Communicating Weather Uncertainty:

An Individual Differences Approach

Margaret A. Grounds

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Susan Joslyn, Chair

John Miyamoto

Chantel Prat

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Communicating Weather Uncertainty:

An Individual Differences Approach

Margaret A. Grounds

Chair of the Supervisory Committee:

Susan Joslyn, PhD

Department of Psychology

Previous research suggests that people make better decisions in weather-related decision tasks when they are given probabilistic uncertainty estimates about the outcomes (Joslyn & Leclerc, 2012; Joslyn & Leclerc, 2013). However, it is unclear whether all users can take advantage of probabilistic forecasts to the same extent. The research reported here assessed various cognitive and demographic factors to explore the extent to which these factors were associated with users' ability to take advantage of probabilistic forecasts. In three studies, participants made decisions about whether to spend limited resources to salt roads to prevent icy conditions. Several expression formats were tested, including numerical uncertainty and explicit advice. Results suggested that increased numeracy was associated with better weather-related decisions when forecasts included numerical uncertainty estimates in all three studies, although no user groups were substantially impaired when given numerical uncertainty. There were also demographic factors such as age and education that explained decision quality, however, their effects were inconsistent or even contradictory between experiments. However, when all other predictors were removed, individuals with higher numeracy made better decisions regardless of the

uncertainty communication method, suggesting that the advantage of numeracy may extend beyond understanding the forecast to larger decision strategy issues. This research adds to a growing body of evidence that numerical uncertainty estimates are an effective way to communicate weather danger. It demonstrates that the advantages of numerical expressions are not seriously limited by individual differences and allow users to better differentiate situations that do and do not require precautionary action and increase trust. Moreover, these results might generalize to other domains in which lay people must make important decisions when faced with uncertainty.

Table of Contents

Foreword: A Note on Risk and Uncertainty

Risk and uncertainty are often used synonymously in everyday language. However, there are several ways to distinguish them, and so I will give an adequate definition of both terms. An early distinction proposed by the economist Frank Knight (1921), was that risk describes decision situations in which the outcome probabilities are objectively known, such as a flip of a fair coin. Uncertainty describes situations in which the outcome probabilities are unknown (i.e., the decision maker must estimate or infer the probabilities), like a game of poker with cards removed from the deck. However, this distinction is not quite complete as it does not cover all the ideas associated with risk, like the fear of loss, as many real-world scenarios do. A different distinction of risk that I think is more fitting to how people make real world decisions comes from the risk communication literature. Here risk can be defined as a function of likelihood and value of a future event (Eiser et al, 2012).

Thus, uncertainty is the lack of complete certainty, that is, the existence of more than one possibility and the true outcome is not known. This uncertainty can be measured however, by a set of probabilities assigned to a set of possible outcomes, e.g., “There is an 80% chance the San Jose Sharks will not win the Stanley Cup this year.” Risk is a state of uncertainty where some of the possibilities involve a loss. It can be measured with quantified probabilities and quantified losses, e.g. “There is an 80% chance the Sharks will not win the Stanley Cup this year, and I will lose my $100 bet.” Thus, one may have an uncertain scenario without risk but risk always comes from uncertainty. I can be uncertain about who will win the Stanley Cup, but unless I have some personal stake in the matter by placing a bet, I have no risk. This is the sense in which the terms risk and uncertainty will be used in the following dissertation.

Chapter 1: General Background

Communicating Uncertainty

Effective uncertainty communication is an important step that must occur for decision makers to make good decisions. People make many important decisions under uncertainty, such as investing money in the stock market, choosing to have a surgery done, or evaluating the potential danger of an upcoming storm. There is often good information available about the uncertainty they face, although exactly how to communicate that uncertainty has been a subject of debate in the scientific community for several decades now. Researchers from different domains understand this importance and several forms of uncertainty expression have been empirically tested to determine whether they lead to accurate understanding and/or better decision making. There are a lot of recommendations, quite a bit of research and some disagreement about how it should be done. I will begin by providing a short review of this research.

Uncertainty information can be communicated as a numerical or verbal statement to the end-user. Verbal expressions of likelihood have been recommended because they are regarded as easier to use (Anker, Senathirajah, Kukafka, & Starren, 2006). For instance, in one study, a seven-point verbally labeled risk scale was rated easier to use than a 100-point numerical scale (Diefenbach, Weinstein & O'Reilly, 1993). Some people, 30% in one sample, prefer verbal expressions because they are seen to be easier to use, as well as more personal (Wallsten, Budescu, Zwick, & Kemp, 1993). Unfortunately, verbal expressions alone are often seen as vague and result in different interpretations between users (Budescu, Por, & Broomell, 2012). The word "likely" for example, could be understood to mean any likelihood from about 50% to 100% and "possible" could mean any degree of likelihood from 0% to 100%, depending on the user (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). Thus, verbal expressions can in fact lead to over or under estimations of risk.

Numerical expressions of risk (e.g. % chance) have several appealing qualities that support their use in uncertainty communication. They are precise and as such lead to more accurate perceptions of risk than the use of verbal phrases (Lipkus, 2007). They can be easily converted from one format to another (i.e. percentages such as 10% can be converted into frequencies such as 1 out of 10), given a large enough sample they can be verified for accuracy, and they can be computed using algorithms or models, often based on clinical and/or other observed data (Lipkus, 2007). Furthermore, numerical uncertainty communication may better address the needs of different user groups with different risk tolerances (Joslyn & Savelli, 2010; Murphy & Winkler, 1979). Numeric probabilities in particular, are precise and clear, and always use the same denominator, making them easy to compare (Lipkus, 2007).

However, information format could influence the extent to which certain user groups understand and utilize uncertainty information. Even at the most basic level, individuals who lack elementary reading comprehension would not be able to utilize written communication. In a study investigating the risks associated with eating contaminated Great Lakes fish, it was concluded that no single communication strategy (simple or sophisticated explanation, verbal or diagram, commanding or suggesting tone) had the same effect on three different target audiences: Hispanic anglers, female anglers of childbearing age, and anglers who ate the most contaminated fish species (Connelly & Knuth, 1998). Thus, it is possible that numerical uncertainty information may not be suitable for all users. Individuals who lack the numerical knowledge to understand probability information may not utilize it fully or in the worst case may seriously misunderstand the risks they are facing. These users may instead benefit from communication that reduces the amount of cognitive effort required to comprehend the information. In fact, studies have shown that individuals with less numerical knowledge have improved comprehension and have better decision quality when less information, both verbal and numerical, is shown (Peters, Dieckmann, Dixon, Hibbard, & Mertz, 2007). Therefore, it may be important to tailor uncertainty communication to different user groups.

Indeed, there is evidence to suggest that numeric probabilities are misunderstood and misused by some people. For instance, many people tend to underestimate low probabilities. Risk information with probabilities close to zero are sometimes interpreted as representing no risk and thus might be ignored (Kahneman, 2011; Lipkus, 2007). Furthermore, mental biases or heuristics, mental rules of thumb, can result in reasoning errors when deciding among risky options given this information. One such bias is known as the effect of framing. Risky options can be framed as either a gain by highlighting the positive outcome, or framed as a loss by highlighting the negative outcome. Studies have shown that people prefer a sure option when it is framed as a gain (i.e. they are risk-averse) but prefer a risk when it is framed as a loss (i.e. they are risk-seeking) even when they are making an error by choosing the suboptimal option from an economic standpoint (Kahneman, 2011). Errors can also be made when people generate probabilities themselves. Reliance on heuristics like the Representativeness heuristic may cause people to ignore base rates of events when judging their likelihood and instead focus on how representativeness of the event (Kahneman & Tversky, 1973). Other biases such as anchoring could also result in self-generated probability errors. People might rely too heavily on some arbitrary initial value and fail to sufficiently adjust their estimate (Tversky & Kahneman, 1974). Therefore, probability information may not be an ideal communication method for all users.

Some argue that some misunderstandings of probability information arise when no reference class is specified (Gigerenzer & Edwards, 2003). For example, the phrase “there is a 40% chance of sexual problems with this medication” does not specify what reference class to which 40% refers. People could interpret the phrase to mean 40% of sexual encounters or 40% of users. This has been demonstrated with probability of precipitation forecasts: 30% chance of rain was interpreted to mean 30% of the time for some people, 30% of the area would receive rain, or 30% of similar days saw rain (Gigerenzer & Edwards, 2003). Similarly, omitting a reference class makes it more difficult to interpret probabilities for single events (Brase, 2002; Gigerenzer & Hoffrage, 1995).

An alternative to probability expression that may address these issues is frequency, for example, “1 out of 10 smokers will develop lung cancer.” These frequencies should theoretically be easier to understand since they explicitly state the reference class (Gigerenzer & Edwards, 2003). In fact, both patients and physicians show better understanding of risk information (in terms of gross comparison and risk assessment tasks) if risks are presented in terms of frequencies (e.g., 5 out of 100 people experience a side effect) rather than in percentages (5%) (Hoffrage & Gigerenzer, 1998). However, much of this research had been conducted using tasks in which risks need to be compared. Conversely, research that involved tasks in which a decision was made based on a single uncertainty expression found equal advantage for frequencies and percentages (Grounds, Joslyn, & Otsuka, in prep; Joslyn & Leclerc, 2012; Savelli & Joslyn, 2013).

Despite the possible drawbacks of numerical expressions of uncertainty, they are clearly more precise than verbal expressions, can address a range of user needs or risk tolerances and are less susceptible to misunderstanding. Moreover, they have been shown to improve decision quality in several previous studies. On the other hand, numerical expressions may be too complex or unsuitable for some users.

Communicating Weather Uncertainty

One interesting domain to study how uncertainty information is communicated and understood is weather. This is a domain in which uncertainty information is available and could make a critical difference in general user decision-making. Where does weather uncertainty come from? Per the National Oceanic and Atmospheric Association (2016), weather forecast uncertainty "can arise due to the complex nature of our atmosphere, which is a chaotic 'fluid' that is sensitive to initial conditions. A slightly inaccurate depiction of the current state of the atmosphere will often result in forecast uncertainty with time in model forecast data. The models themselves are only a simulation of the atmosphere, and their accuracy will be limited by how precisely they can represent complex atmospheric processes. Some processes can be difficult to model perfectly, which can lead to error over time."

Recently, there has been a focused effort by atmospheric scientists to develop predictive models that have improved dramatically over the past few decades to produce better, more reliable probabilistic forecasts (Silver, 2012). This is largely due to improved methods of modeling data (Gneiting & Raftery, 2005). Atmospheric scientists can now use ensemble forecasts that are generated by multiple runs of weather prediction models which differ slightly in initial conditions or representations of the atmosphere (Gneiting & Raftery, 2005). These ensemble forecasts produce reliable probability distributions of future weather events.

However, even though reliable numerical uncertainty estimates are often available, weather communication is frequently achieved using simplified verbal expressions that are categorical in nature, such as, “A high wind warning is in effect today” or “Shelter in place.” For example, the National Weather Service uses categories such as; Outlook, Advisory, Watch, and Warning to communicate possible danger for several weather events. Some field research has suggested that warnings from public officials are an important factor in the decision to take precautionary action (Baker, 1995) and these categorical expressions are advantageous as they provide a summary of the action that should be taken. However, it is not clear how well they are understood by the public or even by emergency managers, the very people who are tasked with preparing a community or organization for an upcoming weather event (Sheperd, 2014).

Moreover, the compliance rate for categorical warnings is unacceptably low. In a study analyzing the tornado seasons of 2009-2011, Nagele and Trainor (2012) found that people within the warning polygon (a visual indication of tornado likelihood) were no more likely to seek shelter than those living outside the polygon but in the same county (approximately 40%). Evidence also suggests that people do not take sufficient precautionary action when flood warnings are issued (Gruntfest, Downing, & White, 1978; Parker, Priest & Tapsell, 2009). Similarly, interviews of residents under mandatory evacuation for Andrew and Hugo, both Category 4 hurricanes, revealed that only 42% evacuated their homes (Riad, Norris, & Ruback, 1999). For Hurricane Floyd, the evacuation rate was only 64% (Dow & Cutter, 2000). For Hurricane Ike, only 60% complied with mandatory evacuation orders and the other 40% of people chose to stay despite being warned of “certain death” (McKinley & Urbina, 2008). A recent report showed 70% of those surveyed in a mandatory evacuation zone for Hurricane Sandy chose not to vacate their homes (Gibbs & Holloway, 2013). These and other equally dramatic examples lead to a growing possibility that categorical communication alone is not sufficient to prompt directed precautionary action for extreme weather events.

One possible explanation for these low compliance rates of precautionary action may be the fact that weather-related decision-making often involves a choice between losses. For instance, one must choose whether to evacuate, involving at minimum travel costs and inconvenience, or stay and risk personal injury or loss of life. An optimal decision-maker would evaluate the expected value of each alternative by multiplying the cost of the possible outcome by its likelihood of occurrence, and select the alternative with the better expected value. Moreover, because of the potential for personal harm (a large loss), it is often beneficial from an economic standpoint to take precautionary action when the probability of observing the event is relatively low. People may be particularly reluctant to do so in such cases, perhaps because they are doubtful that the adverse event will occur.

Nonetheless, experiments have demonstrated that individuals given uncertainty forecasts made better precautionary decisions and had more forecast trust than those who were given a deterministic forecast, (Joslyn & Grounds; 2015; Joslyn & LeClerc, 2012; Joslyn & LeClerc, 2013; Roulston, Bolton, Kleit, & Sears-Collins, 2006). This was true even though probabilistic forecasts increased processing load, adding information about alternative outcomes and their associated probabilities. Moreover, the advantage for probabilistic forecasts increased when forecast error increased (Joslyn & LeClerc, 2012). This is a critical finding, as some of the most important weather forecasts concerning severe weather events must be made far enough in advance to give people time to take precautionary action. With such long lead times, forecast error is high, so including numeric uncertainty estimates can lead to better decisions in these scenarios.

This previous research was conducted using a paradigm I will call the Road Salt Task. Over a series of trials in the Road Salt Task, participants must decide whether they ought to spend a limited monetary resource to apply salt to roads to prevent icing. Or they could take a gamble, saving on costs but subjecting themselves to the potential of a severe penalty if freezing temperatures are observed but no salt has been applied. It is assumed in this task that the value of monetary losses is linear. In these previous studies (Joslyn & Grounds, 2015; Joslyn & Leclerc, 2012; Joslyn & LeClerc, 2013), all participants took precautionary action more as the probability of freezing increased, even when they weren't given probability information, suggesting that they realized that the single value forecast was correlated with the probability of freezing. However, the most common error made in this task was a risk-seeking error, as the cost-loss structure of the task dictated it was optimal to take precautionary action at low probabilities. This low optimal decision threshold was selected because it is representative of real weather events where extreme danger and cost necessitate precautionary action at low probabilities. Importantly, participants given probabilistic forecasts made better decisions than those given a single value (what subsequently will be referred to as deterministic) forecasts, according to several measures of decision quality. Surprisingly, when participants were given explicit decision advice, analogous to an evacuation warning in the real world, it did not help. Their decisions were no different than those of individuals who received just the single value forecast. When the advice was paired with probability information decisions quality improved and risk-seeking errors were reduced.

The advantage for numeric uncertainty forecasts over deterministic forecasts appears to be a robust effect. It has been shown in experimental paradigms in which participants had to explicitly request it, and did so, suggesting that users recognize its usefulness (Nadav-Greenberg & Joslyn, 2009). It has also been shown for scenarios in which the decision complexity was increased by increasing the number of alternatives participants were required to consider

(Grounds & Joslyn, 2015). The latter example was particularly interesting since increasing the number of decision alternatives also increased the processing load with additional outcomes and associated probabilities. Perhaps most strikingly, numerical uncertainty information even makes people more likely to comply with explicit advice (Leclerc & Joslyn, 2015), which suggests that including numeric uncertainty information in evacuation orders may increase compliance rates.

Factors that may influence the ability to use numeric uncertainty estimates

However, numeric uncertainty may not be the best form of communication for all users. Most of the work described above was done using university college students as subjects. Although the decisions made by this population may be representative of the general population, there may be subgroups that would be confused by probabilities and might make better use of a simple and explicit recommendation (e.g. evacuate, shelter in place). Moreover, some people may be confused by probabilities and do worse when they are given numeric uncertainty information. They may not be able to understand, due to specific limitations in mathematical ability or analytic reasoning. Others may simply not have sufficient processing capacity to incorporate additional information especially if it is complex or intellectually demanding. The goal of this dissertation is to test whether specific skill or general processing limitations, are related to users' ability to benefit from explicit numeric uncertainty information. I will begin by providing a theoretical overview of the constructs of interest and how they relate to risk comprehension and decision making. Next I will provide my research questions and overview of experiments. Then I will describe and report each experiment in turn. Finally, I will discuss the major findings of the experiments.

Numeracy. A basic understanding of numerical concepts may be an important part of informed decision making when uncertainty is expressed numerically. Thus, it is critical to understand how numeracy, the mathematical equivalent to literacy, impacts the ability to make use of numerical uncertainty communication. Numeracy can be assessed through objective (Lipkus, Samsa, & Rimer, 2001) and subjective measures (Woloshin, Schwartz, & Welch, 2005) and is not the same as education. Research seems to suggest that higher numeracy leads to better risk comprehension (see Reyna, Nelson, Han & Dieckmann, 2009, for a recent review). This is true in the financial domain where more numerate individuals are more likely to be savvy in choosing mutual funds, including selecting those with lower fees (Hastings & Mitchell, 2011) as well as in medicine where participants higher in numeracy were better able to use risk communication to adjust their personal risk estimates (Schwartz, Woloshin, Black, and Welch, 1997).

However, research shows that the less numerate individuals take more risks and are less sensitive to varying expected value levels than more numerate individuals, especially when it is disadvantageous to take a risk and the choice involves a potential loss (Cokley & Kelley, 2009; Jasper, Bhattacharya, Levin, Jones, & Bossard, 2013). Research also shows that less numerate individuals are more susceptible to framing of percentages (e.g. 25% fat beef is perceived as worse than 75% lean beef), which might cause them to make more decision errors (Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006). Moreover, those with lower numeracy had more trust in the verbal expressions of risk than the numerical, and those with the higher numeracy ability trusted numerical expressions of risk more than the verbal (Gurmankin, Baron, & Armstrong, 2004). Thus, higher numeracy may help people make better decisions when given numeric uncertainty, but low numeracy individuals may better benefit from explicit verbal recommendations.

Deliberate thinking. Limitations in the ability to utilize numerical uncertainty information could also be more general than numeracy, such as the failure to think deliberately about the information. According to the two systems theory, there are two modes of thought; System 1 is rapid, automatic, associative, emotional, and roughly synonymous with intuition, and System 2 is slow, deliberate, rule-based, and effortful (Kahneman, 2003). Appropriate use of explicit numeric uncertainty information requires the decision maker to consider two possible outcomes of varying likelihood throughout the decision process and may well depend on a deliberate, effortful processes. A person who also employs System 2, which is deliberate and systematic, may be more likely to consider all of the relevant factors, including precise numerical uncertainty information, especially when likelihood is low but it is appropriate to act. In other words, they have the mental restraint needed to avoid taking what initially seems like the more attractive option suggested by System 1. This has been demonstrated in the literature. People who engage in System 2 processing are less susceptible to framing effects (McElroy & Seta, 2003). This suggests in the domain of losses, individuals who use more System 2 processing would make fewer risk-seeking errors and thus better decisions. Moreover, individuals who rely on System 1 make worse use of risk information, by choosing a risky gamble with a worse expected value over a sure loss (Frederick, 2005). In this study, Frederick give participants a measure to assess the ability to engage in System 2 processing and separated participants into a group that relied more on System 1 and a group that successfully engaged System 2. Then he gave these participants gambles expressed as probabilities or corresponding sure options to choose from with varying expected values. For loss gambles where the riskier option had a worse expected value, participants who relied more on System 1 chose the riskier option more often.

Thus, failing to engage in System 2 may limit the ability to utilize numerical uncertainty estimates to make good decisions.

Need for cognition. A related factor that may influence the ability to make use of numerical uncertainty estimates is the need for cognition (NFC), which is the extent to which individuals report that they are inclined towards effortful cognitive activities (Petty, Cacioppo, & Kao, 1984). A person who is capable of deliberative thinking but who does not enjoy mental effort may just ignore the uncertainty information provided, which would result in worse decisions. Notice, this may not be the same as deliberate thinking, since it could involve a conscious decision to reject complex thought. The exact relation between deliberate thinking and need for cognition is unclear. Although some factor analyses for measures of these constructs have demonstrated they load onto different latent factors, suggesting that they are distinct (Welsh, Burns, & Delfabbro, 2013), others have demonstrated they load onto the same latent factor, suggesting they both assess deliberate thinking (Svedholm-Häkkinen & Lindeman, 2016).

The literature on the relation between NFC specifically and risk communication is sparse. Some research shows that participants with higher NFC were less susceptible to framing effects (Smith & Levin, 1996), but not in others (LeBoeuf & Shafir, 2003). Other research found no relation between NFC and risk comprehension in the domain of losses (Frederick, 2005). Thus, exploration of the impact of this factor on uncertainty comprehension will make an important contribution to our understanding of how it affects decision making.

Information Processing

Since numeric uncertainty estimates increase processing load, or the mental effort required to make a decision, limitations in overall processing capacity may reduce the effectiveness of numeric uncertainty estimates for some users. Limitations in processing capacity may cause some users to be overwhelmed by the multiple components of the forecast or unable to integrate the information into their decision-making process. Indeed, there is some evidence to suggest that increasing processing load reduces decision quality. For instance, in naturalistic situations, people evaluate very few options when making decisions, often only one (Hepler & Feltz, 2012). Moreover, some evidence suggests that decision quality declines when there are more than as few as two relevant factors to consider (Galotti, 1999), and oftentimes experts only consider a single factor at a time (Klein, 1998). Thus, some people may benefit from a simpler format of communication with less information, similar to the current format preferred by most providers.

Working Memory. Our ability to hold and process information on a conscious level, known as working memory capacity (WMC), may be important when complex information must be applied to the decision at hand. In general, human WMC is limited, but can vary from person to person (Baddeley, 2003). Working memory capacity can be measured reliably (Conway et al., 2005), and many studies have investigated the relationship between this construct and higher order cognition (e.g. Kane et al., 2004) as well as decision making (Romo & Salinas, 2003). Furthermore, some evidence suggests that when people must maintain a large amount of information in working memory, they have a more impulsive, myopic decision-making style, focusing on the short-term outcomes instead of long term outcomes (Hinson, Jameson, & Whitney, 2003). Thus, individuals with lower WMC may find it more difficult to simultaneously maintain the information about the numeric uncertainty associated with each independent possible outcome. They might also lack the executive attention, or the ability to block distracting information, necessary to keep active maintenance of goal-relevant information. As a result, they may make worse decisions, especially when presented with uncertainty information. In contrast, those with higher WMC may be able to simultaneously evaluate outcomes of each option and make better decisions.

However, existing evidence on how WMC interacts with risk understanding is limited and inconclusive. Indeed, some research has shown that individuals with high WMC tend to make more biased decisions when given risk information (Corbin, McElroy, & Black, 2010). However, other evidence suggests that individuals with high WMC can make better use of risk information (Cokley & Kelley, 2009). Consequently, the benefits of probabilistic forecast may be limited of those with high WMC. Individuals with lower WMC may make better use of a forecast that includes explicit instructions instead of one that has uncertainty estimates.

Fluid Intelligence. A psychological construct that many believe to be closely related to working memory is fluid intelligence. Fluid intelligence can include such general abilities as problem-solving, abstract reasoning, pattern recognition, and information integration (Cattell, 1971). Fluid intelligence is in contrast with crystalized intelligence that relies on specific acquired knowledge, experience, or skills (Cattell, 1971). Some researchers believe that fluid intelligence is similar to, or in some cases the same construct as, working memory capacity. The evidence varies considerably, with correlations between measures of fluid intelligence and WMC ranging between .15 to .72 (Conway et al., 2002; Kane, Hambrick, & Conway, 2005), suggesting that there is shared variance between the two factors, but they may not be the same (Ackerman, Beier, & Boyle, 2005). Thus, fluid intelligence could uniquely explain the decisions people make beyond any possible contribution of working memory. There could be some individuals that can maintain numerical uncertainty information in mind (not a limit of WMC), but who are unable to incorporate that information into their decision-making process, thus making worse decisions. Overall the evidence on this topic is mixed. Some research has shown that more fluid intelligence enhances ability to reason with probabilities and is associated with greater modulation of risk-taking (Deakin, Aitken, Robbins, & Sahakian, 2004; Donati, Panno, Chiesi, & Primi, 2014). Other research has found no relation between fluid intelligence and decision quality for decisions under risk (Andersson, Tyran, Wengström, & Holm, 2013; Parker & Fischhoff, 2005; Syngelaki, Moore, Savage, Fairchild, & Van Goozen, 2009). Thus, exploration of the impact of fluid intelligence on uncertainty comprehension may help resolve this discrepancy using a realistic decision task in a domain familiar to all users.

Taken together, this creates a promising set of cognitive factors that could influence how people make decisions when they are given different communication formats, and to potentially identify subgroups of users that make better decisions when given a specific communication format. Higher numeracy could help people make better decisions when given numeric uncertainty, but low numeracy individuals may better benefit from explicit verbal recommendations. Individuals who fail to engage in System 2 may have limited ability to utilize numerical uncertainty estimates to make good decisions. Individuals with lower WMC may make better use of a forecast that includes explicit instructions instead of one that has uncertainty estimates. Greater fluid intelligence could help people make better use of numerical uncertainty information. If so, then these results will indicate how uncertainty communication should be tailored for specific sub-groups.

Research Questions

The goal of this dissertation is to examine how individual difference factors interact with uncertainty communication strategies. Although the factors outlined above may contribute to the decision-making process depending on communication format, the existing research is at times, contrary and inconclusive. Additionally, no research of which I am aware has sought to examine and compare these factors in a single project.

The research presented in this dissertation explores how best to communicate weather uncertainty to specific sub-groups of users. More specifically, it asks the question "Would all users benefit from numerical uncertainty estimates equally, or will certain types of users benefit from a simplified decision recommendation instead?" To answer these questions, I examined decision quality and decision errors for decisions made based of different uncertainty communication strategies.

Overview of Studies

The experiments to follow feature the Road Salt Task, a basic decision-making problem where participants make a series of decisions in which they are faced with a potential threat (as used by Joslyn & Grounds, 2015 and Joslyn & Leclerc, 2012) described above. Over a series of trials, participants must decide whether they ought to spend a limited monetary resource on precautionary action by applying salt to roads to prevent them from freezing. Alternatively, they could take a gamble, saving on precautionary costs but subjecting themselves to the potential of a severe penalty. The decision outcomes are framed entirely as losses, corresponding to parallel real-world weather situations. As the costs and potential losses in the scenario are quantified, a cost-lost ratio can be calculated. This value, the cost divided by the potential loss, serves as an important threshold that identifies economically optimal decisions. Participants are given a monthly budget used to pay for road salt treatment, or from which the penalty is deducted. As it was possible to have a negative budget if too many penalties were assessed, participants are told they can borrow money from the next month's budget to continue the task. Thus, participants should not avoid paying for salt treatment out of concern for a lack of money. To help participants make their decisions, they are given a temperature forecast, but the forecast format is systematically manipulated between participants. Some participants receive a single value forecast that suggests a single deterministic outcome. Other participants also receive numeric probability of a freezing temperature occurring. Additional participants also receive explicit advice recommending the economically optimal decision alternative based on the cost/loss ratio. The most common decision error participants make in this task is a risk-seeking error, that is, failing to take sufficient precautionary action. This is consistent with prospect theory (Kahneman & Tversky, 1979) as participants tended to make more risk-seeking than risk-averse errors with choices framed in terms of losses.

This project marks the first online use of this road salting task using the Mechanical Turk platform. All previous results were collected from the controlled environment of the laboratory setting. Indeed, relative to a traditional laboratory environment, many aspects of the online testing environment are not under the experimenter's control. Participants could be distracted by simultaneously completing other online tasks, attending to other aspects of their environment, or even step away from the task for some period of time. It is unsurprising that unsupervised subjects tend to be less attentive than subjects who are supervised by an experimenter (Oppenheimer, Meyvis, & Davidenko, 2009). Would this affect road salting responses? Would the same effects that have been seen in previous versions of the task (e.g. Joslyn & Leclerc, 2012) be observed?

Experiment 1 examined whether the same pattern of results would be observed in the online version of this task as in the in-person versions. Would participants take the task seriously and salt more often when freezing was more likely, at lower temperatures? Would there be an advantage for explicit numeric uncertainty estimates in the online version of this task, compared to a deterministic forecast and to explicit advice, as there had been when participants were tested in person (Joslyn & Grounds, 2015; Joslyn & Leclerc, 2012).

In addition, there were several changes to the procedure to adapt it to this new online environment. This version of the road salting task was shorter than previous versions tested. All previously published papers using the road salting task had participants complete 120 trials, although only the first 60 were analyzed (e.g. Joslyn & Leclerc, 2012: Joslyn & Grounds, 2015). The subsequent 60 trials provided additional budget increases that were designed to prevent participants from changing decision strategies due to a lack of monetary resources. The version used for this project omitted the last 60 trials to reduce time. Critically, in Experiment 1, participants were not given cash rewards commensurate with performance as had been done in previous studies. Instead, they received the same small reward at the end of the task regardless of how well they did.

In addition Experiment 1 began to explore the impact of individual differences on forecast communication format. A numeracy measure was included to test whether communication format interacted with numeracy in a general online population with participants who were a variety of ages and education levels, using an automated version of the road salting task. I wanted to determine whether the advantage for numeric uncertainty extended to a broader population with potentially lower numeracy skills.

In order to determine whether other cognitive factors are necessary to take advantage of numeric uncertainty estimates, **Experiment 2** measured deliberate thinking, need for cognition, working memory capacity, and fluid intelligence as well as numeracy in an in-person college student sample. This was to determine their relation to decision-making under uncertainty based on communication format. These measures could potentially identify subgroups of users that make better decisions when given a specific communication format. The monetary decision incentive was also reintroduced in Experiment 2.

Finally, **Experiment 3** took a small set of the most promising predictive measures from Experiment 2 back to a general online population. Additional demographic factors were also assessed, such as level of income and more specific educational level (as compared to the education assessment from Experiment 1). This was to determine their relation to decision-making under uncertainty based on communication format. These factors could also potentially identify subgroups of users in a more general population that make better decisions when given a specific communication format and would be more readily available to information providers in a real world setting. The monetary decision incentive was included again in Experiment 3.

Chapter 2: Experiment 1

Introduction

The primary goal of this first experiment was to determine whether numeracy was related to quality decision-making when probability information was added to weather forecasts. It could well be that limitations in numeracy prevent successful use of probability information. Numeracy can either be measured subjectively though self-report or objectively though a math assessment with correct or incorrect answers (see Reyna et al, 2009 for an excellent review). An early three-item numeracy test was developed by Schwartz et al. (1997). It assessed competency with questions about fair coin flips and converting percentages to frequencies. Indeed, scores on this measure were related to understanding risk communication. Women who were better able to answer these questions also had a better understanding of breast cancer risk reduction when it was expressed as a percentage or as a frequency. Later Lipkus, Samsa, & Rimer (2001) expanded this three-item assessment by adding eight additional questions that involved non-specific health risks. Individuals who performed better on this assessment were less likely to overestimate their risk of cancer when it was expressed verbally, as a percentage, or as a fraction (Gurmankin et al., 2004). However, response distributions for these measures were often negatively skewed, suggesting they were too easy for many participants, forcing researchers to preform median splits to classify high and low ability (see Peters et al., 2006).

More recently, Cokely, Galesic, Schulz, Ghazal, and Garcia-Retamero (2012) developed a numeracy assessment that is psychometrically valid and can discriminate abilities well across diverse samples (college students, medical professionals, general populations). In other words, this assessment provides greater psychometric sensitivity among moderate to very highly numerate individuals than previous assessments. Moreover, this measure has been demonstrated to account for unique variance in risk comprehension tasks beyond other cognitive factors (e.g., working memory capacity, fluid intelligence, deliberate thinking; Ghazal, Cokely, & Garcia-Retamero, 2014). I wanted to use an assessment that could discriminate ability across a variety of populations, thus I used this test (Berlin Numeracy Test: Cokely et al., 2012) in Experiment 1, as well as in all subsequent experiments.

However, the level of numeracy skill in a given population is not readily available to those who are tasked with conveying risk information. Demographic information, such as the general age and education level of the population within a specific region is often known though. Demographics might be a useful proxy for an unknown cognitive factor such numeracy by which to tailor risk communication. These factors could also help explain how well people can make use of risk information and some may be correlated with numeracy. To examine this, demographic information such as education level, age, and gender was collected from participants in Experiment 1.

There is mixed evidence for the effect of education on risk understanding. More education does not always help with understanding uncertainty communication (Ibrekk & Morgan, 1987). However other research has shown that less education has been associated with worse health literacy, or worse understanding of health risks (Reyna & Brainerd, 2007). There is also evidence to suggest education plays a role in general risk taking, although this research is also mixed. Higher levels of education have been associated with lower levels of risk taking and increased risk-aversion (Knight, Weir, & Woldehanna, 2003). In direct contradiction, other research has shown higher levels of education associated with higher levels of risk taking (Grable, 2000). Therefore, individuals with higher education could make better use of uncertainty information than those with less education, could perform better (or worse) on the road salt task in general, or there could be no effect.

Age was also recorded in Experiment 1. It is an interesting factor because conflicting evidence suggests that increased age could have either a negative or positive effect on decision-making in the road salt task. Older individuals may not understand uncertainty information well, but may also be risk-averse. Evidence suggests that older individuals (75-93 years) don't understand probability information well, as they both overestimate and underestimate probabilities of simple objective outcomes such as rolling a die (Fuller, Dudley, & Blacktop, 2001). Other studies have also confirmed that older individuals (66-75 years) display worse risk comprehension than younger individuals (18-27 years) in a probability comparison task (Fausset & Rogers, 2012). Much of the literature also supports the idea that decision making effectiveness, decision accuracy, and decision rationale declines as people get older (60+ years; Thornton & Dumke, 2005) and that older (60+ years) individuals use less complex decision-making strategies (Mutter & Pliske, 1996). However older individuals (53-99 years) also have a reduced propensity for risk taking (Deakin, Aitken, Robbins, & Sahakian, 2004; Mata, Josef, & Hertwig, 2016). Thus, older individuals may benefit from a communication format that does not include numerical uncertainty, but also may take fewer risks, resulting in overall better decisions in the road salting task.

There is little reason to believe that men and women should understand uncertainty information differently as gender differences have not typically been reported in studies that test risk communication in both women and men (see Fagerlin, Zikmund-Fisher & Ubel, 2011). However, there is some evidence to suggest gender differences in risk taking. There is some evidence that women are more risk-averse and less risk-seeking (Byrnes, Miller, & Schafer, 1999; Eckel & Grossmann, 2002; Levin, Snyder, & Chapman, 1988; Powell & Ansic, 1997). Furthermore, when asked about their attitudes towards financial risks, women report less risk propensity than men (Barsky, Juster, Kimball, & Shapiro, 1997). Accordingly, women may make better decisions in the road salt task and make fewer risk-seeking errors. It is important to note however, no gender differences have been found in previous experiments using the Road Salt Task in a college undergraduate population.

In sum, present study was an initial test for an online version of the Road Salt Task. I wanted to determine whether the advantage for numeric uncertainty extended to a broader population with potentially lower numeracy skills and whether an online version of the Road Salt Task produced the same pattern of results as previous in person studies. Participants should make more salting decisions at higher probabilities regardless of forecast format and there should be an advantage for probability forecasts over deterministic or recommendation based forecasts.

The online testing environment used for this project was Amazon’s Mechanical Turk (M-Turk). The M-Turk is a platform for thousands of response tasks covering a wide range of topics. Respondents complete tasks for small monetary rewards. Evidence suggests that M-Turk samples are reliable and reasonably representative of the general public and data quality does not seem affected by payment amount (Buhrmester, Kwang, & Gosling, 2011; Mason & Watts, 2009; Paolacci, Chandler, & Ipeirotis, 2010). Moreover, a comparison of M-Turk participants to the typical student population demonstrated that they exhibited the standard decision-making biases: they were risk-averse for gains and risk-seeking for losses (Goodman, Cryder, & Cheema, 2012). However, previous demographic analyses of M-Turk samples have found them to be younger, more educated, lower income, and over represent women than the general US population (Ipeirotis, 2010). Thus, while M-Turk is not a perfect sample of the general population, it is much more representative that a college sample.

In Experiment 1, participants first completed the Road Salt Task. Over 60 trials, participants must decide whether they ought to spend a limited monetary resource to salt the roads to prevent them from freezing. Or they could take a gamble, saving on precautionary costs but subjecting themselves to the potential of a severe penalty. To help make their decisions participants were given a temperature forecast. Some participants were given a single value forecast. Other participants were given a single value forecast with the probability of a freezing temperature occurring. A third set of participants were given a single value forecast with advice about whether or not to salt based on economically optimal decision. Participant numeracy was also assessed and demographic information such as gender, age, and education was collected.

Method

Participants. Participants were recruited from Amazon's Mechanical Turk, an online crowd sourcing service that hires workers for the execution of tasks (called Human Intelligence Task, or HIT). Participants were compensated $1.00 for their responses. The 261 participants who completed the HIT were residents of the U.S. and had a 90% prior "approve rate" (percent of prior HITs accepted by requester). The sample was similar to the 2012 US census (see Table 1) although there were slightly more females (53%), the average age was slightly younger (34 years, range: 18-69), and the sample was better educated (60% of the sample had at least a Bachelor's Degree). An additional 55 participants started the experiment on M-Turk but failed to complete the task.

Procedure. After participants gave informed consent they were asked to provide demographic information. This included their gender (selecting between male and female), age (free response), and highest level of education (selecting between high school or less, some college, or college degree). Then participants completed the road salting task followed by the numeracy assessment. Participants took an average of 20 minutes to complete the experiment, but participants were allowed 60 minutes to complete the experiment.

Road Salt Task. Participants first read task instructions, which included a description of the task and the cost-loss structure. Participants were instructed to take the role of a manager of a road maintenance company, contracted to prevent road icing by appropriately treating the roads for a two-month period in winter. To be effective, salt treatments had to be applied before freezing temperatures were observed (32ºF or less). On each of 60 trials participants decided whether, based on forecasts for the nighttime low temperature, treating the roads with salt was necessary that day. Participants received a monthly budget of $36,000 for each of the two hypothetical months. Each trial consisted of the choice to apply salt to roads costing $1,000 per day or not. If salt was not applied and a freezing temperature was observed, a penalty of $6,000 was deducted from the budget. This was justified as compensation to the city for any accidents or damage as a result of dangerous road conditions.

Because the cost-loss ratio was the same for all trials, the optimal strategy, based on expected value, was to salt whenever the probability of the critical event (temperature $\leq$ 32ºF) was 17% or greater (1000/6000 = .17). Participants were not directly informed of the reason behind this strategy or given the probability of freezing. Instead, they were instructed to attempt to maximize profits by minimizing salting expenses and avoiding penalties.

On each trial, representing one day, a forecast for the next night appeared on the screen. Participants were randomly assigned to one of three forecast formats, all of which included the nighttime low temperature. The first was the control condition that included the nighttime low temperature forecast alone. This is considered a deterministic forecast condition as it implies a single outcome (deterministic condition). The second was an experimental condition that also included the probability of freezing (probability condition). To test whether there were some user groups or situations in which having explicit advice is superior to being given the probabilistic forecast, the last experimental condition included advice from a Decision Support Aid, described as taking into account the temperature forecast, the uncertainty involved, and the cost of salting weighed against the potential penalty of not salting. In fact, it recommended applying salt whenever the probability of freezing was greater than or equal to 17% and not applying salt otherwise. This condition included the single-value forecast and the advice (advice condition). After reading the forecast, participants indicated their salt decision by clicking on one of the decision buttons. Finally, participants indicated what they thought the nighttime low temperature would be, entering a numeric value in a text box. The forecast was visible to participants at all times during a given trial, along with their current budget. Immediately afterward, the observed nighttime low temperature and any balance adjustments appeared on the screen. The salt costs and penalty amounts were shown in red below the balance when they occurred but then disappeared. After each block of 30 trials, a virtual month in the simulation, participants indicated their overall trust in the forecasts on the same scale as described above. After the first block, a break screen was shown and participants clicked "Next" to continue to the second month's trials. At that point, $36,000 was added to the remaining balance. Participants were always able to continue to salt even if their balance dropped below $0 and this was described in the instructions as borrowing against the next month's budget installment.

Salting Task Stimuli. Participants were shown a sequence of 60 forecasts and observations. These were based on archived data from Spokane and Yakima, Washington State, and followed natural weather patterns. Half of all observed temperatures were above their respective single-value temperature forecasts and half were below (M_{Forecast} = 34.33º F, Range $_{\text{Forecast}}$ = 32º F - 37º F; $M_{\text{Observation}}$ = 34.57º F, Range $_{\text{Observation}}$ = 26º F - 42º F). The probability of freezing forecast was reliable: The percentage of observed temperatures at or below freezing was within the range of forecasted probabilities in six probability of freezing (PoF) bins (10-16%, 17-23%, 24-33%, 31-37%, 38-44%, and 45-51%). For example, in the 10-16% bin, freezing temperatures were observed on 2 of 18 days, or 11.1%. There were 18 forecasts in both the 10 – 16% and the 17–23% ranges to allow for decision comparisons in these two categories of particular interest; salting below 17% (risk-averse errors) and failing to salt at low probabilities at or above 17% (risk-seeking errors). The remaining 24 forecasts were separated into subsequent probability rages as follows: seven forecasts in the 24-30% range, six forecasts in the 31-37% range, five forecasts in the 38-44% range, and six forecasts in the 45-51% range.

Numeracy Assessment. The Berlin Numeracy Test (Cokely et al., 2012) assesses knowledge of probability and statistical computation. The test includes questions on probability that increase in difficulty over the series of questions. In the present sample, participants answered a 7-item version of the test that has been calibrated for a general/M-Turk population (Cokely et al., 2012; see Appendix A for questions and answers). Participants answered each question by typing a proportion or percentage into a response box. Each correct answer was awarded one point. A final score was calculated out of seven points. Once both tasks were completed, participants received a completion code that they entered on the M-Turk website indicating they had completed the HIT successfully. Completion codes were verified by the experimenter and then payments were approved.

Design. The road salting task experiment had a single factor (forecast format) between-participants design with three levels: deterministic (single-value forecast), advice, or probability. Participants were randomly assigned to one of these three conditions. The dependent variables were the decision to salt or not, numeric temperature estimate, and trust rating. Individual differences were assessed using the numeracy, age, and level of education as predictors for road salting task dependent variables.

Hypotheses

I hypothesized that participants would take precautionary action more often as the likelihood of freezing temperatures increased, regardless of forecast format, because of the implicit correlation between nighttime low temperature and probability of freezing. I further hypothesized that the information contained in the forecast upon which their decisions were based would matter. Specifically, I predicted that those who were given forecasts that included numerical uncertainty would make better decisions and have greater trust in the forecast, replicating the effects of previous studies using the same task. I also hypothesized that numeracy would predict decision quality for those given numerical uncertainty information, suggesting that those with higher numeracy could make better use of probabilistic information. I thought it was also possible that higher education would lead to better decision quality when given numerical uncertainty information. It was unclear what effect age and gender would have on making use of numerical uncertainty information or following explicit advice.

Results

As this experiment was the first to look at the road salting task for populations beyond college students, and the first to use an online format, the first set of results were conducted to determine whether group-level differences between the forecast format conditions observed in previous studies (Grounds & Joslyn, 2015; Joslyn & Leclerc, 2012; Roulston et al., 2006) were replicated here. The subsequent set of analyses then examined if any of the individual differences measures (numeracy, age and education) predicted road salting task performance within each of the three forecast format conditions.

The first task was to examine participants' numeric temperature estimates from the road salting task to determine whether they understood the stimuli. To summarize across the differing forecasts, the deterministic forecast was subtracted from participants' numeric estimate for each trial and a mean deviance score was calculated for each participant. There were 12 participants (5%) with values at least two standard deviations above the mean standard error in their condition, suggesting that they were not paying attention to the forecast. They were omitted from the analyses below, because this was an indicator that these participants did not understand the stimuli presented or were not paying attention to the task. Thus, there were 82 participants remaining in the deterministic condition, 81 participants in the advice condition, and 86 participants in the probability condition.

Communication Format Comparisons

Decision Quality. Decision quality was examined by calculating the mean expected value of salt decisions for each participant over the 60 trials. For each trial on which participants decided to salt, the cost of salt (i.e., -$1,000) was assigned. For each trial on which participants decided not to salt, the expected value of the decision was the value of the penalty for salt (-

$6,000) multiplied by the probability of freezing on that trial. Then a mean expected value score was calculated for each participant.

An ANOVA was conducted on mean expected value scores with forecast format (deterministic, advice, and probability) as a manipulated independent variable and gender as an observed independent variable. Contrary to the prediction, expected value was not significantly different in the different forecast format conditions (F (2, 243) = 1.90, p = .14; see Figure 1). There was no effect of gender.

Trust. However, the advantage for probability forecasts returned in the analysis for trust. An ANOVA conducted on the average of the end-of-month trust ratings (5-point scale), with forecast format (deterministic, advice, and probability) and gender as the independent variables, showed a main effect for forecast format, F (2, 243) = 4.98, p < .01. Tukey's post hoc tests revealed that mean trust ratings in the probability condition (M = 3.22, SD = 0.99) were significantly higher than in the deterministic condition (M = 2.74, SD = 0.99; p < .01 Cohen's d = .48; see Figure 2). However, the mean trust ratings for the advice condition (M = 3.04, SD = 0.95) were not significantly different from the deterministic (p = .13 Cohen's d = .31) or probability conditions (p = .70, Cohen's d = .19). There was no effect of gender.

Binary Decisions. In order to verify that an M-Turk sample would also chose to apply salt more at colder temperatures (and corresponding higher freeze probabilities), binary decisions were examined more closely. To determine whether participants made different decisions in the ranges of PoF above and below the long-run optional threshold, I calculated the proportion of salt decisions in each of the six probability ranges and then calculated the average proportion above and below the 17% threshold for each participant. Then, a mixed-model ANOVA was conducted on mean proportion of "salt" responses per participant, with PoF range (above and below 17%) as the within-groups variable and uncertainty forecast format (no uncertainty: deterministic and advice combined, uncertainty: probability) and gender as between groups variables. There was a significant main effect for PoF range, F (1, 245) = 904.76, $p < .001$. Participants salted more above the 17% threshold (M = .63, SD = .19) than below it (M = .23, SD = .23; d = 1.90), suggesting that all participants recognized that freezing is more likely with colder temperatures regardless of receiving probability information. There was also a significant interaction with uncertainty forecast format, F (1, 245) = 4.20, $p < .05$, suggesting greater differentiation in salt decisions among those who had the probability of freezing. Those with probability forecast salted less often below the 17% threshold (M = .16, SD = .21) than did those with no uncertainty information (M = .25, SD = .24), and more often above it (M = .67, SD = .17) than those with no uncertainty information (M = .61, SD = .17; see Figure 3). Gender was found to have no significant effects on the results.

Decision Errors. Finally, I examined decision errors defined in terms of the long-run optimal strategy. The previous set of analysis combined deterministic and advice into a no uncertainty forecast condition, but here the three forecast conditions were examined separately again. There were two kinds of errors: Failing to salt when PoF was greater than 17%, incurring more risk than was optimal (risk-seeking), and salting when the PoF was less than 17%, spending more than was necessary to protect against risk (risk-averse). To determine which kind of error was prevalent, analyses were restricted to the range of probabilities of freezing between 10% and 23%, in which there was an equal number of trials above and below 17% PoF. A decision *not to* salt was coded as an error above 17% and decision *to* salt was coded as an error below 17%. Then, a mixed-model ANOVA was conducted on percent “errors,” with error type (risk-seeking and risk-averse) as the within-groups variable and forecast format (deterministic, advice, and probability) and gender as between-groups variables. There was a significant main effect for error type, F (1, 243) = 171.32, $p < .001$. Participants made more risk-seeking (M = .61, SD = .24) than risk-averse errors (M = .23, SD = .23, d = 1.62). There was also a significant forecast format interaction, F (2, 243) = 3.69, $p < .05$. Participants with probability information made fewer risk-averse but more risk-seeking errors than did participants with advice or deterministic forecasts. There was no effect of gender.

Thus, the online version of the road salting task replicated many of the effects found in previous studies. There was greater trust in probability forecasts and all groups salted more above than below the economically optimal threshold suggesting that they were processing the forecast (indeed, single value night time low temperature is correlated with the probability of freezing) and took the task seriously. Fewer risk-averse errors were made with the probability forecast, but risk-seeking errors at low probabilities were increased, as has often been found in previous studies. However, the forecast format comparison analysis for expected value did not reach significance: overall the probability forecast was no different than the advice or the deterministic single value forecasts. There were no effects of gender on the results in any of the analyses.

Individual Differences. All of the analyses previously described examined the differences in decisions based in group level differences (forecasts format and gender). The next set of analyses were conducted on the same responses but instead examined the contribution that a cognitive individual difference factor (numeracy), individual experience as reflected in age, and amount of education can explain person-level variability in responses. The responses of interest were the same dependent variables used in the communication format comparison analyses above.

The following analyses investigated what amount of variance could be explained within each of the forecast format conditions. For each of the dependent variables, a regression analysis was conducted with numeracy, age and education (High School or less, College or more) as independent variables entered in simultaneously as a single step.[1] Regressions were run separately for each forecast format group because the variability in road salting task responses was assumed to be specific to communication format. The first goal was to ensure that there were not any systematic differences in the predictor variables between the forecast format groups and to determine whether numeracy increased with education or age. Indeed, there were no significant differences in numeracy scores [F (2, 246) = 1.09, p = .34], age [F (2, 246) = 1.09, p = .34] or level of education [χ^2 (3, N = 249) = 0.23, p = .89] between forecast format conditions. Moreover, there were also no significant differences in numeracy between participants who had high school or some college education (M = 2.69, SD = 1.45) and those with college and advanced degrees (M = 2.84, SD = 1.90), F (1, 243.57) = 0.68, p = .50. However, participants who only had a high school or some college education (M = 31.92, SD = 11.09) were younger on average than those who had a college degree or more education (M = 36.27, SD = 10.51), F (1, 247) = 3.21, p < .01, d = .40. Mean values of predictor and outcome variables for the three forecast format conditions appear in Table 2. Numeracy and age were not correlated in the deterministic forecast format condition, r(82) = .11, p = .32, however they were moderately correlated in the advice (r(81) = .30, p = .01) and probability (r(82) = .27, p = .02) forecast format conditions suggesting that there might be a slight positive relationship between numeracy and age. However, age and numeracy

[1] An interaction term for age and education was also included in initial analyses for Expected Value (as it was the main variable of interest), but did not significantly improve any of the regression models. Thus, only the individual predictor full models are reported.

were not highly correlated with each other, nor did numeracy differ by level of education suggesting that numeracy is largely an independent predictor.

Decision Quality. In all three forecast format conditions, numeracy explained a significant amount of variance in Expected Value scores, education explained a significant amount of variance in the deterministic and probability conditions, and age explained a significant amount of variance in the deterministic condition alone. Within the deterministic forecast, the model was significant, F (3, 78) = 7.52 $p < .001$, $R^2 = .23$. The results indicated that individuals with higher numeracy had better expected value scores ($t = 3.29$, $p = .002$), older individuals had better expected value scores ($t = 2.23$, $p = .03$), but surprisingly more education led to worse expected value scores ($t = -2.66$, $p = .01$; see Table 3). Within the advice condition, the model was also significant, F (3, 77) = 6.26, $p < .001$, $R^2 = .20$ with higher numeracy leading to better expected value scores ($t = 4.19$, $p < .001$; see Table 3). The model for the probability condition was also significant, F (3, 82) = 8.67, $p < .001$, $R^2 = .24$. The results indicated that individuals with higher numeracy had better expected value scores ($t = 3.76$, $p < .001$), but again more education led to worse expected value scores ($t = -2.73$, $p < .001$; see Table 3). These models explained a substantial proportion of variance because in each case the R^2 value exceeded .14 which corresponds to a Cohen's *d* value of .8, indicating a large effect size (Friedman, 1982).

Trust. A similar set of analyses conducted on mean trust scores indicated that no predictors significantly explained monthly trust ratings in the deterministic, advice, or probability conditions.

Binary Decisions and Decision Errors. Binary decisions were grouped into three separate analyses: decisions to salt below the 17% threshold, failing to salt at low but necessary freeze probabilities, and decisions to salt above the 17% threshold. Decisions to salt below 17%, or spending more than was necessary to protect against risk were considered risk-averse errors.

Failures to apply salt when the probability of freezing was between 17% and 23%, or incurring more risk than was optimal, were considered risk-seeking errors. The final set of analyses investigated decisions to apply salt above the 17% threshold, indicating appropriate precautionary action.

Regression analyses conducted on decisions to salt below the threshold (risk-averse errors) were only explained by numeracy and education in the probability condition. The model for the probability condition was significant, F (3, 82) = 3.39, $p < .05$, $R^2 = .11$, such that individuals with higher numeracy salted less below the threshold ($\beta = -.33$, $t = -3.05$, $p < .01$). There was no effect of age ($\beta = .15$, $t = 1.34$, $p = .18$) or education ($\beta = .07$, $t = 0.68$, $p = .50$). Decisions to salt below the 17% threshold were not explained by the predictors in either the deterministic or advice conditions. Moreover, no predictors significantly explained risk-seeking errors in the range of probabilities immediately above the optimal threshold (17-23% chance of freezing) in any condition.

However, decisions to salt above the 17% threshold in the full range of probabilities (17-51% chance of freezing) were explained by numeracy in both the deterministic and advice forecast format conditions. Within the deterministic forecast, the model was significant, F (3, 78) = 4.00, $p < .05$, $R^2 = .13$. The results indicated that individuals with higher numeracy salted more above the threshold ($t = 2.44$, $p = .02$; see Table 4). A similar pattern was observed for the advice condition (model F (3, 77) = 5.51, $p < .01$, $R^2 = .18$, numeracy $t = 3.80$, $p < .001$; see Table 4). However, no predictors significantly explained decisions to salt above the economically optimal threshold in the probability condition.

Thus, *numeracy* appears to be a good predictor for most dependent variables in the road salting task as the forecast format models explained a substantial amount of variance in expected value as an R^2 value of .14 corresponds to a *d* value of .8, indicating a large effect size (Friedman, 1982). Higher numeracy scores led to higher expected value in all three forecast format conditions (deterministic, advice, and probability) compared to lower numeracy scores. However, the advantage for numeracy appears to be slightly different depending on the forecast format condition. When looking only at the probability condition, higher numeracy scores led to fewer instances of salting below the economically optimal threshold, suggesting fewer risk-averse errors. However, within the deterministic and advice forecast formats, higher numeracy led to more salting above the economically optimal threshold. *Age* and *education* were also predictive in specific conditions. Older participants made better decisions in terms of expected value for the deterministic forecast format as compared to younger individuals. Surprisingly, more education led to worse expected value for the deterministic and probability forecast formats as compared to less education.

Discussion

Experiment 1 replicated many, but not all, of the effects found in previous studies. Thus, an online version of the road salting task seems to be a promising way to collect valid data. As with previous studies, all groups salted more often as temperatures got cooler and the probability of freezing increased, suggesting that they were processing the forecast information and taking the task seriously.

Some of the effects of forecast format observed in previous studies were replicated here as well. Uncertainty estimates led to higher ratings in trust, perhaps because it confirmed users' intuition about weather uncertainty. Participants who had uncertainty estimates also had greater differentiation of salting decisions between low and high probability forecasts than those without uncertainty information, thus taking more appropriate action when it was necessary, but choosing to accept a risk for unlikely events. Moreover, consistent with previous research, participants who had uncertainty estimates were more risk-seeking at low freeze probabilities for which precautionary action was economically necessary, between 17% and 23%, perhaps because they regarded spending money on precautionary action as unneeded at low freeze probabilities.

However, the advantage for uncertainty forecasts seen in previous research in terms of expected value was not observed here. Overall the probability forecast was no different than the advice or deterministic forecasts in terms of decision expected value. This was clearly not due to a lack of power. A power analysis for Joslyn & Leclerc (2012), revealed that the sample size needed for each of the conditions to replicate the advantage for probability was 42 participants in each condition. There were approximately double that number of participants in each condition of Experiment 1 and still the effect was not observed. However, there was one key difference between this study and previous studies (e.g., Joslyn & Grounds, 2015; Joslyn & Leclerc, 2012; Roulston et al., 2006). These previous studies included a monetary incentive to perform well. Participants were paid a bonus based on how much money they had left in their virtual budget at the end of the task[2]. However, the possible bonus based on ending budget may provide motivation to consider all of the relevant information carefully, especially in the probability condition, and make better decisions. The current experiment did not include such incentive, and this fact may explain the lack of the expected format differences. This issue will be addressed in Chapter 3 and Chapter 4.

[2] Ending budget is not used as a dependent variable because it is not a direct measure of their decision quality, as budget also reflects some element of luck (e.g. not salting when it is optimal, but no freeze is observed).

It was somewhat unexpected that older participants were better able to make use of a deterministic forecast as it did not necessarily seem consistent with previous research (e.g., Mutter & Pliske, 1996; Thornton & Dumke, 2005). However, research investigating aging typically assesses respondents above an age threshold, such as 60 years of age or older. If that same threshold was used in this sample, only seven participants would remain (3% of sample collected, range: 60-69). A data set that small is not enough to draw any meaningful conclusions. Thus, the conclusions of this study are limited to middle-aged (and slightly older) participants. Perhaps middle-aged individuals are better at estimating the uncertainty in single value forecasts due to more years of experience using them.

Consistent with the hypothesis, higher numeracy resulted in better expected value scores for participants who were given numerical uncertainty information. While this is not surprising, it supports the idea that a better understanding of numbers and numerical properties may help people incorporate numerical uncertainty information into their decision-making process. However, the effect was not specific to the probability condition, higher numeracy resulted in better expected value scores when participants were given explicit advice as well as when participants were given just a deterministic single value forecast. These communication formats did not include probabilistic information, and yet higher numeracy lead to better decisions. This suggests that higher numeracy not only leads to better understanding of numerical uncertainty information, but that it may also help people estimate uncertainty about a situation when it is not explicitly provided. However, this conclusion should be qualified. Most people have had a great deal of experience with weather forecast uncertainty and research suggests that people already understand that forecasts are uncertain, even when the uncertainty is not explicitly stated (Joslyn & Savelli, 2010; Morss, Demuth, & Lazo, 2008). Thus, better numeracy may help people have more accurate estimates of weather forecast uncertainty. In addition, higher numeracy may help people to appreciate the advantage of precautionary action as a long run strategy by making the cost/loss ratio more salient.

However, as numeracy was the only cognitive factor measured in this experiment, it may be tapping into some larger cognitive factor that improves decision quality. Thus, if numeracy consistently explains decision quality in terms of expected value scores for forecast formats that do not include numerical uncertainty, then the way it improves decision quality may not have anything to do with comprehending the probabilistic information in the forecast. This will continue to be explored in Chapter 3 and Chapter 4.

As has been noted previously, numeracy was not the same as education. In fact, the mean numeracy scores of those with college and advanced degrees was not significantly different than that of those who had not earned a college degree. While those with higher numeracy made better decisions, those with more education made worse decisions.

It was also quite surprising that numeracy did not explain risk-seeking errors (failing to salt in the range immediately above the threshold, 17-23%) in any of the forecast format conditions. There is a strong risk-seeking bias in the domain of losses (Kahneman & Tversky, 1979), so perhaps high numeracy is not enough to overcome this bias. Moreover, individual subject’s risk propensity was not addressed. It was assumed that that the subjective value of money was linear with respect to expenditures in salting or penalties for all participants. However individual risk propensity may result in a willingness to accept a higher than optimal risk that is not due to probability distortion. Risk seeking decisions instead could to due to subjects' convex value function for the monetary losses that result from the decision to salt and the penalty for failing to salt when necessary. And while numeracy was a significant predictor of expected value scores, it did not entirely explain expected value scores, suggesting there are other factors influencing decisions. The contribution of individual risk propensity and other cognitive factors were explored further in Chapter 3 (Experiment 2).

Chapter 3: Experiment 2

Introduction

The driving motivation of this research project was to understand how best to communicate weather uncertainty to specific sub-groups of users. More specifically, should forecast communication be tailored to different user groups based on specific individual differences factors or would one communication method be suitable for all?

Numeracy was a significant predictor of decision quality in Experiment 1, as higher numeracy led to higher expected value for all forecast formats. However, the majority of the variance remained unexplained, suggesting there are other cognitive factors influencing decisions. Working memory and fluid intelligence are important factors that could be influencing decision quality because subjects may be dependent on the ability to simultaneously evaluate and consider all forecast dimensions which could be limited by these factors. In particular, if the forecast included uncertainty information, then people would have to consider both possible outcomes, their implications and the likelihood associated with each throughout the decision process. Thus, working memory capacity or fluid intelligence may be particularly important in the probabilistic condition.

Deliberate thinking and need for cognition are could also influence decision quality as the economically rational approach to the road salting task involves a cost to protect oneself even when the likelihood of freezing temperatures is small. People may be initially reluctant to do so. Perhaps deliberate or effortful thinking is necessary to overcome an automatic response to not apply salt when freezing is unlikely.

In order to determine whether any of these factors predicted road salt performance, in Experiment 2, additional measures to assess working memory capacity, deliberate thinking, need for cognition, and fluid intelligence were included to see whether they influence uncertainty communication comprehension and decision making.

It is also possible that decision quality or decision errors may not be due to misunderstanding of numerical uncertainty or any of the cognitive processes described above, but instead due to individual variation in risk tolerance (Joslyn & Savelli, 2010; Morss et al., 2008). Some individuals may be willing to accept a risk that was greater than economically optimal. However, risk tolerance may be domain-specific, such that a person is not consistently risk-averse or risk-seeking across domains (Blais & Weber, 2006). Thus, risk propensity can vary between people and between a given person’s decisions in different situations. To test for the possibility that decision quality or decision errors could be explained by individual risk tolerance, I included a measure to assess risk propensity, specifically financial risk propensity. I focused on financial risk propensity due to the nature of the Road Salt Task. While it is a task conveying various weather risks, participants must decide whether they ought to spend a limited financial resource on precautionary action. Therefore, a measure that assesses financial risk propensity should capture the risk preferences in the road salting task and was also included.

I collected the data for Experiment 2 using a college student population, as most of the additional measures had been developed using similar populations. However, the direct result of this decision was that the population for Experiment 2 would have the same level of education: some college. Thus, further investigation of possible education differences was not explored again until Chapter 4 (Experiment 3).

In Experiment 2, participants first completed the Road Salt Task. Participants were given a single value forecast alone, a single value forecast with decision advice, or a single value forecast with the probability of a freezing temperature occurring. A fourth forecast format was also included that combined the probability information with the decision advice along with the single value forecast. This combination has been shown to reduce risk-seeking errors (Joslyn & Grounds, 2015; Joslyn & Leclerc, 2012). Thus, a total of four forecast formats were tested: deterministic single value, advice, probability, and probability with advice. A monetary bonus based on ending budged was included in the task. Then participants completed measures to assess numeracy, deliberate thinking, need for cognition, working memory capacity, fluid intelligence, and financial risk propensity.

Method

Participants. The participants were 421 college students from a large northwestern university. The mean age was 19 (range: 18-31) and 223 (53%) were female. They were enrolled in the introductory psychology course and received course credit for participating. The average participant also received $0.72 as bonus compensation, however only 131 participants received a monetary bonus. This resulted in an average of payment of $2.31 for these participants.

Procedure. After participants gave informed consent they filled out some general demographic information. This included their gender (selecting between male and female) and age (free response). Then participants completed the road salting task followed by the set of individual difference measures in a randomized order. The individual difference measures (described in detail below) assessed numeracy, deliberate thinking, need for cognition, working memory capacity, fluid intelligence, and financial risk propensity. It took participants an average of 75 minutes to complete the experiment, but participants were given 90 minutes.

Road Salting Task. Participants completed the same salting task from Experiment 1, however participants were paid commensurate with performance. Participants received $1 for every $2,000 over $12,000 left in their virtual budget at the end of the task.

Numeracy. Participants completed the same numeracy assessment (Berlin Numeracy Test; Cokely et al., 2012) from Experiment 1.

Deliberate thinking. To assess deliberate thinking, participants completed the Cognitive Reflection Test (CRT; Frederick, 2005). The CRT is a series of 3 basic math questions used to assess individuals' ability to suppress an intuitive and spontaneous ("system 1") wrong answer in favor of a reflective and deliberative ("system 2") right answer (see Appendix B).

Need for Cognition. To assess the factor of need for cognition (NFC), participants completed the self-report scale developed by Cacioppo, Petty, & Kao (1984). It is designed to measure the extent to which individuals are inclined towards effortful cognitive activities by asking individuals to rate the extent to which they agree with statements about the satisfaction they gain from thinking using a 9-point scale ranging from -4 (very strong disagreement) to 4 (very strong agreement). There was a total of 18 items, nine of which were reverse scored (see Appendix C).

Working Memory Capacity. WMC was tested with a reading span task requiring participants to perform a semantic judgment while maintaining a memory load (Unsworth, Heitz, Schrock, & Engle, 2005). Participants read a set of unrelated sentences and judged whether each one made sense. Sentences were between 10-15 words in length. Half of them made sense and half did not. After each sentence judgment, a letter appeared on the screen for 800ms that participants were required to memorize. Three to seven sentence-letter pairs were presented in a set. There were three sets of each set size. The order in which set sizes were presented was randomized. After the final sentence-letter pair in each set, participants indicated the letters remembered from the entire set, by clicking each letter in a 4x3 matrix in the order in which they were presented. Then, participants were informed of the number of letters correctly identified in that set. Letters had to be identified in the order in which they were presented to count as "correct". During recall, the percentage of accurate sensibility responses was shown in red in the upper right-hand corner of the screen.

Participants first practiced the letter span task and the sensibility judgment separately, during which a mean reading time was calculated for each participant. Then participants practiced the combined tasks by first making a sensibility judgment and then remembering a letter. Working Memory Capacity scores were the total number of letters recalled in the correct order out of the 75 that were presented overall (Unsworth et al., 2005).

Fluid intelligence. To assess fluid intelligence, participants completed an abbreviated version of the Raven's Advanced Progressive Matrices (APM; Raven, 2003). This task uses visual items and multiple choice responses. The subject is asked to identify the missing item that completes a pattern. This 12-item version (taken from Bors & Stokes, 1998) has a Cronbach's alpha of .73, which is an acceptable measure of reliability, and is highly correlated with the full length APM (r = .92, p < .001), while taking only 15 minutes to complete.

Financial Risk Propensity. To assess the personality factor of risk propensity, participants completed the financial risk portion of the Domain Specific Risk Taking Scale (DOSPERT Scale; Blais & Weber, 2006). This scale measures conventional risk attitudes (i.e. the reported level of risk taking). Respondents rated the likelihood that they would engage in financial risk activities (risk taking) by responding to six items using a 7-point rating scale ranging from 1 (Extremely Unlikely) to 7 (Extremely Likely).

Design. The Road Salt Task had a single factor (forecast format) with four levels: deterministic (single-value forecast), advice, probability, or probability with advice between-participants design was used. Participants were randomly assigned to one of these four conditions. The responses measured were the decision to salt or not, numeric temperature estimate, and trust rating. Individual differences were assessed using the Berlin Numeracy Test, Cognitive Reflection Test, Need for Cognition survey, Automated Reading Span task, abbreviated Raven's Advanced Progressive Matrices, and a financial risk propensity survey as predictors for Road Salt Task dependent variables.

Hypotheses

As the procedure used in Experiment 2 more closely resembled that in previous published studies, I again hypothesized that participants who were given forecasts that included numerical uncertainty would make better decisions (expected value), differentiate to a greater degree when salting was required and have greater trust in the forecast. Additionally, I hypothesized that adding advice to uncertainty estimates would result in fewer risk-seeking errors in the low probability of freezing range above the threshold (17%-23%). I also hypothesized that higher numeracy, deliberate thinking, and need for cognition would result in better decisions. Moreover, these factors may interact with forecast format, perhaps having a greater impact in the condition in which numerical uncertainty information was included. Finally, I predicted that greater working memory capacity and fluid intelligence would also result in the better decisions, particularly in the conditions that included numerical uncertainty information. It was unclear what effect financial risk propensity would have on making use of numerical uncertainty information or decision making, however in a task where participants must decide between accepting a sure loss or the possibility of a large loss, I predicted that more financial risk taking would result in worse decisions and more errors.

Results

The first set of analyses were conducted to determine whether group-level differences between the forecast format conditions observed in previous studies (Grounds & Joslyn, 2015; Joslyn & Leclerc, 2012; Roulston et al., 2006) were replicated here. The subsequent set of analyses then examined if any of the individual differences measures (numeracy, deliberate thinking, need for cognition, working memory capacity, fluid intelligence, and financial risk propensity) predicted road salting task performance in terms of expected value, forecast trust, salting decisions above the optimal threshold, risk-seeking errors, and risk-averse errors within each of the four forecast format conditions.

The same initial process was employed as from Experiment 1 to remove participants who did not understand the forecast stimuli. There was one participant with temperature estimates at least two standard deviations above the mean standard error in their condition who was omitted from the analyses below. An additional 80 participants (19%) did not meet the 85% correct criterion on the semantic judgment in the reading span WMC task[3] and were also removed from all analyses. After all of these were removed there were 90 participants remaining in the deterministic condition, 80 participants in the advice condition, 82 participants in the probability condition, and 87 participants in the probability with advice condition.

---

[3]This threshold is imposed to prevent participants from ignoring the semantic judgement task in favor of rehearsing the letter series. Similar proportions were below 85% in previous research (DeCaro, Thomas, & Beilock, 2008; Heitz & Engle, 2007). When these individuals were included, working memory capacity was a significant predictor of expected value, but this may be due to motivation instead of working memory.

Communication Format Comparisons

Decision Quality. I examined decision quality by calculating the mean expected value of salt decisions for each participant over the 60 trials, as was done in Experiment 1. An ANOVA was conducted on expected value scores with forecast format (deterministic, advice probability, and probability with advice) and gender as independent variables. Consistent with previous findings and contrary to Experiment 1 reported above, there was a main effect of forecast format, F (3, 331) = 18.43 $p < .001$. Tukey's post hoc tests revealed that mean expected value in the probability condition (M = -$1050.83, SD = $71.13) was significantly better than the advice condition (M = -$1095.51, SD = $102.61; $p < .01$, Cohen's d = .51) or the deterministic condition (M = -$1099.97, SD = $85.73; $p < .001$, Cohen's d = .62). Additionally, the mean expected value for the probability with advice condition (M = -$1020.71, SD = $63.75) was also significantly better than the advice ($p < .001$, Cohen's d = .88) or the deterministic condition ($p <$ .001, Cohen's d = 1.05). There was no difference between mean expected value scores between the advice and deterministic conditions. See Figure 4. Once more I saw no effect of gender on the results.

Trust. A similar pattern emerged when looking at the trust ratings. An ANOVA conducted on the average of the end-of-month trust ratings (5-point scale), with forecast format (deterministic, advice, probability, and probability with advice) and gender as independent variables, showed a main effect for forecast format, F (3, 331) = 4.46, $p < .01$. Tukey's post hoc tests revealed that mean trust ratings in the probability with advice condition (M = 2.61, SD = 0.72) were significantly higher than in the deterministic condition (M = 2.23, SD = 0.75; $p < .01$, Cohen's d = .52; see Figure 5). There was no effect of gender.

Binary Decisions. To better understand the decision strategies, the binary decisions were examined more closely. A mixed-model ANOVA was conducted on mean proportion of "salt" responses per participant, with PoF range (above and below 17%) as the within-groups variable and forecast format (no uncertainty: deterministic and advice, uncertainty: probability and probability with advice) and gender as between groups variables. There was a significant main effect for PoF range, F (1, 331) = 1987.72, p < .001. Participants salted more above the 17% threshold (M = .64, SD = .17) than below it (M = .22, SD = .20, d = 2.26), suggesting that all participants recognize the probability of freezing is more likely with colder temperatures regardless of receiving probability information. There was also a significant interaction with uncertainty forecast format, F (1, 337) = 31.53, p < .001, suggesting greater differentiation in salt decisions among those who were given uncertainty information. Those with uncertainty forecasts salted less often below the 17% threshold (M = .19, SD = .18) than did those without uncertainty forecasts (M = .25, SD = .21). Additionally, those with uncertainty forecasts (M = .67, SD = .17) salted more often above it than did those without uncertainty forecasts (M = .62, SD = .21; see Figure 6). Again, there was no effect of gender.

Decision Errors. Finally, decision errors were examined as defined in terms of the long-run optimal strategy. Here the four forecast conditions were examined separately again. There were two kinds of errors: Failing to salt when PoF was greater than 17%, incurring more risk than was optimal (risk-seeking), and salting when the PoF was less than 17%, spending more than was necessary to protect against risk (risk-averse). To determine which kind of error was prevalent, analyses were restricted to the range of probabilities of freezing between 10% and 23%, in which there was an equal number of trials above and below 17% PoF. A decision not to salt was coded as an error above 17% and decision to salt was coded as an error below 17%.

Then, a mixed-model ANOVA was conducted on proportions of "errors," with error type (risk-seeking and risk-averse) as the within-groups variable and forecast format (deterministic, advice, probability, and probability with advice) and gender as between-groups variables. There was a significant main effect for error type, F (1, 331) = 243.33, $p < .001$. Participants made more risk-seeking (M = .55, SD = .24) than risk-averse errors (M = .22, SD = .20, d = 1.49). There was also a significant forecast format interaction, F (3, 331) = 4.71, $p < .01$. Participants with probability information (probability and probability with advice conditions) tended to make fewer risk-averse errors but those with advice (advice and probability with advice conditions) made fewer risk-seeking errors. Gender effects remained insignificant.

Therefore, Experiment 2 replicated all of the between-groups effects found in previous studies using the abbreviated version of the road salt task. Participants made better decisions when given numerical uncertainty information in the form of a probability of freezing. There was greater trust in forecasts that included probability information and greater differentiation of salting decisions above and below the economically optimal threshold. Furthermore, fewer risk-averse errors were made with the probability forecasts, but risk-seeking errors at low probabilities were increased. Adding advice to the probability forecasts reduced risk-seeking errors. There were no effects of gender in any of the analyses.

Individual Differences. The first task again was to determine that there were no differences in any of the individual differences measures between the different forecast format conditions in terms of mean score. Mean values of predictor and outcome variables for the four forecast format conditions appear in Table 5.

Then, for each of the dependent variables, a regression analysis was conducted with numeracy, cognitive reflection (CRT), need for cognition (NFC), working memory capacity

(WMC), fluid intelligence (APM), and financial risk propensity (F-DOSPERT) as predictors entered in a single step. Regressions were run separately for each forecast format groups. Bivariate correlations among the predictors for each forecast format appear in Table 6. The moderate correlations between the cognitive measures suggest that they likely reflect some common factors, but may also measure distinct characteristics.

Decision Quality. Only the model for the probability condition was significant, F (6, 75) = 2.89, p = .01, R^2 = .19. The results indicated that individuals with higher numeracy had better expected value scores (t = 3.38, p = .001; see Table 7). However, within the deterministic, advice, and probability with advice forecast conditions, the model was not significant suggesting that none of the individual differences measured here were related to performance based on these forecast formats.

Trust. No predictors significantly explained monthly trust ratings in the deterministic, advice, probability, or probability with advice conditions.

Binary Decisions and Decision Errors. Binary decisions were grouped into three separate analyses: decisions to salt below the 17% threshold (risk-averse error), failing to salt at low but necessary freeze probabilities (risk-seeking error), and decisions to salt above the 17% threshold (economically rational response).

For decisions to salt below the 17% threshold, only the model for the probability condition was significant, F (6, 75) = 2.33 p < .05, R^2 = .16. The counterintuitive results indicated that individuals with higher cognitive reflection salted more (making more errors) below the 17% threshold (t = 2.48, p = .02; see Table 7). No predictors significantly explained risk-averse errors in the deterministic, advice, or probability with advice conditions.

Cognitive reflection also explained risk-seeking errors. Within the probability forecast condition, the model for risk-seeking errors was significant, F (6, 75) = 5.33, $p < .001$, $R^2 = .30$. The results indicated that higher CRT scores led to fewer risk-seeking errors ($t = -2.18$, $p = .03$; see Table 7). No predictors significantly explained risk-seeking errors in the deterministic, advice, or probability with advice conditions.

For decisions to salt above the 17% threshold (the economically rational choice), only the model in the probability condition was significant, F (6, 75) = 5.63, $p < .001$, $R^2 = .31$. The results indicated that individuals with higher numeracy ($t = 3.09$, $p = .003$) and more need for cognition ($t = 2.22$, $p = .03$) salted more often above the threshold (see Table 7). No predictors significantly explained decisions to salt above the 17% threshold in the deterministic, advice, or probability with advice conditions.

Taken together, it appears that the only forecast format in which individual difference measures predicted responses was the probability alone condition. *Numeracy* predicted expected value and salting decisions above the 17% threshold. *Need for cognition* also predicted salting decisions above the 17% threshold. *Deliberate thinking*, as measured by CRT, predicted both risk-averse errors and risk-seeking errors at low probabilities, but in opposite directions. Higher CRT lead to more salting overall which constituted more errors below the threshold and fewer errors above the threshold. Surprisingly, working memory capacity, fluid intelligence, and financial risk propensity did not predict any of these responses.

Discussion

Experiment 2 replicated all of the effects found in previous studies completed in the laboratory in a version of the Road Salt Task where number of trials was reduced by 50% and the monetary incentive included. Including uncertainty estimates led to better expected value scores and higher ratings in trust. Participants who had uncertainty estimates also had greater differentiation of salting decisions between low and high probability forecasts, thus taking more appropriate action when it was necessary, but choosing to accept a risk for unlikely events when that was the economically rational strategy. Moreover, consistent with previous research, including explicit advice with the numeric uncertainty estimates reduced risk-seeking errors at low probabilities for which salting was the economically rational choice. The benefit of combining the two might be explained by increased forecast trust (including trust in the advice) due to the uncertainty information. Once participants trust the advice, it helps them overcome what might be a natural tendency to not salt at low probabilities.

Consistent with the hypothesis, but contrary to Experiment 1, numeracy was only a significant predictor of expected value scores in the probability alone condition suggesting that higher numeracy leads to better understanding of numerical uncertainty information. Moreover, the amount of variance explained in the probability condition is substantial but similar to the amount of variance explained in Experiment 1 within the deterministic, advice and probability forecast formats. However, the difference between experiments might have been due the facts that numeracy scores were higher in Experiment 2 as compared to Experiment 1 and they had less variation in the deterministic and advice conditions, but not less variation in the probability condition (see Table 2 and Table 5). Perhaps the restricted college student population in Experiment 2 resulted limited ability ranges and the inability to detect any contribution of numeracy for these conditions.

Higher numeracy was also associated with a greater tendency to take precautionary action above the economically optimal threshold for the probability alone condition. This suggests that higher numeracy is beneficial for understanding and utilizing moderately high probability information.

There were two other significant predictors in the probability condition. Need for cognition was also associated with greater precautionary action above the threshold, suggesting that a greater inclination towards effortful thinking is helpful when deciding to pay for protection using probability information. Higher cognitive reflection was associated more salting in the probability ranges above and below the 17% chance of freezing threshold. This led to better decisions between 17% and 23% (e.g. fewer risk-seeking errors) where salting was economically appropriate. However, it also led to more risk-averse errors below the 17% threshold (10-16% chance of freezing). Taken together this suggests that deliberate thinking increased cautiousness, in some cases beyond what was economically rational.

However, several of the factors tested here were not significantly related to performance in any condition including fluid intelligence, financial risk propensity, and working memory capacity. For that reason, all but WMC will be omitted from Experiment 3. However, a caveat must be noted with respect to working memory capacity. It was a significant predictor of expected value scores before participants who made too many semantic sense judgment errors in the working memory task were removed. For example, a participant made this type of error if they deemed “After yelling at the game, I knew I would have a tall voice” to be a semantically correct sentence. These errors were interpreted to indicate that participants were not paying sufficient attention to the task, and thus, were not motivated. This suggests the initial effect was not due to working memory capacity per se, but instead to motivation. However, I am reluctant to conclude that WMC is not relevant to this task because the lack of effect in Experiment 2 could be due to the restricted range of abilities in the population sampled in this experiment.

College students may have mainly high working memory capacities. Thus, working memory capacity might be associated with better task performance in a more diverse population with greater variation in working memory abilities. Furthermore, much of the literature has suggested a robust relationship between working memory and higher order cognition (e.g., Kane et al., 2004; Unsworth & Engle, 2007; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). This suggests working memory could still have important theoretical implications for explaining precautionary decision-making, in a broader sample. This issue will be addressed in Chapter 4 (Experiment 3).

Chapter 4: Experiment 3

Introduction

The main goal of this research project was to understand how best to communicate weather uncertainty to users. More specifically, should forecast communication be tailored to different user groups or would one communication method be suitable for all? The last experiment in this dissertation combined what was learned from Experiment 1 and Experiment 2 to test that question. This experiment assessed the most promising individual difference measures in a broad population and included a decision incentive.

The results of Experiment 1 and Experiment 2 demonstrated the continued advantage of numeric uncertainty estimates on task performance and higher trust, yet Experiment 1 saw no benefit in terms of expected value. This was not due to lack of power, but could possibly be explained by a lack of decision incentive in the form of a bonus payment commensurate with performance, as participants were not paid bonuses based on their decisions in Experiment 1 but were in Experiment 2. Alternatively, perhaps there is no real advantage for numeric uncertainty in decision quality when looking at a broad population. I explored these possibilities by including bonus payments on M-Turk with a broader population.

The results of Experiment 1 and Experiment 2 also demonstrated that higher numeracy is associated with better performance overall in Experiment 1 and better understanding of numerical uncertainty information in Experiment 2. Accordingly, numeracy was assessed again in Experiment 3. However, with the exception of Working Memory Capacity, in the interest of reducing the total length of time of the experiment, measures of factors that were not significant in Experiment 2 as well as the Need for Cognition measure which had limited impact, were not included in Experiment 3. To my knowledge, the working memory abilities of the general population have not been assessed on a large scale. Most previous studies that have measured working memory capacity have used college student or "moderately diverse" samples that included a small subset of community participation (see Kane et al., 2004). It would be valuable to understand how representative a college sample is to reflect the abilities of the general population. Furthermore, only one study looked at working memory capacity on M-Turk, but used a small sample of participants and a much more difficult version of the task where participants had to maintain 7 letters each trial and did not compare their performance to a college sample (Gray & Gallo, 2016). Thus, it was possible that working memory capacity could explain how people understand numerical uncertainty in a large online population. Accordingly working memory capacity was assessed in Experiment 3.

Furthermore, I was still interested in determining whether it was possible to tailor communication to specific communities based on some demographic factor to which atmospheric scientists or officials would have access. Thus, gender, age and level of education were collected. A final factor that was included in Experiment 3 to explain decision quality or numerical uncertainty understanding was participant level of income. Some theories predict that wealthier individuals would make more risky choices (Cohn, Lewellen, Lease, & Schlarbaum, 1975) which would result in worse decision quality, in the road salt task. However, empirical research has suggested that individuals who are very wealthy are typically risk-averse while those who are less wealthy prefer taking risks (Gregory, 1980). To assess this, participants provided their yearly income.

Method

Participants. Eight hundred and twenty-eight participants who were residents in the U.S. and had a 90% prior "approve rate" were recruited from M-Turk. Participants were compensated $1.00 for their responses, plus a bonus commensurate with the money remaining in their virtual budget at the end of the road salting task. The sample was similar to the 2012 US census (see Table 8) although the average age was slightly younger (35 years, range: 18-75), and the sample was better educated (57% of the sample had a Bachelor's degree or higher education). The average participant received $0.80 as bonus compensation, however only 389 participants received a monetary bonus. This resulted in an average of payment of $1.70 for these participants. An additional 87 participants started the experiment on M-Turk but failed to complete the task.

Procedure. After participants gave informed consent they filled out some general demographic information. This included their gender (selecting between male and female), age (free response), current yearly income (free response), and highest education attained (8 categories ranging from high school or less up to doctoral degree, taken from the US Census Bureau categories). Then participants completed the Road Salt Task followed by the set of individual difference measures in a randomized order. These measures assessed numeracy, working memory capacity and cognitive reflection. Participants took an average of 45 minutes to complete the experiment, but participants were allowed two hours to complete the experiment.

Road Salt Task. Participants completed the same salting task from Experiment 1 and Experiment 2, however unlike the M-Turk participants in Experiment 1, they were paid commensurate with performance. Participants received $1 for every $2,000 over $12,000 left in their virtual budget at the end of the task.

Individual Difference Measures. Participants also completed the same numeracy assessment (Berlin Numeracy Test; Cokely et al., 2012) from Experiment 1 and Experiment 2, the same working memory capacity task (Automated RSPAN; Unsworth, Heitz, Schrock, & Engle, 2005) from Experiment 2, and the same Cognitive Reflection Test (CRT; Frederick, 2005) from Experiment 2.

Design. For the Road Salt Task, a single factor (forecast format) with four levels: deterministic (single-value forecast), advice, probability, or probability with advice between-participants design was used. Participants were randomly assigned to one of four conditions. The dependent variables were the decision to salt or not, numeric temperature estimate, and trust rating. Individual differences were assessed using the numeracy, working memory capacity and cognitive reflection as predictors for Road Salt Task dependent variables.

Hypotheses

I hypothesized again that participants who were given forecasts that included numerical uncertainty would make better decisions and have greater trust in the forecast, replicating the effects observed in Experiment 2. Moreover, I hypothesized that adding advice to uncertainty estimates would result in fewer risk-seeking errors, replicating the effects observed in Experiment 2. I also hypothesized that greater numeracy would again result in better decisions when given numerical uncertainty information as was observed in Experiment 1 and 2, and may improve performance in general as was observed in Experiment 1 with a similar M-Turk population. Furthermore, I hypothesized that deliberate thinking, and working memory capacity would also result in better decisions when given numerical uncertainty information. Age, and income may also improve decision quality in the condition in which numerical uncertainty was included. It was unclear what effect education would have, since Experiment 1 saw worse decision quality with higher education individuals. However, those individuals were not given any decision incentive, so higher education could result in better decisions when the incentive is included.

Results

The first set of analyses were conducted to determine whether group-level differences between the forecast format conditions observed in previous studies (Grounds & Joslyn, 2015; Joslyn & Leclerc, 2012; Roulston et al., 2006) were replicated here. The subsequent set of analyses then examined if any of the individual differences measures [numeracy, deliberate thinking (CRT), working memory capacity, age, education, and income] predicted road salting task performance in terms of expected value, forecast trust, salting decisions above the optimal threshold, risk-seeking errors, and risk-averse errors within each of the four forecast format conditions.

Using the same initial process from Experiment 1 and Experiment 2, I removed 27 participants (3%) from the analyses below who did not understand the forecast stimuli. An additional 106 participants (13%) did not meet the 85% correct criterion on the semantic judgment in the reading span WMC task and were also removed from all analyses. The remaining 695 were included with 161 participants in the deterministic condition, 163 participants in the advice condition, 195 participants in the probability condition, and 176 participants in the probability with advice condition.

Communication Format Comparisons

Decision Quality. An ANOVA was conducted on expected value scores with forecast format (deterministic, advice probability, and probability with advice) and gender as independent variables. Consistent with the previous experiment, there was a main effect of forecast format, F (3, 687) = 11.94, p < .001. Tukey's post hoc tests revealed that mean expected value in the probability condition (M = -\$1020.23, SD = \$48.93) was significantly better than the deterministic condition (M = -\$1052.53, SD = \$71.60; p < .001, Cohen's d = .53). Additionally, the mean expected value for the probability with advice condition (M = -\$1012.18, SD = \$74.27) was significantly better than the deterministic condition (p < .001, Cohen's d = .55) and the advice alone condition (M = -\$1031.92, SD = \$67.60; p < .05, Cohen's d = .28; see Figure 7). There was again no effect noted for gender.

Trust. A similar pattern emerged when looking at the trust ratings. An ANOVA conducted on the average of the end-of-month trust ratings (5-point scale), with forecast format (deterministic, advice, probability, and probability with advice) and gender as the independent variables, showed a main effect for forecast format, F (3, 687) = 9.14, p < .001. Tukey's post hoc tests revealed that mean trust ratings in the probability condition (M = 3.02, SD = 0.82) were significantly higher than the trust ratings in the deterministic condition (M = 2.66, SD = 0.89; p < .001, Cohen's d = .42) and the advice condition (M = 2.77, SD = 0.89; p < .05, Cohen's d = .29). Moreover, the mean trust ratings in the probability with advice condition (M = 3.08, SD = 0.81) were also significantly higher than the trust ratings in the deterministic condition (p < .001, Cohen's d = .49) and the advice condition (p < .01, Cohen's d = .36; see Figure 8). Gender continued to have no effect.

Binary Decisions. To better understand the decision strategies, the binary decisions were examined more closely. A mixed-model ANOVA was conducted on mean proportion of "salt" responses per participant, with PoF range (above and below 17%) as the within-groups variable and uncertainty forecast format (no uncertainty: deterministic and advice, uncertainty:

probability and probability with advice) and gender as between groups variables. There was a significant main effect for PoF range, F (1, 691) = 5704.95, $p < .001$. Participants salted more above the 17% threshold (M = .66, SD = .17) than below it (M = .17, SD = .20, d = 2.64), suggesting that all participants recognize the probability of freezing is more likely with colder temperatures regardless of receiving probability information. There was also a significant interaction with forecast format, F (1, 691) = 23.84, $p < .001$, suggesting greater differentiation in salt decisions among those who had uncertainty information. Those with uncertainty information (M = .16, SD = .19) salted less often below the 17% threshold than did those with no uncertainty information (M = .20, SD = .21). Additionally, those uncertainty information (M = .67, SD = .16) salted more often above it than did those with no uncertainty information (M = .61, SD = .18; see Figure 9). In this case, there was a between subjects effect of gender, F (1, 691) = 6.90, $p < .01$, suggesting men salted more often below 17% (M = .20, SD = .21) and above 17% (M = .67, SD = .16) than women (M = .16, SD = .20 and M = .65, SD = .17, respectively).

Decision Errors. Finally, decision errors defined in terms of the long-run optimal strategy were examined. A mixed-model ANOVA was conducted on proportions of "errors," with error type (risk-seeking and risk-averse) as the within-groups variable and forecast format (deterministic, advice, probability, and probability with advice) as the between-groups variable. There was a significant main effect for error type, F (1, 687) = 689.20, $p < .001$. Participants made more risk-seeking (M = .60, SD = .26) than risk-averse errors (M = .18, SD = .20, d = 1.81). There was also a significant forecast format interaction, F (3, 687) = 8.06, $p < .001$. Participants with probability information (probability and probability with advice conditions) tended to make fewer risk-averse but those with advice (advice and probability with advice conditions) made fewer risk-seeking errors. There was also a significant gender interaction, F (3, 687) = 8.30, p < .01, suggesting that men made more risk-averse errors than women but women made more risk-seeking errors than men.

These results show that Experiment 3 replicated all of the between-groups effects found in previous studies with an online version of the task in an M-Turk population. Participants made better decisions when given numerical uncertainty information in the form of a probability of freezing. There was greater trust in forecasts that included probability information and greater differentiation of salting decisions above and below the economically optimal threshold. Furthermore, fewer risk-averse errors were made with the probability forecasts, but risk-seeking errors at low probabilities were increased, except when advice was added.

The one difference noted is that for the first time there was an effect of gender when looking at risk propensity. Contrary to all previous studies conducted in this experimental paradigm (where no gender effects were noted) and previous research (in which women tend to be more risk-averse), women made more risk-seeking errors and men made more risk-averse errors. Moreover, men salted more above the economically optimal threshold than women.

Individual Differences. There were no differences in means of any of the individual differences measures between the forecast format conditions. Mean values of predictor and outcome variables for the three forecast format conditions appear in Table 9. For each of the dependent variables, a regression analysis was conducted with numeracy, WMC, CRT, highest education attained[4], age, and current income as predictors in a single step[5]. Regressions were run separately for each forecast format group. Bivariate correlations among the predictors for each forecast format appear in Table 10. The correlation matrices are separated by forecast format as all regressions were conducted separately for each forecast format group. The moderate correlations between the cognitive measures suggest that they likely reflect partially common factors, but may also measure distinct characteristics.

Decision Quality. There were several significant predictors of decision quality, however they varied by forecast format presentation. Numeracy explained a significant amount of variance in expected value scores in all but the deterministic condition (advice, probability, and probability with advice). Income negatively predicted expected value in the two conditions that included advice (advice and probability with advice). Age and education positively predicted expected value in the probability condition and age alone predicted expected value in the probability with advice conditions.

Within the advice condition, the model was significant, F (6, 156) = 2.31, $p < .05$, R^2 = .08. More numeracy led to better expected value scores (t = 2.37, p = .02), and less income led to better expected value scores (t = -2.62, p = .01; see Table 11). The model for the probability condition was also significant, F (6, 188) = 4.30, $p < .01$, R^2 = .16. The results indicated that individuals with higher numeracy had better expected value scores (t = 3.33, p = .001), older individuals had better expected value scores (t = 2.97, p = .003), and more educated individuals had better expected value scores (t = 1.97, p = .04; see Table 11). And finally, the model for the

[4] Highest education was coded as follows: 1 = Some High School or less, 2 = High School Diploma, 3 = Some College, 4 = Associates Degree, 5 = Bachelor Degree, 6 = Masters Degree, 7= Professional Degree, 8 = Doctoral Degree

[5] Interaction terms were included for age, education, and income in initial analyses for Expected Value (as it was the main variable of interest), but did not significantly improve any of the regression models. Thus, only the individual predictor full models are reported.

probability with advice condition was also significant F (6, 169) = 9.74, $p < .001$, $R^2 = .26$. The results indicated that individuals with higher numeracy had better expected value scores ($t = 3.86$, $p < .001$), older individuals had better expected value scores ($t = 2.38$, $p = .02$), and less income led to better expected value scores ($t = -5.17$, $p < .001$; see Table 11). Within the deterministic forecast condition, the overall model was significant, F (6, 154) = 3.10 $p < .01$, $R^2 = .11$. While none of the predictors had a significant effect, both CRT ($t = 1.94$, $p = .054$) and education ($t = 1.92$, $p = .06$) were marginal positive predictors (see Table 11).

Trust. None of the models for the forecast formats significantly explained any variance in trust ratings.

Binary Decisions and Decision Errors. Binary decisions were grouped into three separate analyses: decisions to salt below the 17% threshold (risk-averse error), failing to salt at low (17-23%) probabilities (risk-seeking error), and decisions to salt above the 17% threshold (correct response).

No predictors significantly explained decisions to salt below the economically optimal threshold (risk-averse errors) in any condition. However, risk-seeking errors were significantly explained by the predictors in the conditions that included probability information. Within the probability condition (F (6, 188) = 3.78, $p < .001$, $R^2 = .11$), the results indicated that individuals with higher numeracy made fewer risk-seeking errors ($t = -3.81$, $p < .001$; see Table 13). Within the probability with advice condition (F (6, 169) = 3.90, $p < .001$, $R^2 = .12$), the results indicated that individuals with higher numeracy made fewer risk-seeking errors ($t = -3.23$, $p = .001$) and wealthier individuals made more risk-seeking errors ($t = 2.16$, $p = .03$; see Table 12). No predictors significantly explained risk-seeking errors in the deterministic or advice conditions.

Decisions to salt above the 17% threshold were also significantly explained by the predictors in the conditions that included probability information. Within the probability condition (F (6, 188) = 5.26, p < .001, R^2 = .14), the results indicated that individuals with higher numeracy salted more above the threshold (t = 4.26, p < .001), older individuals salted more above the threshold (t = 2.35, p = .02), and more educated individuals salted more above the threshold (t = 2.15, p = .03; see Table 12). Within the probability with advice condition (F (6, 169) = 7.90, p < .001, R^2 = .22), individuals with higher numeracy salted more above the threshold (t = 4.08, p < .001), and less wealthy individuals salted more above the threshold (t = -3.89, p < .001; see Table 13). No predictors significantly explained decisions to salt above the economically optimal threshold the deterministic or advice conditions.

Thus, in Experiment 3 there were four significant predictors of Road Salting Task performance: numeracy, age, education, and income. *Higher numeracy* led to better decisions in terms of expected value for three forecast format conditions (advice, probability, and probability with advice) as compared to lower numeracy. Higher numeracy also led to more salting above the economically optimal threshold and fewer risk-seeking errors at low probabilities within the probability and probability with advice forecast formats as compared to lower numeracy. More experience in terms of *age* was also associated with greater expected value within the probability and probability with advice forecast formats as compared to less experience (age). More experience in terms of age also led to more salting above the economically optimal threshold within the probability forecast format. More *education* led to better expected value and more salting decisions above the economically optimal threshold in the probability forecast format as compared to less education. *Lower income* led to better expected value for the advice and probability with advice forecast formats, as well as more salting decisions above the economically optimal threshold as compared to higher income. However, it also led to more risk-seeking errors at low probabilities within the probability with advice forecast format.

Discussion

Unlike Experiment 1, which was also conducted on an M-Turk population, Experiment 3 replicated *all* of the effects found in previous studies completed in the laboratory. Including uncertainty estimates led to better decision quality in terms of expected value and higher ratings in trust. Participants who had uncertainty estimates also had greater differentiation of salting decisions between low and high probability forecasts, thus taking more appropriate action when it was necessary, but choosing to accept a risk for unlikely events when it was economically appropriate to do so. Consistent with previous research, participants in the probability condition made more risk-seeking errors at low freeze probabilities for which precautionary action was economically necessary, between 17% and 23%, including explicit advice with the numeric uncertainty estimates reduced these errors. The benefit of combining the two might be explained by increased forecast trust (including trust in the advice) due to the uncertainty information. Once participants trust the advice, it helps them overcome the natural tendency to not salt at low probabilities.

Experiment 3 marks the first instance of observing all of the effects in an online format. In contrast, Experiment 1 did not include a monetary bonus, and it resulted in no difference in decision expected value between probability, advice, and deterministic forecast formats. This suggests that the tendency to utilize numerical uncertainty information in the decision-making process may be contingent on extrinsic motivation: in this case, a monetary bonus for preforming well. This is an issue that deserves further research.

Again, consistent with my hypothesis, higher numeracy resulted in better decisions in terms of expected value when participants were given numeric uncertainty information, as higher numeracy was associated with better expected value scores for the probability and the probability with advice conditions. Higher numeracy also led to more salting above the economically optimal threshold and fewer risk-seeking errors at low probabilities within the probability and probability with advice forecast formats. This contributes to the growing body of empirical evidence that an understanding of numbers and numerical information is associated with making good use of numerical uncertainty information. Moreover, the amount of variance explained in the probability inclusive conditions was substantial and similar to the amount of variance explained within all forecast formats in Experiment 1 and within the probability forecast format in Experiment 2, which was the only significant forecast format that had significant regression models.

In addition, higher numeracy was associated with better decision quality for the advice condition as was observed in Experiment 1. This suggests that higher numeracy may have a more general benefit, also helping people realize and understand the benefit of the advice in the long run. However, the amount of variance explained within the advice condition was less than the probability inclusive conditions in this experiment or previous experiments, as the amount of variance explained translates into a medium effect size. This suggests the contribution of individual predictors was less meaningful in the advice condition.

Greater experience (in terms of age) had a similar benefit. It resulted in better decision quality (expected value) for the probability and probability with advice forecast conditions and more salting decisions above the 17% threshold in the probability forecast condition. This could perhaps be explained with the accumulation of experience with weather forecasts and outcomes over a longer lifetime. However, as mentioned in the discussion of Experiment 1, research investigating aging typically assesses respondents above an age threshold, such as 60 years of age or older. If I were to use that same threshold in this sample, I would only have 30 participants (4% of sample collected, range: 60-75). A data set that small is not enough to draw any meaningful conclusions. Thus, the benefit of age is limited to middle-aged (and slightly older) participants.

The results of Experiment 3 also identified a sub-group of users that benefited from a combination of probability and explicit advice. Lower income individuals had better decision quality (expected value) when they were given the advice and the probability with advice forecast formats. Lower income individuals also made more salting decisions above the economically optimal threshold and fewer risk-seeking errors at low probabilities when the advice was paired with probability information. Perhaps the advice alerted them to situations in which they would not normally take precautionary action or the potential penalty associated with choosing not to salt loomed larger for these lower income individuals so they followed advice that often indicated precautionary action was necessary.

The present study was also the first experiment to find gender effects for measures of the road salting task. Women made more risk-seeking errors by failing to salt at low freeze probabilities above the 17% chance of freezing threshold in the 17-23% probability range and men made more risk-averse errors by salting at unnecessarily low freeze probabilities below the threshold. Moreover, men salted more above the economically optimal threshold (17-51% probability range) than women. This is surprising as gender differences have not typically been reported in studies that test risk communication in both women and men as one review noted (Fagerlin, Zikmund-Fisher & Ubel, 2011) and gender effects have never been observed in weather related decision research from this lab (Grounds, Joslyn, & Otsuka, in prep; Savelli & Joslyn, 2013, Simianu et al., under review). This is also surprising because when gender differences are noted, women tend to be more risk-averse then men (Byrnes, Miller, & Schafer, 1999; Eckel & Grossmann, 2002). However, the gender differences observed in Experiment 3 did not lead to overall differences in decision quality, as there were no gender differences for expected value analyses. Perhaps this is simply an anomaly or perhaps there are some situations in which women adopt slightly different strategies than men, but the strategies have no meaningful impact on overall decision quality, as has been demonstrated in other financial decision-making research (Powell & Ansic, 1997). Women took a riskier approach at low probabilities, which resulted in better decision quality below the economically optimal threshold. However, the riskier strategy led to worse decision quality for probabilities just above the optimal threshold. This could be further explored in future research.

Chapter 5: Final Research Questions

This chapter includes two sets of analyses that attempt to examine the mutual implications of the three studies reported here. The research focus thus far has been to ask if any factors predict *better* decision making. The direct consequent of this research question is that once a factor has been identified (e.g. numeracy), then lower scores on that factor are associated with *worse* decision making within a particular forecast format. However, from a practical perspective, what I would really like to know is whether one communication format is better or worse for this group than other communication formats. This question is explored using lower numeracy, younger, lower education, and higher income user groups. I focused on these specific user groups because previous regression analyses indicated they made worse decisions with numerical uncertainty or advice, as compared to higher numeracy, older, higher education, and lower income users.

Are Any Groups Harmed by Uncertainty Information?

Higher numeracy was associated with better expected value within the deterministic condition (Experiment 1), advice condition (Experiment 1 and Experiment 3), probability condition (Experiment 1, Experiment 2, and Experiment 3), and the probability with advice condition (Experiment 3). Thus it is possible that probability information or advice could cause low numerate individuals to make worse decisions than if they were given a single value forecast. To explicitly test this possibility, the numeracy scores from all three experiment populations were combined into one global data set. The lowest 25% of numeracy scores in the global set was calculated to determine a low numeracy population (numeracy scores less than or equal to 2). A one-way ANOVA was then conducted on expected value scores with forecast format as the between subjects variable for this subset of participants in each experiment. The forecast formats tested in Experiment 1 were deterministic, advice, and probability. Experiments 2 and 3 tested the same formats but also included probability with advice. Low numeracy individuals had statistically equivalent expected value scores regardless of communication format in all three experiments (Experiment 1: F (2, 147) = 0.84, p = .43, Experiment 2: F (3, 54) = 1.80, p = .16, Experiment 3: F (3, 233) = 2.54, p = .06). No pairwise comparisons between forecast formats were significant within the three experiment, however within Experiment 3, low numeracy individuals who had the probability alone forecast (M = -\$1035.41, SD = \$49.27) had marginally better expected value scores than those who had the deterministic forecast (M = -\$1070.48, SD = \$85.32; p = .06, Cohen's d = .50). This suggests that although explicit numeric probabilities do not necessarily help those with low numeracy (although the trend is in that direction) as they do the larger group, they do not hurt expected value scores either (see Figure 10).

Greater experience as measured by age was also associated with better expected value scores within the deterministic condition in Experiment 1 and within the probability and probability with advice conditions in Experiment 3. Therefore, it is also possible that younger individuals would not benefit from probability information and instead would make better decisions when given advice. This was tested by calculating the lowest 25% of age in a global set of Experiment 1 and Experiment 3 to determine a younger population (age less than or equal to 26). A one-way ANOVA was conducted on expected value scores with forecast format as the between subjects variable for this subset of participants in each experiment. Younger individuals had statistically equivalent expected value scores regardless of communication format in both experiments (Experiment 1: F (2, 68) = 0.31, p = .74, Experiment 3: F (3, 167) = 0.79, p = .50), suggesting that although explicit numeric probabilities do not necessarily help younger individuals in these experiments, as they did the larger group, they did not hurt expected value scores either (see Figure 11).

Greater education was also associated with better expected value scores within the deterministic and probability conditions in Experiment 3, thus it is possible that less educated individuals would benefit from a communication method that includes advice. To directly test this, a one-way ANOVA was conducted on expected value scores with forecast format as the between subjects variable for participants who had a high school diploma or less education. Less educated individuals had better expected value scores when given additional information than just the single value forecast (F (3, 55) = 4.11, p = .01). Less educated individuals who were given a deterministic forecast (M = -\$1115.31, SD = \$123.69) made worse decisions than those given advice (M = -\$1018.56, SD = \$42.51; p < .05, Cohen's d = 1.05), probability information (M = -\$1034.47, SD = \$56.99; p < .05, Cohen's d = .84), or probability information with advice (M = -\$1029.75, SD = \$67.05; p < .05, Cohen's d = .86). This suggests that numeric probabilities or explicit advice (or a combination of the two) can help less educated individuals (see Figure 12). Thus, the advantage of probability inclusive forecasts is consistent for lower levels of education as well as all levels of education, but less educated also benefit from advice alone.

A separate group of users were identified who made worse decisions when they were given an explicit recommendation (advice condition). Higher income participants in Experiment 3 had worse expected value scores (as compared to lower income participants) in terms of expected value when they were given advice, even if the advice also included numerical uncertainty information. Thus, it is also possible that receiving advice could cause higher income individuals to make worse decisions. To directly test this possibility, the highest 25% reported income level of participants from Experiment 3 was calculated (yearly income greater than or equal to \$50,000). A one-way ANOVA was then conducted on expected value scores with forecast format as the between subjects variable for this subset of participants. Although the results were borderline significant (F (3, 185) = 2.60, p = .054), there were no differences in expected value scores between the advice condition (M = -$1029.26, SD = $61.24) and any other forecast format. This suggests that advice was not detrimental to high income participants. Moreover, high income participants who were given probability information alone (M = -$1016.28, SD = $37.78) had better expected value scores than high income participants who were given a deterministic forecast (M = -$1054.54, SD = $72.02; p < .05), suggesting numerical uncertainty is still beneficial to these users (see Figure 13).

Taken together, these results suggest that decision quality as measured by expected value is not impaired when given probability information, as compared to less complex forms of communication. This is true for users who have low scores on standard numeracy tests or younger individuals on M-Turk. In fact, in many cases having explicit uncertainty information helps, including those who lack higher education, as well as users who have relatively higher income. Additionally, while lower education users also benefited from explicit advice, no groups did worse when advice was paired with probability information. This suggests the best communication format for all users is one that provides both probability information and advice.

Furthermore, the consistent finding across the body of research covered in this dissertation was that higher numeracy helped individuals make better decisions which resulted in better expected value when probability information was include (Experiment 1, Experiment 2, and Experiment 3), but also at times when it was not (Experiment 1 and Experiment 3). However, other predictor factors were also included in these models and were often not significant. What I wanted to know next was whether the contribution of numeracy to expected value was greater when decisions were based numerical uncertainty information than in other formats. In other words, is numeracy as important when given advice or a single value forecast as when given probabilistic information? This question is explored by conducting regression analyses for expected value scores with numeracy as the only predictor for each experiment data set.

Is Numeracy More Important When Given Probabilistic Information?

Better numeracy predicted better decision quality when given numerical uncertainty forecast in all three experiments, which suggests that better numeracy promotes better understanding of numerical uncertainty. However, better numeracy also predicted better expected value scores in communication formats that did not include numerical uncertainty in some experiments, although not consistently. This suggests that the contribution of numeracy might be more than just better understanding of a probabilistic forecast. It is possible that more numeracy might help individuals better estimate the inherent uncertainty in weather forecasts regardless of the additional information provided or be better understand the implications of the cost loss ratio to employ an economical rational salting strategy and thus, might explain expected value scores to the same degree regardless of forecast format.

To address this issue, the contribution of numeracy to expected value scores across conditions in each experiment was compared. Regression analyses were conducted for each experiment again, using numeracy as the only predictor of expected value. The unstandardized coefficients, 95% confidence intervals, standardized coefficients, and t-values for numeracy appear in Table 14. In Experiment 1, numeracy significantly predicted expected value to a similar degree for all three forecast formats tested (Deterministic: F (1, 80) = 12.51, p = .001, R^2 = .14; Advice: F (1, 79) = 18.38, p < .001, R^2 = .19; Probability: F (1, 84) = 16.08, p < .001, R^2 = .16).

The confidence intervals for numeracy overlap between forecast formats, suggesting numeracy is predicting expected value to a similar extent for all forecast formats.

A similar pattern emerged for Experiment 2, however the deterministic forecast model was not significant (Deterministic: F (1, 88) = 0.01, p = .92, R^2 = .00; Advice: F (1, 78) = 4.72, p = .03, R^2 = .06; Probability: F (1, 80) = 14.83, p < .001, R^2 = .17; Probability with Advice: F (1, 85) = 6.24, p = .01, R^2 = .07). As these confidence intervals also overlap, this also suggests numeracy alone predicts expected value to a similar extent for all conditions except the deterministic condition. Note here numeracy alone predicted expected value within the advice and probability with advice conditions, but the corresponding models for expected value reported in Chapter 3 above were not significant when other predictors were included (cognitive reflection, working memory capacity, fluid intelligence, need for cognition, and financial risk propensity). In these analyses, the regression coefficient for numeracy was somewhat near significance. Thus, the removal of the additional predictors pushed the model into significance.

Returning to the numeracy only regression analyses, similar pattern also emerged for Experiment 3; numeracy predicted expected value scores to a similar degree for all forecast formats tested (Deterministic: F (1, 159) = 7.14, p < .01, R^2 = .04; Advice: F (1, 161) = 5.80, p = .02, R^2 = .03; Probability: F (1, 193) = 12.77, p < .001, R^2 = .06; Probability with Advice: F (1, 174) = 21.67, p < .001, R^2 = .11). Note here numeracy alone predicted expected value in the deterministic condition but did not when cognitive reflection, working memory capacity, education, age, and income were included (reported in Chapter 4 above), which had cognitive reflection and education as the only marginally significant predictors. However, numeracy and cognitive reflection were correlated in the deterministic condition sample, r =.47, p <.001 (see

Table 10). This suggests that numeracy alone was accounting for some variance that was better explained by cognitive reflection once it was included in the regression model.

Taken together, these results demonstrate that the contribution of numeracy to decision quality is not unique to forecasts formats that include probability. This suggests that the advantage of higher numeracy is more than just better understanding of numerical uncertainty. Thus, higher numeracy might also help individuals better estimate the inherent uncertainty in weather forecasts regardless of the additional information provided or appreciate the cost loss ratio to employ a more economically rational decision strategy that allows them to take precautionary action in a situation when such action is required in low likelihood situations.

Chapter 6: General Discussion

The main goal of this research project was to understand how best to communicate weather uncertainty to users. In particular, I wanted to know whether it was necessary to tailor the message to specific user groups. This set of experiments generated compelling evidence for the overall advantage of forecasts that include numerical uncertainty, irrespective of the user group. However, this is not to say individual differences do not matter. These experiments also demonstrated that individuals with higher numeracy make better decisions in a loss-based situation, but that forecast format did not interact with this finding.

First, results replicated previous research by demonstrating that inclusion of numerical uncertainty estimates promotes better decision quality as compared to providing explicit decision advice or a deterministic forecast as a means to promote better decision quality for a general user group. The inclusion of numerical uncertainty also leads to greater trust in the information communicated and greater differentiation between likely and unlikely outcomes as compared to both advice and deterministic forecasts. Additionally, numeric estimates combined with decision advice leads to better decisions for low probability events for which precautionary action is economically optimal and the inclusion of advice is not detrimental in any situations or for any user groups. Numeric uncertainty estimates, therefore, are an effective way to communicate weather uncertainty and when combined with advice, they lead to better decisions overall.

However, not all users can benefit from numerical uncertainty forecasts equally. Low numeracy individuals did not benefit from forecasts including numerical uncertainty, as expected value scores were no different for low numeracy individuals based on forecast format. This suggests that perhaps there is a minimum amount of numeracy that is required to see the benefit for probabilistic information. However less educated and higher income individuals were helped by numerical uncertainty suggesting that it is advantageous for many user groups.

One reason numerical uncertainty information is often not provided with weather forecasts is that many atmospheric scientists think that it is too confusing for certain parts of the population, perhaps those who are less numerate or those who have less education. The fear is that this subset of users would make worse decisions if they had to deal with numerical uncertainty. However, it is most encouraging that individuals who score low on standard numeracy tests are not negatively impacted by numerical uncertainty. The quality of decisions made by less numerate individuals is no different whether probabilistic information is included or excluded. Thus, this research suggests that there is no harm in including numerical uncertainty in forecasts, even if it doesn't benefit all users equally. And while it was not part of the scope of this project, numeracy skills can be improved. Much attention has been focused on interventions to improve numeracy skills of children that focus on progress monitoring, number comparison, and understanding proportions (e.g. Bryant et al., 2011; Holmes & Dowker, 2013; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009). However, the numeracy skills of adults can be improved as well (Coben, 2003; Warburton & Kahn, 2007). Some of the techniques advocated for adult learning are learning to recognize when and where mathematical concepts can help problem solving or decision making, and applying or adapting mathematical concepts to novel situations (Coben, 2003). Thus, it is possible that adults can improve their mathematical abilities and make better use of numerical uncertainty.

In addition, higher numeracy helped people make better decisions when they were given probability information in all three experiments. However higher numeracy also led to better decisions for people given explicit advice, although this effect was not consistent across the three experiments. It was observed in Experiment 1 and Experiment 3, the experiments that used an M-Turk sample, but not Experiment 2 with college students. This suggests that the benefit of numeracy might extend beyond the ability to reason with probability information. Perhaps, it helps people recognize the usefulness of a long-run optimal strategy, because even though the advice may not match the observed outcome every time, a small chance of a big loss is enough to necessitate action. However, when all other predictors were removed, numeracy also significantly predicted expected value scores for the deterministic forecast format in Experiment 1 and Experiment 3. This suggests the benefit of numeracy could be something even more general, such as helping individuals better appreciate the cost/loss ratio of the task to employ a more economically rational decision strategy. This would allow them to take precautionary action in a situation when such action is required in low likelihood situations. In addition, increased numeracy could also help individuals better estimate the inherent uncertainty in weather forecasts allowing them to estimate the probability of freezing when it is not provided by realizing that it is closely related to nighttime low temperature. This should be further explored to determine why it was observed in a general population but not in a more numerate college sample (see Tables 2, 5, and 9 for mean numeracy scores of dissertation participants).

While numeracy was consistently a factor that explained expected value scores, there were demographic factors such as education and age that also explained expected value scores. However, they were inconsistent or even contradictory between the different experiments. More education led to worse decisions when there was no decision incentive in Experiment 1 in the deterministic and probability forecast format conditions. Though when the incentive was reinstated in Experiment 3, more education led to better decisions in the deterministic and probability forecast format conditions. Perhaps the decision incentive was necessary to motivate quality task performance for these individuals. Most real-world weather related decisions include some sort of decision incentive, and in extreme cases, the incentive can be life or death. Since more education led to better decisions in the task version that is more closely aligned with real world scenarios, it seems that is a better depiction of the relation between education and decision quality. However, the bottom line is that uncertainty communication does not need to be tailored to different regions or groups of people, as providing numeric uncertainty information results in better decisions or no difference in decision quality than when it is omitted.

There were inconsistent effects regarding the influence of age on expected value scores across experiments. Older individuals made better decisions than younger individuals with only a single value forecast when no decision incentive was offered in Experiment 1, however older individuals made better decisions than younger individuals with numerical uncertainty information when the decision incentive was present in Experiment 3. Although, what was considered “older” in these studies was quite a bit younger than most age-related decision research (e.g., Mutter & Pliske, 1996; Thornton & Dumke, 2005). However, this research was not specifically investigating aged or elderly decision-making. Only 3% of participants in Experiment 1 and 4% of participants in Experiment 3 were 60 years of age or older, and there were no participants over the age of 75 in either study. Consequently, while these results are encouraging, they cannot be extended to elderly populations.

However, younger individuals did not receive the same benefit from probability forecasts as older individuals, as expected value scores were no different for younger M-Turk individuals based on forecast format. This in in contrast with the advantage of probability forecasts in the young college student of Experiment 2. Perhaps the advantage of education helped college students benefit from probability forecasts, although there was no significant interaction between age and education for models of expected value in Experiment 1 or Experiment 3. Perhaps instead a specific *type* of education that is focused on analytical reasoning, as the psychology major is, can help younger individuals make better decisions with probability forecasts. Therefore, future studies should also identify the major or focus of study along with level of education to help resolve this discrepancy.

This research also identified a group of users that benefited from explicit advice, although it was only observed in one experiment, and thus should be replicated to allow for strong conclusions. In Experiment 3, I found that lower income was associated with increased advice compliance compared to higher income and resulted in better expected value scores. This is in contrast with field research which suggests that greater income is positively associated with greater compliance for hurricane evacuation orders (Peacock, Morrow, & Gladwin, 1997). However, this evidence based on field work may be due to the fact that wealthier individuals in the real world have more financial resources available to them that can facilitate evacuation, whereas in the experimental task, all participants were given the same initial financial resources. Further research should be conducted to determine the reliability of income effect. It is important to note however, that higher income participants who were given advice had similar decision quality in terms of expected value scores as compared to higher income participants with the deterministic forecast, suggesting advice was not detrimental to these users. Furthermore, higher income participants who were given probability information had better expected value scores than higher income participants who were given a deterministic forecast, suggesting numerical uncertainty is also beneficial to these users. Perhaps higher income individuals were reluctant to comply with advice that recommended additional spending but were able to use the probability forecasts to make better decisions based on their own risk tolerance.

Some results were inconsistent with past research however. In Experiment 3, women were more risk-seeking and men were more risk-averse. These results are largely inconsistent with past research suggesting that women are more risk-averse and men more risk-seeking, regardless of content frames or concreteness of situation (Byrnes, Miller, & Schafer, 1999; Eckel & Grossmann, 2002). This has also been observed in weather field studies. Women are more likely to comply with hurricane evacuations than men (Peacock, Morrow, & Gladwin, 1997). Few studies have found women to be more risk-seeking, and that was in the domain of health decisions such as smoking and sharing needles (Wayment, Newcomb, & Hannemann, 1993; Zuckerman, Ball, & Black, 1990). I do not have a good explanation for why women were more risking seeing in the road salting task in only one experiment, and thus believe the effect should be replicated again before interpretation.

Somewhat surprisingly, none of the factors assessed across this dissertation explained forecast trust within any of the forecast formats tested. In every experiment reported here, participants had greater trust in forecasts that included uncertainty information. This suggests that trust in weather forecasts is not dependent on cognitive factors such as numeracy, working memory capacity, cognitive reflection, fluid intelligence, or need for cognition. It also suggests forecast trust is not dependent on the specific demographic factors measured here: age, education, or level of income. Other demographic factors such as political ideology or beliefs about science may influence forecast trust but they were not assessed here.

It was also surprising that there were factors measured that did not seem to predict decision quality and decision errors. More specifically, working memory capacity and fluid intelligence did not predict any measures of decision quality for any of the uncertainty formats tested. This suggests that reasoning with uncertainty information, at least in the simplified tasks used in this experimental paradigm, is not overly taxing for users. In other words, the multiple components of the forecast did not prevent users who scored lower on measures to assess working memory capacity and fluid intelligence from making good use of uncertainty information as compared to users who scored higher on those same measures. However, it is important to note the decisions asked of participants were simplified versions of the real-world task; they only had to choose between two alternatives and they had to evaluate only one variable, temperature. An actual road-salting decision would include numerous factors besides the temperature forecast, not the least of which is the forecast for air moisture, as there cannot be ice on roads without moisture to freeze. Thus, factors such as working memory capacity or fluid intelligence may influence decision quality in more complex situations in which there are more alternatives or multiple variables must be compared. Hence, a limitation of the present research is that the road salting task is a simplified version of the decision it conceptually represents.

Additionally, the consequences faced by real-world decision makers might not be represented adequately by the decision incentive of the road salting task. The possibility of payment for good performance was intended to represent a realistic consequence, but it is quite possible that such small rewards did not represent the gravity of the real-world decision consequences. Whereas the real consequence to participants in the road salting task was the possibility of winning money, the real consequence to decision makers in weather emergencies is personal safety. Thus, another limitation of the present research is the degree to which the results can be generalized to real-world situations. While personal safety might represent greater incentive, it may be more difficult to take precautionary action in real situations as other factors besides monetary resources must be considered. People may be reluctant to leave behind a treasured pet and instead choose to take a risk. Nevertheless, the results reported here replicate patterns of behavior demonstrated in actual weather emergencies (e.g. not taking precautionary action when it is advised), suggesting results are applicable to real-life scenarios.

An unanticipated benefit of this project was to highlight the importance of decision incentives commensurate with performance to motivate quality task performance. When participants were given a flat fee payment for their participation (Experiment 1), there was no difference in decision quality in terms of expected value scores between the different forecast communication formats. However, when participants were paid contingent on their final balance (Experiment 2 and Experiment 3), participants provided with probabilistic information made better decisions, suggesting that incentives are important in encouraging participants to take full advantage of what may be more challenging decision relevant information. Thus, decision incentives may be important in this line of research because they represent the motivation real-world decision makers experience, influencing them to make the best decision amongst alternatives in a situation. Performance based incentives may also be important for more basic psychological research as they are more likely to reveal optimal performance in reasoning with probabilities and frequencies as compared to course credit or flat fee payment (Brase, 2009). Further research along these lines should, if possible, include performance based incentives to elicit quality participant performance.

Generalizability can be a major concern for many researchers. If results are only valid for the population sampled, then the theoretical importance of any effect can be diminished. This research project provided a unique opportunity to assess the same set of cognitive abilities in a very common research population (college students) and a general population of all age, education, and experience levels (M-Turk). I was then able to use this same set of cognitive predictors (numeracy, working memory capacity and need for cognition) to determine how they relate to decision quality in terms of expected value scores for each population (see Appendix D). Interestingly, there was no difference in the amount of variance explained between the two populations or between individual regression coefficients. Moreover, the predictive power of each predictor variable was not significantly different between populations. This is a hopeful result, as it suggests that data collected using a college student sample is similar to a broader population. However, M-Turk, while more representative of a general population, is not sampling a truly general population. The workers on M-Turk have self-selected into this participant pool, have the financial means to utilize a computer with internet access, and are generally more educated than the national average (see Table 1 and Table 9) and younger. Nevertheless, these participants also replicate patterns of behavior demonstrated in actual weather emergencies, suggesting they are representative enough.

Overall, this research, and hopefully more like it in the future, provides evidence for the way humans respond to different forecast formats. This evidence is clearly relevant to how weather uncertainty is communicated to the public. This type of research is critical as high-quality information is increasingly available, but without a framework for communicating it effectively to users its value cannot be realized.

Much of this research is pioneering in the field. Although the topic of effective risk communication for weather-related decision making is of utmost importance to saving lives there is little experimental evidence of the person-level factors affecting risk comprehension. Moreover, very little past research has systematically measured and manipulated these factors in an experimental setting. The research reported here represented a substantial contribution to our understanding of how numerous factors, such as numeracy, age, education, and income affect the decisions people make when faced with weather uncertainty. Increased numeracy was associated with better weather-related decisions when forecasts included numerical uncertainty estimates in all three studies, although no user groups were substantially impaired when given numerical uncertainty. Increased numeracy was also associated with better expected value scores when forecasts included advice or just the deterministic forecast. This suggests that the benefit of numeracy is more than a better understanding of probability information. Perhaps numeracy helps users make better decisions in this challenging situation where precautionary action is necessary at low probabilities. Further research and replication of the results reported herein is necessary, but the evidence from this set of experiments is a meaningful start.

The results of this research lead to some recommendations that should be useful to weather forecasters and emergency managers who need to communicate weather uncertainty and risk. It suggests that weather forecasts should include numerical estimates of uncertainty, even for scenarios that involve a large potential loss. This way many users will be better able to understand their risk, trust the forecast and hopefully take necessary precautionary action more often. Users will also continue to trust forecasts that include numerical uncertainty, even when a forecasted large loss does not occur. Likewise including explicit advice is useful for situations wherein precautionary action should be taken, but the likelihood of an adverse event is quite low. The advice should help users by highlighting the importance of precautionary action, which is particularly important for some groups, while acknowledging forecast uncertainty. And even though not all users will benefit equally from numerical uncertainty information, it doesn't make their decisions worse than if they hadn't received any numerical uncertainty information.

An interesting extension of the present research would be to explore the influence of cognitive and demographic factors on decisions made using representations of uncertainty beyond categorical and numerical. Graphical representations can provide a quick "gist" summary of uncertainty or risk for decision makers and are especially useful for individuals with lower numeracy (Galesic, Garcia-Retamero, & Gigerenzer, 2009). However, some recent research suggests that users have a tendency to misinterpret a visual representation of a predictive interval (Grounds, Joslyn, & Otsuka, in prep) that may extend to uncertainty information in general. However, at this point it is unclear how factors such as numeracy and working memory capacity would influence interpretation of visual expressions of uncertainty.

Another interesting extension of the current research would be to change the decision structure of the task. The Road Salt Task is a loss scenario, so the advantage of numeracy for all forecast formats may be confined to the domain of losses, where numeracy helped overcome the tendency to be risk-seeking. Alternatively, numeracy could also help overcome the tendency to be risk-averse in a gain scenario, but it is unclear if numeracy would help to the same extent. Future research, therefore, should systematically measure these variables to provide further clarity to decision making under uncertainty using other task frames.

In conclusion, the research reported herein offers an optimistic view of our ability to make decisions when faced with uncertainty. It provides a clear line of evidence that adding numeric uncertainty estimates to weather forecasts helps most people to make better decisions. Moreover, tailoring communication to different user groups may not be necessary, as no groups did worse when given numeric uncertainty information. Moreover, I have shown that numeric uncertainty is best when paired with decision advice that promotes action at low, but necessary, probabilities and potentially helps some user groups, such as low income users. Ultimately, my research shows that emergency managers could better communicate weather-related uncertainty with numeric estimates. Many people would have greater trust in the forecast, make better decisions, and lives may be saved. Furthermore, the pattern of results demonstrated might be applicable to other domains, such as finance or medicine. Professionals from both domains could benefit from increased user trust associated with numerical expressions of risk, and users themselves could benefit from better differentiation between likely and unlikely outcomes.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs?. *Psychological bulletin*, *131*(1), 30-60.

Andersson, O., Tyran, J. R., Wengström, E., & Holm, H. J. (2013). Risk aversion relates to cognitive ability: Fact or fiction?. Working Paper No. 2013:9. Department of Economics, Lund University.

Anker, J., Senathirajah, Y., Kukafka, R., Starren, J. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*, 608-618.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829-839.

Baker, E. J. (1995). Public response to hurricane probability forecasts. *The Professional Geographer, 47*, 137-147.

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics, 112,* 537-579.

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1), 33-47.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*(3), 382-398.

Brase, G. L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, *15*(5), 381-401.

Brase, G. L. (2009). How different types of participant payments alter task performance. *Judgment and Decision Making*, *4*(5), 419-428.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, *78*(1), 7-23.

Budescu, D. V., Por, H. H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic change*, *113*(2), 181-200.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science*, *6*(1), 3-5.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological bulletin*, *125*(3), 367-383.

Cacioppo, J. T., Petty, R. E., & Sidera, J. A. (1982). The effects of a salient self-schema on the evaluation of proattitudinal editorials: Top-down versus bottom-up message processing. *Journal of Experimental Social Psychology*,*18*(4), 324-338.

Cattell, R. B. (1971). Abilities: their structure, growth, and action. Oxford: Houghton Mifflin.

Coben, D. (2003). *Adult numeracy: review of research and related literature.* London: National Research and Development Centre for Adult Literacy and Numeracy.

Cohn, R. A., Lewellen, W. G., Lease, R. C., & Schlarbaum, G. G. (1975). Individual investor risk aversion and investment portfolio composition. *The Journal of Finance*, *30*(2), 605-620.

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.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*(1), 20-33.

Connelly, N. A., & Knuth, B. A. (1998). Evaluating risk communication: Examining target audience perceptions about four presentation formats for fish consumption health advisory information. *Risk Analysis*, *18*(5), 649-659.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Corbin, J., McElroy, T., & Black, C. (2010). Memory reflected in our decisions: Higher working memory capacity predicts greater bias in risky choice.*Judgment and Decision Making*, *5*(2), 110-115.

Deakin, J., Aitken, M., Robbins, T., & Sahakian, B. J. (2004). Risk taking during decision-making in normal volunteers changes with age. *Journal of the International Neuropsychological Society*, *10*(04), 590-598.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*(1), 284-294.

Diefenbach, M. A., Weinstein, N. D., & O'Reilly, J. (1993). Scales for assessing perceptions of health hazard susceptibility. *Health education research*, *8*(2), 181-192.

Donati, M. A., Panno, A., Chiesi, F., & Primi, C. (2014). A mediation model to explain decision making under conditions of risk among adolescents: the role of fluid intelligence and probabilistic reasoning. *Journal of clinical and experimental neuropsychology*, *36*(6), 588-595.

Dow, K., & Cutter, S. L. (2000). Public orders and personal opinions: Household strategies for hurricane risk assessment. *Global Environmental Change Part B: Environmental Hazards*, *2*(4), 143-155.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*,*23*(4), 281-295.

Eiser, J. R., 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.

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: ten steps to better risk communication. *Journal of the National Cancer Institute*, *103*(19), 1436-1443.

Fausset, C. B., & Rogers, W. A. (2012, September). Younger and older adults’ comprehension of health risk probabilities: Understanding the relationship between format and numeracy.

In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1),120-124.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25-42.

Friedman, H. (1982). Simplified determinations of statistical power, magnitude of effect and research sample sizes. *Educational and Psychological Measurement*, *42*(2), 521-526.

Fuller, R., Dudley, N., & Blacktop, J. (2001). Risk communication and older people—understanding of probability and risk information by medical inpatients aged 75 years and older. *Age and Ageing*, *30*(6), 473-476.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*,*28*(2), 210-216.

Galotti, K. M. (1999). Making a" major" real-life decision: College students choosing an academic major. *Journal of Educational Psychology*, *91*(2), 379-387.

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and decision making*, *9*(1), 15-34.

Gibbs, L., & Holloway, C. (2013). Hurricane Sandy after action: report and recommendations to Mayor Michael R. Bloomberg. *Hurricane Sandy After Action: Report and Recommendations to Mayor Michael R. Bloomberg, The City of New York, New York, NY*, *36*.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: from innumeracy to insight. *BMJ: British Medical Journal*, *327*, 741-744.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*(4), 684-704.

Gneiting, T., & Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science*, *310*(5746), 248-249.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213-224.

Grable, J. E. (2000). Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of Business and Psychology*,*14*(4), 625-630.

Gray, S. J., & Gallo, D. A. (2016). Paranormal psychic believers and skeptics: a large-scale test of the cognitive differences hypothesis. *Memory & cognition*, *44*(2), 242-261.

Gregory, N. (1980). Relative wealth and risk taking: A short note on the Friedman-Savage utility function. *The Journal of Political Economy*, 1226-1230.

Grounds, M. A., Joslyn, S., & Otsuka, K. (2016). *Understanding and using probabilistic forecasts: individual differences and education.* Manuscript in preparation.

Gruntfest, E., Downing, T., & White, G. F. (1978). Big Thompson Flood. Working Paper No. 32, Institute of Behavioral Science, Univ. of Colorado, Boulder, CO.

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.

Hastings, J. S., & Mitchell, O. S. (2011). How financial literacy and impatience shape retirement wealth and investment behaviors. Working Paper No. w16740. National Bureau of Economic Research.

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*(2), 217-240.

Hepler, T. J., & Feltz, D. L. (2012). Take the first heuristic, self-efficacy, and decision-making in sport. *Journal of Experimental Psychology: Applied*, *18*(2), 154-161.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 298-306.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic medicine*, *73*(5), 538-540.

Holmes, W., & Dowker, A. (2013). Catch up numeracy: a targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education*, *15*(3), 249-265.

Hubbard, D. W. (2007). *How to measure anything: Finding the value of" intangibles" in business*. John Wiley & Sons.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to nontechnical people. *Risk analysis*, *7*(4), 519-529.

Ipeirotis, P. (2010). Demographics of Mechanical Turk. *CeDER-10–01 working paper*, New York University.

Jasper, J. D., Bhattacharya, C., Levin, I. P., Jones, L., & Bossard, E. (2013). Numeracy as a predictor of adaptive risky decision making. *Journal of Behavioral Decision Making*, *26*(2), 164-173.

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.

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.

Joslyn, S., & LeClerc, J. (2013). Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, *22*(4), 308-315.

Joslyn, S., & Savelli, S. (2010). Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*,*17*(2), 180-195.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697-720.

Kahneman, D. (2011). Thinking, fast and slow. New York: Farrar, Straus, and Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, *47*(2), 263-291.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66-71.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189-217.

Klein, G. (1999). *Sources of power: How people make decisions*. Boston: MIT press.

Knight, F. H. (1921). *Risk, uncertainty and profit*. New York: Hart, Schaffner and Marx.

Knight, J., Weir, S., & Woldehanna, T. (2003). The role of education in facilitating risk-taking and innovation in agriculture. *The Journal of Development Studies*, *39*(6), 1-22.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*,*16*(2), 77-92.

LeClerc, J., & Joslyn, S. (2015). The Cry Wolf Effect and Weather-Related Decision Making. *Risk analysis*, *35*(3), 385-395.

Levin, I. P., Snyder, M. A., & Chapman, D. P. (1988). The interaction of experiential and situational factors and gender in a simulated risky decision-making task. *The Journal of Psychology*, *122*(2), 173-181.

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making*, *27*(5), 696-714.

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.

Mason W., & Watts D. J. (2009). Financial incentives and the performance of crowds. *Proceedings of the Human Computation Workshop: Knowledge Discovery and Data Mining Conference*, Paris, France.

Mata, R., Josef, A. K., & Hertwig, R. (2016). Propensity for risk taking across the life span and around the globe. *Psychological Science*, *27*(2), 1-13

McElroy, T., & Seta, J. J. (2003). Framing effects: An analytic–holistic perspective. *Journal of Experimental Social Psychology*, *39*(6), 610-617.

McKinley, J. C., & Urbina, I. (2008). A million flee as huge storm hits Texas coast. *The New York Times*.

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the US public. *Weather and Forecasting*, *23*(5), 974-991.

Murphy, A. H., & Winkler, R. L. (1979). Probabilistic temperature forecasts: The case for an operational program. *Bulletin of the American Meteorological Society*, *60*(1), 12-19.

Mutter, S. A., & Pliske, R. M. (1996). Judging event covariation: Effects of age and memory demand. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *51*(2), 70-80.

Nagele, D. E., & Trainor, J. E. (2012). Geographic specificity, tornadoes, and protective action. *Weather, Climate, and Society*, *4*(2), 145-155.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209-227.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

NOAA (2016). Weather Forecast Uncertainty Fact Sheet. Accessed online September, 10, 2016, at http://www.crh.noaa.gov/images/lmk/pdf/WeatherForecastUncertainty.pdf

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411-419.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, *18*(1), 1-27.

Parker, D. J., Priest, S. J., & Tapsell, S. M. (2009). Understanding and enhancing the public's behavioural response to flood warning information. *Meteorological applications*, *16*(1), 103-114.

Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, *64*(2), 169-190.

Peacock, W. G., Morrow, B. H., & Gladwin, H. (1997). *Hurricane Andrew: Ethnicity, gender, and the sociology of disasters*. Gainesville: University Press of Florida.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological science*, *17*(5), 407-413.

Petty, R. E., Cacioppo, J. T., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306-307.

Powell, M., & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of economic psychology*, *18*(6), 605-628.

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*(4), 450-472.

Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223-237). Springer US.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, *17*(2), 147-159.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological bulletin*, *135*(6), 943-973.

Riad, J. K., Norris, F. H., & Ruback, R. B. (1999). Predicting evacuation in two major disasters: risk perception, social influence, and access to Resources1.*Journal of Applied Social Psychology*, *29*(5), 918-934.

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience*,*4*(3), 203-218.

Roulston, M. S., Bolton, G. E., Kleit, A. N., & Sears-Collins, A. L. (2006). A laboratory study of the benefits of including uncertainty information in weather forecasts. *Weather and Forecasting*, *21*(1), 116-122.

Savelli, S., & Joslyn, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, *27*(4), 527-541.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*, *127*(11), 966-972.

Sheperd, M. (2014). An open thank you to meteorologists in Atlanta. Accessed online on February 16, 2014, at http://meterologistandtheatlantasnow2014.blogspot.com/2014/01/an-open-thank-you-letter-to-atlanta.html

Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. New York: The Penguin Press.

Simianu, V. V., Grounds, M. A., Joslyn, S. L., Leclerc, J. E., Ehlers, A. E., Agrawal, N., Alfonso, Cristancho, R., Flaxman, A. D., Flum, D. R. (2016). Understanding clinical and non-clinical decisions under uncertainty: A scenario-based survey. *BMC Medical Informatics and Decision Making.*

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, *9*(4), 283-290.

Svedholm-Häkkinen, A. M., & Lindeman, M. (2016). Intuitive and deliberative empathizers and systemizers. *Journal of Personality*. Online early access. doi:10.1111/jopy.12263

Syngelaki, E. M., Moore, S. C., Savage, J. C., Fairchild, G., & Van Goozen, S. H. (2009). Executive functioning and risky decision making in young male offenders. *Criminal Justice and Behavior*, *36*(11), 1213-1227.

Szrek, H., Chao, L. W., Ramlagan, S., & Peltzer, K. (2012). Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. *Judgment and decision making*, *7*(6), 716-727.

Thornton, W. J., & Dumke, H. A. (2005). Age differences in everyday problem-solving and decision-making effectiveness: a meta-analytic review. *Psychology and aging*, *20*(1), 85-99.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157): 1124–1131.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038-1066.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635-654.

Warburton, P., & Kahn, P. (2007). Improving the numeracy skills of nurse prescribers. *Nursing Standard*, *21*(28), 40-43.

Wayment, H. A., Newcomb, M. D., & Hannemann, V. L. (1993). Female and male intravenous drug users not-in-treatment: Are they at differential risk for AIDS?. *Sex Roles*, *28*(1-2), 111-125.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The Cognitive Reflection Test: how much more than numerical ability. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1587-1592). Austin, TX: Cognitive Science Society.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*(4), 348.

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.

Woloshin, S., Schwartz, L. M., & Welch, H. G. (2005). Patients and medical statistics. *Journal of General Internal Medicine, 20*(11), 996-1000.

Zuckerman, M., Ball, S., & Black, J. (1990). Influences of sensation seeking, gender, risk appraisal, and situational motivation on smoking. *Addictive behaviors*, *15*(3), 209-220.

Tables And Figures

Table 1

Demographic information for participants in Experiment 1 compared to 2012 US Census

	M-turk sample	National population (2012 census data)
Age	Mean 34.17, Median 32 years	Median 37.2 years
Gender	53.3% female, *n* = 139	50.8% female
Education	60.2% college educated, *n* = 157	39% college educated

Table 2

Mean numeracy, age, expected decision value, mean monthly trust rating, proportion of risk-averse errors (salt decisions below 17%), proportion of salt decisions above 17%, and proportion of risk-seeking errors by forecast format for Experiment 1. Standard deviations are in parentheses.

	Deterministic N = 82	**Advice N = 81**	**Probability N = 86**
Predictor Variables			
Numeracy	2.79 (1.60)	2.98 (1.94)	2.58 (1.62)
Age	33.98 (11.05)	36.04 (11.04)	33.74 (10.76)
Outcome Variables			
Expected Value	-$1071.48 ($106.59)	-$1043.74 ($93.24)	-$1048.17 ($98.00)
Average Trust	2.74 (.99)	3.04 (.95)	3.22 (.99)
Salt Decisions 10-16% / Risk-Averse Error Proportion	.30 (.28)	.21 (.20)	.16 (.20)
Salt Decisions 17-51% Proportion	.62 (.21)	.66 (.20)	.62 (.16)
Risk-Seeking Error Proportion	.61 (.28)	.56 (.28)	.67 (.23)

Table 3

Regression Analyses of Expected Value for Forecast Format Presentations in Experiment 1

	Deterministic		Advice		Probability	
Predictor	β	t	β	t	β	t
Numeracy	.33	3.29**	.45	4.19**	.38	3.76**
Age	.24	2.23*	-.03	-.23	.15	1.50
Education	-.28	-2.66**	-.08	-.78	-.27	-2.73**

(* p<.05, ** p<.01)

Table 4

Regression Analyses of Decisions to salt above 17% Probability for Forecast Format Presentations in Experiment 1.

	Deterministic		Advice		Probability	
Predictor	β	t	β	t	β	t
Numeracy	.26	2.44*	.41	3.80**	.13	1.15
Age	.19	1.70	-.02	-.16	.20	1.81
Education	-.20	-1.78	-.15	-1.40	-.14	-1.27

(* p<.05, ** p<.01)

Table 5

Mean numeracy, working memory capacity, cognitive reflection, advanced progressive matrices, financial risk propensity, need for cognition, expected decision value, mean monthly trust rating, proportion of risk-averse errors (salt decisions below 17%), proportion of salt decisions above 17%, and proportion of risk-seeking errors by forecast format for Experiment 2. Standard deviations are in parentheses.

	Deterministic **N= 90**	**Advice** **N = 82**	**Probability** **N= 82**	**Probability with Advice** **N = 87**
Predictor Variables				
Numeracy	3.81 (1.51)	4.21 (1.46)	4.41 (1.74)	4.37 (1.74)
WMC	39.14 (15.51)	39.61 (16.37)	38.18 (16.94)	40.67 (16.38)
CRT	1.11 (1.13)	1.41 (.99)	1.50 (1.12)	1.60 (1.08)
APM	6.66 (3.21)	6.21 (3.53)	7.15 (3.25)	6.70 (3.30)
F-DOSPERT	2.97 (1.20)	3.02 (.95)	2.98 (.90)	3.08 (1.09)
NFC	9.87 (17.29)	12.06 (17.62)	14.04 (15.96)	15.26 (14.60)
Outcome Variables				
Expected Value	-$1099.97 ($85.73)	-$1095.51 ($102.61)	-$1050.83 ($71.13)	-$1020.71 ($63.75)
Average Trust	2.22 (.74)	2.35 (.77)	2.50 (.77)	2.61 (.72)
Salt Decisions 10-16% / Risk-Averse Error Proportion	.27 (.22)	.22 (.19)	.17 (.18)	.18 (.19)
Salt Decisions 17-51% Proportion	.59 (.17)	.65 (.18)	.71 (.16)	.65 (.17)
Risk-Seeking Error Proportion	.58 (.24)	.46 (.22)	.64 (.22)	.50 (.25)

Table 6

Bivariate correlations between cognitive measures in forecast format conditions used in Experiment 2.

Forecast Format	Measure	Numeracy	WMC	CRT	APM	F-DOSPERT	NFC
Deterministic	Numeracy	1	.18	.46**	.32**	-.01	.12
	WMC		1	.24*	.20	-.01	.16
	CRT			1	.41**	.15	.21
	Ravens				1	-.06	.18
	F-DOSPERT					1	-.01
Advice	Numeracy	1	.30**	.57**	.55**	.21	.13
	WMC		1	.24*	.23*	-.09	.31**
	CRT			1	.57**	.06	.20
	Ravens				1	-.04	.26*
	F-DOSPERT					1	.13
Probability	Numeracy	1	.25*	.52**	.47**	.04	.21
	WMC		1	.23	.29**	.01	.05
	CRT			1	.22*	.09	.40**
	Ravens				1	-.07	.04
	F-DOSPERT					1	-.18
Probability with Advice	Numeracy	1	.18	.40**	.38**	.07	.18
	WMC		1	.25*	.36**	.03	.24*
	CRT			1	.31**	.16	.19
	Ravens				1	.03	.32**
	F-DOSPERT					1	.02

(* *p*<.05, ** *p*<.01)

Table 7

Regression analyses of dependent variables in the probability forecast format condition in Experiment 2.

	Expected Value		Salting Above 17%		Salting Below 17%		Risk-seeking Errors	
Predictor	β	t	β	t	β	t	β	t
Numeracy	.46	3.38**	.39	3.09**	.05	.33	-.22	-1.77
WMC	-.09	-.79	-.18	-1.73	-.07	-.64	.19	1.86
CRT	-.11	-.83	.06	.50	.30	2.48*	-.24	-2.18*
Ravens	-.01	-.07	.09	.82	.08	.67	-.12	-1.10
F-DOSPERT	-.07	-.63	-.03	-.32	.06	.50	-.03	-.31
NFC	.11	.97	.24	2.22*	.13	.96	-.23	-1.77

(* p<.05, ** p<.01)

Table 8

Demographic information for participants in Experiment 3 compared to 2012 US Census

	M-turk sample	National population (2012 census data)
Age	Mean 35.45, Median 33 years	Median 37.2 years
Gender	51.1% female, *n* = 423	50.8% female
Education	57.2% college educated, *n* = 474	39% college educated

Table 9

Mean numeracy, working memory capacity, cognitive reflection, age, income, expected decision value, mean monthly trust rating, proportion of risk-averse errors (salt decisions below 17%), proportion of salt decisions above 17%, and proportion of risk-seeking errors by forecast format for Experiment 3. Standard deviations are in parentheses.

	Deterministic **N = 161**	**Advice** **N = 163**	**Probability** **N = 195**	**Probability with Advice** **N = 176**
Predictor Variables				
Numeracy	3.29 (1.90)	3.17 (1.70)	3.43 (1.66)	3.45 (1.76)
WMC	45.19 (20.76)	47.98 (19.50)	45.43 (20.20)	43.22 (20.31)
CRT	1.35 (1.18)	1.37 (1.21)	1.35 (1.13)	1.51 (1.17)
Age	36.31 (11.52)	35.69 (12.50)	36.21 (12.36)	35.89 (11.26)
Income	$40304.93 ($33255.60)	$40916.99 ($80178.76)	$35410.23 ($35069.26)	$36022.03 ($34655.08)
Outcome Variables				
Expected Value	-$1052.52 ($71.60)	-$1031.92 ($67.57)	-$1020.23 ($48.93)	-$1012.18 ($74.27)
Average Trust	2.66 (.89)	2.77 (.89)	3.02 (.82)	3.08 (.81)
Salt Decisions 10-16% / Risk-Averse Error Proportion	.21 (.20)	.20 (.21)	.15 (.19)	.16 (.19)
Salt Decisions 17-51% Proportion	.62 (.18)	.68 (.18)	.70 (.14)	.70 (.17)
Risk-Seeking Error Proportion	.62 (.25)	.53 (.27)	.68 (.23)	.57 (.26)

Table 10

Bivariate correlations among predictor variables by forecast format conditions in Experiment 3.

Forecast Format	Measure	Numeracy	WMC	CRT	Education	Age	Income
Deterministic	Numeracy	1	.14	.47**	.17*	-.02	.11
	WMC		1	.12	-.04	-.20*	-.12
	CRT			1	.17*	.03	.04
	Education				1	.09	.25**
	Age					1	.19*
Advice	Numeracy	1	.09	.40**	.08	-.03	-.02
	WMC		1	.07	-.03	-.01	.09
	CRT			1	-.01	-.05	.02
	Education				1	.29**	.15
	Age					1	.08
Probability	Numeracy	1	.13	-.09	.20**	-.03	.21**
	WMC		1	.01	.13	-.17*	.05
	CRT			1	.01	-.04	.01
	Education				1	-.04	.29**
	Age					1	.08
Probability with Advice	Numeracy	1	.08	.41**	.11	.01	.01
	WMC		1	.13	.04	-.09	-.04
	CRT			1	.14	-.02	-.03
	Education				1	-.12	.26**
	Age					1	.14

(* p<.05, ** p<.01)

Table 11

Regression Analyses of Expected Value for Forecast Format conditions in Experiment 3

	Deterministic		Advice		Probability		Probability with Advice	
Predictor	β	t	β	t	β	t	β	t
Numeracy	.11	1.26	.20	2.37*	.24	3.33**	.28	3.86**
WMC	-.02	-.20	.03	.34	.01	.18	.08	1.17
CRT	.17	1.94	-.04	-.45	.05	.70	.09	1.20
Age	-.13	-1.69	.03	.38	.21	2.97**	.16	2.38*
Education	.15	1.92	-.04	-.49	.14	1.97*	.12	1.72
Income	-.07	-.89	-.21	-2.62*	-.04	-.55	-.36	-5.17**

(* p<.05, ** p<.01)

Table 12

Regression Analyses for Risk-seeking Errors for Probability Inclusive Forecast Formats in Experiment 3

	Probability		Probability with Advice	
Predictor	β	t	β	t
Numeracy	-.28	-3.81**	-.26	-3.23**
WMC	.05	.69	-.04	-.61
CRT	.03	.45	-.08	-.94
Age	-.09	-1.27	-.08	-1.07
Education	-.14	-1.83	-.06	-.58
Income	.05	.68	.16	2.16*

(* p<.05, ** p<.01)

Table 13

Regression Analyses for Salting Above 17% Probability for Probability Inclusive Forecast Formats in Experiment 3

	Probability		Probability with Advice	
Predictor	β	t	β	t
Numeracy	.30	4.26**	.30	4.08**
WMC	-.04	-.52	.08	1.19
CRT	.01	.01	.09	1.18
Age	.16	2.35*	.13	1.87
Education	.16	2.15*	.08	1.11
Income	-.05	-.65	-.28	-3.89**

(* p<.05, ** p<.01)

Table 14

Regression analyses of expected value with numeracy as the only predictor variable for all forecast formats separated by Experiment.

	B	**SE**	**95% CI**	**β**	**t**
Experiment 1					
Deterministic	24.49	6.93	10.71-38.29	.37	3.54**
Advice	20.85	4.86	11.17-30.53	.43	4.29**
Probability	24.15	6.02	12.18-36.13	.40	4.01**
Experiment 2					
Deterministic	.60	6.04	-11.41-12.60	.01	.10
Advice	16.84	7.75	1.40-32.27	.24	2.17*
Probability	16.14	4.19	7.80-24.49	.40	3.85**
Probability with Advice	9.59	3.84	1.96-17.22	.26	2.50*
Experiment 3					
Deterministic	7.82	2.93	2.04-13.60	.21	2.67**
Advice	7.43	3.08	1.34-13.51	.19	2.41*
Probability	7.34	2.05	3.29-11.38	.25	3.57**
Probability with Advice	14.03	3.01	8.08-19.97	.33	4.66**

(* *p*<.05, ** *p*<.01)

Figure 1

Mean Expected Value scores by forecast format and education level for Experiment 1. Note: values are negative, thus a longer bar indicates worse performance.

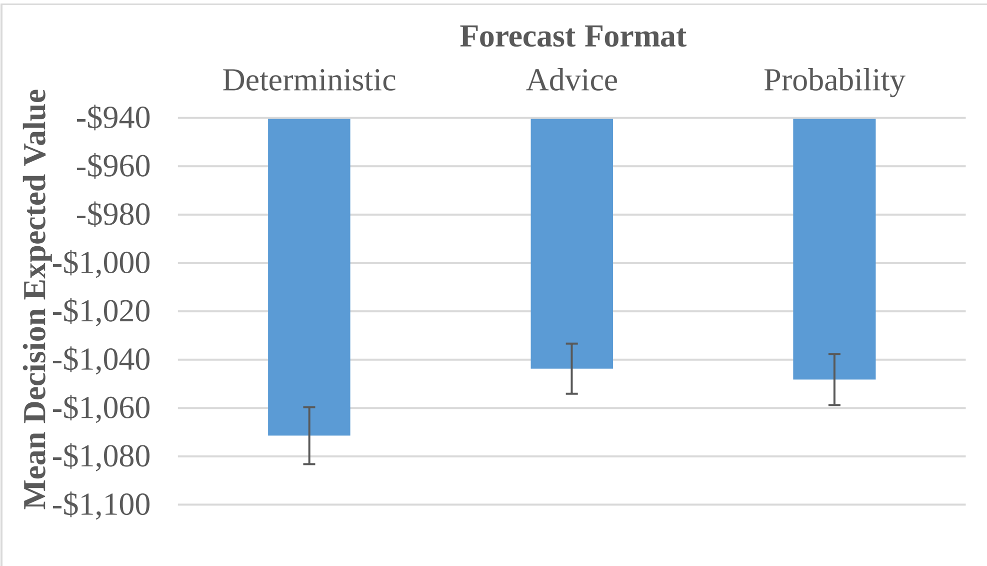


Figure 2

Mean monthly trust ratings by forecast format for Experiment 1.

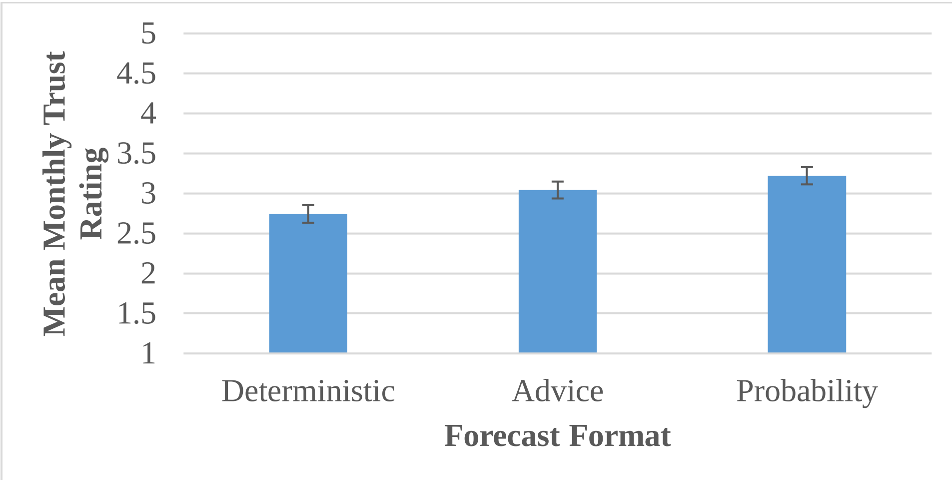


Figure 3

Decisions to apply salt over freeze probability ranges by uncertainty condition and education for Experiment 1.

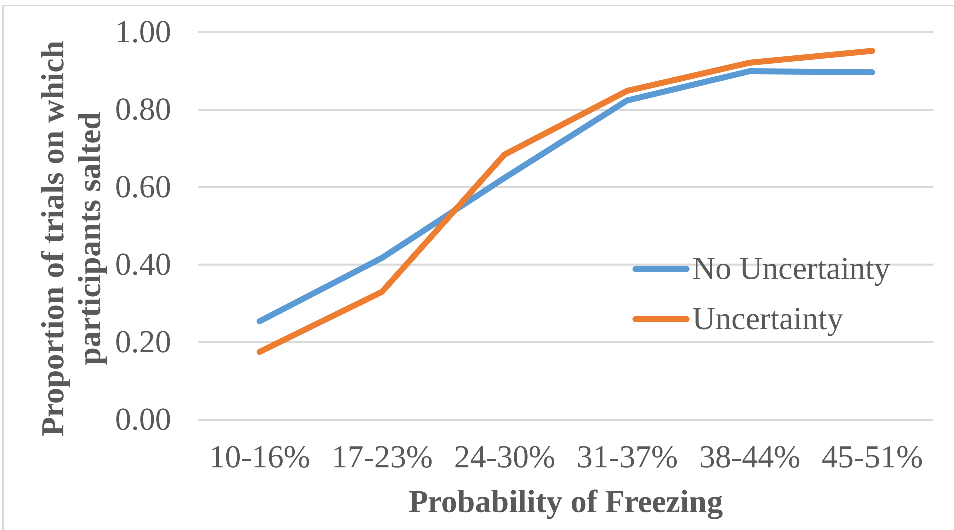


Figure 4

Mean Expected Value scores by forecast format for Experiment 2. Note: values are negative, thus a longer bar indicates worse performance.

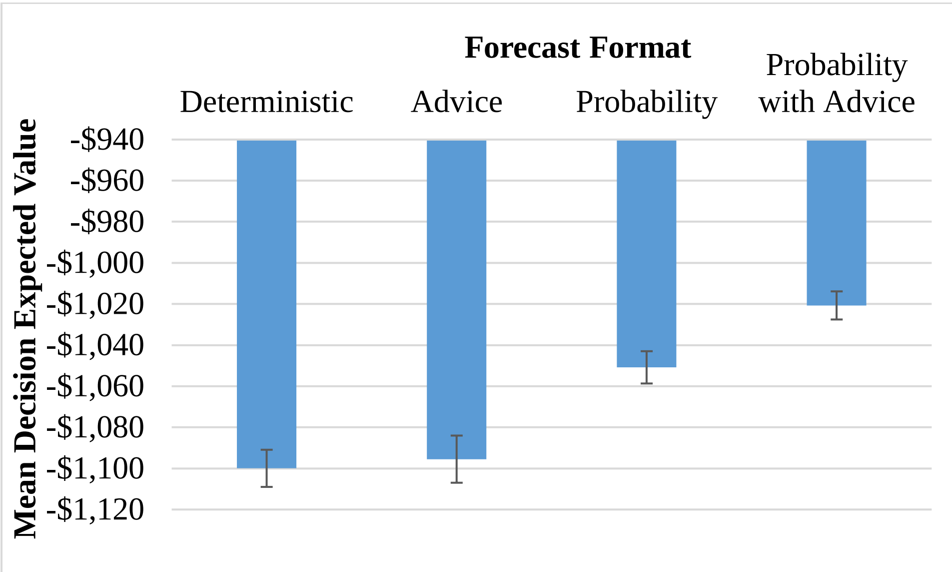


Figure 5

Mean monthly trust ratings by forecast format for Experiment 2.

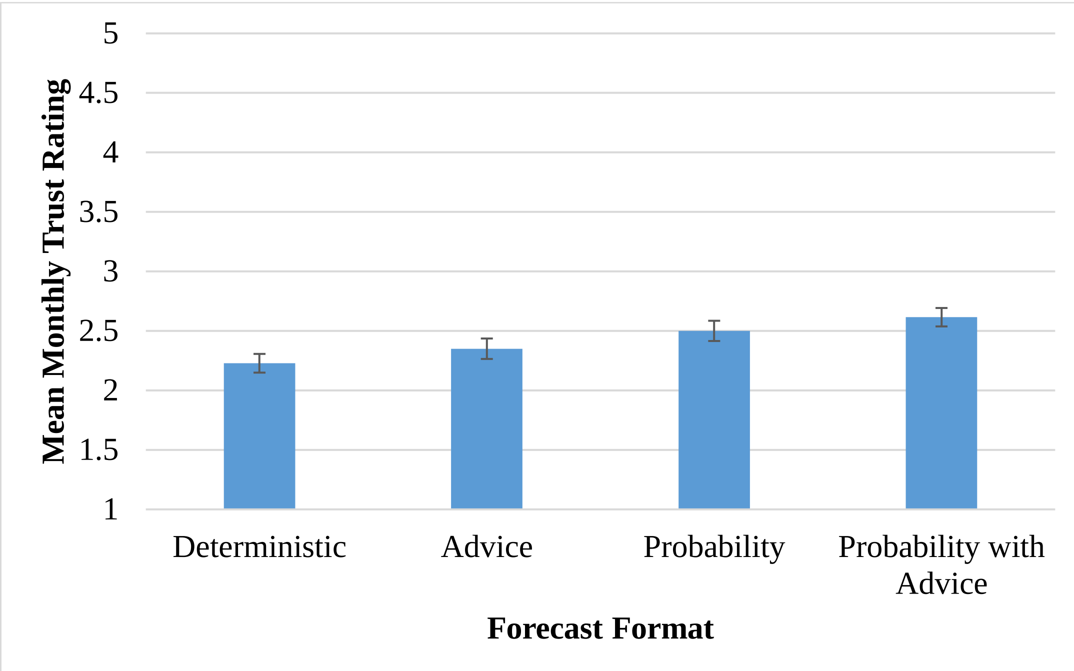


Figure 6

Decisions to apply salt over freeze probability ranges by uncertainty condition for Experiment 2.

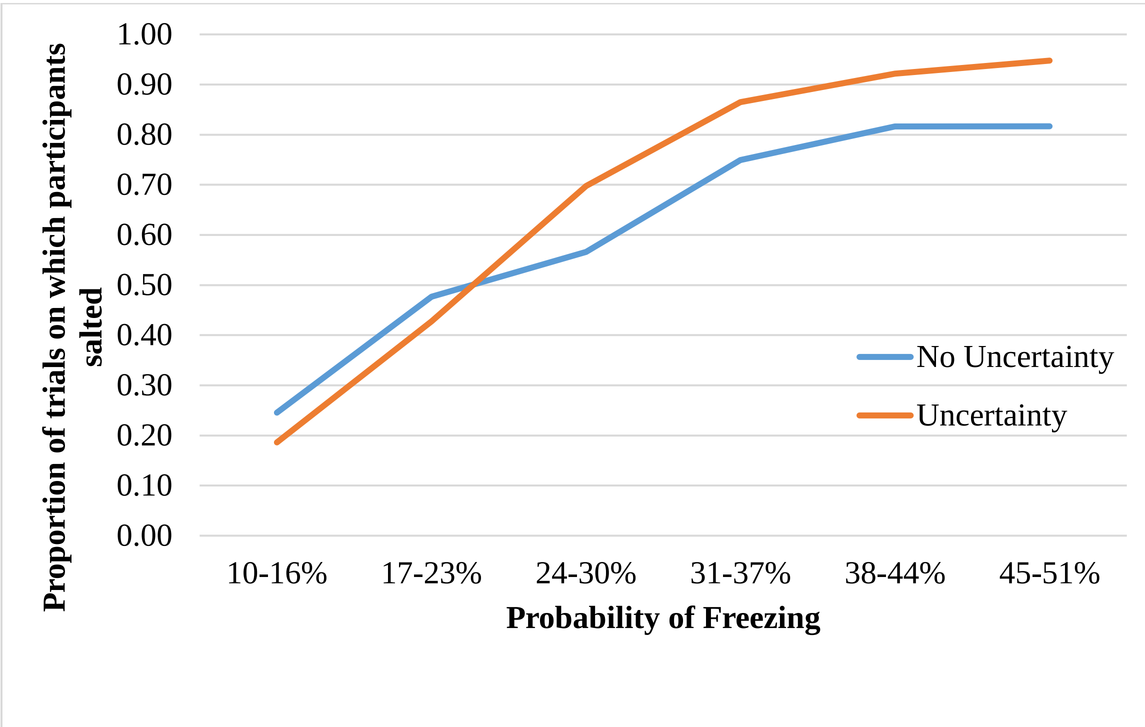


Figure 7

Mean Expected Value scores by forecast format for Experiment 3. Note: values are negative, thus a longer bar indicates worse performance.

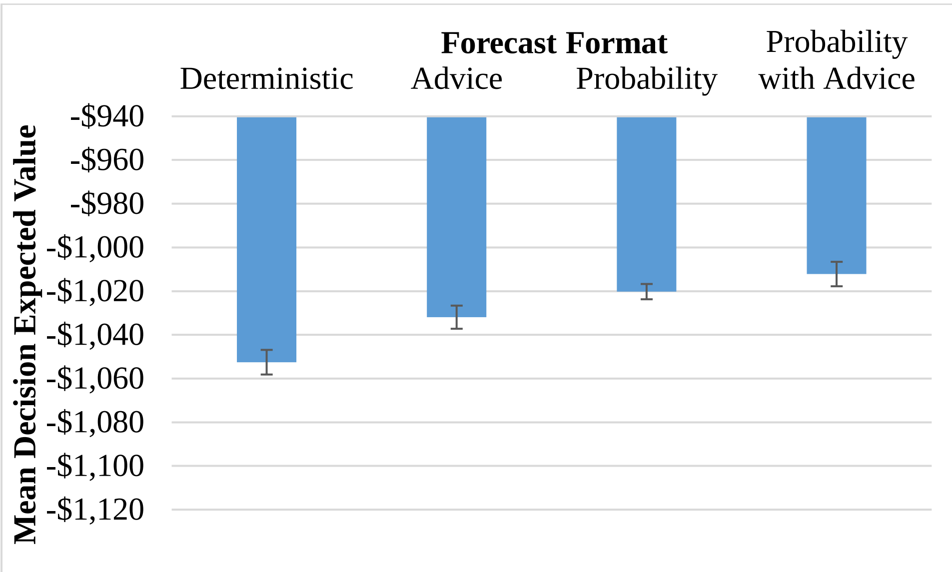


Figure 8

Mean monthly trust ratings by forecast format for Experiment 3.

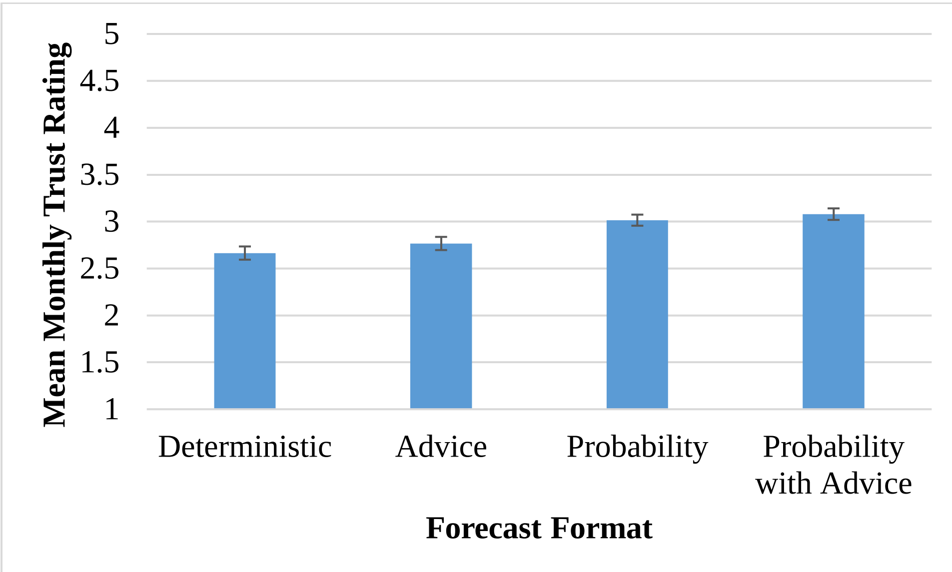


Figure 9

Decisions to apply salt over freeze probability ranges by uncertainty condition for Experiment 3.

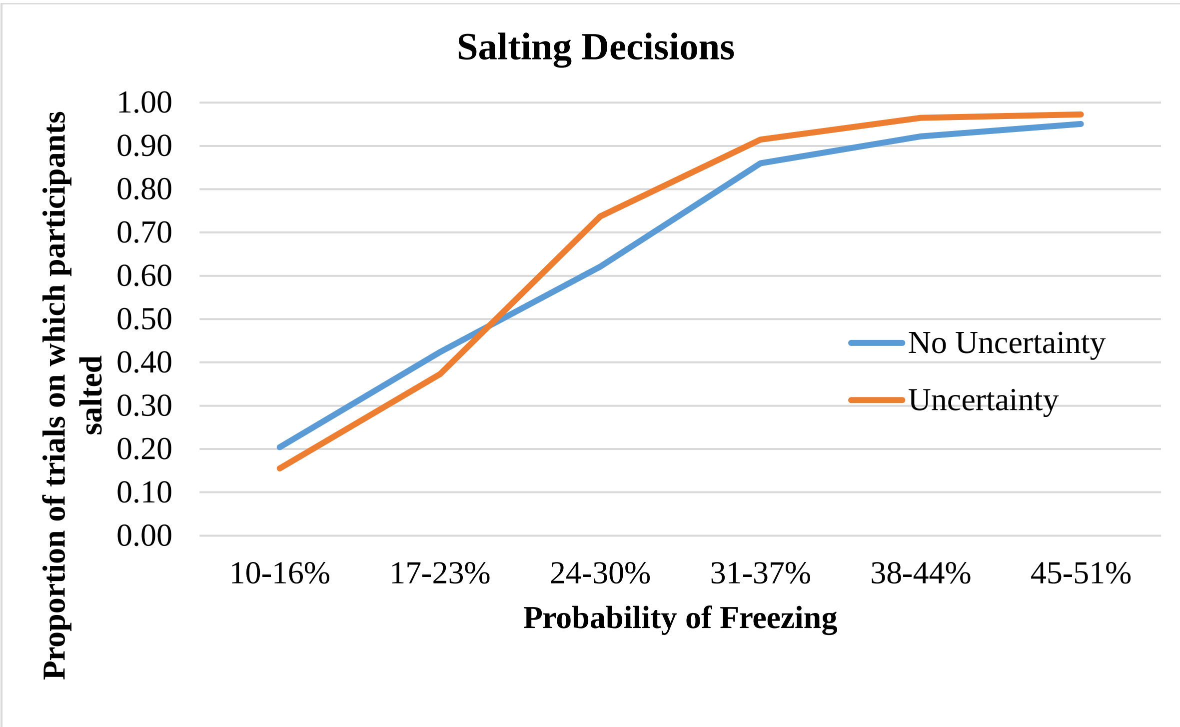


Figure 10

Mean Expected Value scores for low numeracy individuals (numeracy scores ≤ 2) for each experiment by forecast format.

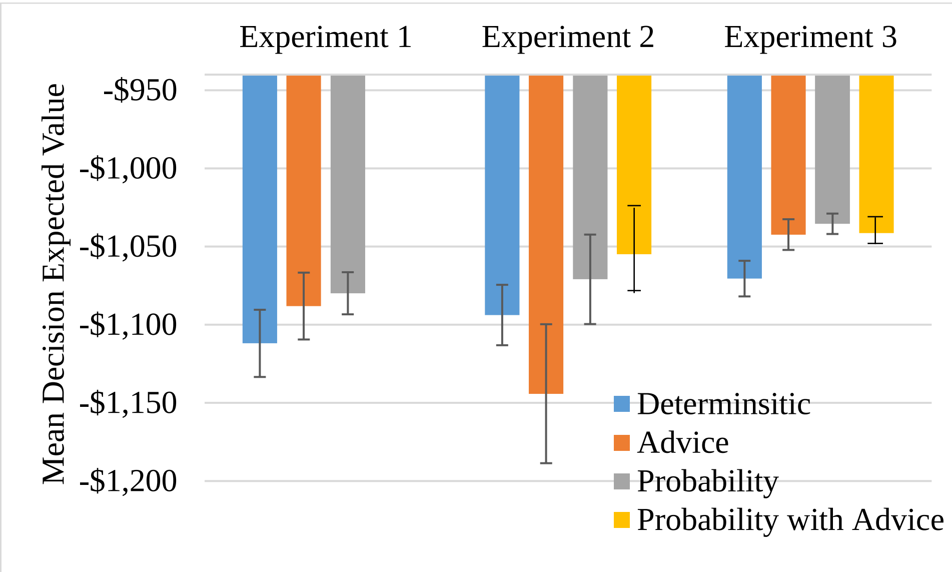


Figure 11

Mean Expected Value scores for younger individuals (age ≤ 26) for Experiment 1 and Experiment 3 by forecast format.

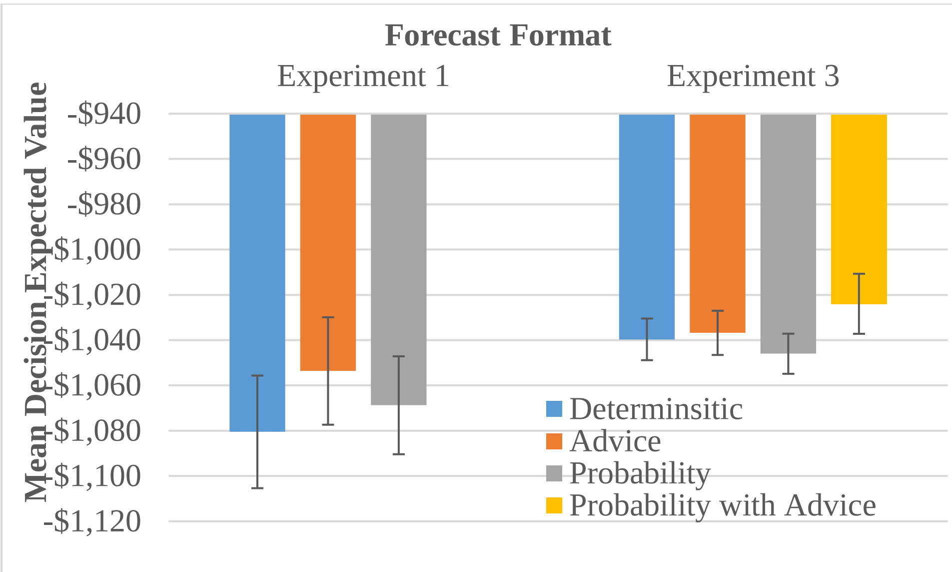


Figure 12

Mean Expected Value scores for less educated individual (high school diploma or less) by forecast format for Experiment 3.

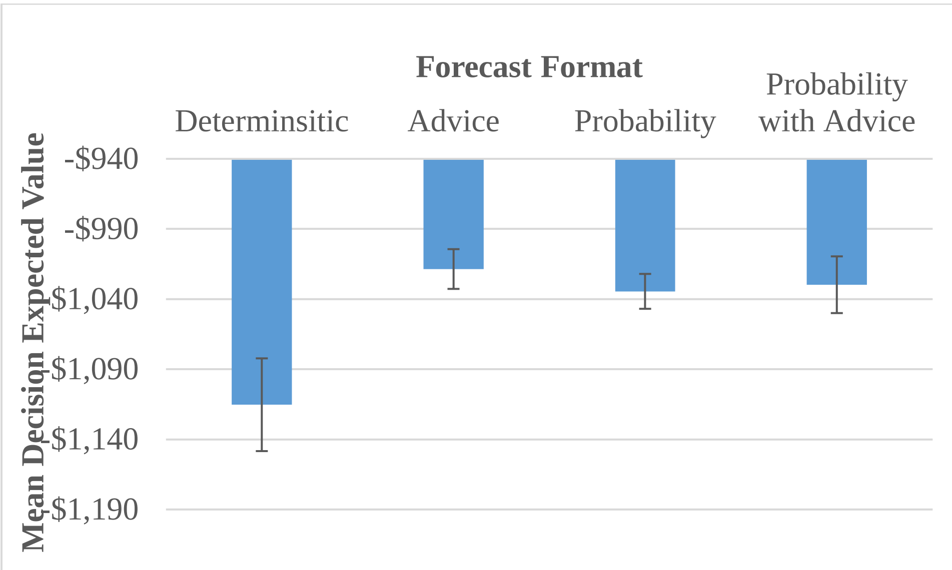
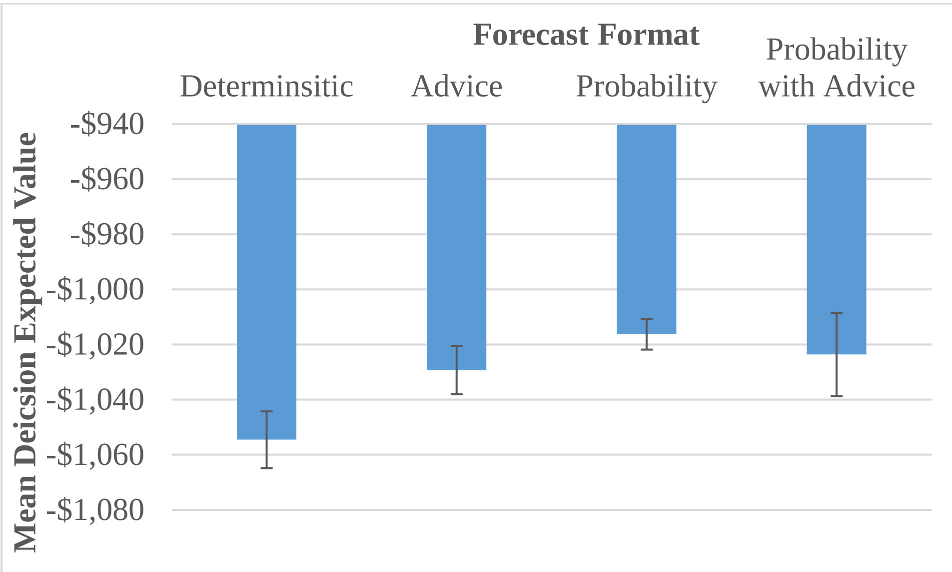


Figure 13

Mean Expected Value scores for high income individuals (income ≥ $50,000) by forecast format for Experiment 3.



Appendices

Appendix A

Complete Numeracy Assessment used in all Experiments (Cokely et al., 2012)

1. Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?

A: 500

2. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

A: 30

3. In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS?

A: 10

4. In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?

A: .1%

5. Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in a choir 100 are men. Out of the 500 inhabitants that are not in a choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.

A: 25%

6. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?

A: 20

7. In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?

A: 50%

Appendix B

Cognitive Reflection Test (Frederick, 2005)

Instructions: Below are several problems that vary in difficulty. Try to answer as many as you can.

1. A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?

System 1 Answer: 10 cents

System 2 (Correct Answer): 5 cents

2. If it takes five machines five minutes to make five widgets, how long does it take 100 machines to make 100 widgets?

System 1 Answer: 100 minutes

System 2 (Correct Answer): 5 minutes

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

System 1 Answer: 24 days

System 2 (Correct Answer): 47 days

Appendix C

Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984)

Asterisks designate the items that are reverse scored.

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.*
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.*
5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.*
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to.*
8. I prefer to think about small, daily projects to long-term ones.*
9. I like tasks that require little thought once I've learned them.*
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn't excite me very much.*
13. I prefer my life to be filled with puzzles that I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.*
17. It's enough for me that something gets the job done; I don't care how or why it works.*
18. I usually end up deliberating about issues even when they do not affect me personally

Appendix D

Direct comparison between predictors of Expected Value used in both Experiment 2 and Experiment 3 (Numeracy, WMC, and Cognitive Reflection). The confidence intervals for the percent of variance explained and individual regression coefficients overlap across experiments, suggesting the amount of variance explained is not significantly different between populations.

		Experiment 2				**Experiment 3**			
Forecast Format	Predictor	B (95% CI)	β	t	% Variance Explained (95% CI)	B (95% CI)	β	t	% Variance Explained (95% CI)
Deterministic	Numeracy	-1.41 (-14.94-12.13)	-.03	-.21	3% (-4%-10%)	4.63 (-1.89-11.15)	.12	1.40	7%* (0%-14%)
	WMC	-.83 (-2.04-.37)	-.15	-1.37		.04 (-.49-.57)	.01	.14	
	CRT	10.05 (-8.22-28.33)	.13	1.09		10.81 (.34-21.27)	.18	2.04*	
Advice	Numeracy	9.45 (-9.68-28.58)	.13	.98	9% (-2%-20%)	8.09 (1.41-14.77)	.20	2.39*	4% (-2%-10%)
	WMC	1.02 (-.42-2.46)	.16	1.41		.04 (-.50-.57)	.01	.13	
	CRT	10.11 (-17.53-37.75)	.10	.73		-2.49 (-11.86-6.89)	-.04	-.52	
Probability	Numeracy	18.56 (8.61-28.50)	.46	3.72**	17%** (3%-31%)	7.47 (3.35-11.59)	.25	3.58**	6%** (0%-12%)
	WMC	-.387 (-1.28-.51)	-.09	-.86		-.02 (-.35-.32)	-.01	-.09	
	CRT	-4.37 (-19.63-10.88)	-.07	-.57		1.79 (-4.24-7.81)	.04	.59	
Probability with Advice	Numeracy	7.08 (-1.24-15.39)	.19	1.69	10* (-1%-21%)	11.85 (5.37-16.84)	.28	3.61**	13%** (4%-22%)
	WMC	-.24 (-1.07-.60)	-.06	-.57		.29 (-.23-.81)	.08	1.12	
	CRT	11.68 (-1.87-25.23)	.20	1.71		7.03 (-2.79-16.84)	.11	1.41	

(* *p*<.05, ** *p*<.01)