L.-J., Bürgelt, P. T., & Gregg, C. E. (2010). Making sense of natural hazard mitigation: Personal, social and cultural influences. *Environmental Hazards*, *9*(2), 183–196. https://doi.org/10.3763/ehaz.2010.0039

Patt, A. G., & Dessai, S. (2005). Communicating uncertainty: Lessons learned and suggestions for climate change assessment. *Comptes Rendus Geoscience*, *337*(4), 425–441. https://doi.org/http://doi.org/10.1016/j.crte.2004.10.004

Patt, A. G., & Schrag, D. P. (2003). Using specific language to describe risk and probability. *Climate Change*, *61*, 17–30. https://doi.org/10.1023/A:1026314523443

Peachey, J. A., Schultz, D. M., Morss, R. E., Roebber, P. J., & Wood, R. (2013). How forecasts expressing uncertainty are perceived by UK students. *Weather*, *68*(7), 176–181. https://doi.org/10.1002/wea.2094

Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality*, *8*(1), 95–101. https://doi.org/10.1016/0092-6566(74)90049-X

Peters, E. M. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*(1), 31–35. https://doi.org/10.1177/0963721411429960

Peters, E. M., Dieckmann, N., Västfjäll, D., Mertz, C. K., Slovic, P., & Hibbard, J. H. (2009). Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied*, *15*(3), 213–227. https://doi.org/10.1037/a0016978

Peters, E. M., Hart, P. S., & Fraenkel, L. (2011). Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making*, *31*(3), 432–436. https://doi.org/10.1177/0272989X10391672

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. https://doi.org/10.1177/0272989X13511705

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

Peters, R. G., Covello, V. T., & McCallum, D. B. (1997). The determinants of trust and credibility in environmental risk communication: An empirical study. *Risk Analysis*, *17*(1), 43–54. https://doi.org/10.1111/j.1539-6924.1997.tb00842.x

Peterson, D. W., & Tilling, R. I. (1993). Interactions between scientists, civil authorities and the public at hazardous volcanoes. In C. R. J. Kilburn & G. Luongo (Eds.), *Active Lavas: Monitoring and Modelling* (pp. 339–365). London: UCL Press.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847–855. https://doi.org/10.1037/0022-3514.41.5.847

Pidgeon, N. F. (1998). Risk assessment, risk values and the social science programme: why we do need risk perception research. *Reliability Engineering & System Safety*, *59*(1), 5–15. https://doi.org/10.1016/S0951-8320(97)00114-2

Piercey, M. D. (2009). Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organizational Behavior and Human Decision Processes*, *108*(2), 330–341. https://doi.org/10.1016/j.obhdp.2008.05.004

Pighin, S., Bonnefon, J.-F., & Savadori, L. (2011). Overcoming number numbness in prenatal risk communication. *Prenatal Diagnosis*, *31*(8), 809–813. https://doi.org/10.1002/pd.2771

Plough, A., & Krimsky, S. (1987). The emergence of risk communication studies: Social and political context. *Science, Technology, & Human Values*, *12*(3/4), 4–10.

Politi, M. C., Clark, M. A., Ombao, H., Dizon, D., & Elwyn, G. (2011). Communicating uncertainty can lead to less decision satisfaction: A necessary cost of involving patients in shared decision making? *Health Expectations*, *14*(1), 84–91. https://doi.org/10.1111/j.1369-7625.2010.00626.x

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. https://doi.org/10.1177/0272989X07307270

Pornpitakpan, C. (2004a). The effect of celebrity endorsers' perceived credibility on product purchase intention. *Journal of International Consumer Marketing*, *16*(2), 55–74. https://doi.org/10.1300/J046v16n02_04

Pornpitakpan, C. (2004b). The persuasiveness of source credibility: a critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Portnoy, D. B., Han, P. K. J., Ferrer, R. A., Klein, W. M. P., & Clauser, S. B. (2013). Physicians' attitudes about communicating and managing scientific uncertainty differ by perceived ambiguity aversion of their patients. *Health Expectations*, *16*(4), 362–372. https://doi.org/10.1111/j.1369-7625.2011.00717.x

Rapoport, A., Wallsten, T. S., & Cox, J. A. (1987). Direct and indirect scaling of membership functions of probability phrases. *Mathematical Modelling*, *9*(6), 397–417. https://doi.org/10.1016/0270-0255(87)90506-9

Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, *74*(3), 433–442. https://doi.org/10.1037/0021-9010.74.3.433

Regan, Á., McConnon, Á., Kuttschreuter, M., Rutsaert, P., Shan, L., Pieniak, Z., Barnett, J., Verbeke, W., & Wall, P. (2014). The impact of communicating conflicting risk and benefit messages: An experimental study on red meat information. *Food Quality and Preference*, *38*, 107–114. https://doi.org/10.1016/j.foodqual.2014.05.019

Reichelt, J., Sievert, J., & Jacob, F. (2014). How credibility affects eWOM reading: The influences of expertise, trustworthiness, and similarity on utilitarian and social functions. *Journal of Marketing Communications*, *20*(1–2), 65–81. https://doi.org/10.1080/13527266.2013.797758

Reimer, T., Jones, C., & Skubisz, C. (2014). Numeric communication of risk. In *The SAGE Handbook of Risk Communication* (pp. 166–179). Thousand Oaks, California: SAGE Publications, Inc. https://doi.org/10.4135/9781483387918.n20

Renn, O. (1992a). Concepts of risk: a classification. In S. Krimsky & D. Golding (Eds.), *Social Theories of Risk* (pp. 53–79). Westport, CT: Praeger. https://doi.org/10.18419/opus-7248

Renn, O. (1992b). Risk communication: Towards a rational discourse with the public. *Journal of Hazardous Materials*, *29*(3), 465–519. https://doi.org/10.1016/0304-3894(92)85047-5

Renn, O. (1998). Three decades of risk research: accomplishments and new challenges. *Journal of Risk Research*, *1*(1), 49–71. https://doi.org/10.1080/136698798377321

Renn, O. (2005). Risk perception and communication: Lessons for the food and food packaging industry. *Food Additives and Contaminants*, *22*(10), 1061–1071. https://doi.org/10.1080/02652030500227792

Renn, O., & Levine, D. (1991). Credibility and trust in risk communication. In R. E. Kasperson & P. J. Stallen (Eds.), *Communicating Risks to the Public: International Perspectives* (pp. 175–218). Dordrecht, Netherlands: Kluwer. https://doi.org/10.1007/978-94-009-1952-5

Renooij, S., & Witteman, C. L. M. (1999). Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, *22*(3), 169–194. https://doi.org/10.1016/S0888-613X(99)00027-4

Reyna, V. F. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory & Cognition*, *9*(6), 642–650. https://doi.org/10.3758/BF03202359

Reyna, V. F., Nelson, W. L., Han, P. K. J., & Dieckmann, N. (2010). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327.How

Reynolds, B. J. (2011). When the facts are just not enough: Credibly communicating about risk is riskier when emotions run high and time is short. *Toxicology and Applied Pharmacology*, *254*(2), 206–214. https://doi.org/10.1016/J.TAAP.2010.10.023

Richards, L., & Heath, A. (2015). *CSI 15: The Uneven Distribution and Decline of Social Capital in Britain*. Retrieved from Understanding Society website: https://www.understandingsociety.ac.uk/research/publications/523399

Rieh, S. Y. (2009). Credibility and cognitive authority of information. In M. J. Bates & M. Niles Maack (Eds.), *Encyclopedia of Library and Information Sciences* (Third Ed, pp. 1337–1344). Boca Raton, FL: CRC Press. https://doi.org/10.1081/E-ELIS3-120044103

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, *41*(1), 307–364. https://doi.org/10.1002/aris.2007.1440410114

Risbey, J. S., & Kandlikar, M. (2007). Expressions of likelihood and confidence in the IPCC uncertainty assessment process. *Climatic Change*, *85*(1–2), 19–31. https://doi.org/10.1007/s10584-007-9315-7

Rød, S. K., Botan, C., & Holen, A. (2012). Risk communication and the willingness to follow evacuation instructions in a natural disaster. *Health, Risk & Society*, *14*(1), 87–99. https://doi.org/10.1080/13698575.2011.641522

Rowe, G., & Frewer, L. J. (2000). Public participation methods: A framework for evaluation. *Science, Technology, & Human Values*, *25*(1), 3–29. https://doi.org/10.1177/016224390002500101

Samson, S., Reneke, J. A., & Wiecek, M. M. (2009). A review of different perspectives on uncertainty and risk and an alternative modeling paradigm. *Reliability Engineering & System Safety*, *94*(2), 558–567. https://doi.org/10.1016/J.RESS.2008.06.004

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

Sapir, D. G., & Lechat, M. F. (1986). Reducing the impact of natural disasters: why aren't we better prepared? *Health Policy and Planning*, *1*(2), 118–126. https://doi.org/10.1093/heapol/1.2.118

Savadori, L., Savio, S., Nicotra, E., Rumiati, R., Finucane, M. L., & Slovic, P. (2004). Expert and public perception of risk from biotechnology. *Risk Analysis*, *24*(5), 1289–1299. https://doi.org/10.1111/j.0272-4332.2004.00526.x

Scheer, D., Benighaus, C., Benighaus, L., Renn, O., Gold, S., Röder, B., & Böl, G. F. (2014). The distinction between risk and hazard: Understanding and use in stakeholder communication. *Risk Analysis*, *34*(7), 1270–1285. https://doi.org/10.1111/risa.12169

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. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024187

Self, C. C. (2009). Credibility. In D. W. Stacks & M. B. Salwen (Eds.), *An Integrated Approach to Communication Theory and Research* (Second Ed, pp. 435–456). Abingdon, England: Routledge.

Sellnow, D. D., Iverson, J., & Sellnow, T. L. (2017). The evolution of the operational earthquake forecasting community of practice: the L'Aquila communication crisis as a triggering event for organizational renewal. *Journal of Applied Communication Research*, *45*(2), 121–139. https://doi.org/10.1080/00909882.2017.1288295

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447. https://doi.org/10.1111/j.1467-7687.2012.01135.x

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*(2), 252–266. https://doi.org/10.1016/0749-5978(92)90064-E

Sharma, U., & Patt, A. (2012). Disaster warning response: the effects of different types of personal experience. *Natural Hazards*, *60*(2), 409–423. https://doi.org/10.1007/s11069-011-0023-2

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

Sher, S., & McKenzie, C. (2008). Framing effects and rationality. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 79–96). Oxford, England: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216093.003.0004

Siegrist, M. (1999). A causal model explaining the perception and acceptance of gene technology. *Journal of Applied Social Psychology*, *29*(10), 2093–2106. https://doi.org/10.1111/j.1559-1816.1999.tb02297.x

Siegrist, M. (2000). The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk Analysis*, *20*(2), 195–204. https://doi.org/10.1111/0272-4332.202020

Siegrist, M., & Cvetkovich, G. (2000). Perception of hazards: The role of social trust and knowledge. *Risk Analysis*, *20*(5), 713–719. https://doi.org/10.1111/0272-4332.205064

Siegrist, M., Cvetkovich, G., & Roth, C. (2000). Salient value similarity, social trust, and risk/benefit perception. *Risk Analysis*, *20*(3), 353–362. https://doi.org/10.1111/0272-4332.203034

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. https://doi.org/10.1080/1366987032000105315

Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science*, *25*(4), 400–414. https://doi.org/10.1177/0963662516629749

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

Sirota, M., & Juanchich, M. (2015). A direct and comprehensive test of two postulates of politeness theory applied to uncertainty communication. *Judgment and Decision Making*, *10*(23), 232–240.

Sjöberg, L. (2000). Factors in risk perception. *Risk Analysis*, *20*(1), 1–11. https://doi.org/10.1111/0272-4332.00001

Sjöberg, L. (2001). Limits of knowledge and the limited importance of trust. *Risk Analysis*, *21*(1), 189–198. https://doi.org/10.1111/0272-4332.211101

Skubisz, C. (2010). Perceptions of risk evidence: Are all statistics considered equal? In D. S. Gouran (Ed.), *The Functions of Argument and Social Context* (pp. 476–484). Washington, DC: National Communication Association.

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. https://doi.org/10.1016/j.pec.2016.01.009

Slater, M. D., & Rouner, D. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly*, *73*(4), 974–991. https://doi.org/10.1177/107769909607300415

Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Slovic, P. (1992). Perception of risk: Reflections on the psychometric paradigm. In D. Golding & S. Krimsky (Eds.), *Theories of Risk* (pp. 117–152). New York: Praeger.

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

Slovic, P. (1999). Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Analysis*, *19*(4), 689–701. https://doi.org/10.1111/j.1539-6924.1999.tb00439.x

Slovic, P. (2010). *The Feeling of Risk: New Perspectives on Risk Perception*. London: Routledge. https://doi.org/10.4324/9781849776677

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty* (pp. 463–492). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.034

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24*(3), 271–296. https://doi.org/10.1023/A:1005595519944

Slovic, P., & Weber, E. U. (2002). Perception of risk posed by extreme events. In *Conference on Risk Management Strategies in an Uncertain World* (pp. 1–21). Palisades, New York. https://doi.org/10.1017/CBO9781107415324.004

Smithson, M. (2012). Psychology's ambivalent view of uncertainty. In G. Bammer & M. Smithson (Eds.), *Uncertainty and Risk: Multidisciplinary Perspectives* (pp. 205–217). London: Routledge. https://doi.org/10.4324/9781849773607

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. https://doi.org/10.1016/j.ijar.2012.06.019

Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition*. Oxford, England: Blackwell Publishing.

Starr, C. (1969). Social benefit versus technological risk. *Science*, *165*(3899), 1232–1238. https://doi.org/10.1126/science.165.3899.1232

Sternthal, B., Phillips, L. W., & Dholakia, R. (1978). The persuasive effect of source credibility: a situational analysis. *Public Opinion Quarterly*, *42*(3), 285–314. https://doi.org/10.1086/268454

Sword-Daniels, V., Eriksen, C., Hudson-Doyle, E. E., Alaniz, R., Adler, C. E., Schenk, T., & Vallance, S. (2018). Embodied uncertainty: living with complexity and natural hazards. *Journal of Risk Research*, *21*(3), 290–307. https://doi.org/10.1080/13669877.2016.1200659

Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science*, *63*(2), 497–518. https://doi.org/10.1287/mnsc.2015.2344

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*(1), 27–38. https://doi.org/10.1016/0001-6918(88)90043-1

Teigen, K. H. (2008). More than X is a lot: Pragmatic implicatures of one-sided uncertainty intervals. *Social Cognition*, *26*(4), 379–400. https://doi.org/10.1521/soco.2008.26.4.379

Teigen, K. H. (2014). When very likely is not so likely. *Nature Climate Change*, *4*(6), 421–422. https://doi.org/10.1038/nclimate2256

Teigen, K. H. (2015). *Framing of numerical quantities*. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333

Teigen, K. H., Andersen, B., Alnes, S. L., & Hesselberg, J.-O. (2019). Entirely possible overruns: How people think and talk about probabilistic cost estimates. *International Journal of Managing Projects in Business*, Advance online publication. https://doi.org/10.1108/IJMPB-06-2018-0114

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

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

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., & Filkuková, P. (2013). Can > will: Predictions of what can happen are extreme, but believed to be probable. *Journal of Behavioral Decision Making*, *26*(1), 68–78. https://doi.org/10.1002/bdm.761

Teigen, K. H., Filkuková, P., & Hohle, S. M. (2018). It can become 5 °C warmer: The extremity effect in climate forecasts. *Journal of Experimental Psychology: Applied*, *24*(1), 3–17. https://doi.org/10.1037/xap0000149

Teigen, K. H., Halberg, A.-M., & Fostervold, K. I. (2007). Single-limit interval estimates as reference points. *Applied Cognitive Psychology*, *21*(3), 383–406. https://doi.org/10.1002/acp.1283

Teigen, K. H., Juanchich, M., & Filkuková, P. (2014). Verbal probabilities: An alternative approach. *The Quarterly Journal of Experimental Psychology*, *67*(1), 124–146. https://doi.org/10.1080/17470218.2013.793731

Teigen, K. H., Juanchich, M., & Riege, A. H. (2013). Improbable outcomes: Infrequent or extraordinary? *Cognition*, *127*(1), 119–139. https://doi.org/10.1016/j.cognition.2012.12.005

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts. *Cognition*, *25*(4), 291–293. https://doi.org/10.1016/S0

Teigen, K. H., & Løhre, E. (2017). Expressing (un)certainty in no uncertain terms: reply to Fox and Ülkümen. *Thinking and Reasoning*. https://doi.org/10.1080/13546783.2017.1314965

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. https://doi.org/10.1080/13669870110038179

Thon, F. M., & Jucks, R. (2017). Believing in expertise: How authors' credentials and language use influence the credibility of online health information. *Health Communication*, *32*(7), 828–836. https://doi.org/10.1080/10410236.2016.1172296

Timmermans, D. (1994). The roles of experience and domain of expertise in using numerical and verbal probability terms in medical decisions. *Medical Decision Making*, *14*(2), 146–156. https://doi.org/10.1177/0272989X9401400207

Timmermans, D., & Oudhoff, J. (2011). Different formats for the communication of risks: Verbal, numerical, and graphical formats. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, & J. C. Smith (Eds.), *Wiley Encyclopedia of Operations Research and Management Science* (pp. 1–11). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9780470400531.eorms0954

Treise, D., Walsh-Childers, K., Weigold, M. F., & Friedman, M. (2003). Cultivating the science internet audience. *Science Communication*, *24*(3), 309–332. https://doi.org/10.1177/1075547002250298

Trettin, L., & Musham, C. (2000). Is trust a realistic goal of environmental risk communication? *Environment and Behavior*, *32*(3), 410–426. https://doi.org/10.1177/00139160021972595

Trumbo, C. W., & McComas, K. A. (2003). The function of credibility in information processing for risk perception. *Risk Analysis*, *23*(2), 343–353. https://doi.org/10.1111/1539-6924.00313

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.

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, *145*(10), 1280–1297. https://doi.org/10.1037/xge0000202

van Asselt, M., & Rotmans, J. (2002). Uncertainty in integrated assessment modelling. From positivism to pluralism. *Climatic Change*, *54*(1–2), 75–105. https://doi.org/10.1023/A:1015783803445

van Asselt, M., & Vos, E. (2006). The precautionary principle and the uncertainty paradox. *Journal of Risk Research*, *9*(4), 313–336. https://doi.org/10.1080/13669870500175063

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, *6*(5), 181870. https://doi.org/10.1098/rsos.181870

van der Linden, S. (2014). On the relationship between personal experience, affect and risk perception:The case of climate change. *European Journal of Social Psychology*, *44*(5), 430–440. https://doi.org/10.1002/ejsp.2008

van Kleef, E., Houghton, J. R., Krystallis, A., Pfenning, U., Rowe, G., Van Dijk, H., Van der Lans, I.A., & Frewer, L. J. (2007). Consumer evaluations of food risk management quality in Europe. *Risk Analysis*, *27*(6), 1565–1580. https://doi.org/10.1111/j.1539-6924.2007.00989.x

Verbeke, W., Frewer, L. J., Scholderer, J., & De Brabander, H. F. (2007). Why consumers behave as they do with respect to food safety and risk information. *Analytica Chimica Acta*, *586*, 2–7. https://doi.org/10.1016/j.aca.2006.07.065

Verplanken, B. (1997). The effect of catastrophe potential on the interpretation of numerical probabilities of the occurrence of hazards. *Journal of Applied Social Psychology*, *27*(16), 1453–1467. https://doi.org/10.1111/j.1559-1816.1997.tb01608.x

Villejoubert, G., Almond, L., & Alison, L. (2009). Interpreting claims in offender profiles: The role of probability phrases, base-rates and perceived dangerousness. *Applied Cognitive Psychology*, *23*(1), 36–54. https://doi.org/10.1002/acp.1438

Viscusi, W. K., Magat, W. A., & Huber, J. (1991). Communication of ambiguous risk information. *Theory and Decision*, *31*(2–3), 159–173. https://doi.org/10.1007/BF00132991

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. https://doi.org/10.1111/j.1539-6924.2012.01942.x

Walaski, P. F. (2011). *Risk and Crisis Communications*. Hoboken, NJ, USA: John Wiley & Sons. https://doi.org/10.1002/9781118093429

Walker, J. (1823). *A Rhetorical Grammar or Course of Lessons in Elocution*. Oxford, England: Oxford University Press.

Walker, W., Rotmans, J., van der Sluijs, J. P., van Asselt, M., Janssen, P., & Krayer von Krauss, M. (2003). Defining uncertainty: A conceptual basis for uncertainty management. *Integrated Assessment*, *4*(1), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466

Wallsten, T. S. (1990). Measuring vague uncertainties and understanding their use in decision making. In G. M. von Furstenberg (Ed.), *Acting under Uncertainty: Multidisciplinary Conceptions* (pp. 377–398). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-7873-8_15

Wallsten, T. S., Budescu, D. V, Rappaport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*(4), 348–365. https://doi.org/10.21236/ADA196944

Wallsten, T. S., Budescu, D. V, Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, *31*(2), 135–138. https://doi.org/10.3758/BF03334162

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, *25*(5), 571–587. https://doi.org/10.1016/0749-596X(86)90012-4

Walton, D. (1997). *Appeal to Expert Opinion: Arguments from Authority*. University Park, PA: Pennsylvania State University Press.

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, *53*(2), 134–144. https://doi.org/10.1002/asi.10016

Watts, S., & Zhang, W. (2008). Capitalizing on content: information adoption in two online communities. *Journal of the Association for Information Systems*, *9*(2), 73–94. https://doi.org/10.17705/1jais.00149

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

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. https://doi.org/10.1037/0096-1523.16.4.781

Welkenhuysen, M., Evers-Kiebooms, G., & D'Ydewalle, G. (2001). The language of uncertainty in genetic risk communication: Framing and verbal versus numerical information. *Patient Education and Counseling*, *43*(2), 179–187. https://doi.org/10.1016/S0738-3991(00)00161-0

Wiedemann, P. M., Schtz, H., & Thalmann, A. (2008). Perception of uncertainty and communication about unclear risks. In P. M. Wiedemann & H. Schütz (Eds.), *The Role of Evidence in Risk Characterization* (pp. 163–183). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527622351.ch14

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

Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, *21*(2), 101–112. https://doi.org/10.1007/BF02894421

Windschitl, P. D., Martin, R., & Flugstad, A. R. (2002). Context and the interpretation of likelihood information: The role of intergroup comparisons on perceived vulnerability. *Journal of Personality and Social Psychology*, *82*(5), 742–755. https://doi.org/10.1037//0022-3514.82.5.742

Windschitl, P. D., Smith, A. R., Scherer, A. M., & Suls, J. (2017). Risk it? Direct and collateral impacts of peers' verbal expressions about hazard likelihoods. *Thinking & Reasoning*, *23*(3), 259–291. https://doi.org/10.1080/13546783.2017.1307785

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70%--right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory and Cognition*. https://doi.org/10.1037/0278-7393.25.6.1514

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology*, *2*(4), 343–364. https://doi.org/10.1037/1076-898X.2.4.343

Wintle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E., & Fidler, F. (2019). Verbal probabilities: Very likely to be somewhat more confusing than numbers. *PLOS ONE*, *14*(4), e0213522. https://doi.org/10.1371/journal.pone.0213522

Witteman, C. L. M., & Renooij, S. (2003). Evaluation of a verbal-numerical probability scale. *International Journal of Approximate Reasoning*, *33*(2), 117–131. https://doi.org/10.1016/S0888-613X(02)00151-2

Witteman, C. L. M., Renooij, S., & Koele, P. (2007). Medicine in words and numbers: a cross-sectional survey comparing probability assessment scales. *BMC Medical Informatics and Decision Making*, *7*(1), 13. https://doi.org/10.1186/1472-6947-7-13

Witteman, H. O., Zikmund-Fisher, B. J., Waters, E. A., Gavaruzzi, T., & Fagerlin, A. (2011). Risk estimates from an online risk calculator are more believable and recalled better when expressed as integers. *Journal of Medical Internet Research*, *13*(3), e54. https://doi.org/10.2196/jmir.1656

Woo, G. (2008). Probabilistic criteria for volcano evacuation decisions. *Natural Hazards*, *45*(1), 87–97. https://doi.org/10.1007/s11069-007-9171-9

World Economic Forum. (2019). *The Global Risks Report 2019*. Retrieved from World Economic Forum http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf

World Health Organization. (2014). An Introduction to Risk Communication. Retrieved from Retrieved from World Health Organization website: https://www.who.int/risk-communication/introduction-to-risk-communication.pdf?ua=1&ua=1

Wynne, B. (1992). Uncertainty and environmental learning. Reconceiving science and policy in the preventive paradigm. *Global Environmental Change*, *2*(2), 111–127. https://doi.org/10.1016/0959-3780(92)90017-2

Yang, Q., & Beatty, M. (2016). A meta-analytic review of health information credibility. *Health Information Management Journal*, *45*(2), 80–89. https://doi.org/10.1177/1833358316639432

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, *124*(4), 424–432. https://doi.org/10.1037/0096-3445.124.4.424

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

Yoo, K. H., & Gretzel, U. (2008). The influence of perceived credibility on preferences for recommender systems as sources of advice. *Information Technology & Tourism*, *10*(2), 133–146. https://doi.org/10.3727/109830508784913059

Young, S., & Oppenheimer, D. M. (2009). Effect of communication strategy on personal risk perception and treatment adherence intentions. *Psychology, Health & Medicine, 14*(4), 430–442. https://doi.org/10.1080/13548500902890103

Zhang, Y., & Schwarz, N. (2012). How and why 1 year differs from 365 days: A conversational logic analysis of inferences from the granularity of quantitative expressions. *Journal of Consumer Research*, *39*(2), 248–259. https://doi.org/10.1086/662612

Zhang, Y., & Schwarz, N. (2013). The power of precise numbers: A conversational logic analysis. *Journal of Experimental Social Psychology*, *49*(5), 944–946. https://doi.org/10.1016/j.jesp.2013.04.002

Zikmund-Fisher, B. J. (2012). The right tool is what they need, not what we have: a taxonomy of appropriate levels of precision in patient risk communication. *Medical Care Research and Review*, *70*(Supplement 1), 37-49S. https://doi.org/10.1177/1077558712458541

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probabilities. *Advances in Psychology*, *16*, 159–182. https://doi.org/10.1016/S0166-4115(08)62198-6

Appendix

Appendix A. Natural Hazard Vignettes

Volcano.

Mount Ablon has a history of explosive eruptions forming lava flows. An eruption has been predicted; the figure below shows the model's predictions of the distance extended by lava flows for this eruption, given the volcano's situation and recent scientific observations.

Flooding.

The Wayston flood plain has a history of flooding due to its flat terrain and proximity to the east side of the River Wayston. A flood has been predicted; the figure below shows the model's predictions of the distance extended by floodwater for this flood, given the river's situation and recent scientific observations.

Earthquake.

The uninhabited area of Griffinton lies on an active fault line and has a history of experiencing earthquake activity, resulting in tremors which can be felt in various areas. A large earthquake has been predicted; the figure below shows the model's predictions of the distance extended by the tremors for this earthquake, given the fault line's situation and recent scientific observations.

Landslide.

A site on the Copeland Pass has been identified by geologists as susceptible to debris flow, with evidence of past flow of material occurring in certain weather conditions. A landslide has been predicted; the figure below shows the model's predictions of the length extended by the debris flow for this landslide, given its situation and recent scientific observations.

Appendix B. Experiments 1 and 2: Geological Scenario Figure Examples

(Correct answer in brackets)

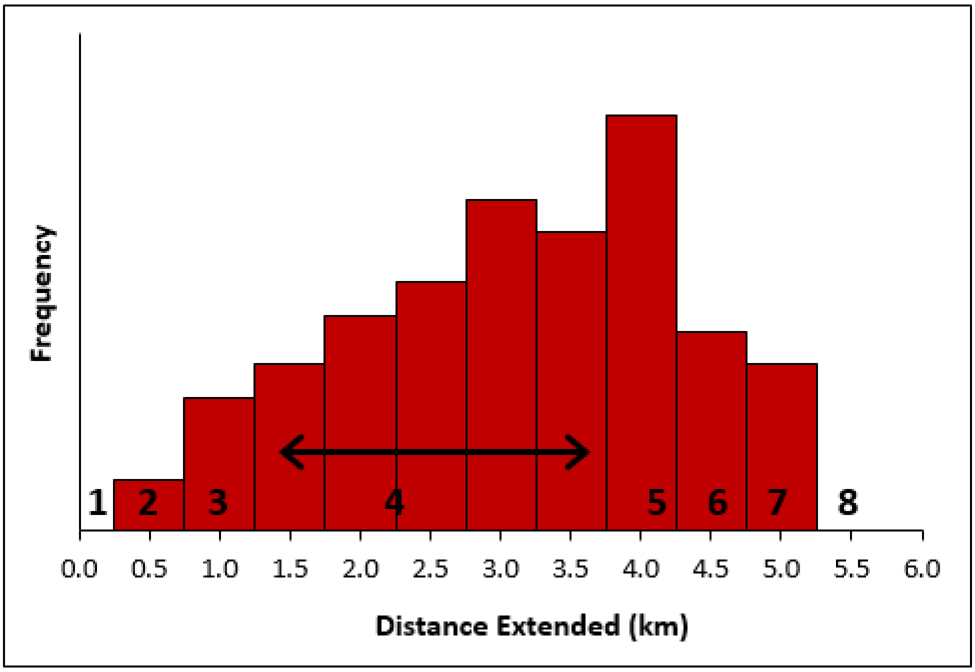


Figure B1. Volcano – No salient site (4 km). Bold numbers represent outcome bin.

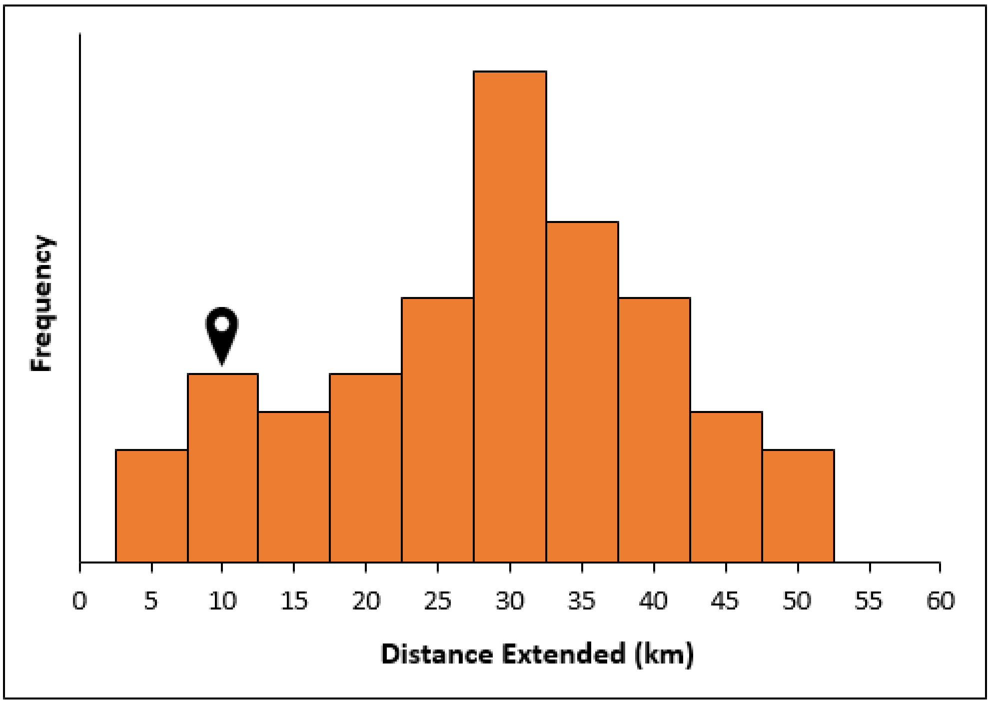


Figure B2. Earthquake – Close salient site (40 km).

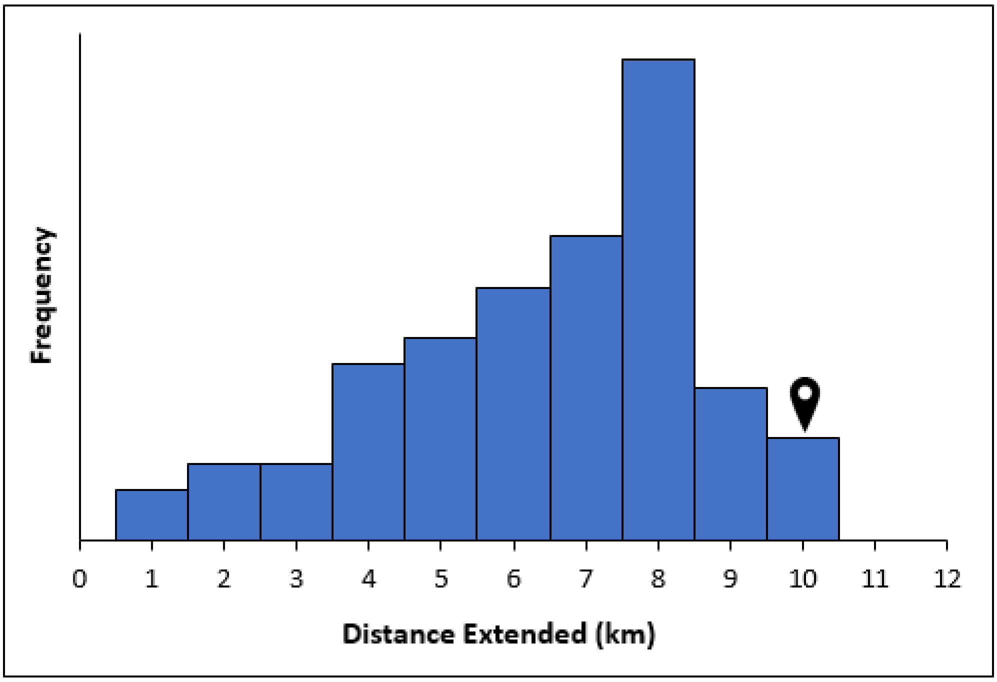


Figure B3. Flooding – Far salient site (8 km).

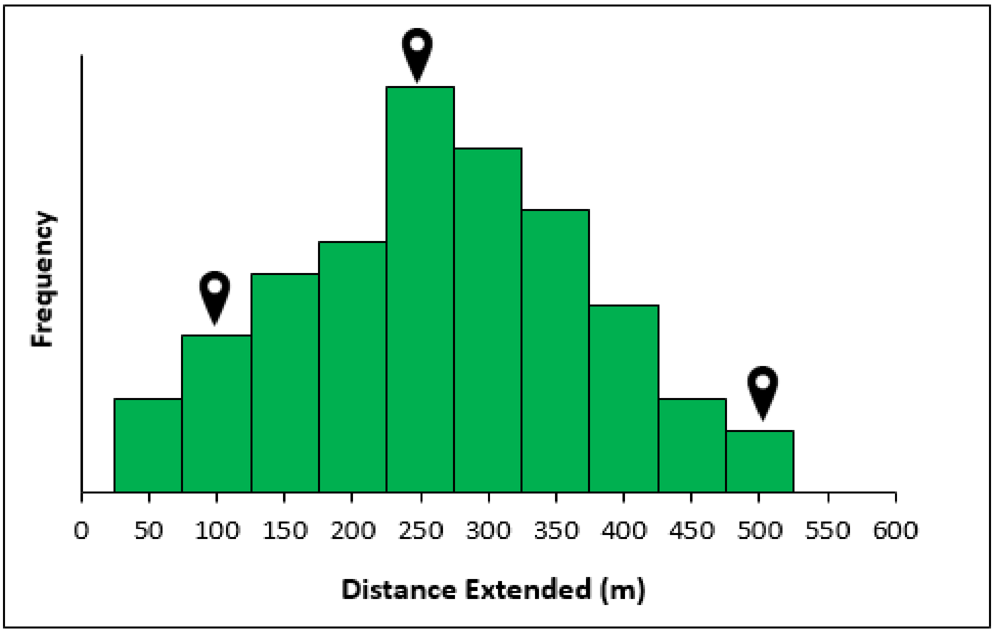


Figure B4. Landslide – Multiple salient sites (350 m).

Appendix C. Experiment 3: Battery Histograms

(Correct answer in brackets)

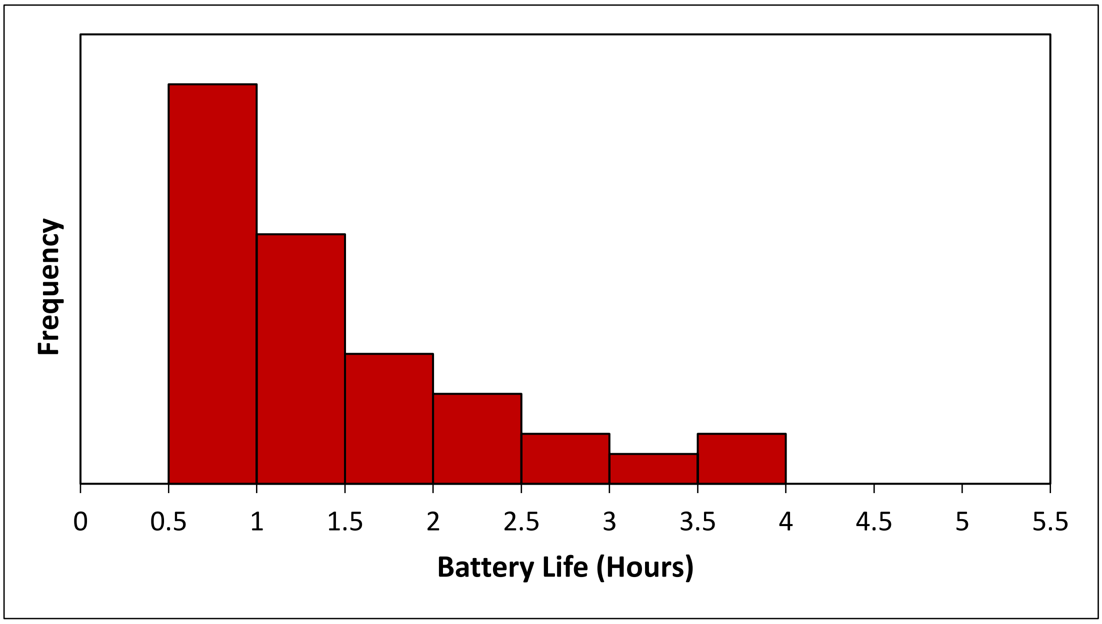


Figure C1. Positive (2.25 hours).

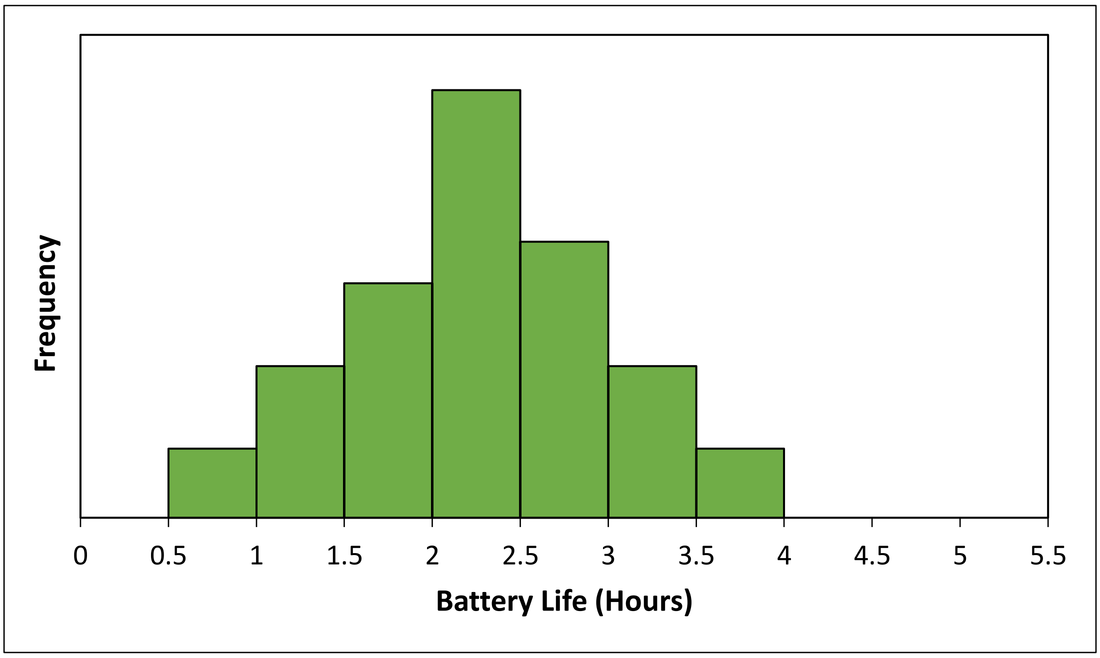


Figure C2. Normal (2.75 hours).

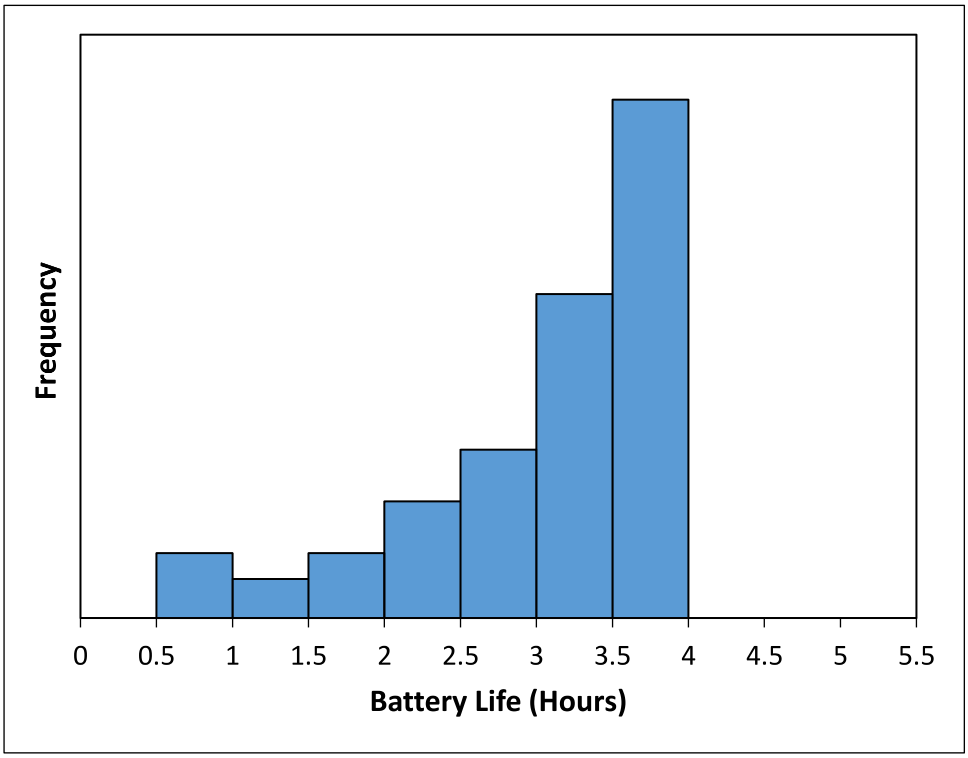


Figure C3. Negative (3.75 hours).

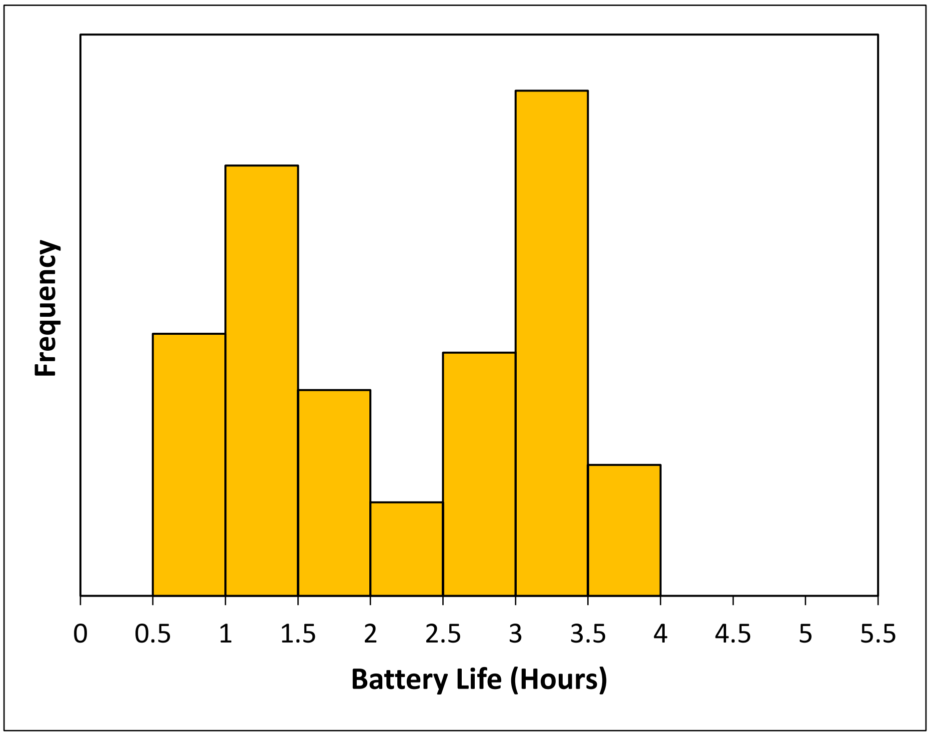


Figure C4. Bi-Modal (3.25 hours).

Appendix D. Experiment 3: Cumulative Distribution Graphs

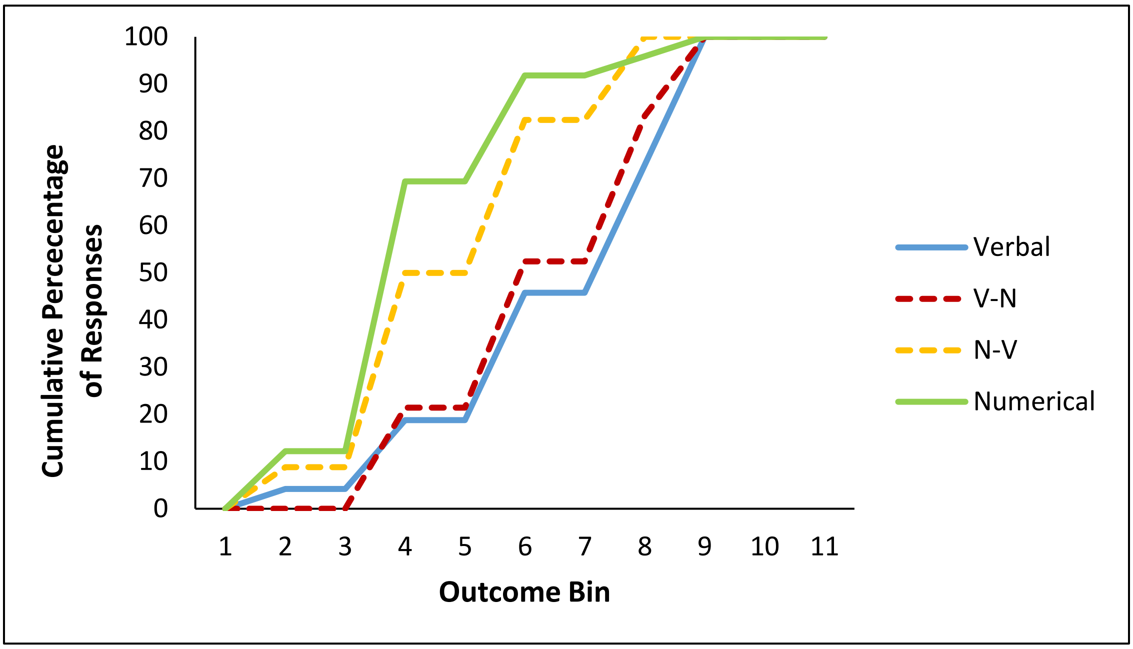


Figure D1. Cumulative distribution of responses by communication format – Positive distribution.

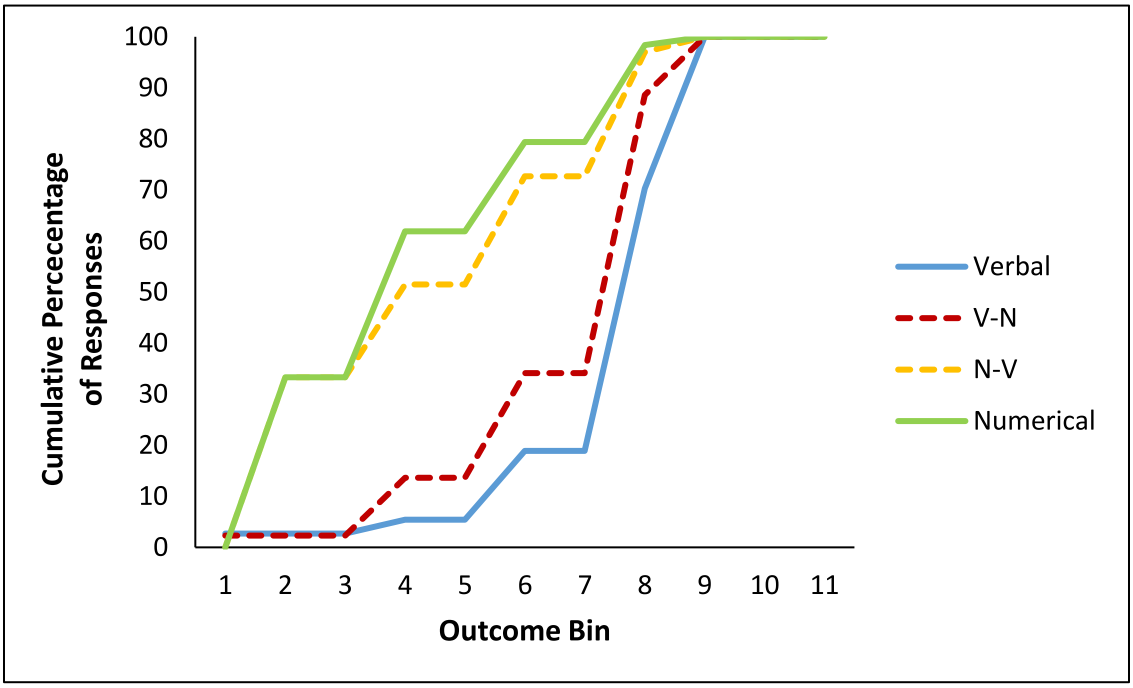


Figure D2. Cumulative distribution of responses by communication format – Normal distribution.

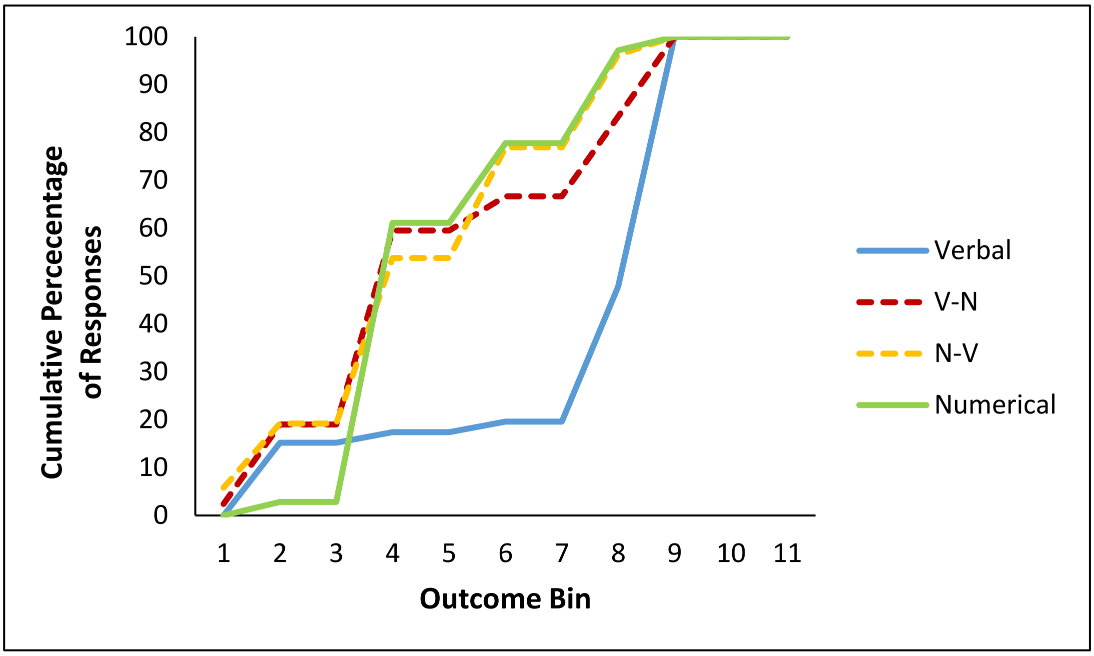


Figure D3. Cumulative distribution of responses by communication format – Negative distribution.

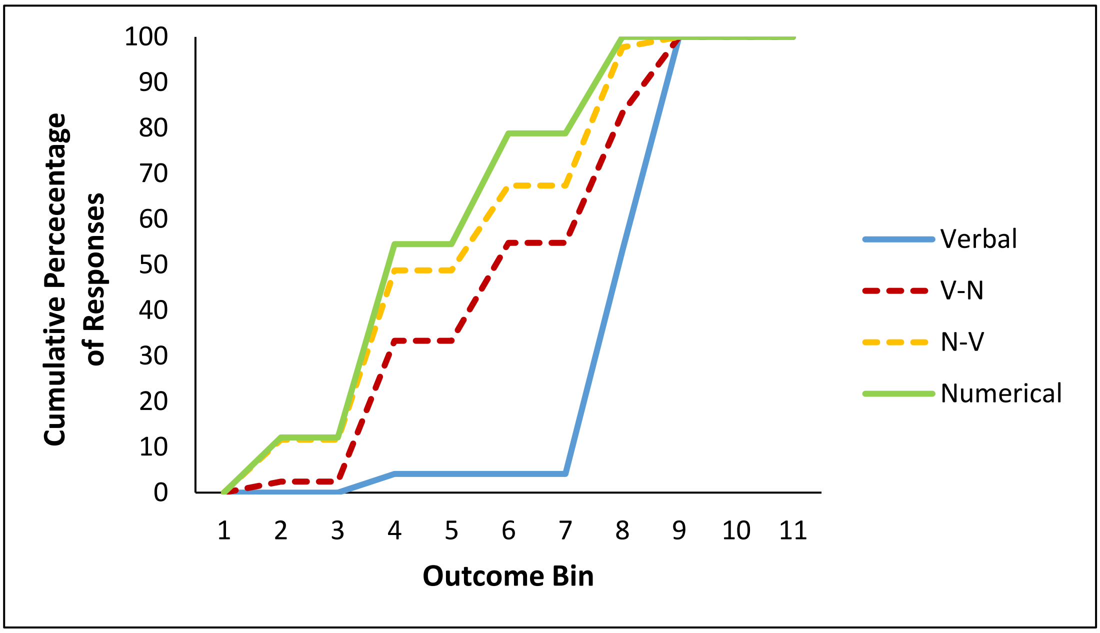


Figure D4. Cumulative distribution of responses by communication format – Bi-modal distribution.

Appendix E. Experiment 6: Natural Hazard Histograms

(Correct answer in brackets).

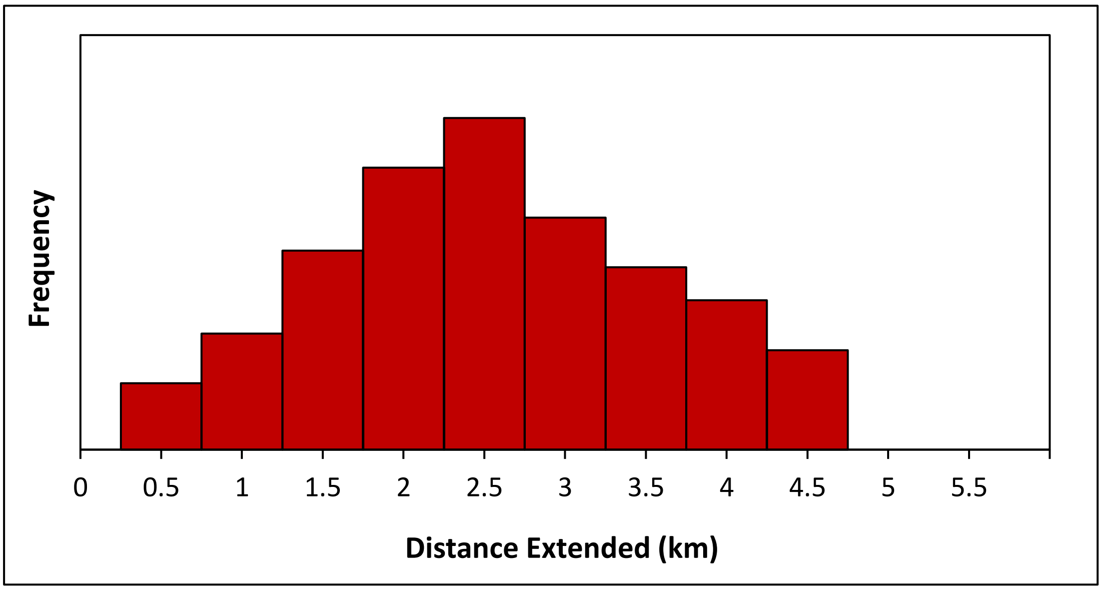


Figure E1. Volcano (3.5 km).

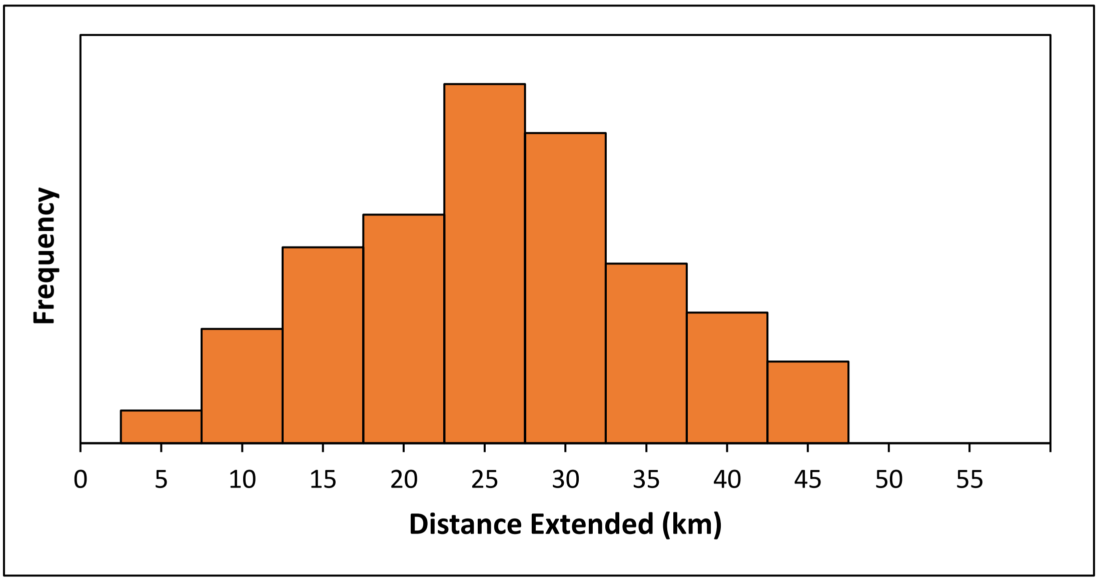


Figure E2. Earthquake (35 km).

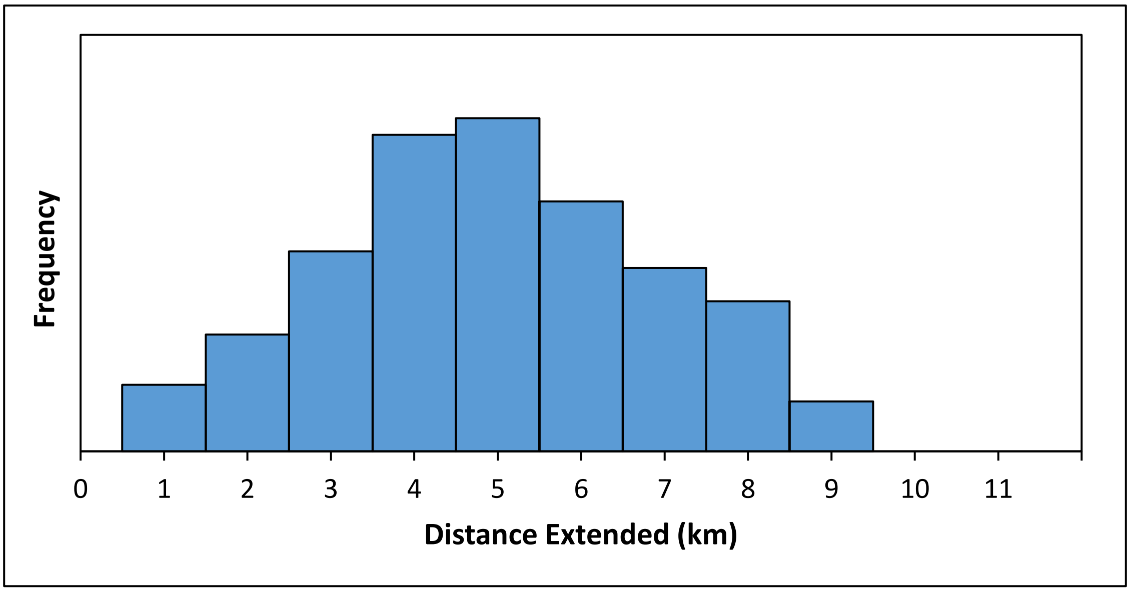


Figure E3. Flood (7 km).

Appendix F. Experiments 7 and 10: Volcano Vignette

Mount Ablon has a history of explosive eruptions forming 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 situation and recent scientific observations, it is **unlikely** that the lava flow will extend 3.5km.

[NEW SCREEN]

The capital city Ablon lies within 3km of the volcano. At least 48 hours is required to marshal resources, order and execute a mass evacuation of the capital city Ablon. A mass evacuation is very expensive and extremely disruptive to residents.

[NEW SCREEN]

Mount Ablon **did** in fact erupt and the lava flow extended beyond 3.5km.

[NEW SCREEN]

TWO YEARS LATER...

In 2019, after months of showing signs of unrest, Mount Ablon is currently erupting and lava flows are expected. The same volcanologists from Ablon Geological Centre are communicating information about the volcano. A volcanologist has suggested that, given the volcano's situation and recent scientific observations, it is **unlikely** that the lava flow will extend 3.5km. (Verbal format, low probability domain, emphasis in original).

Appendix G. Chapter 3: Analyses of Pre-Outcome Ratings

G.1. Experiment 7.

(All 4 × 2 ANOVAs).

Credibility.

There was no significant effect of communication format, $F(3, 282) = 0.23$, $p = .88$ or numeracy $F(1, 282) = 1.20$, $p = .28$ on pre-outcome credibility ratings, nor was there a significant interaction between the two factors, $F(3, 282) = 0.68$, $p = .56$.

Decisions to evacuate.

There was a significant effect of format, $F(3, 282) = 3.23$, $p = .03$, with a post-hoc Gabriel test indicating those in the numerical condition were less certain about evacuating than those in the verbal condition ($p = .02$).[45] No other differences between formats were significant (all *p*s > .27). There was a marginally significant effect of numeracy $F(1, 282) = 3.85$, $p = .05$, with those higher in numeracy more certain about evacuating (M= 2.79, SE= 0.09) than the low numeracy group (M= 2.51, SE = 0.11). There was no significant interaction between communication format and numeracy, $F(3, 282) = 1.27$, $p = .29$.

G.2. Experiment 8.

(All 4 × 2 × 2 ANOVAs).

Credibility.

There was no significant effect of communication format, $F(2, 231) = 2.41$, $p = .09$, numeracy, $F(1, 231) = 0.08$, $p = .79$, or scenario, $F(1, 231) = 0.93$, $p = .34$, on pre-outcome credibility ratings. There were no two- or three- way interactions (all *p*s > .13).

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

Insurance decision

There was no significant effect of communication format, *F(*2, 221) = 0.76, p = .47, nor a significant effect of scenario, *F(*1, 221) = 0.47, p = .50 or numeracy, *F(*1, 221) = 0.06, p = .80 on the initial insurance decision. There were no two- or three- way interactions (all *p*s > .26).

G.3. Experiment 9.

(All 4 × 2 × 2 ANOVAs).

Credibility.

There was no significant effect of communication format *F(*2, 234) = .90, p = .41, scenario *F(*1, 234) = 1.07, p = .30, or numeracy, *F(*1, 234) = 0.25, p = .62 on pre-outcome credibility ratings. There were no significant two- or three-way interactions (all *p*s > .12).

Insurance decision.

There was no significant effect of communication format, *F(*2, 225) = 0.94, p = .39, scenario, *F(*1, 225) = 0.23, p = .63, or numeracy, *F(*1, 225) = 0.23, p = .63, on the initial insurance decision. All two-way interactions were non-significant (all *p*s > .27), though the three-way interaction was significant, $F(2, 225) = 4.00$, p = .02, η_p^2= .03. Simple effects analysis revealed there was an effect of numeracy, but only in the less predictable (fire) scenario when a large range format was used, $F(1, 225) = 6.92$, p = .009, η_p^2= .03. Here, the high numeracy group were willing to pay a significantly higher price for insurance (M = 976.00, SE = 39.02) compared to the low numeracy group (M = 795.83, SE = 56.32).

G.4. Experiment 10.

(All 4 × 2 ANOVAs).

Credibility.

There was no effect of communication format, $F(3, 273) = 1.14$, p = .34, or numeracy, $F(1, 273) = 0.59$, p = .44 observed on pre-outcome credibility ratings. The interaction was not significant, $F(3, 273) = 1.10$, p = .35.

Decisions to evacuate.

There was no effect of communication format on initial decision to evacuate, $F(3, 273) = 1.44$, $p = .23$. There was an effect of numeracy, $F(1, 273) = 6.72$, $p = .01$, $\eta_p^2 = 0.02$, with the low numeracy group more certain about evacuating (M= 2.38, SE = 0.11) than the high numeracy group (M= 1.96, SE= 0.12). The interaction was not significant, $F(3, 273) = 1.62$, $p = .19$.

G.5. Experiment 11.

(All 4 × 2 × 2 ANOVAs).

Credibility.

No effect of communication format was observed on pre-outcome credibility ratings, $F(3, 229) = 0.50$, $p = .68$. There was a trend towards a format × scenario interaction, though this did not reach conventional levels of significance, $F(3, 229) = 2.41$, $p = .07$, $\eta_p^2 = 0.03$. Follow up simple effect tests indicated that a communicator who used a narrow range expression to describe the extent of floodwaters was perceived as more credible (M= 4.28, SE= 0.12) than when one used a narrow range expression to describe the extent of forest fires (M= 3.89, SE = 0.14), $F(1, 229) = 4.79$, $p = .03$, $\eta_p^2 = 0.02$. These follow up tests also demonstrated a trend for a communicator who used a numerical point expression to describe the extent of floodwaters was perceived as more credible (M= 4.18, SE = 0.13) than when one used a point expression to describe the extent of forest fires (M= 3.86, SE = 0.15), $F(1, 229) = 2.84$, $p = .09$, $\eta_p^2 = 0.01$. No other effects or interactions were observed (all *p*s > .20).

Insurance decision.

There was no effect of communication format observed on the initial insurance decision, $F(3, 208) = 0.15$, $p = .93$, nor an effect of scenario, $F(1, 208) = 0.02$, $p = .88$, or numeracy, $F(1, 208) = 0.98$, $p = .32$. No two- or three- way interactions were significant (all *p*s > .22).

Appendix H. Experiment 8: Analyses of Tolerable Risk Scale

Background.

The tolerable risk scale (Haynes et al., 2008b) was developed following fieldwork with scientists, government officials and the public during a period of volcanic crisis in Montserrat. The researchers observed a disparity between the level of risk tolerated by the authorities and scientists compared to the public. As the former had more responsibility for the islanders, it was felt they were less tolerant to risk, whereas the public were more tolerant, often feeling the benefits outweighed the risks of ignoring expert advice. The fieldwork identified nine risky activities which, contrary to expert advice, the public engaged in and these form the items in the scale. The scale was included as an alternative behavioural measure to the 'decision to evacuate' measure featured in Experiment 7.

Procedure.

Before being presented the items, participants were informed that an area relevant to the scenario had been evacuated and anyone who remained in this area would be killed:

> "Imagine that an area in Redmill has been evacuated and if the fire reaches Redmill. It is forecast anyone in the area **will be killed.** Based on the geologist's prediction that there is a **10-30% likelihood** that the forest fire will extend 80km from its origin, how acceptable do you rate the activities below, thinking about the danger you think is involved and the likely benefits that people get?" (Forest fire, numerical range condition, emphasis in original).

Following this, participants were asked how acceptable activities featured in the tolerable risk scale (see Table H1) were, rating answers on a scale from 1 (*Not at all acceptable*) to 5 (*Completely acceptable*). A total was calculated (out of a possible 45) for each participant and then averaged to produce a tolerable risk score.

Table H1. Adapted items from tolerable risk scale

Item
Going into the zone to save a life
Going into the zone once to pick up property
Standing on the edge of the exclusion zone and looking in
Going into the exclusion zone every day for 3 h to look after your house
Taking a walk into the zone
Going into the exclusion zone every day to look after livestock and crops
Going into the exclusion zone all day every day to live in your house (more than 3h)
Staying overnight from time to time in your house in the exclusion zone
Staying in your house full time

Results.

Pre-outcome, there was no significant effect of communication format, $F(2, 231) = 1.62$, $p = .20$, numeracy, $F(1, 231) = 1.11$, $p = .29$, or scenario, $F(1, 231) = 0.01$, $p = .91$ on perceptions of credibility, nor were there any significant interactions (all *p*s > .19).

Post-outcome, there was a reduction in perceived tolerable risk but again there was no significant difference between formats, $F(2, 231) = 0.39$, $p = .68$. There was also no significant effect of scenario, $F(1, 231) = 0.43$, $p = .51$, nor numeracy, $F(1, 231) = 0.23$, $p = .63$, nor any significant two- or three-way interactions (all *p*s > .29).

Appendix I. Experiments 8, 9, 11, 12, 13: Forest Fire/Flood Vignettes

Forest fire.

A site in South Redmill has a history of forest fires due to a lack of rainfall in the area. A forest fire is currently blazing. Geologists from Redmill Geological Centre are communicating information about the forest fire. A geologist has suggested that, given the recent weather, it is **unlikely** that the forest fire will extend 80km from its origin. (Verbal condition, low probability domain, emphasis in original).

Flood.

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. Hydrologists from Wayston Geological Centre are communicating information about the flood. A hydrologist has suggested that given the river's situation and recent weather, there is a **10–30% likelihood** that the floodwater will extend 7km. (Wide range condition, low probability domain, emphasis in original).

Appendix J. Chapter 4: Translations of VPEs from Pilot Studies

Table J1. High Probability Expressions

Directionality	VPE	*n*	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 J2. Low Probability Expressions

Directionality	VPE	*n*	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

Appendix K. Chapter 4: Analyses of Pre-Outcome Ratings

K.1. Experiment 12.

Credibility.

A 3 (communication format – congruent, incongruent, numerical) × 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, η_p^2= .03 on pre-outcome credibility ratings (see Figure K.1). There was a significant interaction between the two factors, *F(*2, 430) = 11.72, *p* < .001, η_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, η_p^2= .08, but not for the low probability expressions, *F(*2, 430) = 0.29, *p* = .75, η_p^2= .001.

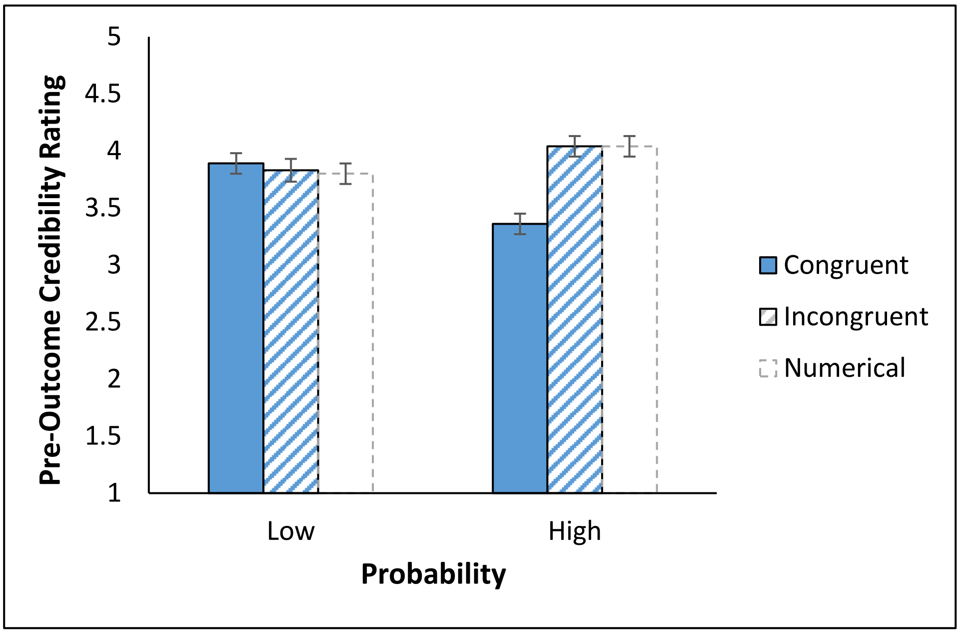


Figure K.1. Pre-outcome credibility ratings by probability – Experiment 12. (Error bars represent ±1 SE).

K.2. Experiment 13.

Credibility.

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, η_p^2= .02 on pre-outcome credibility ratings (see Figure K.2). There was a significant interaction between communication format and probability, *F(*2, 411) = 7.40, *p* = .001, η_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, η_p^2= .04, but not for the low probability expressions, *F(*2, 411) = 1.91, *p* = .15.

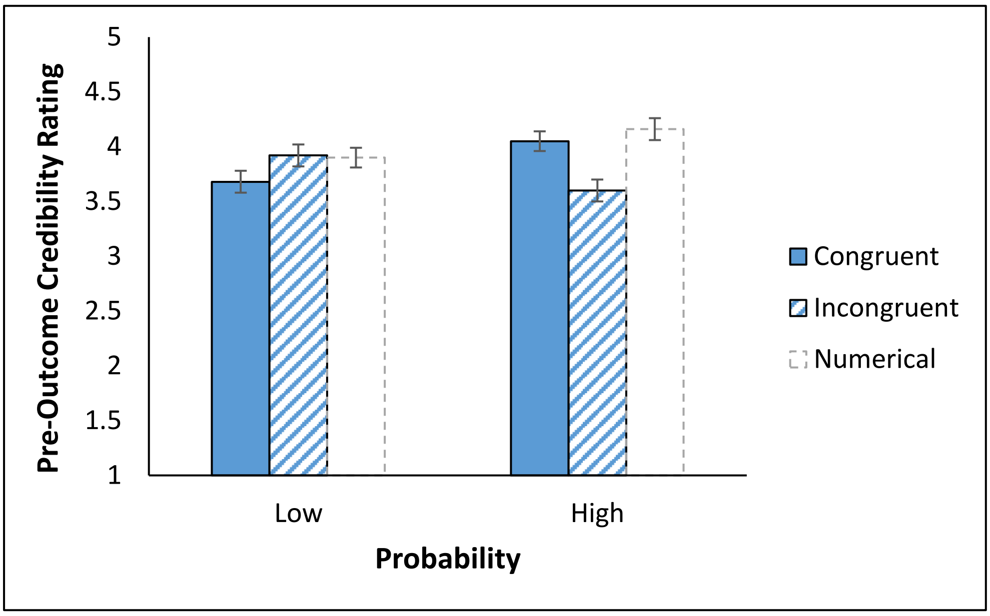


Figure K.2. Pre-outcome credibility ratings by probability – Experiment 13. (Error bars represent ±1 SE).