Sources of Bias in Naturalistic Decision Making Under Risk

Chao Qin

A dissertation submitted

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Susan Joslyn, Chair

John Palmer

Andrea Stocco

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Sources of Bias in Naturalistic Decision Making Under Risk

Chao Qin

Chair of the Supervisory Committee:

Susan Joslyn

Psychology

Severe weather situations such as tornadoes and droughts require people to take protective action even when the probability of the severe weather is low because the consequence of not protecting is very serious. In naturalistic decision experiments based on these situations, people are risk-seeking such that they often do not take protective actions when it is economically rational to do so. This project studied this phenomenon from a signal detection theory perspective. A random likelihood model was introduced to estimate the subjective criterion which is the likelihood above which one takes protection actions. This model separates the subjective criterion from subjective likelihood, which is participants' perception of the probability of the weather event. Three experiments manipulated the gain-loss framing and the economically rational criterion (the criterion based on expected value theory) to examine their effect on the subjective criterion. When the gain-loss framing was manipulated, the subjective criterion was higher in a loss frame than a gain frame. When the economically rational criterion was manipulated, the subjective criterion was between the economically rational criterion and the center of the possible likelihood range (50%). Neither manipulation affected subjective likelihood. In addition, participants showed an overestimation in subjective likelihood in all conditions. The shifted subjective criterion overcame this overestimation and resulted in risk-seeking decisions in some conditions. Thus, the random likelihood model analysis suggests that shift of the subjective criterion is the source of risk-seeking decisions in naturalistic decision tasks. Potential interventions are discussed with the aim to improve the placement of the subjective criterion

Acknowledgments

I am grateful for the support of my advisor, Dr. Susan Joslyn, and my co-advisor, Dr. John Palmer, throughout my time in the program and the dissertation work. I want to thank Susan for introducing me to the field of judgement and decision making and then later took me on as a graduate student. I also want to thank her for guiding me in my journal of becoming an independent researcher and a teacher. I want to thank John for introducing me to signal detection theory and guided me every step of the way in my dissertation project.

I also want to thank my committee members, Dr. Andrea Stocco, Dr. Ann Bostrom, and Dr. Sonia Savelli. My dissertation project greatly benefited from your thoughtful advice and feedback, especially during the planning phase.

Finally, I want to thank my fellow Decision Making with Uncertainty Lab member, Jee Hoon Han, for your perspective and feedback in this project and support in many other projects I worked on.

Table of Contents:

Introduction....7

Reanalyses of Previous Experiments....28

Experiment 1....50

Experiment 2....80

Experiment 3....97

General Discussion....123

References....135

Introduction

In real world decision making under risk situations such as severe weather events, people often need to make decisions for an uncertain future. For example, when facing a possible tornado, people must decide whether to take protective actions such as taking shelter. Taking protective actions costs time and resources but can protect them from harm. On the other hand, not taking protective actions might expose decision-makers to potential harm but can save time and resources if a tornado does not materialize. Due to the potential serious harm of these events, people are advised to take protective actions even when the probability is low. For example, tornado warnings are issued by the NWS when the probability of a tornado is 10% or higher (Qin et al., 2024). Due to their low probability, these weather events often fail to occur at the residents' location, making protective actions seem like a waste of time. Indeed, research on people's response to forecasts and warnings in severe weather events, such as floods, tornadoes, and hurricanes, showed that people often failed to take protective actions, a risk-seeking tendency (Baker, 1995; Joslyn & LeClerc, 2013; Atreya et al., 2015; LeClerc & Joslyn, 2015; Qin et al., 2024).

However, many warnings do not currently provide people with an estimated probability of the event (e.g., there is a 20% chance of tornado), although such information (e.g., there is 30% chance of tornado) is increasingly available (Joslyn & Savelli, 2021; Gulacsik et al., 2022; Qin et al., 2024). There is evidence that providing such information could help people better understand the likelihood of the weather event, increase trust in the forecast, and allow people to make better decisions (Ash et al., 2014; Joslyn & LeClerc, 2013; Klockow-McClain et al., 2020; Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). However, people's decisions still often showed a risk-seeking tendency (often not taking protective actions when they were warranted) even though their perception of the likelihood of the weather event measured by self-reporting was accurate or even slightly overestimated (Grounds & Joslyn, 2018; Burgeno & Joslyn, 2023; Gulacsik et al., 2022; Qin et al., 2024). Therefore, additional biases might have contributed to their risk-seeking decisions.

This dissertation project explores the cognitive processes that might lead to risk-seeking decisions in an experimental setting modeled on these real-world decisions. Experimental settings like this are considered naturalistic to distinguish them from experiments based on simple artificial gambling tasks, like drawing cards or blackjack games, upon which the majority of research on decision biases were based (Kahneman & Tversky, 1979; Krizan & Windschitl, 2007; Windschitl et al., 2010). There are several differences between naturalistic decision tasks and artificial gambling tasks that might affect people's judgement and decision-making and thereby influence the generalizability of the research. First, naturalistic settings usually center around a real-life situation where people might have prior experience or pre-conceived notions. Therefore, the context might affect people's judgement. For example, when participants were told that they had a 20% chance of contracting malaria when traveling to a location, participants felt that they were less likely to contract malaria if the location was Hawaii compared to India (Windschitl & Weber, 1999). Therefore, people's perception about the likelihood of an outcome can be affected by the context. In simple artificial gambling tasks such as a coin flip, context information is not expected to bias people's belief about the likelihood away from the objective probability of an outcome (e.g., 50% chance of heads; Kahneman & Tversky, 1979). However, many original experiments using artificial gambling tasks did not directly measure people's perception of the likelihood via self-reporting measures (Kahneman & Tversky, 1979; Krizan & Windschitl, 2007). In the cases where such likelihood perception was measured, the participants usually reported likelihood ratings similar to the probabilistic information that was available to them (e.g., being told a 25% chance or playing a fair coin toss; Krizan & Windschitl, 2007). This meant that people's perception of the likelihood of the event might shift away from the probabilistic information due to the context in naturalistic settings but not in artificial gambling tasks. In this dissertation project, people's perception about the likelihood of an outcome, usually measured by direct solicitation such as self-reporting a percent chance, is defined as the subjective likelihood.

The next difference is that when there is probabilistic information for an outcome in the task (e.g., there is a 30% chance of tornado; 50% chance of heads in a fair coin toss), such information is generally not perceived as unreliable by people in artificial gambling tasks but might be open to interpretation in naturalistic tasks. This might be because in artificial gambling tasks, people have no reason to doubt the probabilistic information. On the other hand, in naturalistic settings such as weather forecasting, the probabilistic information might not be completely reliable or precise due to real world constraints (like technology limitations). People might have intuitions about this unreliability and imprecision (called second-order uncertainty; Morss et al., 2008; Joslyn & Savelli, 2021). This second-order uncertainty might also lead to variability in interpretation and in turn allow systematic bias in subjective likelihood (Lench et al., 2014). Indeed, when people were told that the probabilistic information was open to interpretation, they considered the information ambiguous and became prone to systematic biases in their subjective likelihood judgments, even in artificial gambling tasks (Lench et al., 2014). In summary, systematic biases in the subjective likelihood that can manifest in real life situations can be better studied in naturalistic experimental settings modelled after these situations rather than artificial gambling tasks.

Therefore, experiments using naturalistic tasks can contribute to better understanding of people's perception of the weather event's likelihood and their decision-making in similar real-life situations. This better understanding can in turn help to develop interventions that may improve people's decision-making. In short, studying naturalistic tasks, compared to studying artificial gambling tasks, should allow better examination of: 1) Decision making under risk in context rich scenarios that are more applicable to similar real-life situations; 2) Systematic bias in subjective likelihood, which might in turn affect decision-making. Therefore, this dissertation project used a naturalistic weather decision task, rather than an artificial gambling task.

Naturalistic Tasks Requiring Decision Making under Risk

There have been several studies examining people's behavior in naturalistic tasks requiring decision making under risk with numeric probabilistic information (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2017; Demnitz & Joslyn, 2020; Klockow-McClain et al., 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). In these tasks, participants were given numeric probabilistic information for a possible adverse weather event (e.g., 30% chance of tornadoes) and decided between a safe option or a risky option. These tasks were usually framed in a loss frame such that participants could only lose points but not gain points, similar to real world severe weather situations. In the usual safe option, participants paid a small point cost to protect themselves from any harm the weather event might cause. In the usual risky option, participants did not take any protective action and the weather event incurred a greater point penalty if it occurred. The participants' goal was to have as many points as possible after a series of trials. These studies usually had two dependent measures on each trial: 1) Self-reported subjective likelihood: Participants' perception of the likelihood of the weather event, usually measured using a continuous scale from 0% to 100% (reported in Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024); 2) Binary decisions: The decision between the safe and the risky option (reported in all studies).

Both measures were compared to a correct or rational standard in these studies (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2018; Demnitz & Joslyn, 2020; Klockow-McClain et al., 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). Self-reported subjective likelihood was compared to objective probabilities, which were the probabilities of the weather event on each trial and calibrated to be roughly reliable[1]. Objective probabilities were provided to the participants in experimental conditions as probabilistic information. On the other hand, binary decisions were compared to economically rational decisions based on expected value theory. According to expected value theory, if people want to maximize their gain or minimize their loss, they should choose the option with the best expected value (sum of the option's outcome values multiplied by the probability; Tversky & Fox, 1995). For example, in experiment 1 of Qin et al., 2024, the fixed cost of the safe option was 90 points regardless of a tornado while the risky option had a penalty of 1000 points if there was a tornado and 0 points if there was no tornado. The breakeven probability of tornado, for which the expected value of the risky option and the cost of the safe option were equal, was 9%. Therefore, when the probability of a tornado was higher than 9%, the safe option had the best expected value (lowest expected loss) and should always be chosen as the economically rational decision. This 9% is considered the economically rational probability threshold above which one should choose the save option. As participants' goal was to have as many points as possible, they should make decisions using this economically rational probability threshold. Decisions made with this probability threshold

[1] Usually rounded up or down to a number divisible by 5 (e.g., 33.33% rounded to 30%).

therefore were an economically rational standard with which participants' binary decisions could be compared.

The results of these studies were twofold. First, participants' self-reported subjective likelihood was close to but consistently slightly overestimated compared to the objective probabilities (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). Next, participants' decisions almost always failed to adhere to the economically rational standard (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2018; Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). This happened even when their self-reported subjective likelihood of the weather event was close to the objective probabilities (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). Most tasks used a loss frame like the example above in which participants could only lose points. Participants were overall risk-seeking, a decision bias where they chose the risky option more often than the economically rational decisions (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2018; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). On the other hand, some of the tasks involved a mixed gamble in which it was possible to both gain and lose points from the options. In these tasks, participants were overall risk-averse, a decision bias where they chose the safe option more often than the economically rational decisions (Demnitz & Joslyn, 2020).

Several accounts might explain these results. First, participants' overall risk-seeking tendency in a loss frame and risk-aversion in a mixed gamble is consistent with a gain-loss framing effect that people tend to take more risk than is economically rational for losses and less risk than is rational for gains (Kahneman & Tversky, 1979). The gain-loss framing effect can be explained by the utility function of prospect theory (see Figure 1 for a typical utility function; Kahneman & Tversky, 1979). The theory suggests that people translate the value of an outcome (e.g., $100) into a utility (how much $100 is worth to them). Prospect theory further suggests that people translate values into utility relative to a status quo reference point, and that the increase/decrease in value has diminishing return in utility as it moves away from the reference point. For example, gaining or losing $200 is assumed to have less than twice the absolute utility of gaining or losing $100. Because of the decrease in utility from the reference point when it comes to expected utility, a 50% chance of gaining $200 (risky option) has a worse expected utility than gaining $100 for sure (safe option) while a 50% chance of losing $200 (risky option) has a better expected utility than losing $100 for sure (safe option). In this example, assume gaining or losing $200 translates to gaining or losing 150 units of utility while $100 translates to 100 units of utility[2]. In the gain frame, the expected utility of the risky option was +75 (+150*0.5)[3] which is worse than the utility of the safe option (+100). In the loss frame, the expected utility of the risky option was -75 (-150*0.5), which was better than the utility of the safe option (-100). Therefore, people tend to be risk-averse in a gain frame while risk-seeking in a loss frame.

[2] For simplicity reasons, here the same utility was assigned regardless of whether the value is a gain or a loss. Prospect theory suggests that a loss is assigned greater utility than the same value of a gain.
[3] For simplicity reasons, this 50% here is the objective probability of the outcome. In prospect theory, the expected utility was calculated by *utility * Decision Weight*. See the paragraph below for the explanation of decision weights.

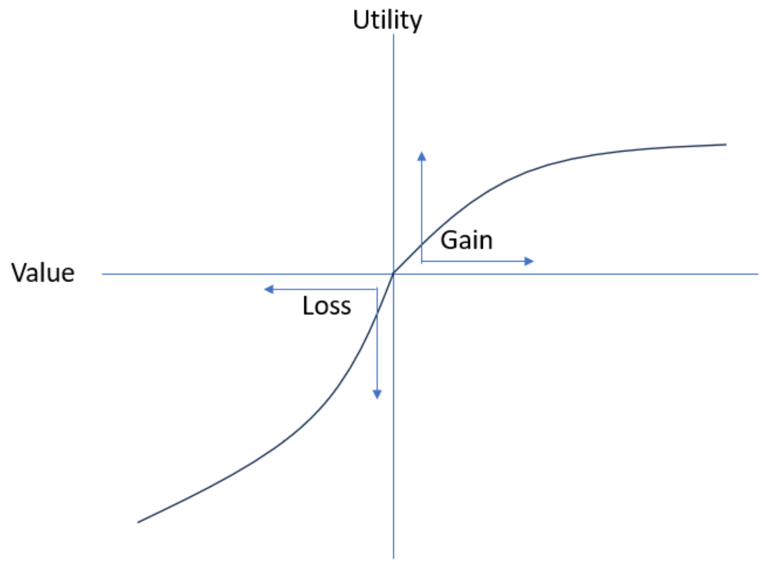


Figure 1
The Typical Gain/loss Utility Function
The x axis indicates the objective values of gains and losses. The y axis indicates the utility of these gains and losses to a decision-maker. The center is the reference point. The increase/decrease in values in relation to the reference point have diminishing return on the utility. The function is concave for gains and convex for losses. The slope is greater in the loss scenario than in the gain scenario. This means that people judged loss more heavily than gain.

While prospect theory is able to explain participants' risk-seeking decision patterns in these naturalistic tasks requiring decision making under risk, it does not provide an explanation for the overestimated self-reported subjective likelihood observed in the naturalistic studies reviewed above. In prospect theory, the expected utility was determined by *utility * decision weight*. Decision weights represented how much weight was assigned to an outcome based on its probability, called the probability weighting function (Tversky & Kahneman, 1992). The general pattern of the probability weighting function is that it is inverted-S shaped, concave at low probabilities while convex at high probabilities (See Figure 2 for a typical pattern). This pattern describes that at extreme probabilities (like 5%- or 95%+) compared to the middle of the range, every unit change in the probability can elicit a large change in decision weights. This pattern also describes that people tend to overweight small probabilities while underweighting large probabilities. The decision weights were estimated in the original prospect theory experiments based on people's decisions and although they were not assumed to be identical to objective probabilities, they were also not thought to represent self-reported subjective likelihood (Kahneman & Tversky, 1979; Weber 1994). In a study, participants provided self-reported subjective likelihood and made decisions on a rare event with no provided probabilistic information (Barron & Ursino, 2013). Participants had to infer the probability of the event by experiencing a number of trials first (inferring from experience), different from common prospect theory experiments where they were provided probabilistic information (inferring from description). For these rare events, their decisions showed underweighting, but their self-reported subjective likelihood showed overestimating. This suggests that underweighted decision weights and overestimated subjective likelihood can occur at the same time and thus are not necessarily the same. This is not to say, however, that subjective likelihood, usually elicited directly, has no relationship to decision weights calculated from decisions. Indeed, subjective likelihood has been proposed to contribute to decision weights as well (Weber 1994). Subjective likelihood is similarly thought to be among the components leading to people's decisions (Ferrell & McGoey, 1980; Krizan & Windschitl, 2007). In summary, prospect theory does not directly explain overestimated self-reported subjective likelihood observed in some naturalistic studies.

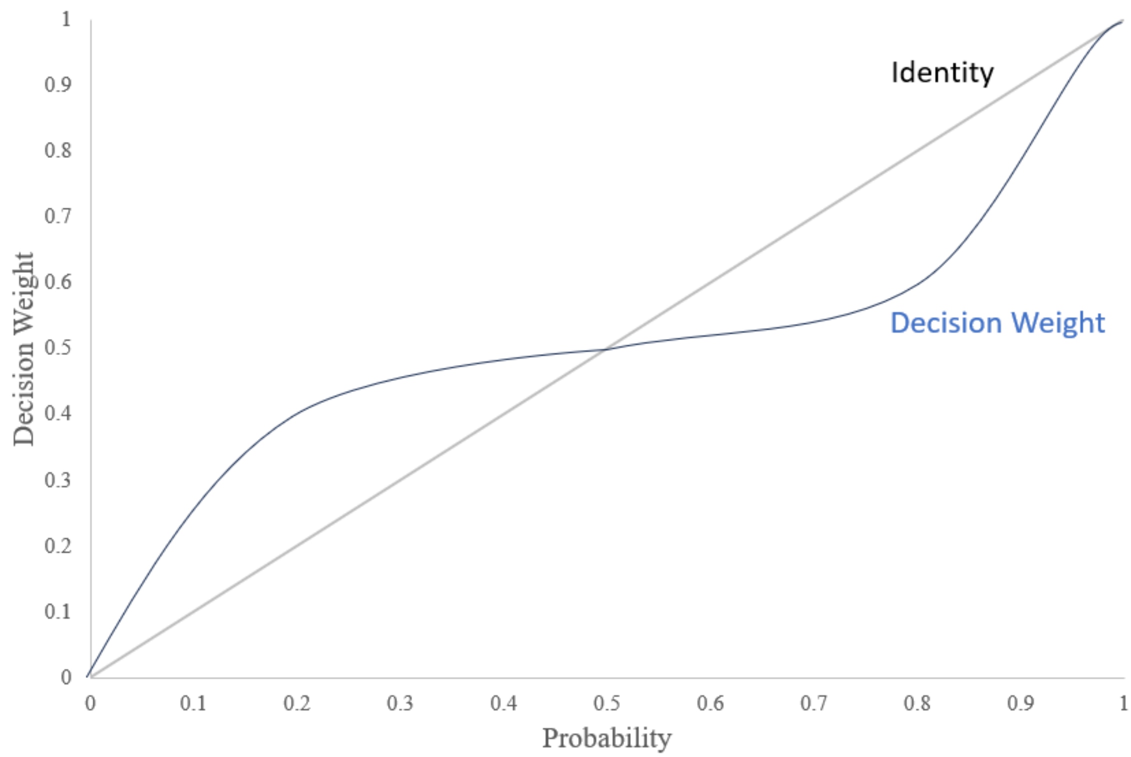


Figure 2
The Typical Probability Weighting Function
The x axis indicates the objective probability of an outcome. The y axis indicates the decision weights of this outcome in a decision. The blue curve represents the decision weight as a function of the probability. At extreme probabilities (like 5%- or 95%+) compared to the middle of the range, every unit change in the probability can elicit a large change in decision weights. People also tend to overweight small probabilities while underweight large probabilities.

The result of overestimated self-reported subjective likelihood and risk-seeking decisions in a loss frame from previous experiments (Gulacsik et al., 2022; Qin et al., 2024) presents a curiosity if considered together: When all else is equal and unbiased, an overestimated self-reported subjective likelihood should have led to risk-averse decisions. If people thought the likelihood of the weather event was higher than it actually was, they should have taken the safe option more often than economically rational. Nonetheless, they were risk-seeking in their decisions. Solving this curiosity is important for studies using naturalistic tasks like weather related settings. These studies aimed to improve people's perception of the likelihood of the event, as represented by self-reported subjective likelihood and with the idea that it might, in turn, improve their decision making (Baker, 1995; Morss et al., 2008; Ash et al., 2014; Lindell et al., 2016; Gulacsik et al., 2022; Qin et al., 2024). With some exceptions (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024), most studies that measured self-reported subjective likelihood did not use a decision task with an economically rational standard (Baker, 1995; Morss et al., 2008; Ash et al., 2014; Lindell et al., 2016; Ash et al., 2014). These studies made the assumption that improving subjective likelihood can lead to better decision-making. While this is valid overall in some cases (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024), the result of risk-seeking decisions and overestimated self-reported subjective likelihood indicates that this is not the whole picture. There must be another component in the decision-making process that leads to the risk seeking bias described above.

This dissertation project examined the curiosity of risk-seeking decisions and overestimated self-reported subjective likelihood in naturalistic weather forecasting tasks using an alternative perspective: Signal detection theory. This alternative perspective assumes that a decision to choose the safe option is made when the subjective likelihood is above a certain likelihood threshold called a criterion. For economically rational decisions, the economically rational probability threshold is the economically rational criterion. It is assumed that people choose the safe option whenever their subjective likelihood is above their subjective criterion, which is not necessarily the economically rational criterion. This dissertation project proposes that a biased subjective criterion is the component in the decision-making process, other than the subjective likelihood, that leads to biased, risk-seeking decisions.

Biased subjective criterion was used to explain biased reasoning in the motivated reasoning literature such that people require more evidence to be convinced that their preferred conclusion is false than nonpreferred conclusion (Kunda, 1990). In outcome prediction artificial gambling tasks with probabilistic information, biased subjective criterion was considered one account for why people had accurate self-reported subjective likelihood but were more likely to predict their desired outcome to happen than the undesirable outcome (Windschitl et al., 2010). The account suggests that while people's subjective likelihood remained the same, they had a lower subjective criterion to predict a desirable outcome to happen than to predict an undesirable outcome. The same component of subjective criterion might also be biased in naturalistic weather tasks requiring decision making under risk. When subjective likelihood is well calibrated to objective probabilities, a higher subjective criterion can lead to a decision bias towards the risky option, as people require higher subjective likelihood in order to choose the safe option. Theoretically a high enough subjective criterion can counteract overestimated subjective likelihood (detected in previous similar research) and lead to risk-seeking decisions. For example, if the objective probability is 30% and the economically rational criterion is 20%, one should choose the safe option. However, this person chooses the risky option because their subjective likelihood is 40% and their subjective criterion is 50%. In this case, despite the overestimated subjective likelihood, this person shows a decision bias towards the risky option.

Signal Detection Theory and Random Likelihood and Fixed Criterion Model

In perceptual experiments, signal detection theory concerns two separate variables: people's *internal representation of the stimulus* and their *subjective criterion* (See Figure 3, Macmillan & Creelman, 2005). Separate internal representations are generated, one representing that the stimulus is present and the other that it is absent. They are assumed to be noisy and follow a normal distribution. This means that the same external signal might be internally represented differently each time. The ability to differentiate these two internal representations representing that the stimulus is present and absent is the person's *sensitivity*. The sensitivity is jointly determined by the distance between the two internal representation distributions and how noisy they are. The further apart they are, the less the two distributions overlap, and in turn the greater the sensitivity. The less noisy the two distributions are, the less they overlap, and in turn the greater the sensitivity. The subjective criterion is the strength of the internal representation above which the presence of the stimulus is reported. The subjective criterion indicates whether there is a decision bias towards reporting a presence or an absence of the stimulus.

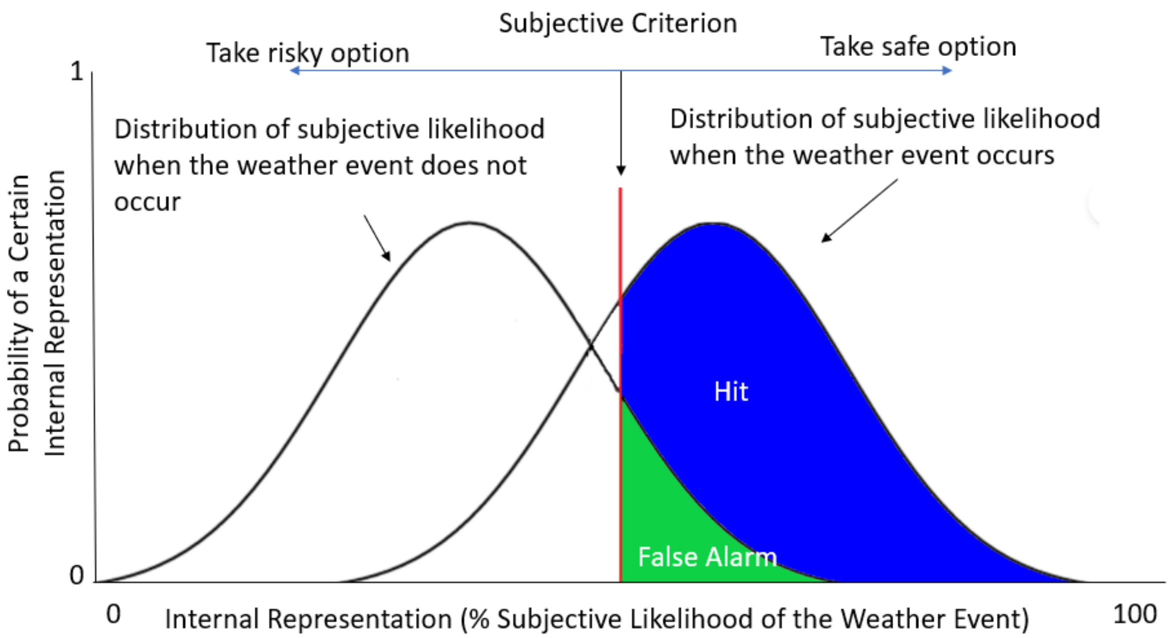


Figure 3
Random Likelihood Model Example Based on Signal Detection Theory
Note. The random likelihood model does not assume the shape of the internal representation distribution. Normal distributions are used in the graph as an example.

Given the same sensitivity, the noisy internal representation might fall below the subjective criterion when the stimulus is present, or it might fall above the subjective criterion when the stimulus is absent. Both instances lead to errors. Therefore, there are four possible outcomes in a stimulus detection task, two correct outcomes and two errors: Hit, false alarm, miss, and correct rejection (See Table 1). A hit is when people correctly report the presence of the stimulus when it is present. A false alarm is when people erroneously report the presence of the stimulus when it is absent. A miss is when people erroneously report the absence of the stimulus when it is present. A correct rejection is when people correctly report the absence of the stimulus when it is absent. The theory suggests that under the same sensitivity, a lower subjective criterion (a decision bias towards reporting a presence of the stimulus) leads to an increase in both hits and false alarms while a high subjective criterion (a decision bias towards reporting an absence) leads to a decrease in both these outcomes. In short, a change in subjective criterion can lead to a behavioral change even if the internal representation distributions stay the same.

	Stimulus Present	Stimulus Absence
Reporting a presence	Hit	False Alarm
Reporting an absence	Miss	Correct Rejection

Table 1
Stimulus-Response Table Used in Signal Detection Theory

This dissertation project applied signal detection theory to naturalistic weather decision tasks requiring decision-making under risk similar to those described above. The four outcomes in these tasks aligned with outcomes in signal detection theory: 1) Hit: Participants chose the safe option and the weather event occurred; 2) False alarm: Participants chose the safe option and the weather event did not occur; 3) Miss: Participants chose the risky option and the weather event occurred: 4) Correct rejection: Participants chose the risky option and the weather event did not occur. This allows the application of the theory to participants' behavior in these tasks. Subjective likelihood is considered the internal representations of whether or not the weather event will occur, using a similar method to a previous application of signal detection theory (Ferrell & McGoey, 1980). The subjective criterion can be expressed as a subjective likelihood threshold at which the expected utility of the options breakeven. This is parallel to how the economically rational criterion is expressed as an objective probability at which the expected value of the options breaks even. Participants choose the safe option whenever their subjective likelihood is above their subjective criterion. Therefore, the subjective likelihood of the weather event and the subjective criterion jointly determine the decision.

Signal detection theory is applied in this project by assuming a random likelihood model (Figure 3), analog to the random utility model where the given value is translated to a noisy utility (Bockenholt, 2006). The random likelihood model assumes that people's subjective likelihood of the weather event and their subjective criterion are the sole determents of their decision. In this model, the subjective likelihood of the weather event has variability while the subjective criterion has no variability. This model predicts that the subjective likelihood varies over the trials even when given the same external predictive information (objective probability in this case). The model does not assume a specific shape of distribution for this variability. This is because in this project subjective likelihood was elicited directly and therefore its actual distribution was known. For example, in Figure 3, some part of the subjective likelihood distribution falls higher than the subjective criterion even when the center of the subjective likelihood distribution is lower than the subjective criterion, resulting in false alarms. This assumed variability is consistent with signal detection theory where the internal representation is noisy (Macmillan & Creelman, 2005). This assumption is also consistent with previous studies with naturalistic decision-making tasks where the subjective likelihood was different in the same participant in different trials with the same displayed probabilistic information (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). In this project, this random variability of subjective likelihood is considered the result of an amalgamation of many factors. For example, experience of multiple weather events in preceding trials was found to affect the behavior in the following trials, perhaps due to the availability heuristic, where an event is deemed more likely if an event of the same type is experienced more recently and therefore, more likely to come to mind (Kahneman, 2003; Demnitz & Joslyn, 2020). A learning effect might also have taken place to vary the subjective likelihood from trial to trial.

On the other hand, the random likelihood model assumes that the subjective criterion has no variability, consistent with the typical application of signal detection theory (Macmillan & Creelman, 2005). This fixed subjective criterion is a simplifying assumption used in the model. People might change their subjective criterion between decisions. One example was that when people experienced adverse weather in recent trials, they tended to choose the safe option more often despite the unchanged self-reported subjective likelihood (Demnitz & Joslyn, 2020). With this assumption, as the subjective likelihood of the weather event given the same information can fluctuate above or below the subjective criterion, this model predicts that one's decision might not always be the same when given the same information.

This model allows systematic bias in the subjective likelihood or the subjective criterion. Systematic biases in either of these two can shift one's decision away from the economically rational decision. This model allows the subjective likelihood and the subjective criterion to affect the decision in three separate ways. 1) A higher subjective criterion can lead to a decision bias towards the risky option while a lower subjective criterion can lead to a bias towards the safe option. 2) A bias in the subjective likelihood can also lead to a shift of both the internal representation of the weather occurring and that of the weather not occurring (See Figure 3). Higher subjective likelihood shifts the internal representations right, effectively leading to a decision bias towards the safe option. Lower subjective likelihood shifts the internal representations left, leading to a bias towards the risky option. This bias is assumed to not affect the sensitivity as both internal representation distributions are moved equally. 3) A greater sensitivity either resulting from increased distance between the distributions or reduced noise can lead to more hits and less false alarms regardless of the decision bias. This dissertation project focused on the subjective criterion while the sensitivity and subjective likelihood were also examined.

As mentioned above, for naturalistic tasks requiring decision making under risk, prior studies have found that there is a small but consistent trend of overestimated self-reported subjective likelihood from the objective probability (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). This does not explain the tendency toward risk-seeking decisions in loss frame tasks. Therefore, there might also be a bias in the subjective criterion. However, with one exception with a self-report measurement in a decision task (Joslyn & Grounds, 2015), the subjective criterion was not examined in similar studies. This dissertation project used the random likelihood model to examine this possible systematic bias in the subjective criterion in naturalistic tasks requiring decision making under risk. This project proposes that biases in the subjective criterion can lead to biased decisions in naturalistic weather tasks requiring decision making under risk. This is called the subjective criterion hypothesis.

Possible Sources on the Subjective Criterion

Two sources that might shift the subjective criterion from the economically rational criterion were the gain-loss framing effect and the centering effect. It is possible that the decision biases in the gain and loss frame, as explained by prospect theory, manifest as biased subjective criteria. Risk-aversion in the gain frame might correspond to a lower subjective criterion while risk-seeking tendency in the loss frame might correspond to a higher subjective criterion. This can potentially explain the risk-seeking decisions observed in loss frame tasks, despite the overestimated self-reported subjective likelihood (Gulacsik et al., 2022; Qin et al., 2024). In these cases, the higher subjective criterion due to the loss frame might have counteracted the overestimated self-reported subjective likelihood.

On the other hand, the subjective criterion might also be affected by a centering effect (also called central tendency bias; Poulton, 1979; Olkkonen et al., 2014). This effect suggests that when judging a quantity, people tend to bias their judgement towards the center of the quantity's range (Poulton, 1979). This effect traced its roots in perceptual experiments and has been observed in many different settings, including noise volume, distance, color perception, and estimation of opposing bidders' bid in an auction (Poulton, 1979; Radvansky et al., 1995;

Olkkonen et al., 2014). It is possible that a similar effect is also present in the subjective criterion and subjective likelihood in decision-making tasks.

It was also found that when the information regarding the quantity in question was available, the centering effect was reduced (Radvansky et al., 1995). The economically rational criterion in the naturalistic weather decision tasks were not readily available unless calculated by participants (which they might or might not do; Joslyn & LeClerc, 2013; Grounds & Joslyn, 2017; Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). This left room for centering. In contrast, while not examined in the original studies, centering did not seem to affect the self-reported subjective likelihood in these studies because a general overestimation was observed, rather than overestimation at lower likelihoods and underestimation at higher likelihood that one might expect if centering was influencing participants responses (Figure 4 & Figure 8 from Qin et al., 2024). The lack of centering might have been due to the probabilistic information provided to participants in these studies. This dissertation project proposes this centering as another effect shifting the subjective criterion. The subjective criterion might be affected by this bias such that it shifts towards 50% (the center of the range) from the lower-than-50% economically rational criterion, potentially raising it to lead to risk seeking decisions.

Reanalysis of Previous Experiments

This dissertation project first analyzed the subjective likelihood, subjective criterion, and the sensitivity of two experiments from Qin et al., 2024. The purpose of these analyses was to examine the random likelihood model method of calculating the subjective criterion. The analysis further considered whether the gain-loss framing effect and the centering effect might explain the results in existing experiments with a naturalistic decision making under risk task.

Method of Tornado Experiments

In these two experiments, participants made decisions on whether to take shelter from possible tornadoes (safe option) or not (risky option) based on a tornado warning. These two experiments were conducted online in 2018 and 2019 with US participants recruited from Amazon Mturk, an online crowdsourcing platform for participant recruitment. The format of the tornado warning was manipulated between groups. The analysis here only included two *probabilistic formats (red format and tabular format)* that were present in both experiments. In these formats, participants were given reliable percent chance information about the tornado (e.g., 30% chance of tornado) either in the form of a color-coded visualization (red format) or a numeric percentage (tabular format). As the difference between these conditions was not the concern of this dissertation project, they were combined. For these conditions, the number of participants was 83 for tornado experiment 1 and 85 for tornado experiment 2.

The procedure of the two experiments was identical. There were 68 trials in total per participant. On each trial, participants saw a tornado warning, rated how likely they thought the tornado would be (likelihood ratings), and decided between the safe option and the risky option (binary decision). At the end of each trial, they were told whether the tornado occurred. Four outcomes were possible: 1) They chose the safe option and the tornado occurred (a hit); 2) They chose the safe option and the tornado did not occur (a false alarm); 3) They chose the risky option and the tornado occurred (a miss); 4) They chose the risky option and the tornado did not occur (a correct rejection). The outcomes of their decisions were represented with a point structure (see the section below). The experiment was framed in a loss frame such that the risky option had a large point penalty while the safe option mitigated the loss for a smaller fixed-point cost. Participants' goal was to lose as few points as possible by the end of the experiment. They were paid a monetary bonus commensurate with their ending point balance.

The main difference between the two experiments was the point structure (see Table 2). In both experiments, the risky option had no cost but would incur a 1,000-point penalty if there was a tornado on that trial. Participants could mitigate this loss completely by choosing the safe option for a fixed cost. In tornado experiment 1, this cost was 90 points. In tornado experiment 2, this cost was 270 points. Therefore, the probability of the tornado where the expected value of the risky option (probability of the tornado * 1,000) broke even with the cost of the safe option was 9% in tornado experiment 1 and 27% in tornado experiment 2. This provided an economically rational standard for participants' decisions based on expected value. When the probability of a tornado in a trial was above this standard, the expected value of risky option was worse than the cost of the safe option. Therefore, it was economically rational to choose the safe option. On the other hand, when the probability was below the standard, it was economically rational to choose the risky option. The 9% and 27% were hence considered the economically rational criterion of tornado experiment 1 and 2 respectively. Other than this key difference, another difference was that in tornado experiment 1 the overall proportion of trials with a tornado was 23.5% while in tornado experiment 2 it was 38.2%. As the objective probabilities (provided to participants) were mostly reliable (M = 24.4% for tornado experiment 1 and 37.4% for tornado experiment 2), the mean objective probability and the proportion of drought trials were considered the same (23.5% for tornado experiment 1 and 38.2% for tornado experiment 2) in the analyses.

Experiment 1	Safe Option	Risky Option
Tornado Occurred	-90 points	-1000 points
Tornado Did Not Occur	-90 points	0 points

Experiment 2	Safe Option	Risky Option
Tornado Occurred	-270 points	-1000 points
Tornado Did Not Occur	-270 points	0 points

Table 2
Point Structure of the Two Tornado Experiments

In summary, in both tornado experiments, participants received probabilistic information about the tornado. Their likelihood ratings and binary decisions were measured. The main difference was that the economically rational criterion was 9% in tornado experiment 1 and 27% in tornado experiment 2.

Results

Three sets of analyses were conducted on each dataset to examine the three possible ways the decisions can be affected: Mean calculated subjective criterion for each experiment, mean likelihood ratings, and Receiver Operating Characteristic (ROC) plots constructed from likelihood ratings and decisions. ROC plots are used in signal detection theory to indicate the sensitivity of participants for predicting a tornado, in this case. The greater the area under the curve the greater is sensitivity (see Figure 4). The ROC curve also indicates how participants' hit and false alarm probability would change given different subjective criterion. Inferential statistics in the form of t-tests were conducted with an alpha of .05.

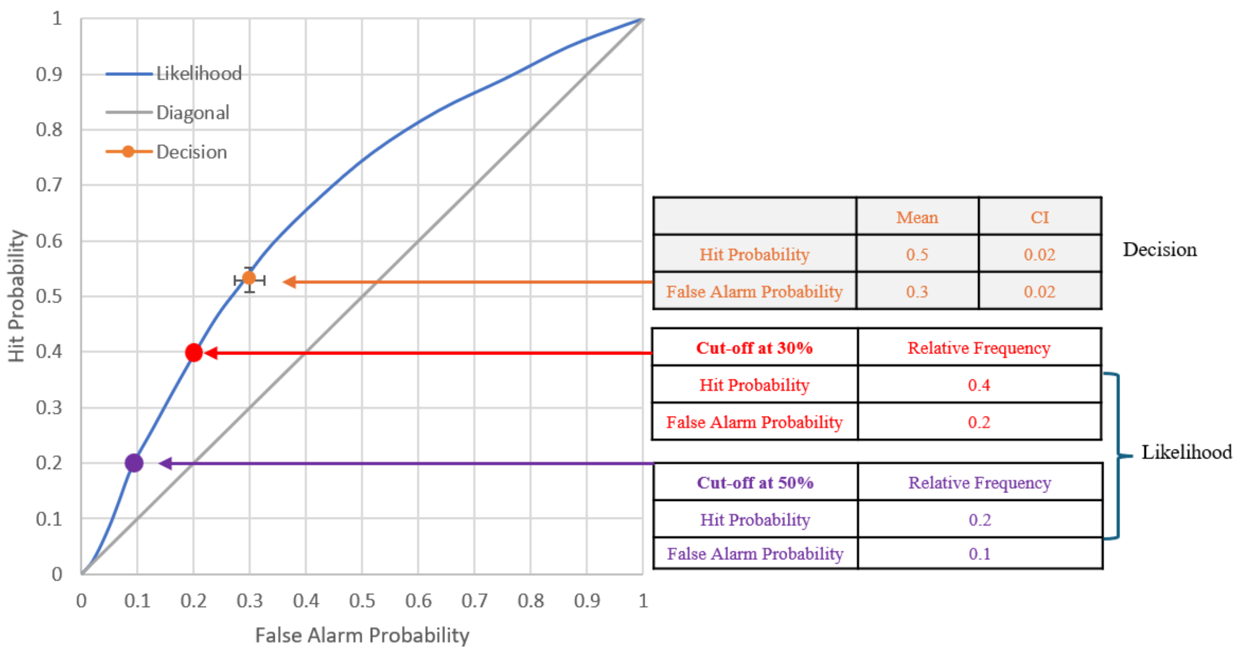


Figure 4

Calculation of the ROC plots with Hypothetical Data

For each hypothetical cut-off point, a pair of hit and false alarm probabilities was obtained from the relative frequency of likelihood ratings above the cut-off. For example, if the cut-off is at 30%, 40% of the tornado trials had a rating above the cut-off (hits) while 20% of the no tornado trials had a rating above the cut-off (false alarms). There were 20 possible cut-offs and thus 20 possible pairs of hit and false alarm probabilities. These pairs were plotted to form the ROC curve. The orange point was calculated from participants' decisions. A pair of hit and false alarm probabilities of actual decisions were calculated for each participant. The mean hit and false alarm probability were then plotted as the orange point.

Calculated Subjective Criterion

The mean calculated subjective criterion was estimated using each participant's tornado likelihood ratings and their shelter option decisions in each experiment. Its calculation was based on the assumption of the random likelihood model that participants chose the safe option whenever their subjective likelihood (operationalized here as their tornado likelihood rating) was higher than their subjective criterion and that the subjective likelihood followed a distribution while the subjective criterion had no variability. Therefore, an assumption of the random likelihood model was that participants' subjective criterion and subjective likelihood were the sole determinants of their decision.

The subjective criterion was calculated from their decisions and the distribution of likelihood ratings. For each participant, the cumulative proportion of likelihood ratings falling between X% chance and 100% chance, such that the proportion of trials in between matched the proportion of trials on which the participant chose the safe option, was calculated. This point was called the *calculated subjective criterion* because if the participant always chose the safe option when their likelihood rating was above this point, their proportion of trials with a likelihood rating higher than their subjective criterion would be the same as the observed proportion of trials in which they chose the safe option. The calculated subjective criterion was expressed as a percent with a possible range from 0 to 100%. For example, say a participant chose the safe options on 50% of the trials, then a number was located on their likelihood rating distribution such that on 50% of the trials had a likelihood rating higher than this number (See Figure 5). This number, say 40% likelihood rating, was regarded as the calculated subjective criterion for this participant. With this method, a calculated subjective criterion was obtained for each participant and a mean was calculated for each experiment.

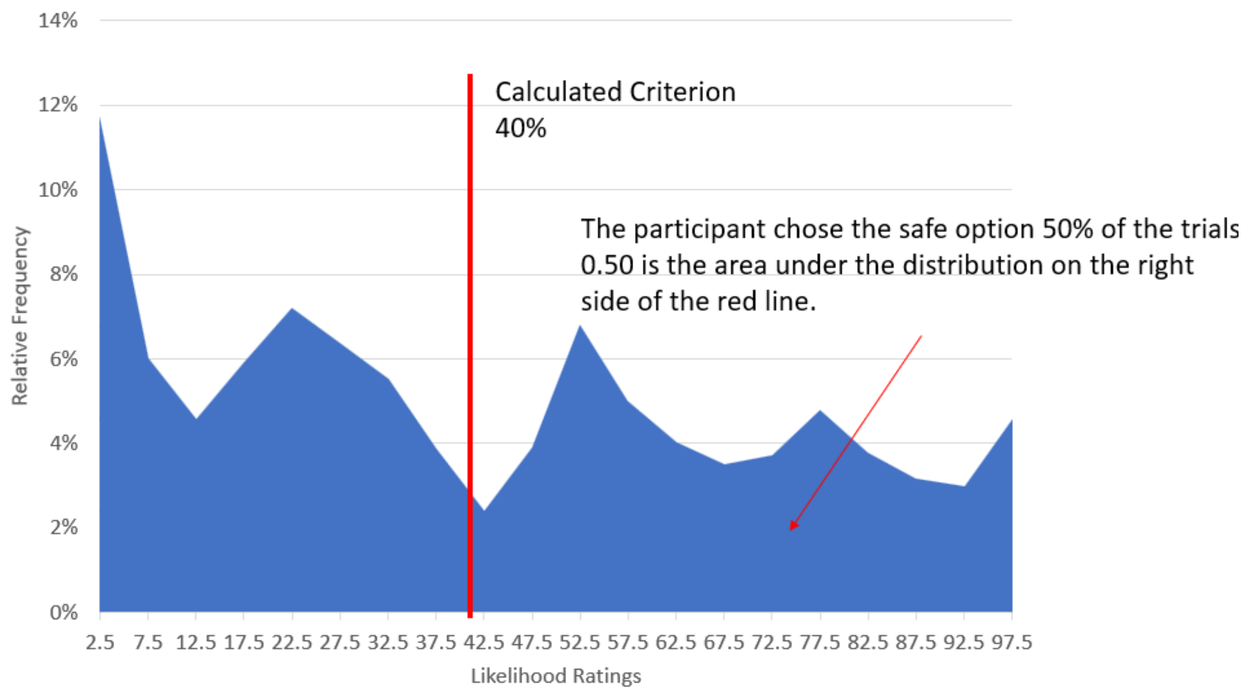


Figure 5
Example of Calculation of Subjective Criterion
The blue area is the likelihood rating distribution of a hypothetical participant. This distribution and the relative frequency of choosing the safe option was known. A calculated criterion was then placed at the line of 40% where 50% of the trials were on the right side of the line. This meant that if participants used a fixed subjective criterion of 40% and always chose the safe option when the likelihood rating was above the subjective criterion, they would have chosen the safe option in 50% of the trials.

In Figure 6, the calculated criterion is shown for tornado experiments 1 and 2. The x axis shows the economically rational criterion, and the y axis shows the calculated criterion. The diagonal line represents when the calculated criterion matches the economically rational criterion. If the calculated criterion is above this line, it is considered biased towards the risky option. The dashed line represents a calculated criterion of 50% - the center of the range. An independent t-test revealed that the mean calculated criterion in tornado experiment 1 ($M = 31.0\%$, $SD = 18.6\%$) was significantly lower than in tornado experiment 2 ($M = 41.2\%$, $SD = 13.8\%$) with a difference of -10.2% ($t(148.6) = 4.07$, $p < .001$). In addition, the mean calculated criterion in each tornado experiment was compared to the respective economically rational criterion in two one-sample t-tests. In tornado experiment 1, the calculated criterion was significantly higher than the economically rational criterion of 9% with a difference of 22% ($t(82) = 10.77$, $p < .001$). In tornado experiment 2, the calculated criterion was significantly higher than the economically rational criterion with a difference of 14.2% ($t(84) = 9.81$, $p < .001$). The calculated criterion was higher than the economically rational criterion in both experiments, but more so in the experiment with the lower economically rational criterion. Therefore, there is evidence of a biased subjective criterion.

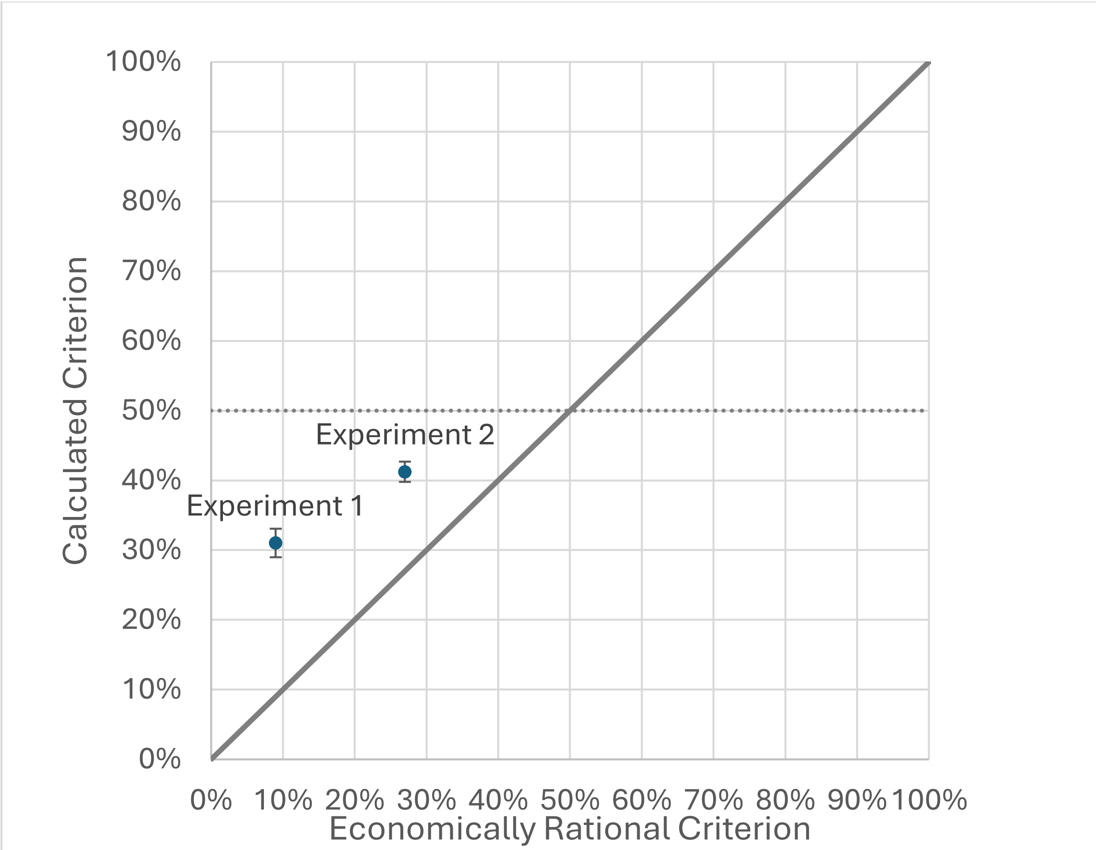


Figure 6
Calculated Subjective Criterion in the Two Tornado Experiments
The difference of 10.2% between the two experiments was significant

Likelihood Ratings

The analysis of mean likelihood ratings indicated whether there was a bias in subjective likelihood. Figure 7 shows the likelihood ratings for tornado experiment 1 and 2 as a function of objective probabilities. The mean likelihood rating in each experiment was compared to the respective actual proportion of tornado trials with two one-sample t-tests. In tornado experiment 1, the mean likelihood rating ($M = 33.7\%$, $SD = 10.0\%$) was significantly higher than the proportion of tornado trials of 23.5% with a difference of 10.2% ($t(82) = 9.33$, $p < .001$). In tornado experiment 2, the mean likelihood rating ($M = 42.7\%$, $SD = 7.5\%$) was 4.5% higher than the proportion of tornado trials of 38.2% ($t(84) = 5.48$, $p < .001$). In addition, Figure 7 shows that the overestimation was observed at most objective probability levels except for the extreme high end.

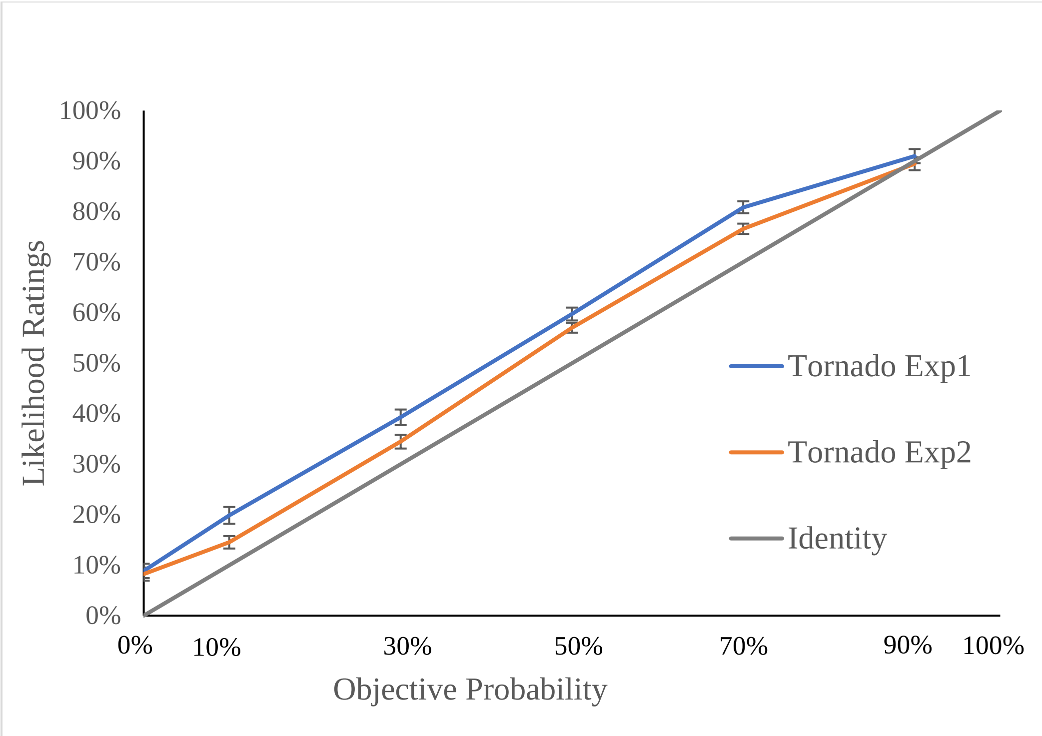


Figure 7

Likelihood Ratings as a Function of Objective Probabilities in Tornado Experiment 1 and 2

The blue line represents the likelihood ratings of tornado experiment 1. The orange line represents the likelihood ratings of tornado experiment 2. There was no observed centering effect.

Receiver Operating Characteristic Plots

The ROC plots analysis had two aims: 1) Illustrating the random likelihood method of calculating the subjective criterion, 2) Examining whether the sensitivity was different between experiments by calculating the area under ROC curves. The sensitivity analysis can reveal whether participants' ability to predict the tornado was the same in the two experiments.

The ROC plots were composed of two parts: 1) Receiver Operating Characteristic curves, based on signal detection theory, were created from participants' likelihood ratings and actual tornado occurrence; 2) Points on the plots representing the outcomes of participants' binary decisions (safe option/risky option; See Figure 4 for a hypothetical ROC plot with both the curve (blue) and the point (orange)) were also created.

The ROC curves were created by estimating hit and false alarm rates based on the actual tornado occurrence, the likelihood rating distribution, and a varying hypothetical cut-off on the distribution with a method similar to Ferrel & McGoey (1980). Participants should have high likelihood ratings on tornado trials and low ratings on no tornado trials. This was estimated by a varying hypothetical cut-off on the likelihood rating. A hit was defined as a trial on which the likelihood rating was above the cut-off and a tornado occurred. This indicated that the likelihood rating was high on a tornado trial. The hit probability was the relative frequency of tornado trials above the cut-off. A false alarm was a trial on which the likelihood rating was above the hypothetical cut-off and a tornado did not occur. This indicated that the likelihood rating was high on a no tornado trial. The false alarm probability was the relative frequency of no tornado trials above the cut-off. By varying the hypothetical cut-off at 5% steps from 0% to 100% on the likelihood rating frequency distribution, a pair of hit probability and the false alarm probability of likelihood ratings at each step (20 in total) was calculated (See Figure 8 for an example). The 20 pairs were plotted as the ROC curve (the blue curve in Figure 4).

Table 3 shows an example of two pairs of hit and false alarm probabilities used in Figure 4 with hypothetical data. In the example, if the hypothetical cut-off was at 30% likelihood rating, in 40% of tornado trials the likelihood rating was above the cut-off (hit probability) and in 20% of no tornado trials the ratings were above the cut-off (false alarm probability). If the hypothetical cut-off was increased to 50%, the hit probability became 20% and the false alarm probability became 10%. In summary, points on the ROC curves indicated the hit and false alarm probabilities of likelihood ratings above 20 different hypothetical cut-offs in tornado trials and no tornado trials.

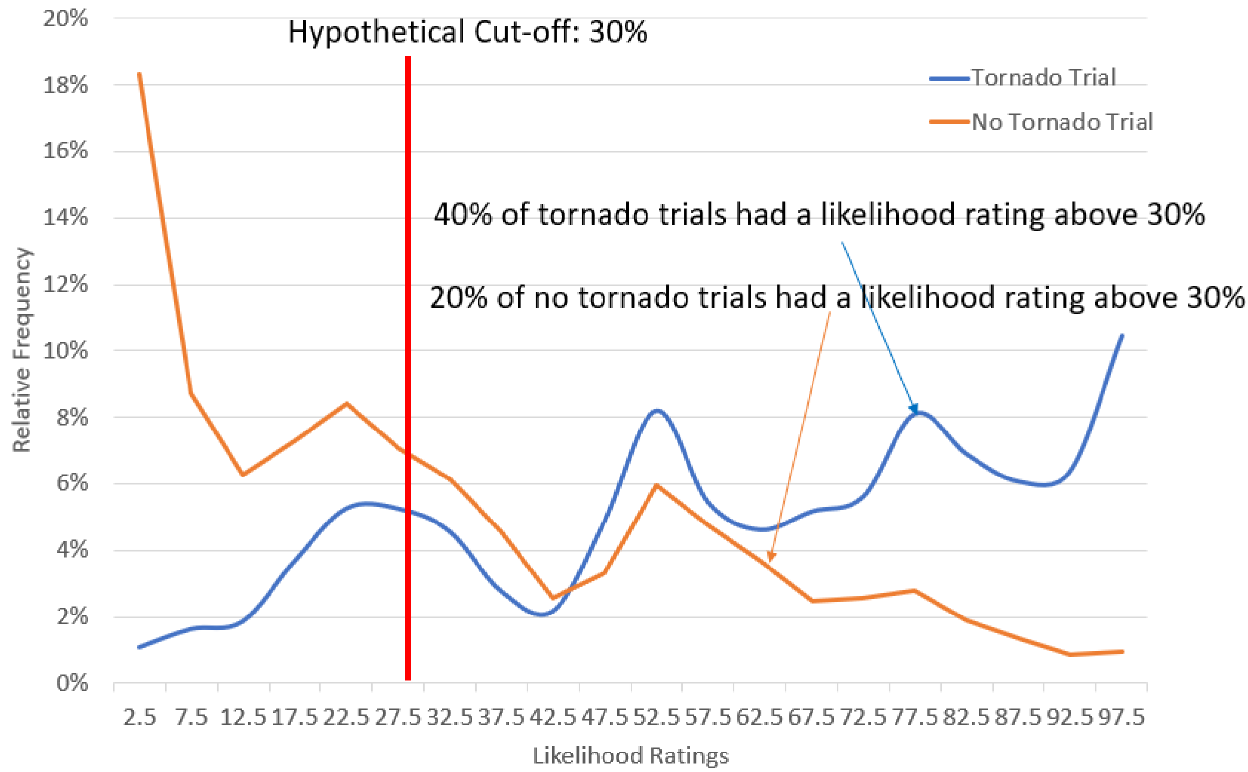


Figure 8
Example of Hit and False Alarm Probability Calculation with Hypothetical Data
The hypothetical cut-off is placed at 30%. Using this cut-off, 40% of tornado trials (hits) had a likelihood rating above 30%. The hit probability is therefore 40%. 20% of no tornado trials (false alarms) had a likelihood rating above 30%. The false alarm probability is therefore 20%.

Hypothetical Cut-off at 30% likelihood rating

	Relative Frequency of Trials with Likelihood Rating Below the Cut-off	Relative Frequency of Trials with Likelihood Rating Above the Cut-off
Tornado Trial	60% (Miss Rate)	40% (Hit Rate)
No Tornado Trial	80% (Correct Rejection Rate)	20% (False Alarm Rate)

Hypothetical Cut-off at 50% likelihood rating

	Relative Frequency of Trials with Likelihood Rating Below the Cut-off	Relative Frequency of Trials with Likelihood Rating Above the Cut-off
Tornado Trial	80% (Miss Rate)	20% (Hit Rate)
No Tornado Trial	90% (Correct Rejection Rate)	10% (False Alarm Rate)

Table 3
An Example of Hit and False Alarm Rate of Likelihood Ratings with Different Hypothetical Cut-offs
Note. This table used hypothetical data.

Next, a point representing the proportion of hits and false alarms of participants' actual binary decisions was added to the ROC plot of each condition (the orange dot in Figure 4). For this point, a hit was when the participant chose the safe option and the tornado occurred. A false alarm was when the participant chose the safe option and no tornado occurred.

The first goal of the ROC plot analysis was to test the random likelihood method of calculating the subjective criterion. If participants always chose the safe option whenever their likelihood rating was above their subjective criterion, then the hit and false alarm probabilities of their likelihood ratings above a hypothetical cut-off would be the same as the hit and false alarm probabilities of binary decisions with a subjective criterion of the same number. This suggests that the decision was solely determined by a likelihood rating and a subjective criterion. If the mean binary decision point had its 95% CI overlapping with the likelihood rating ROC curve, then the mean binary decision point was considered consistent with the ROC curve.

In the next ROC plot analysis, the mean percent area under the ROC curve was measured as the sensitivity (ability to predict the tornado). The greater the area under the curve the greater the sensitivity or the ability to distinguish signal from noise. Figures 9 & 10 show the ROC plots for tornado experiment 1 and 2. In both experiments, the ROC curve and the 95% CI of the decision point overlapped. This indicates that the decision point was consistent with the ROC curve. An independent t-test revealed that the mean percent area under the ROC curve was significantly lower in tornado experiment 1 ($M = 64.0\%$, $SD = 6.8\%$) than in tornado experiment 2 ($M = 78.1\%$, $SD = 6.8\%$) with a difference of -14.1% ($t(165.86) = 13.41$, $p < .001$). This indicates a worse sensitivity in experiment 1 than experiment 2.

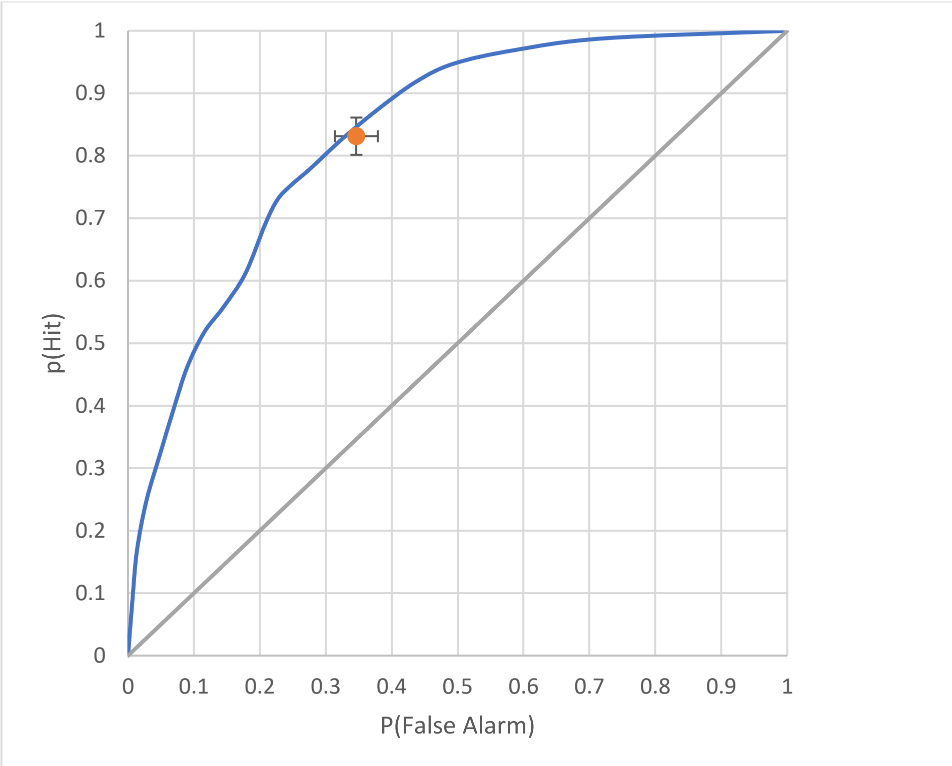


Figure 9
ROC Plot for Tornado Experiment 1
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The mean percent area under curve was 64.0%.

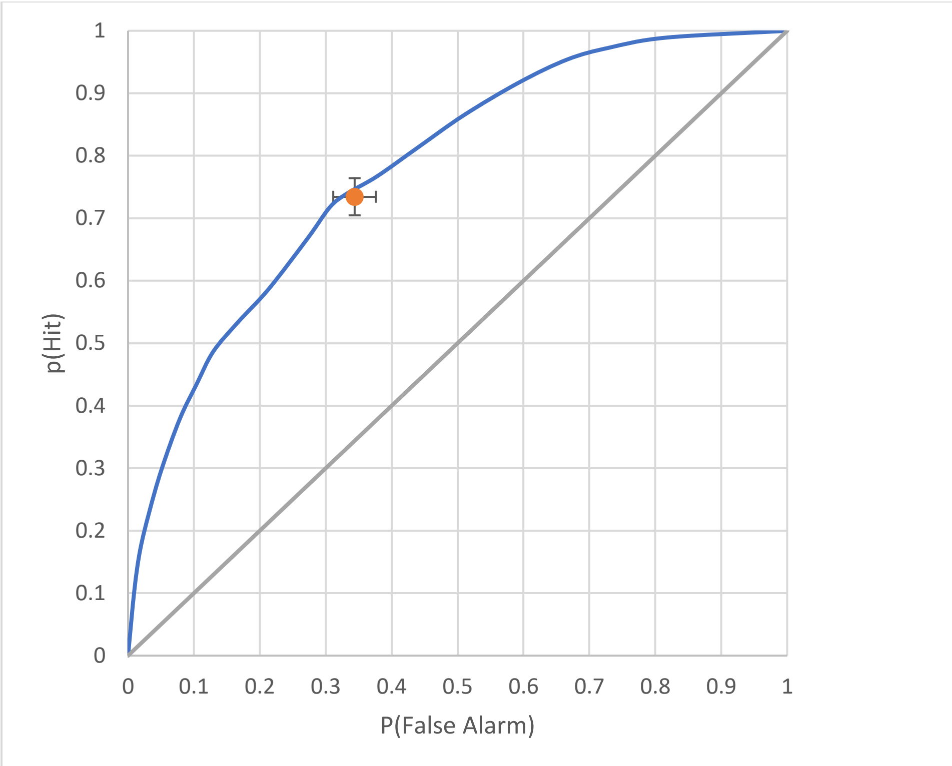


Figure 10
ROC Plot for Tornado Experiment 2
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The mean percent area under curve was 78.1%

Discussion

The reanalysis of the tornado experiments indicates that participants' subjective criterion was higher than the economically rational criterion in both experiments, which could lead to a risk-seeking tendency. This is consistent with the loss framing leading to risk-seeking behavior. In addition, the subjective criterion was closer to 50% than the economically rational criterion in both experiments. This result is consistent with the centering effect. Because the tornado experiments did not have a gain frame condition nor a condition where the economically rational criterion was higher than 50%, it is not possible to distinguish between these two effects. Hence in the new experiments conducted as part of this dissertation project and reported below, these conditions were added to distinguish the gain-loss framing effect and the centering effect.

The mean likelihood rating analysis showed that the likelihood ratings were significantly overestimated in both experiments. The likelihood ratings also showed no centering effect as there was similar amount of overestimation across all objective probability levels except for 90%. When all else is equal, overestimated likelihood ratings should have led to risk-averse decisions. However, the higher-than-rational subjective criterion counteracted the overestimated likelihood ratings and led to risk-seeking decisions instead. On the other hand, as the trial composition (e.g., proportion of trials with a tornado) in the two experiments was different, their mean likelihood ratings were not comparable with each other. In the new experiments in this project, the proportion of trials with a weather event was held constant among conditions to allow for that comparison.

In addition, participants had different mean sensitivity in the experiments. However, as the two experiments had different trial compositions (e.g., the number of trials at each objective probability level and the proportion of tornado trials), it cannot be inferred that this difference stemmed from the change of the economically rational criterion between the two experiments. This is because participants could predict whether a tornado would happen better in trials with extreme (e.g., 10% or 90%) than mid-range (e.g., 50%) objective probabilities. The two experiments differed in the proportion of these trials. Hence in the experiments of this dissertation project, the trial composition was held constant across the conditions to remove it as a potential confound.

More importantly, in the ROC plots of both of the reanalyses of the tornado experiments, the 95% CI of the binary decision point overlapped with the ROC curve. This indicates that the proportion of hits and false alarms of participants' actual binary decisions was consistent with the ROC curve created with likelihood ratings and actual tornado occurrence. This in turn suggests that the subjective criterion and subjective likelihood (likelihood ratings) were the sole determinants of the decisions. This supports the random likelihood method of calculating the subjective criterion as a way of operationalizing the subjective criterion.

One limitation of the reanalysis was that an analysis of the relative frequency of choosing the safe option could not be conducted in the reanalysis. This analysis was often used in previous naturalistic weather decision tasks to examine the decision bias (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2017; Demnitz & Joslyn, 2020; Klockow-McClain et al., 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). It could serve as another, non-signal detection theory analysis of the decision bias. However, the relative frequency can be affected by trial composition or sensitivity, in addition to the decision bias. Therefore, a conclusion about decision bias (e.g., more risk seeking or risk averse) is not possible based on a difference in relative frequency of choosing the safe option when the sensitivity and the trial compositions are different. This was another reason why the trial composition was kept constant among conditions in the new experiments in this project. If the sensitivity was also the same in the new experiments, then the analysis of the relative frequency could be used to examine the decision bias.

Overall, the reanalyses indicated a bias in the subjective criterion that overcame the bias in subjective likelihood and led to risk-seeking decisions in the two tornado experiments. It left the door open to a possible gain-loss framing effect and a centering effect.

Overview of Experiments

This project examined the subjective criterion hypothesis and possible effects leading to risk-seeking decisions and overestimated subjective likelihood in a naturalistic weather decision task from a signal detection theory perspective. A drought task based on Demnitz & Joslyn (2020) was used. It was chosen as it was easier to frame as a gain or a loss compared to the tornado setting in the reanalyses. The focus was to discriminate a bias in subjective criterion from a bias in subjective likelihood and/or a difference in sensitivity. Two manipulations were examined: 1) Manipulation of the gain-loss framing. According to prospect theory, people are usually risk-averse in a gain frame while risk-seeking in a loss frame. If this were the case, a lower subjective criterion in a gain frame compared to a loss frame was expected. To rule out the possibility of this manipulation affecting subjective likelihood or sensitivity, they were also examined. 2) Manipulation of the economically rational criterion to expose the centering effect. According to the centering effect, when people provide a numeric rating with a finite range, they tend to steer their rating towards the center of this range. Therefore, if this were the case the subjective criterion was expected to shift towards 50%. Again, to rule out the possibility of this manipulation affecting subjective likelihood or sensitivity, these variables were also examined.

This project had three experiments. Experiment 1 examined the manipulation of gain-loss framing on the shift of the subjective criterion, subjective likelihood, and sensitivity. Experiment 2 manipulated the economically rational criterion to examine centering on the same three variables. Experiment 3 combined both manipulations to allow a comparison between the two effects within a single experiment.

Experiment 1

Experiment 1[4] focused on the framing effect on subjective criterion by manipulating the economically rational criterion. The effect of this manipulation on subjective likelihood and sensitivity was also examined. Experiment 1 used a drought task where the framing was between-group manipulated in which the outcomes of participants' decisions were framed as either gains or losses. This manipulation was expected to affect only the subjective criterion and not the subjective likelihood or sensitivity.

Method

Participants

A total of 110 participants from the US were recruited from Prolific Academic in December 2023, a crowdsourcing platform for online research. Each participant was paid $4 for participation plus a performance based monetary bonus. After an elimination process, 96 participants were used in the analysis. 11 were eliminated for having a lower than 0.7 ReCAPTCHA score, a bot detection system used by Qualtrics survey platform. Three were eliminated for failing the comprehension check (see below). The mean age was 38 years (SD =

[4] A pilot study with near identical procedure using university students was conducted before experiment 1 showing a significant gain-loss framing effect.

12.49, range 19 to 68 years). There were 37 (39%) females, 58 (60%) males, and 1 (1%) who preferred not to say.

Procedure and Stimuli

The experiment was hosted on Qualtrics. The experiment information and link to the Qualtrics survey were posted on Prolific inviting potential participants who were residents of the US to participate. The informed consent form was displayed on the first page of the Qualtrics survey. Participants were instructed to click next and continue if they consented, or to close the program and cancel their participation on Prolific if they did not consent.

After providing informed consent, participants were provided instructions to a task in which they decided which crop to plant based on climate projections concerning possible droughts. This task is henceforth called the *drought task*. Figure 11 to 16 illustrate the procedure of the drought task, including the instructions shown to the participants and the questions they answered. In the task, participants played the role of an agricultural consultant who advised farmers on whether to plant a drought resistant crop (safe option) or a regular crop (risky option; See Figure 11). The outcomes of their decisions were tracked with a point structure (see the section below). Their goal was to have as many points as possible by the end of the experiment. They were paid a monetary bonus commensurate with their point balance at the end of the experiment.

At the beginning of the task, participants read background information on the threat of drought to farmers' crops and their own role in the task (See Figure 11). There were two between-groups conditions based on the framing of the task as a gain or a loss. In the gain frame condition, participants were told that the potential drought might reduce the amount of profit gained from the crops. They achieved their goal by gaining as many points as possible. In the loss frame condition, participants were told that the potential drought might incur a loss, compared to regular, non-drought seasons. They achieved their goal by losing as few points as possible. They were then introduced to the point structure of the task and then performed a practice trial.

Background (Gain)

Farmers are faced with many tough decisions. One is deciding what type of crop to plant when the future climate is uncertain. Drought conditions (extended periods of below-normal precipitation, like rain) pose a significant threat to many farmers. Droughts have the potential to drastically reduce farmers' crop yields, thereby reducing farmers' monetary profit.

Fortunately, some precautions can be taken. For example, instead of the regular crop, farmers can plant crops that are more resistant to drought than conventional crops, but such drought-resistant crop often cost more. Farmers want to **maximize their profit** despite the irregular seasons. Therefore they hired you from an international agricultural consulting firm to advise them which crop to plant. You must decide which crop to plant based only on a climate prediction which will be provided to you.

Background (Loss)

Farmers are faced with many tough decisions. One is deciding what type of crop to plant when the future climate is uncertain. Drought conditions (extended periods of below-normal precipitation, like rain) pose a significant threat to many farmers. Droughts have the potential to drastically reduce farmers' crop yields, thereby reducing farmers' monetary profit.

Fortunately, some precautions can be taken. For example, instead of the regular crop, farmers can plant crop that are more resistant to drought than conventional crop, but such drought-resistant crop often cost more. Farmers want to **minimize their losses** compared to the regular crop in non-drought seasons. Therefore they hired you from an international agricultural consulting firm to advise them which crop to plant. You must decide which crop to plant based only on a climate prediction which will be provided to you.

Figure 11
Scenario Background Description in the Drought Task
Note: The colored texts are to indicate condition and the highlights were used to show the difference between the two conditions. They were not shown to the participants.

Task (Gain)

Your clients are all faced with possible drought conditions in the upcoming season. Based on climate projections (e.g., there is a 20% chance of drought), you will advise your clients to plant either the regular crop or the drought resistant crop. You will earn points based on your advice.

The regular crop has a low cost, but the yield depends on the possible drought.

- During non-drought seasons, all of the crop will survive. This is the best possible outcome with a profit of 400 points).
- During drought seasons, much of the crop will not survive. The yield will be reduced to breakeven with the cost, resulting in a profit of 0 points.

The drought resistant crop has a high cost but the same yield regardless of whether or not there is a drought. It will survive regardless of drought. This means a net profit of 300 points regardless of drought.

As a consultant, you have 50 farmer-clients each year. For each farmer-client, you will get a climate prediction (e.g., there is a 30% chance of drought) for the farmer-client's area. Then you will select the crop you think will result in the greatest profit. The profits made by farmer-clients determine how much consultants get paid (your points), so you want to make your decisions carefully.

You will start with a virtual budget of 0. When a farmer makes a profit, your bonus points will be added to your budget. You will be paid with real money based on your ending balance. You will get paid $1 per 1,000 points above 15,000 points. Your goal is to finish the task gaining as many points as possible.

A)

Task (Loss)

Your clients are all faced with possible drought conditions in the upcoming season. Based on climate projections (e.g., there is a 20% chance of drought), you will advise your clients to plant either the regular crop or the drought resistant crop. You will earn points based on your advice.

The regular crop has a low cost, but the yield depends on the possible drought.

- During non-drought seasons, none of the crop will die. This is the best possible outcome with a loss of 0 points.
- During drought seasons, much of the crop will die. The yield will be reduced, resulting in a loss of 400 points.

The drought resistant crop has a high cost but the same yield regardless of whether or not there is a drought. It will not die regardless of drought. This means a loss of 100 points regardless of drought.

As a consultant, you have 50 farmer-clients each year. For each farmer-client, you will get a climate prediction (e.g., there is a 30% chance of drought) for the farmer-client's area. Then you will select the crop you think will result in the smallest loss relative to regular crop in a non-drought season. The losses avoided by farmer-clients determine how much consultants get paid (your points), so you want to make your decisions carefully.

You will start with a virtual budget of 20,000. When a farmer sustains a loss, your bonus points will be deducted from your budget. You will be paid with real money based on your ending balance. You will get paid $1 per 1,000 points above 15,000 points. Your goal is to finish the task losing as few points as possible.

B)

Figure 12
Task Description in the Drought Task
Note: The colored texts are to indicate condition and the highlights were used to show the difference between the two conditions. They were not shown to the participants. A) shows the gain frame condition. B) shows the loss frame condition.

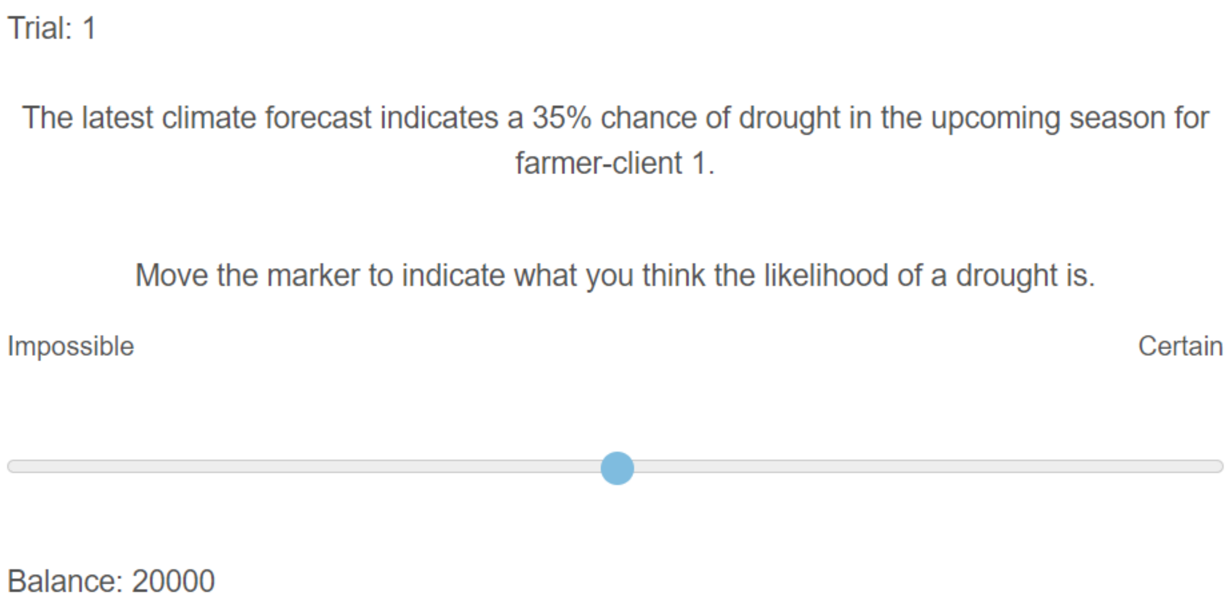


Figure 13
Likelihood Rating Screen of a Trial in the Drought Task

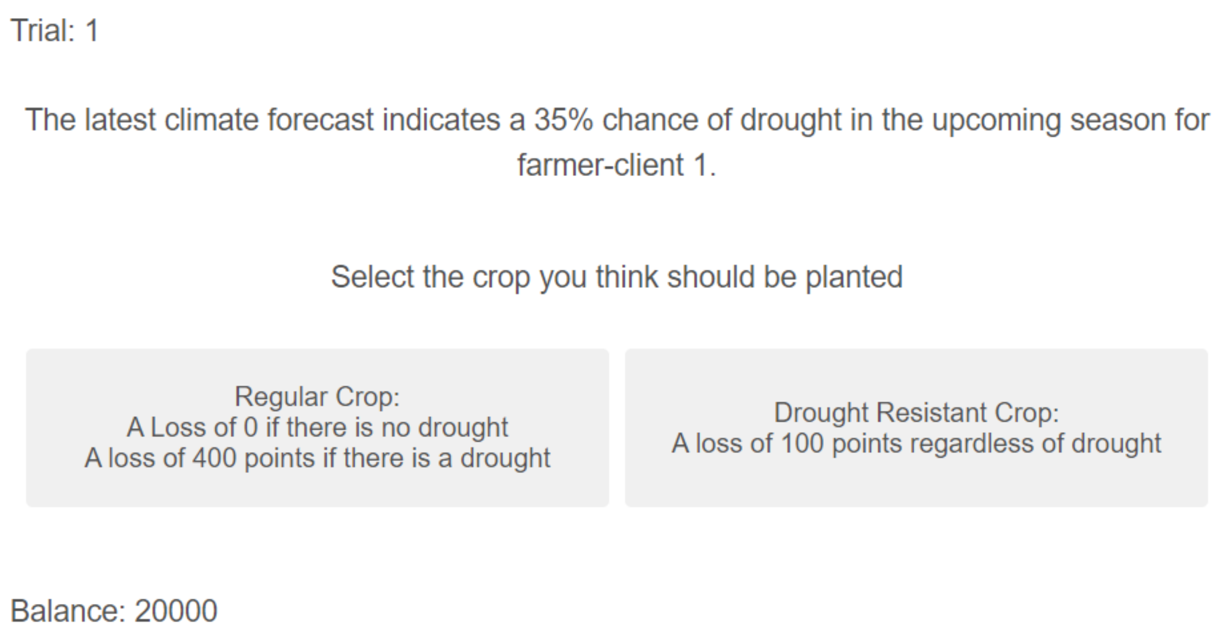


Figure 14
Decision Screen of a Trial of the Loss Condition in the Drought Task

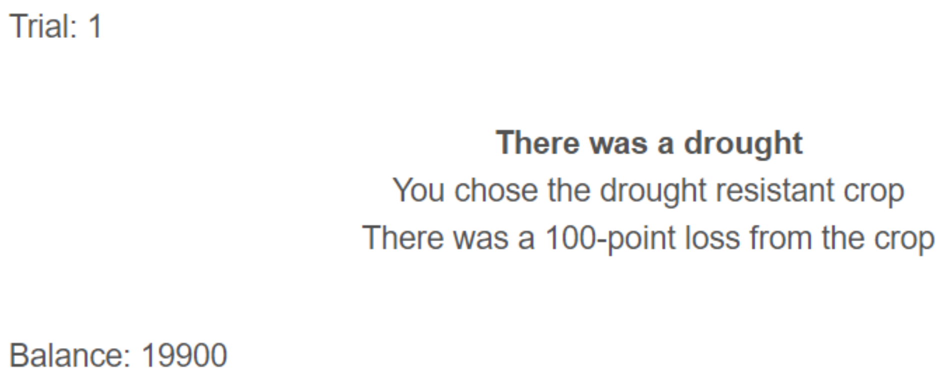


Figure 15
Outcome Screen of a Trial in the Loss Condition in the Drought Task

1. Comprehension check:
   What is the adverse climate condition in this task? [Text Entry, correct answer "drought"]
2. Self-reported subjective criterion:
   Was there a probability above which you generally chose the drought resistant crop in this task? For example, you might have generally chosen the drought resistant crop whenever the probability of a drought was above XX%
   [Multiple Choice:
   Yes. If so, what was it? Please enter a number without % sign [Text Entry],
   No]
3. Scenario description difficulty question:
   How hard do you think the task was for you to understand? [VAS with anchor points: Not hard at all – Very hard]
4. Scenario description difficulty question:
   Optional: If there is part of the task that you found particularly hard to understand, please tell us what it is. [Text box]

Figure 16
Post-Task Questions

Point Structure. In order to simulate real life decisions with consequences, and to encourage participants to put forth their best effort, the point structure, mentioned above, was used in which participants gained or lost points based on their decision (See Table 4). An equivalent point structure was used in the gain and loss conditions. In the gain frame condition, the drought resistant crop (safe option) provided a sure gain of 300 points regardless of drought. The regular crop (risky option) provided a gain of 400 points if there was no drought and a gain of 0 points if there was a drought. In the loss frame condition, the drought resistant crop (safe option) provided a sure loss of 100 points regardless of drought. The regular crop (risky option) provided a loss of 0 points if there was no drought and a loss of 400 points if there was a drought. The starting balance was 0 points in the gain condition and 20,000 points in the loss condition so that participants in both conditions would end with the same point balance if they chose the safe option on every trial. In both conditions, participants were paid $1 for every 1,000 points in their balance above a payment threshold of 15,000 points. This payment threshold was set up to prevent participants from taking the simplistic approach of choosing the safe option in every single trial. If they chose the safe option in all 50 trials, they would end up with 15,000 points (20,000 – 50 * 100 (Loss) = 0 + 50 * 300 (Gain) = 15,000).

Gain Condition	Safe Option	Risky Option
Drought Occurred	300 Points	0 Points
Drought Did Not Occur	300 Points	400 Points

Loss Condition	Safe Option	Risky Option
Drought Occurred	-100 points	-400 points
Drought Did Not Occur	-100 points	0 points

Table 4
Point Structure of the Two Conditions in Experiment 1

This point structure also provided an economically rational standard for participants' decisions based on expected value. The fixed value of the safe option could be compared to the expected value of the risky option in each trial. The safe option always had a gain of 300 points or a loss of 100 points. The risky option had an expected value of (Probability of No Drought * 400 points gain or Probability of Drought * 400 points loss*)*. Therefore, when the probability of drought was 25%, the gain/loss of the safe option was equal to the expected value of the risky option (0.75 * 400 = 300 points gain or 0.25 * 400 = 100 points loss). At 25% objective probability of drought, choosing either the safe or the risky option was equivalent in terms of expected value. This 25% was considered the economically rational criterion. When the probability of drought was above this criterion, the risky option had smaller expected gain/greater expected loss than the safe option. It was therefore economically rational to choose the safe option. When the probability of drought was below this criterion, the risky option had greater expected gain/smaller expected loss than the safe option. Hence the economically rational choice was the risky option.

Trial Structure. After reading through the background information, point structure, and going through a practice trial, participants began the 50 experimental trials. In each trial, participants saw three screens. On the first screen, participants saw a forecast which described the probability that drought would occur (e.g., The latest climate forecast indicates a 35% chance of drought in the upcoming season for farmer-client 1; See Figure 13). This percent chance will be referred to as the objective probability. It was calibrated to be roughly reliable. For example, there were eight trials where the objective probability was 35%. There are three droughts among them (37.5%), close to the objective probability of 35%. Participants then moved the slider on a visual analog scale (VAS) with anchor points *impossible* and *certain* to answer the question "Move the marker to indicate what you think the likelihood of a drought is." This likelihood rating was the operationalization of subjective probability. On the second screen, participants saw the same forecast. This time they chose which crop they wished to plant by pressing one of the two buttons (See Figure 14). In the gain condition, the two buttons were "Regular Crop: A gain of 400 points if there is no drought; A gain of 0 if there is a drought" and "Drought Resistant Crop: A gain of 300 points regardless of drought." In the loss condition, the two buttons were "Regular Crop: A loss of 0 points if there is no drought; A loss of 400 if there is a drought" and "Drought Resistant Crop: A loss of 100 points regardless of drought." After making their decision, the third screen showed the outcome of the trial. Based on whether a drought occurred and participants' decisions, the appropriate number of points were added or deducted from the participants' balance (See Figure 15).

The 50 trials varied in their objective probabilities and occurrence of drought (see Table 5). Objective probability was shown to participants as part of the drought forecast. It had five within-subject levels: 20%, 35%, 50%, 65%, and 80%. The trial order was randomized for each participant.

Trial	Objective Prob	Drought	Trial	Objective Prob	Drought
1	20	1	26	20	0
2	20	1	27	35	1
3	20	1	28	35	1
4	20	1	29	35	1
5	20	1	30	35	0
6	20	0	31	35	0
7	20	0	32	35	0
8	20	0	33	35	0
9	20	0	34	35	0
10	20	0	35	50	1
11	20	0	36	50	1
12	20	0	37	50	1
13	20	0	38	50	0
14	20	0	39	50	0
15	20	0	40	50	0
16	20	0	41	65	1
17	20	0	42	65	1
18	20	0	43	65	1
19	20	0	44	65	0
20	20	0	45	65	0
21	20	0	46	80	1
22	20	0	47	80	1
23	20	0	48	80	1
24	20	0	49	80	1
25	20	0	50	80	0

Table 5
Trial Structure of the Experiments

Post-Task Questions. After completing all 50 trials, a summary of participants' decisions and their outcomes across the trials was shown along with several questions (See Figure 16). Participants were first asked a comprehension check question "What is the adverse climate condition in this task?" for which the correct answer was *drought* or any related phrase. Participants who answered this question incorrectly were not included in the data analysis (3 participants were eliminated using this question). They then were then asked to indicate their (self-reported) criterion of their decision: "Was there a percent chance likelihood above which you generally chose the drought resistant crop? For example, you might have generally chosen the drought resistant crop whenever the chance of a drought was above XX%." If participants chose yes, they were prompted to provide a number indicating their self-reported criterion. This was an alternative way to the calculated criterion (described in the reanalyses of previous experiments) of getting the subjective criterion. This self-reported criterion was asked at the end of the experiment and thus reflected the decision criterion based on experience from the 50 trials, similar to a previous study (Joslyn & Grounds, 2015). This self-reported criterion was also conscious. As such, it might not match participants' actual decision-making process which might also be affected by unconscious processes. Next, participants rated how difficult the task was on a VAS with anchor points *not hard at all* and *very hard*: "How hard do you think the task was for you to understand?" This question was followed by the open-ended question: " If there is part of the task that you found particularly hard to understand, please tell us what it is," These two questions were meant to check whether the descriptions of the goal in the gain or loss frame conditions had a different difficulty which would have been a confound. Participants reported no difference in the difficulty of understanding the two conditions. Finally, participants answered an open-ended question: " Did you encounter any bug or glitch?" This question allowed participants to report any problem in the experimental program. Participants reported that they encountered no bugs or glitches with the program. After completion of all questions, participants were thanked and provided with a unique completion code to enter into Prolific to verify their participation and receive payment.

Design

Experiment 1 used a 2 x 6 mixed design. There was one between-group independent variable: Framing with two levels: Gain vs. loss framing. There was one within-group independent variable: Objective probability of drought with levels: 20%, 35%, 50%, 65% and 80%. Other parameters were fixed: The economically rational criterion was 25% and the proportion of drought trials (trials with a drought out of all trials) was 36%. As the objective probabilities were reliable (M = 36.5%), the mean objective probability and the proportion of drought trials were considered the same (36%) in the analyses.

In terms of dependent variables, participants' likelihood ratings were measured using a VAS and their decisions using a binary choice in each trial. At the end of the experiment, participants answered a comprehension check question, reported their self-reported criterion, difficulty of understanding the task, and any glitches they encountered.

Results

Analysis Overview

Based on the gain-loss framing effect (Tversky & Kahneman, 1979), the prediction was that the participants' subjective criterion should account for risk seeking decisions with losses and risk averse decisions with gains. If the gain-loss framing effect was the sole source of bias, based on the subjective criterion hypothesis, participants would have a higher-than-rational subjective criterion in the loss frame condition and a lower-than-rational subjective criterion in the gain frame condition. However, as other sources of bias (e.g., a centering effect) might also shift the subjective criterion away from the economically rational criterion, the above prediction might not be observed even when there was a gain-loss framing effect. Therefore, a more simplified prediction that the subjective criterion would be higher in the loss frame condition than in the gain frame condition was used in this experiment. Two operationalizations of the subjective criterion (calculated criterion and self-reported criterion) were examined in the analyses below. In addition to these two operationalizations of subjective criterion, participants' relative frequency of choosing the safe option was also analyzed as an alternative way to examine whether participants had a decision bias (risk seeking/risk averse) in either condition. If the gain-loss framing effect took place and given the same sensitivity, participants would choose the safe option more often in the gain frame than in the loss frame condition. In addition, this analysis allowed the comparison between participants' decision pattern in this experiment to similar previous studies (Joslyn & LeClerc, 2013; Grounds & Joslyn, 2017; Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). Next, an analysis examined the likelihood ratings of the two conditions to detect any bias in subjective likelihood. It was not expected to be affected by the framing and was expected to be slightly overestimated like in previous studies with naturalistic weather decision tasks. After that, another analysis examined participants' sensitivity or their ability to predict the drought regardless of their decision bias. It was not expected that the gain-loss frame effect could affect the sensitivity. Finally, if participants' sensitivity was the same between the two conditions but the relative frequency of taking the safe was different, it can be inferred that this difference was due to a decision bias.

A series of t-tests were conducted with an alpha of .05. Both planned and post hoc t-tests were corrected for familywise errors with Holm-Bonferroni method (Holm, 1979). This method corrected the alpha to which p values were compared. In this method, the smallest p value in a family was compared to the most stringent Bonferroni corrected alpha based on the number of t-tests. For example, in a family of three t-tests, the smallest p value was compared to a corrected alpha of .05/3 = .017. If significant, the next smallest p value was compared to the next most stringent corrected alpha (.05/2 = .025). The procedure continued until the first insignificant comparison appeared or all p values were compared. P values already higher than the uncorrected alpha of .05 did not need to be compared with a corrected alpha. T-tests on the same despondent variable in the same experiment were considered a family.

Decision Bias

The first set of analyses examined the degree to which the above-described biases could be detected in the two conditions. Three dependent variables were examined here: Calculated subjective criterion, self-reported subjective criterion, and the relative frequency of choosing the safe option.

Calculated Subjective Criterion. The calculated criterion was an operationalization of the subjective criterion. It was calculated for participants using their likelihood rating distribution and binary decisions (See reanalyses of previous experiments section). In Figure 17, the mean calculated criterion is shown with blue bars for the gain and loss frame conditions. As predicted, the mean calculated criterion in the loss frame (M = 33.8%, SD = 11.7%) was higher than in the gain from (M = 31.0%, SD = 14.2%). However, the difference of 2.9% failed to reach significance in a planned independent t-test ($t(94.0) = 1.01$, $p = .32$). In addition, the calculated criterion in each condition was compared to the economically rational criterion of 25% in two post hoc one-sample t-tests. In the gain frame condition, the calculated criterion was significantly higher than the economically rational criterion of with a difference of 6.0% ($t(49) = 2.94$, $p = .005$, corrected alpha = .05). In the loss frame condition, the calculated criterion was significantly higher than the economically rational criterion with a difference of 8.0% ($t(45) = 4.44$, $p < .001$, corrected alpha = .025). Thus, the calculated criterion was biased but not by the gain-loss framing effect. Thus, there was a consistent bias in the calculated criterion in the gain and the loss frame condition.

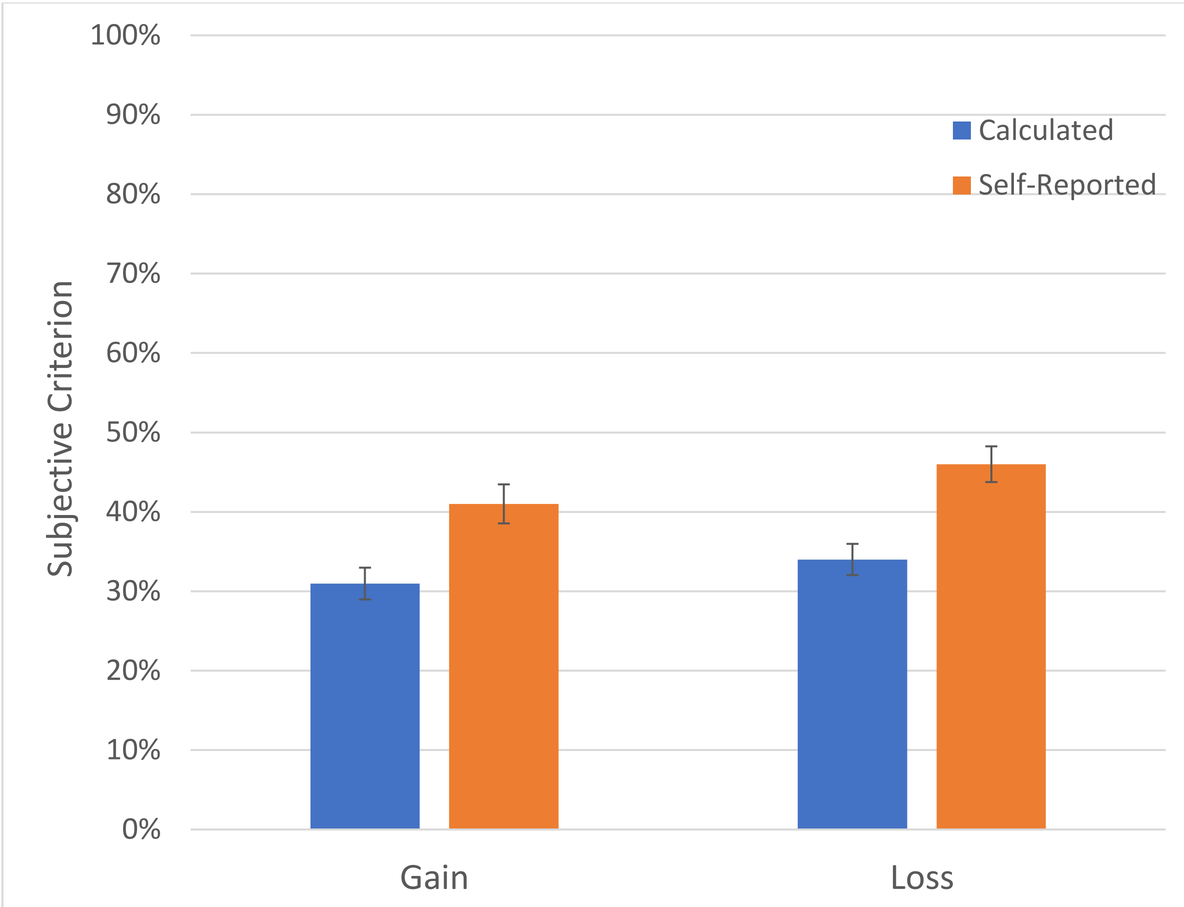


Figure 17
Calculated and Self-Reported Criterion in the Two Conditions of Experiment 1
For the calculated criterion, the difference of 2.9% between the gain and the loss frame condition was not significant. The economically rational criterion was 25%.
For the self-reported criterion, the difference of 5.6% between the gain and the loss frame condition was not significant.

Self-Reported Criterion. The self-reported criterion was an alternative operationalization of the subjective criterion in which participants were asked to report the probability above which they took the safe option at the end of the experiment. Unlike the calculated subjective criterion, which might be affected by unconscious processes, the self-reported criterion captured conscious processes and memory over the entire sequence of trials and therefore might differ from the former.

In Figure 17, the self-reported criterion is shown with orange bars for the gain and the loss frame conditions. Among the 96 participants, 88 (92%) indicated that they had used a criterion for their decision and 8 (8%) indicated that they did not. The following analyses included only these 88 participants. The mean self-reported criterion in the gain frame condition ($M = 40.6\%$, $SD = 16.3\%$) was lower than in the loss frame ($M = 46.2\%$, $SD = 14.9\%$). The difference of -5.6% failed to reach significance in a post hoc independent t-test ($t(85.3) = 1.69$, $p = .094$). The trend of this result was consistent with the prediction of the gain-loss framing effect but failed to reach significance. In addition, the self-reported criterion was compared to the calculated criterion in each condition in two post hoc paired t-tests. In the gain frame condition, the self-reported criterion was significantly higher than the calculated criterion ($M = 32.5\%$, $SD = 14.1\%$) with a difference of 8.1% ($t(43) = 3.12$, $p = .003$, corrected alpha = .050). In the loss frame condition, the self-reported criterion was significantly higher than the calculated criterion ($M = 34.5\%$, $SD = 13.5\%$) with a difference of 11.7% ($t(43) = 3.94$, $p < .001$, corrected alpha = .025). Thus, there was a consistent bias in the self-reported criterion in the gain and the loss frame condition.

Relative Frequency of Choosing the Safe Option. The participants' relative frequency of choosing the safe option in the gain and loss frame conditions is shown in Figure 18. As expected, a post hoc independent t-test revealed that the mean relative frequency of choosing the safe option was significantly higher in the gain frame ($M = 53.2\%$, $SD = 20.7\%$) than in the loss frame ($M = 43.9\%$, $SD = 16.8\%$) with a difference of 9.3% ($t(92.62) = 2.42$, $p = .02$). Next, two post hoc one-sample t-tests compared the relative frequency in each condition to 48%, the relative frequency of choosing the safe option if participants had chosen the safe option whenever the objective probability was above the economically rational criterion of 25%. In the gain frame condition, although higher, the relative frequency was not significantly different from 48%, with a difference of 5.2% ($t(49) = 1.76$, $p = .085$). Similarly, in the loss frame condition, although lower, the relative frequency was not significantly different from 48%, with a difference of -4.1% ($t(45) = 3.61$, $p = .10$). This suggests a bias that participants were more risk-seeking in the loss frame than in the gain frame condition.

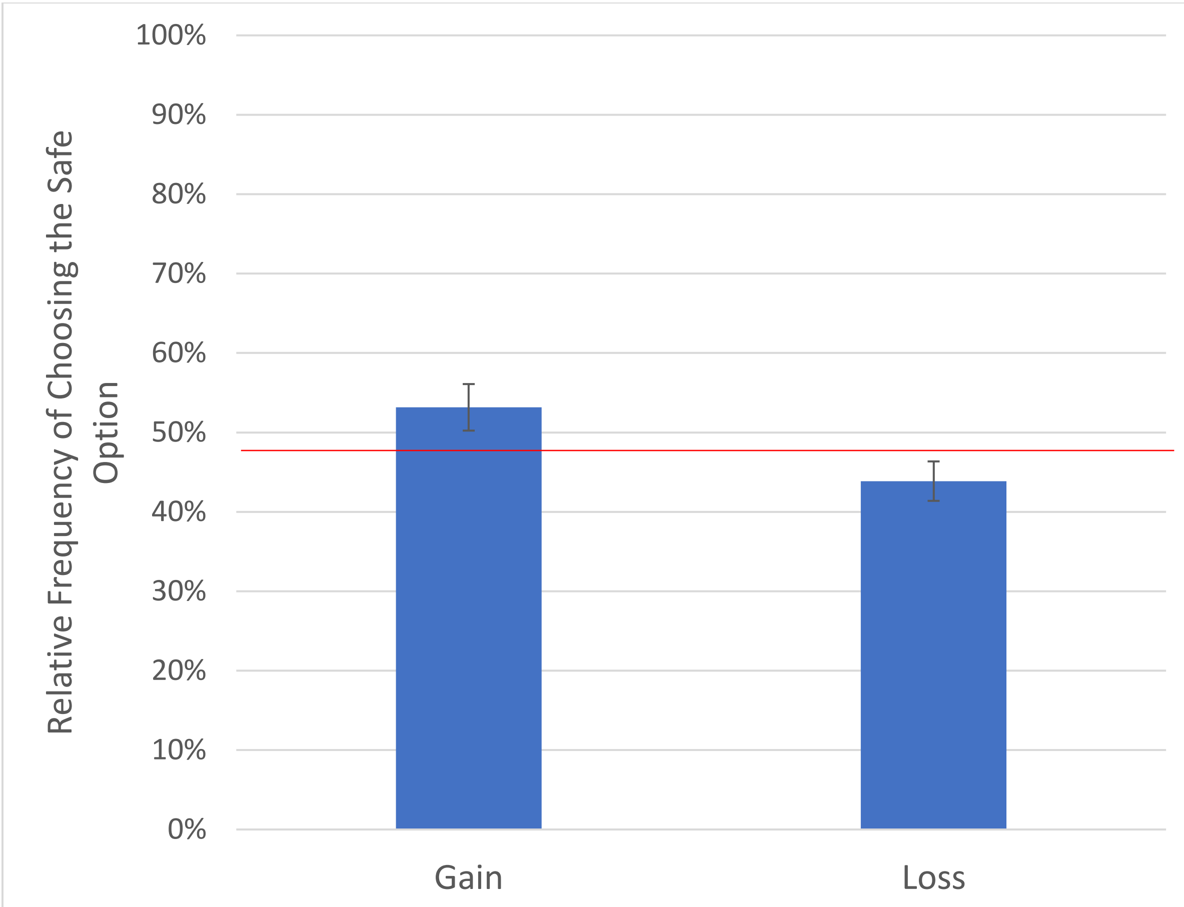


Figure 18
Relative Frequency of Choosing the Safe Option in the Two Conditions of Experiment 1
The red line shows the relative frequency of 48%. This is the relative frequency the participants would get if they always chose the safe option when the objective probability was above the economically rational criterion. The difference of 9.3% between the gain and the loss frame condition was significant.

Likelihood Ratings

Next, an analysis was conducted in order to determine whether there is a bias in likelihood ratings in the two conditions based on the same forecasts. Figure 19 shows the likelihood ratings as a function of objective probabilities in the gain frame and loss frame conditions. As expected, no systematic bias due to the manipulation was detected. The mean likelihood rating in the loss frame (M = 38.5%, SD = 8.1%) was 2.4% higher than in the gain frame (M = 36.1%, SD = 5.8%). However, the difference was not significant in a post-hoc independent t-test ($t(89.1) = 1.30$, $p = .20$). In addition, the mean likelihood rating in each condition was compared to the proportion of drought trials of 36% in two post-hoc one-sample t-tests. The mean likelihood rating in the loss frame was not significantly different from the proportion of drought trials of 36%, with a difference of 2.5% ($t(49) = 2.16$, $p = .036$, corrected alpha = .025). Similarly, the mean likelihood rating in the gain frame was not significantly different from the proportion of drought trials of 36%, with a difference of 0.1% ($t(45) = .071$, $p = .48$). This suggests that the likelihood of drought ratings was close to the proportion of drought trials (relative frequency of drought). The two conditions had similar likelihood rating patterns. Likelihood ratings were close to the objective probability at all objective probability levels except for 35%.

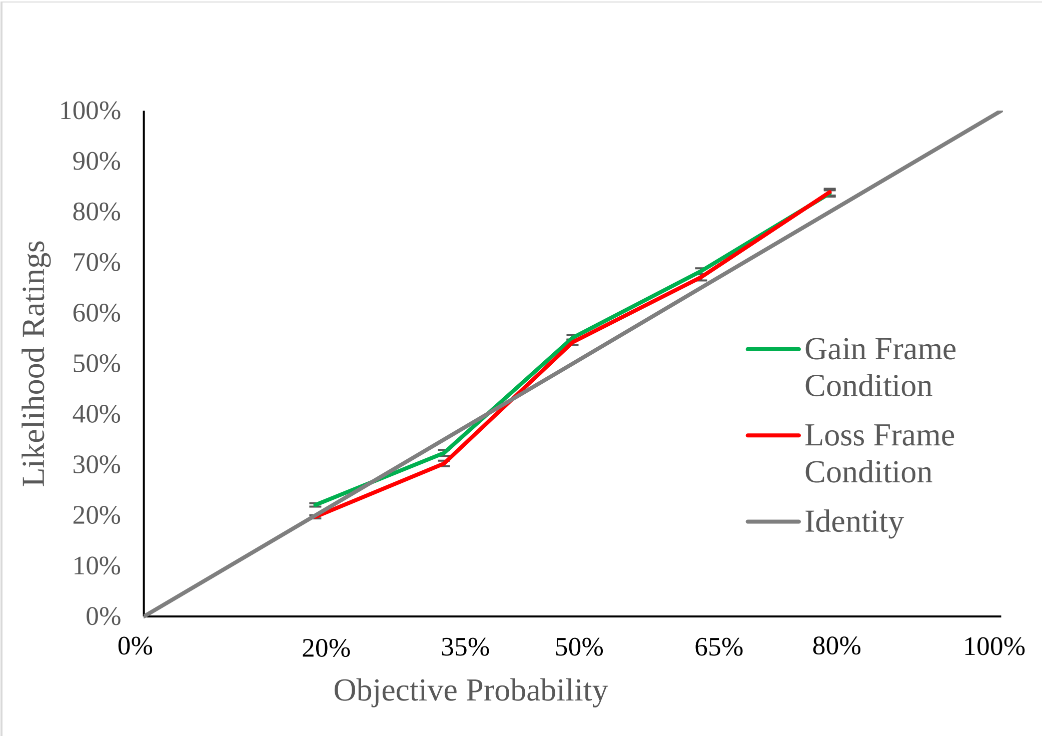


Figure 19
Likelihood Ratings in the Gain and Loss Frame Conditions as a Function of Objective Probability in Experiment 1
The green line represents the gain frame condition. The red line represents the loss frame condition. There was no observed centering effect.

Sensitivity

The next set of analyses examined the sensitivity or participants' ability to predict the drought as measured by the area under the ROC curve, (See the reanalyses of previous experiments section). Figure 20 and 21, ROC plots are shown for the gain frame and loss frame condition respectively. The ROC curves of the gain and loss frame conditions were similar, indicating that both conditions had similar sensitivity or the ability to predict drought based on provided drought forecasts. The mean percent area under ROC curve was 71.4% ($SD = 5.7\%$) in the gain condition and 71.2% ($SD = 5.7\%$) in the loss condition. The difference of 0.2% was not significant in a post hoc independent t-test ($t(82.3) = 0.19$, $p = .85$). Thus, there is no sign of an effect on the sensitivity.

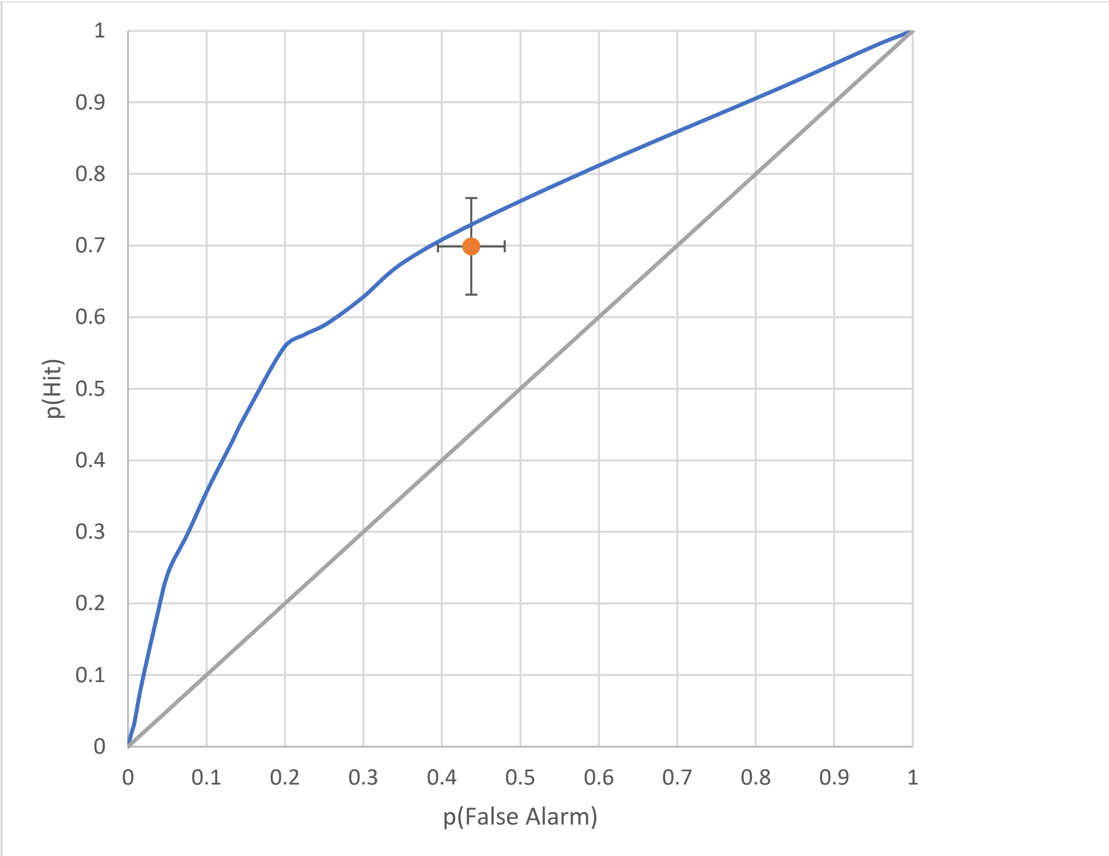


Figure 20
ROC Plot for the Gain Condition in Experiment 1
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 71.4%.

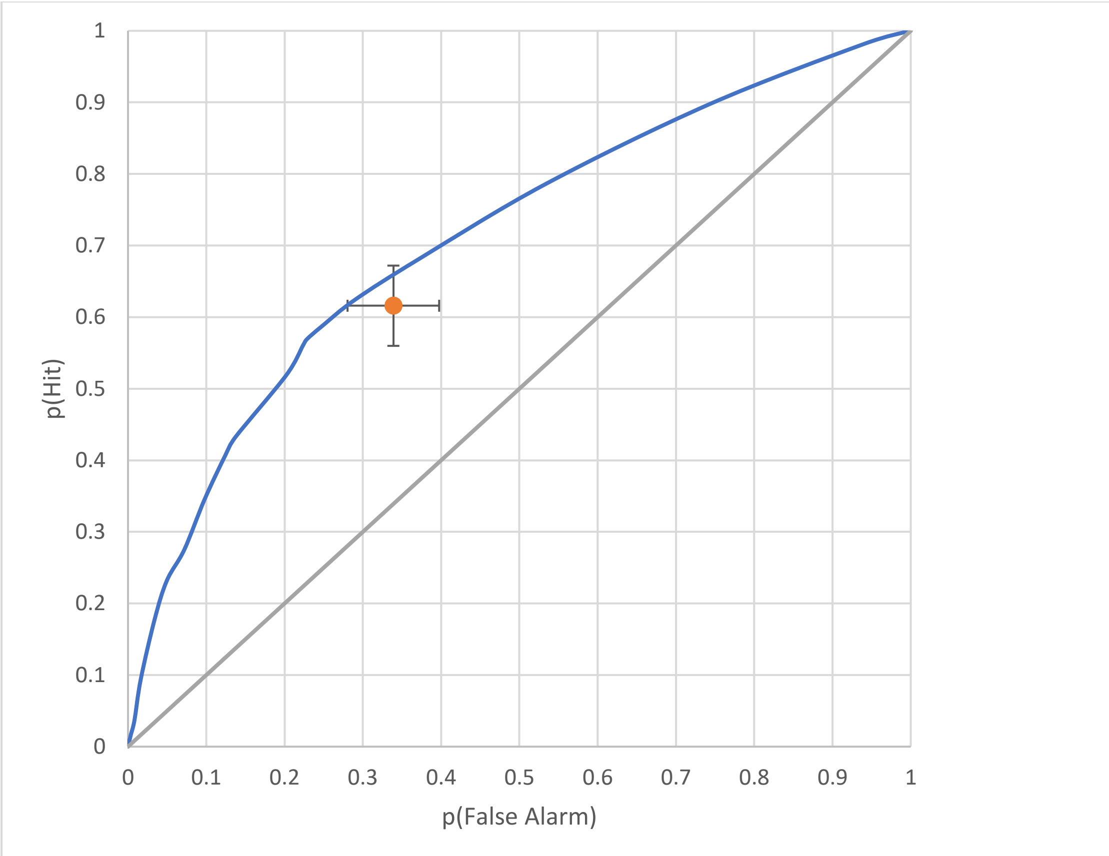


Figure 21
ROC Plot for the Loss Condition in Experiment 1
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 71.2%. This was not significantly different from the area under curve of the gain frame condition.

Discussion

The prediction of experiment 1 was that, based on the subjective criterion hypothesis, participants in a loss frame would adopt a higher subjective criterion than those in a gain frame, consistent with a greater risk-seeking tendency in a loss frame than in a gain frame as suggested by prospect theory. No effect of framing on likelihood ratings and sensitivity was expected. The results showed that although both the calculated criterion and the self-reported criterion were higher in the loss frame condition than in the gain frame condition, the difference failed to reach significance in either dependent variable. This fails to provide consistent support for the prediction that the subjective criterion would be higher in the loss frame than in the gain frame. On the other hand, neither sensitivity nor likelihood ratings were different between the two conditions, suggesting that framing did not affect how people perceive and understand the probability of the drought.

Next, the analysis of the relative frequency of choosing the safe option indicated that participants in the gain frame condition chose the safe option significantly more often than those in the loss frame condition. As the sensitivity and trial composition did not differ between the two conditions, it can be inferred that the cause of this difference in decision came from a decision bias towards the safe option in the gain frame compared to the loss frame condition. This was consistent with a higher, more risk-seeking subjective criterion in the loss frame condition. Although analyses of both operationalizations of the subjective criterion failed to reach significance, the results suggested a trend in the expected direction in the present study. This, coupled with the significant difference in the relative frequency of choosing the safe option analysis, suggests that the possibility of a gain-loss framing effect on the subjective criterion is not ruled out. In addition, a null effect on the sensitivity and likelihood ratings was observed. If gain-loss framing indeed had no effect on the sensitivity or likelihood while having an effect on participants' decisions, it perhaps did not act through participants' biased subjective likelihood or sensitivity.

A possible reason that the analysis of the calculated criterion failed to reach significance while the analysis of the relative frequency did, might be that the calculation of the calculated criterion incorporated the variance from both the likelihood ratings and binary decisions. Thereby this analysis required higher statistical power to detect a difference compared to the analysis of the relative frequency of choosing the safe option which only had variance from binary decisions. Overall, while the analyses of the subjective criterion were in the predicted direction but failed to reach significance, the relative frequency of choosing the safe option analysis provided some support for participants being biased towards the safe option in the gain compared to the loss frame condition.

The results also yielded other interesting findings. First, the self-reported criterion was significantly higher than the calculated criterion in the gain and the loss frame conditions among those who reported that they used a criterion. For these participants, the mean self-reported criterion was between 40% and 50% in both conditions while the calculated criterion was under 40%. Two possible explanations might account for this difference. First, it is possible that this measurement only captured the conscious components of participants' decision-making strategy while their actual decisions were made with some unconscious components. Next, the self-reported criterion was measured at the end of the experiment once instead of measured in every trial. Therefore, this measurement perhaps captured participants' hindsight about their decision-making strategy which might not accurately reflect how they made decisions in the heat of a trial.

The results also suggest that, in both the gain and the loss frame conditions, the calculated and self-reported criterion were higher than the economically rational criterion of 25%. This suggests a higher subjective criterion than the economically rational criterion in both conditions rather than a lower-than-rational subjective criterion in the gain condition as one might expect from risk aversion. A higher-than-rational subjective criterion in both conditions is consistent with a centering effect where the subjective criterion shifts towards 50% from the economically rational criterion.

Furthermore, there was a trend of overestimation in likelihood ratings in both conditions, consistent with previous studies (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). This overestimation was shown at all objective probability levels except for 35%, where the likelihood ratings were underestimated, and 80%, where the likelihood ratings were calibrated. In addition, as seen in Figure 19, the lines showing the likelihood ratings as a function of objective probability levels did not show a sign of a centering effect. If there was a centering effect, the overestimation at low objective probability levels should be greater than the rest of the range while there should be underestimation at high objective probability levels. As objective probabilities were provided to participants, this result supports the notion that if information about a quantity is provided, the centering effect on this quantity is reduced (Radvansky et al., 1995).

Next, in the ROC plots, the decision points overlapped with the respective ROC curves just like in the reanalyses of previous experiments. This suggests that the decision points were consistent with the ROC curves. This further suggests that the decision was solely determined by a likelihood rating and a subjective criterion. This in turn supports the use of the random likelihood model method to calculate the subjective criterion from likelihood ratings and binary decisions.

In conclusion, experiment 1 found modest evidence supporting the gain-loss framing effect in subjective criterion and in turn binary decisions. Likelihood ratings and the sensitivity were unaffected. The higher-than-rational subjective criterion was consistent with the subjective criterion hypothesis.

Experiment 2

Experiment 2 focused on the centering effect on subjective criterion by manipulating the economically rational criterion. The idea was that the shift of the subjective criterion towards 50% due to centering can be observed by varying the economically rational criterion below and above 50%. The effect of this manipulation on subjective likelihood and sensitivity was also examined. Experiment 2 used the same task as experiment 1, using only the loss frame. The economically rational criterion was manipulated instead of gain-loss framing. The economically rational criterion was manipulated to be higher, the same as, or lower than 50% to expose a centering effect such that participants' subjective criterion would shift towards 50%. This manipulation was expected to affect only the subjective criterion and should not affect the subjective likelihood or sensitivity.

Method

Participants

A total of 160 participants from the US were recruited from Prolific Academic in January 2024, a crowdsourcing platform for online research. After an elimination process, 157 participants were used in the analysis. Three were eliminated for failing the comprehension check (same question as experiment 1). Each participant was paid $4 for participation plus a performance based monetary bonus. The mean age was 40 (SD = 13.91, range 20 to 80 years). There were 80 (51%) females, 75 (48%) males, 1 (1%) who preferred not to say, and 1 where Prolific did not provide data.

Procedure and Stimuli

The procedure was mostly identical to experiment 1. Participants performed a drought decision task as described in experiment 1. The outcomes of their decisions were tracked with a point structure mostly identical to the loss frame condition of experiment 1. As with Experiment1 participants' goal was to have as many points as possible by the end of the experiment and they were paid a monetary bonus commensurate with their point balance.

Point Structure. In Experiment 2, only the loss frame was used. The point structure used in experiment 2 was mostly unchanged from the loss frame condition of experiment 1 except that the economically rational criterion was manipulated between groups. There were three economically rational criterion conditions (erc): 25% (25erc), 50% (50erc), and 75% (75erc). The 25erc condition was identical to the loss frame condition of experiment 1: the drought resistant crop (safe option) provided a sure loss of 100 points regardless of drought. The regular crop (risky option) provided a loss of 0 points if there was no drought and a loss of 400 points if there was a drought. In the 50erc and 75erc condition, the drought resistant crop (safe option) provided a sure loss of 200 and 300 points respectively (See table 6). All conditions had the same starting balance of 20,000 points like the loss condition in experiment 1. Changes were made to the bonus payment structure to make it roughly equivalent in the three conditions. In the 25erc condition, participants were paid $1 for every 1,000 points in their balance above a payment threshold of 15,000 points. In the 50erc condition, participants were paid $1 for every 2,000 points in their balance above 10,000 points. In the 75erc condition, participants were paid $1 for every 3,000 points in their balance above 5000 points.

Trial Structure. The trial structure was mostly the same as experiment 1 with one difference. After reading through the background information, point structure, and going through a practice trial, participants saw two attention check questions created per Prolific attention check policy (See table 7). Participants then began the 50 experimental trials whose composition was identical to experiment 1. The post-survey questions were identical to experiment 1.

25erc Condition		
	Safe Option	Risky Option
Drought Occurred	-100 points	-400 points
Drought Did Not Occur	-100 points	0 points
50erc Condition		
	Safe Option	Risky Option
Drought Occurred	-200 points	-400 points
Drought Did Not Occur	-200 points	0 points
75erc Condition		
	Safe Option	Risky Option
Drought Occurred	-300 points	-400 points
Drought Did Not Occur	-300 points	0 points

Table 6
Point Structure of the Three Conditions in Experiment 2

Earthquakes are a very devastating natural disaster. Powerful earthquakes have the potential to level entire city blocks. Earthquakes are prevalent on the west coast of North America. For this question, please choose tornado. **A. Tornado** B. Earthquake C. Volcano D. Hurricane E. Thunderstorm
The surface of the moon has no air and is blasted with solar radiation. How much do you agree with this statement: I can walk on the moon without a spacesuit or any kind of shelter/protection. **A. Strongly disagree** B. Somewhat disagree C. Neutral D. Somewhat agree E. Strongly Agree
Cars run on gasoline, jets run on kerosine, steamships run on coal, trucks run on diesel, plants run on solar power. What fuel does the famous British ocean liner Titanic, a steamship, use? **A. Coal** B. Gasoline C. Diesel D. Kerosine E. Solar power
The distance between Hawaii and the US mainland is over 2,400 miles. It can take a week for a cruise ship to travel from Hawaii to the US west coast. How much do you agree with this statement: I cannot swim from the US west coast to Hawaii in 2 days non-stop. A. Strongly disagree B. Somewhat disagree C. Neutral D. Somewhat agree **E. Strongly Agree**

Table 7
Additional Attention Check Questions in Experiment 2
Correct answers are bolded. The order of answers was randomized. Participants saw only two out of four questions.

Design

Experiment 2 used a 3 x 6 mixed design. There was one between-group independent variable: Economically rational criterion with three levels: 25erc, 50erc, and 75erc. There was one within-group independent variable: Objective probability of a drought with levels: 20%, 35%, 50%, 65% and 80%. Other parameters were fixed: The frame was the loss frame from experiment 1 and the proportion of drought was 36% across 50 trials. Like in experiment 1, as the objective probabilities were reliable (M = 36.5%), the mean objective probability and the proportion of drought trials were considered the same (36%) in the analyses.

In terms of dependent variables, as with experiment 1, participants reported likelihood ratings using a VAS and binary decisions on each trial. At the end of the experiment, participants answered the comprehension check question, reported their self-reported criterion, difficulty of understanding the task (no difference in the difficulty between conditions reported), and any glitches they encountered as with experiment 1.

Results

Analysis Overview

The hypothesis of experiment 2 was that the mean subjective criterion would shift towards 50%. This meant that in the 25erc condition, the subjective criterion would be between 25% and 50% while in the 75erc condition, the subjective criterion would be between 50% and 75%. In the 50erc condition, the subjective criterion would be close to 50%. The manipulation of the economically rational criterion should not affect subjective likelihood or sensitivity. In addition, the centering effect in likelihood ratings regardless of the manipulation of economically rational criterion was also examined. It was expected that, based on previous studies (Qin et al., 2024), the reanalyses, and experiment 1, there would be no centering effect in likelihood ratings.

The same set of dependent measures were used in experiment 2. To measure the decision bias, the calculated criterion and the self-reported criterion were analyzed. In addition, the relative frequency of choosing the safe option was analyzed as an alternative way to examine the decision bias. If a centering effect was present, participants would choose the safe option less often (consistent with a subjective criterion shifted from 25% to 50%) in the 25erc condition than the economically rational decisions. They would also choose the safe option more often (consistent with a subjective criterion shifted from 75% to 50%) in the 75erc condition than they should. They would be comparatively risk-neutral (consistent with a subjective criterion near 50%) in the 50erc condition. As with Experiment 1, likelihood ratings were analyzed to detect any bias in subjective likelihood. Finally, to measure the sensitivity, the area under the ROC curve was analyzed. The analysis of sensitivity tested whether it was possible to rule out the manipulation of economically rational criterion affecting it. A series of ANOVAs and t-tests were conducted. Holm-Bonferroni Method was used for planned and post hoc t-tests as week as planned pairwise comparisons under omnibus ANOVAs. Tukey method was used for post hoc pairwise comparisons under omnibus ANOVAs.

Decision Bias

The first set of analyses examined the degree to which participants biased their decisions towards the safe option in the two conditions. As with Experiment 1, three dependent variables were examined here: Calculated subjective criterion, self-reported subjective criterion, and the relative frequency of choosing the safe option. The former two variables were operationalizations of the subjective criterion of the random likelihood model and signal detection theory. The latter was an alternative way to examine the decision bias not under the purview of the model and the theory.

Calculated Subjective Criterion. In Figure 22, the calculated criterion is shown as blue dots for 25erc, 50erc, and 75erc conditions. The mean calculated criterion was 35.0% (SD = 13.3%) in the 25erc condition, 48.4% (SD = 15.9%) in the 50erc condition, and 58.1% (SD = 15.3%) in the 75erc condition. An ANOVA with the economically rational criterion manipulation (25erc, 50erc, and 75erc) as the independent variable on the calculated subjective criterion showed a main effect of this manipulation ($F(2,154) = 32.53$, $p < .001$). Two planned pairwise comparisons were conducted. The 50erc condition had a significantly higher calculated criterion than the 25erc condition with a difference of 13.4% ($t(154) = 4.54$, $p < .001$, corrected alpha = .025). The 75erc condition had a significantly higher calculated criterion than the 50erc condition with a difference of 9.7% ($t(154) = 3.30$, $p = .001$, corrected alpha = .05). In summary, the economically rational criterion manipulation had a main effect on the calculated criterion.

In addition, three planned one-sample t-tests compared the calculated criterion in each condition with the respective economically rational criterion. In the 25erc condition, the calculated criterion was significantly higher than 25% with a difference of 10% ($t(53) = 5.47$, $p < .001$, corrected alpha = .017). In the 50erc condition, the difference of 1.6% between the calculated criterion and 50% was not significant($t(46) = 1.40$, $p = .17$). In the 75erc condition, the calculated criterion was significantly lower than 75% with a difference of -11.9% ($t(53) = 8.06$, $p < .001$, corrected alpha = .025). This result was consistent with the centering effect such that the calculated criterion shifted towards 50% from the economically rational criterion.

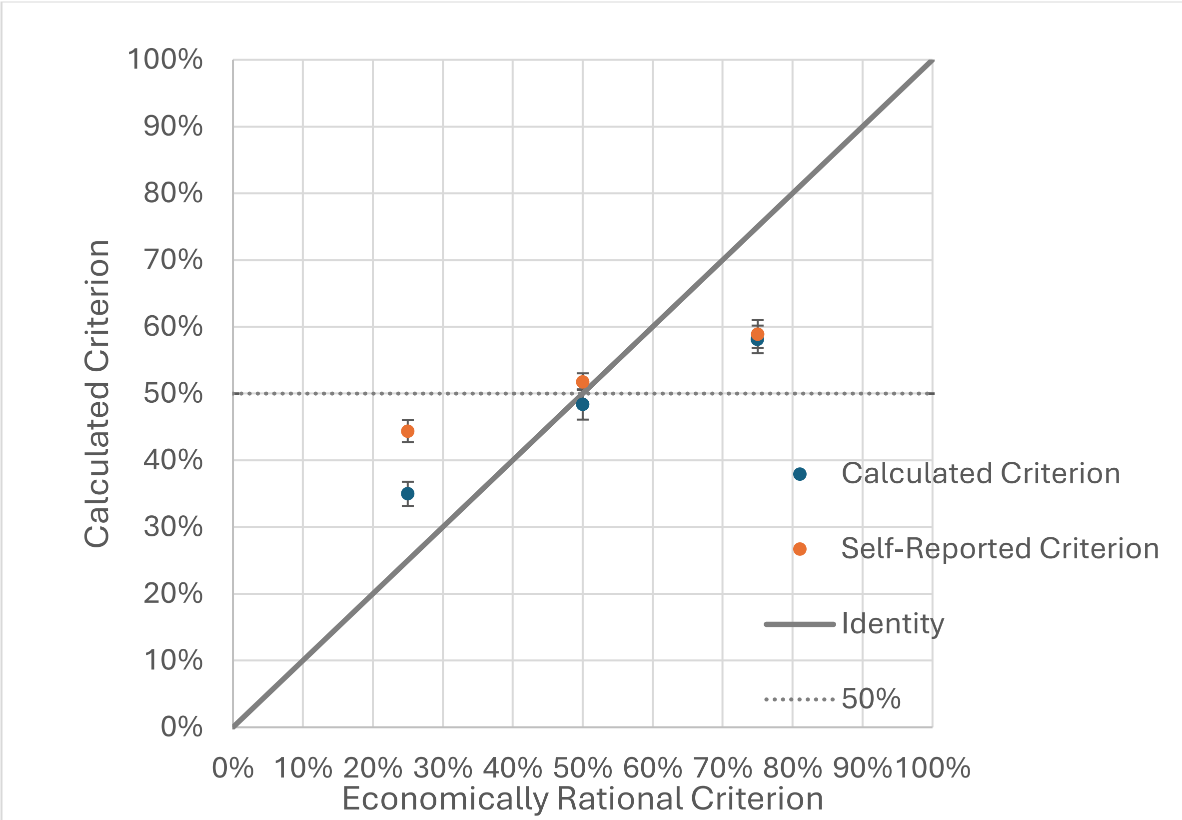


Figure 22
Calculated and Self-Reported Subjective Criterion in the Three Conditions of Experiment 2
In the 25% economically rational criterion condition, the calculated criterion was significantly higher than 25%. In the 50% economically rational criterion condition, the calculated criterion was not significantly different from 50%. In the 75% economically rational criterion condition, the calculated criterion was significantly lower than 75%.
In the 25% economically rational criterion condition, the self-reported criterion was significantly higher than 25%. In the 50% economically rational criterion condition, the self-reported criterion was not significantly different from 50%. In the 75% economically rational criterion condition, the self-reported criterion was significantly lower than 75%.

Self-Reported Criterion. In Figure 22, the self-reported criterion is shown as orange dots for the 25erc, 50erc, and 75erc conditions. Among the 157 participants, 146 (93%) indicated that they had used a criterion for their decision and 11 (7%) indicated that they did not. The following analyses included only these 146 participants. The mean self-reported criterion was 44.4% (SD = 12.0%) in the 25erc condition, 51.8% (SD = 8.6%) in the 50erc condition, and 58.9% (SD = 14.4%) in the 75erc condition. An ANOVA with the economically rational criterion manipulation (25erc, 50erc, and 75erc) as the independent variable on the self-reported subjective criterion showed a main effect of this manipulation ($F(2, 143) = 18.45$, $p < .001$). Two planned pairwise comparisons were conducted. The 50erc condition had a significantly higher self-reported criterion than the 25erc condition with a difference of 7.4% ($t(143) = 3.09$, $p = .002$, corrected alpha = .025). The 75erc condition had a significantly higher self-reported criterion than the 50erc condition with a difference of 7.2% ($t(143) = 2.91$, $p = .004$, corrected alpha = .05). In summary, the economically rational criterion manipulation had a main effect on the self-reported criterion.

In addition, three planned one-sample t-tests compared the self-reported criterion in each condition with the respective economically rational criterion. In the 25erc condition, the self-reported criterion was significantly higher than 25% with a difference of 19.4% ($t(51) = 11.62$, $p < .001$, corrected alpha = .017). In the 50erc condition, the difference of 1.8% between the self-reported criterion and 50% was not significantly different ($t(46) = 1.40$, $p = .17$). In the 75erc condition, the self-reported criterion was significantly lower than 75% with a difference -16.1% ($t(46) = 7.68$, $p < .001$, corrected alpha = .025). This result was consistent with the centering effect.

Finally, three post hoc paired t-tests compared the self-reported criterion and the calculated criterion in the 25erc, 50erc, and 75erc conditions. The self-reported criterion was significantly higher than the calculated criterion ($M = 34.5$, $SD = 13.4$) in the 25erc condition with a difference of 9.9% ($t(51) = 4.26$, $p < .001$, corrected alpha = .017). The difference of 3.2% between the self-reported criterion and the calculated criterion ($M = 48.6$, $SD = 16.3$) in the 50erc condition was not significantly different ($t(46) = 1.36$, $p = .18$). Similarly, the difference of 2.6% between the self-reported criterion and the calculated criterion ($M = 56.3$, $SD = 14.7$) in the 75erc condition was not significantly different ($t(46) = 1.38$, $p = .17$). In summary, the self-reported criterion was higher than the calculated criterion in the 25erc condition but not different from it in the 50erc or 75erc conditions.

Relative Frequency of Choosing the Safe Option. As with Experiment 1, participants' relative frequency of choosing the safe option was also analyzed. In Figure 23, the relative frequency is shown for the 25erc, 50erc, and 75erc conditions. The mean relative frequency was 41.0% ($SD = 11.1\%$) in the 25erc condition, 32.4% ($SD = 11.1\%$) in the 50erc condition, and 22.9% ($SD = 11.4\%$) in the 75erc condition. An ANOVA with the economically rational criterion manipulation (25erc, 50erc, and 75erc) on the relative frequencies showed a main effect of this manipulation ($F(2,154) = 35.40$, $p < .001$). Three post-hoc pairwise comparisons were conducted. The 50erc condition had a significantly higher relative frequency than the 25erc condition with a difference of 8.6% ($t(154) = 3.87$, $p < .001$ with Tukey correction). The 75erc condition had a significantly higher relative frequency than the 25erc condition with a difference of 18.1% ($t(154) = 8.41$, $p < .001$, with Tukey correction). The 75erc condition had a significantly higher relative frequency than the 50erc condition with a difference of 9.6% (t(154)

= 4.33, $p < .001$, with Tukey correction). In summary, the manipulation had an effect such that participants chose the safe option most often in the 75erc condition and least in the 25erc condition.

In addition, the relative frequency of choosing the safe option in each condition was compared to the relative frequency the participants would get if they used objective probabilities and the economically rational criterion. The economically rational based relative frequency would have been 48% in the 25erc condition, 20% to 32% in the 50erc condition, and 10% in the 75erc condition. The economically rational based relative frequency in the 50erc condition was a range because the expected value of the safe and risky option was the same in trials with 50% objective probability. Therefore, in these trials, either option was economically rational. Therefore, if the relative frequency lay between 20% and 32%, it was considered consistent with the economically rational standard.

Three post hoc one-sample t-tests compared the relative frequency of choosing the safe option in each condition with the respective economically rational based relative frequency. In the 25erc condition, the relative frequency was significantly lower than 48% with a difference of -7.0% ($t(53) = 4.65$, $p < .001$, corrected alpha = .025). In the 50erc condition, the relative frequency was higher than the range of 20% to 32%. Therefore, it was compared to 32%. The difference of 0.4% between the relative frequency and 32% was not significantly different ($t(48) = .026$, $p = .80$). In the 75erc condition, the relative frequency was significantly higher than 10% with a difference of 12.9% ($t(53) = 8.30$, $p < .001$, corrected alpha = .017). This pattern of results suggests that participants were risk-seeking in the 25erc condition, risk-neutral in the 50erc condition, and risk-averse in the 75erc condition.

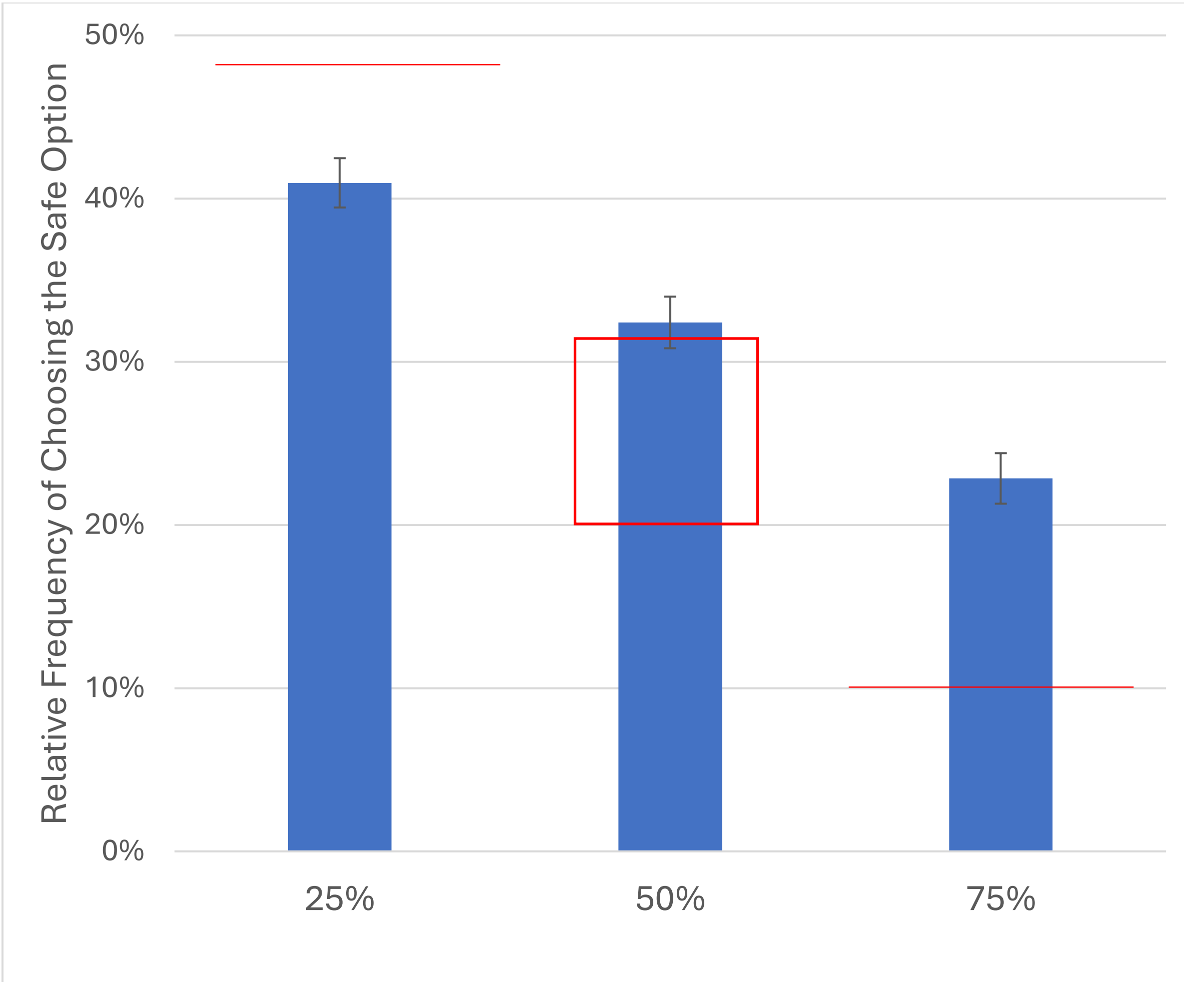


Figure 23
Relative Frequency of Choosing the Safe Option in the Three Conditions of Experiment 2
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red line in the 75% economically rational criterion condition shows the relative frequency of 10%. The red box in the 50% economically rational criterion condition shows the relative frequency between 20% to 32%. These lines are the relative frequency the participants would get if they always chose the safe option when the objective probability was above the economically rational criterion in their respective conditions.

Likelihood Ratings

Figure 24 shows the likelihood ratings as a function of objective probabilities in the three conditions. The mean likelihood rating was 37.1% (SD = 7.0%) in the 25erc condition, 39.3% (SD = 11.9%) in the 50erc condition, and 37.8% (SD = 7.1%) in the 75erc condition. An ANOVA with the economically rational criterion manipulation (25erc, 50erc, and 75erc) on the mean likelihood ratings showed no significant differences among the conditions ($F(2,154) = 0.81$, $p = .45$). In addition, three post hoc one-sample t-tests compared the mean likelihood rating in each condition with the proportion of drought trials of 36%. In the 25erc condition, the difference of 1.1% between the mean likelihood rating and the proportion of drought trials was not significant ($t(53) = 1.16$, $p = .25$). In the 50erc condition the difference of 4.3% between the mean rating and 36% was also not significant ($t(48) = 1.94$, $p = .06$, corrected alpha = .017). In the 75erc condition from the difference of 1.8% between the mean rating and 36% was not significant ($t(53) = 1.88$, $p = .07$, corrected alpha = .025). This suggests that the likelihood of drought ratings was close to the proportion of drought trials (relative frequency of drought) and not affected by the manipulation of economically rational criterion. In addition, as seen in Figure 24, the three conditions had similar likelihood rating patterns. Likelihood ratings were slightly overestimated at all objective probability levels except for 35%.

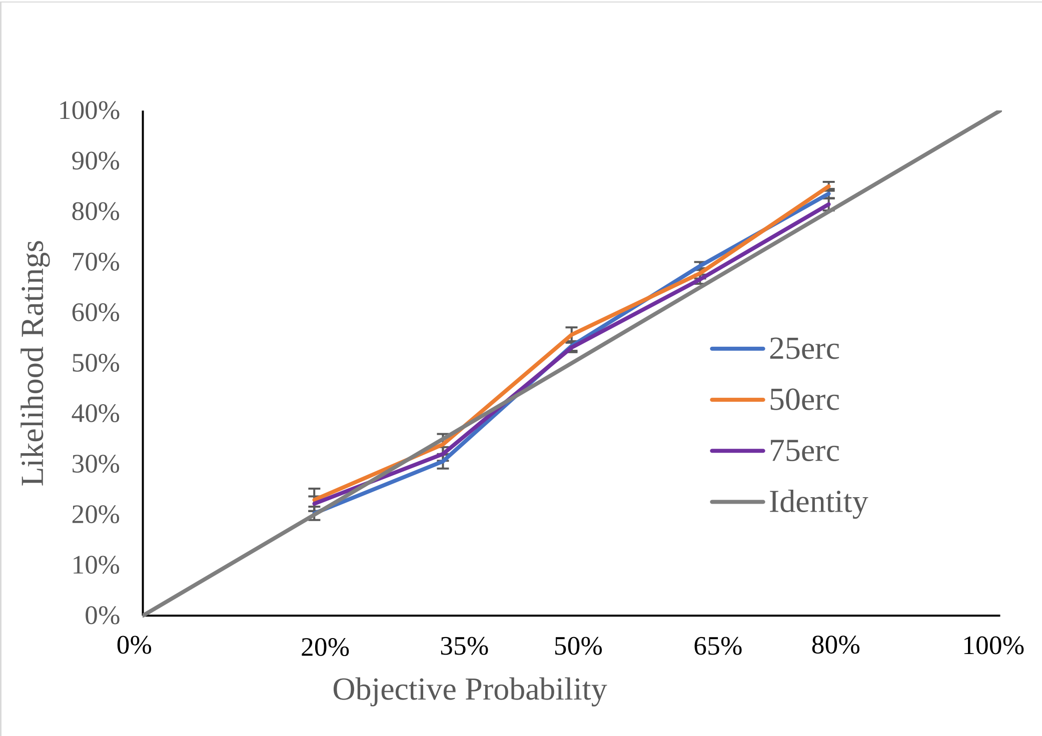
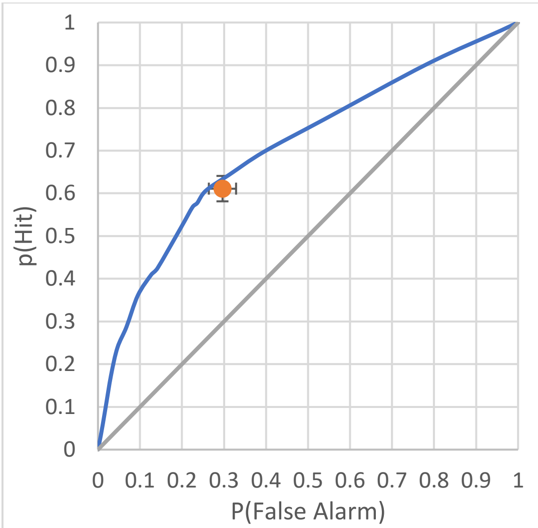


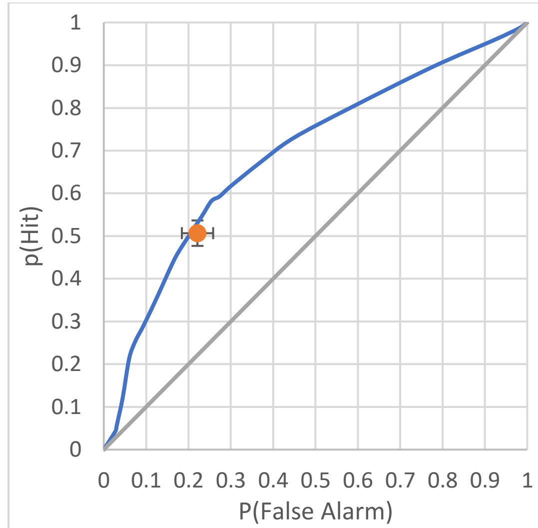
Figure 24
Likelihood Ratings in the 25%, 50%, and 75% Economically Rational Criterion Conditions as a Function of Objective Probability in Experiment 2
The blue line represents the 25 economically rational criterion condition. The orange line represents the 50 economically rational criterion condition. The purple line represents the 75 economically rational criterion condition. There was no observed centering effect.

Sensitivity

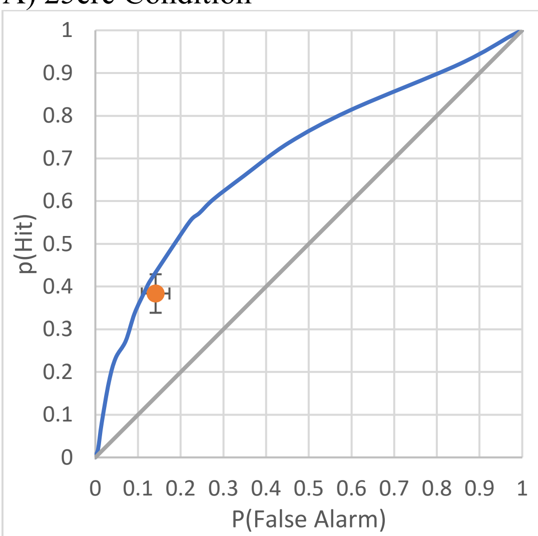
The next analysis examined the sensitivity or participants' ability to predict the drought, as measured by the area under the ROC curve. In Figure 25, ROC plots are shown for the 25erc, 50erc, and 75erc conditions respectively. The ROC curves of the three conditions were similar, indicating that all conditions had similar sensitivity or the ability to predict drought based on provided drought forecasts. The mean percent area under ROC curve was 71.2% (SD = 5.0%) in the 25erc condition, 70.3% (SD = 5.5%) in the 50erc condition, and 70.6% (SD = 5.4%) in the 75erc condition. An ANOVA with the economically rational criterion manipulation (25erc, 50erc, and 75erc) as the independent variable on the mean percent area under curve revealed that the differences among the conditions were not significant ($F(2,154) = .043$, $p = .65$). This suggests that the economically rational criterion manipulation had no effect on the sensitivity.



A) 25erc Condition



B) 50erc Condition



C) 75erc Condition

Figure 25
ROC Plot for the 25% Economically Rational Criterion Condition in Experiment 2
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 71.2% for the 25erc condition, 70.4% for the 50erc condition, and 70.6% for the 75erc condition.

Discussion

Experiment 2 had two predictions: 1) Participants would shift their subjective criterion towards 50%; 2) The likelihood ratings and sensitivity were not expected to be affected by this manipulation. The results showed both the calculated and the self-reported criterion showed the shift towards 50% in the 25erc and 75erc condition, while there was little shift in the 50erc condition. This result supports the centering effect on the subjective criterion. In addition, the analysis of the relative frequency of choosing the safe option indicated that participants were risk-seeking in the 25erc condition, risk-neutral in the 50erc condition, and risk-averse in the 75erc condition. Together with the same trial composition and sensitivity among the conditions, this difference in relative frequency of choosing the safe option indicated a higher subjective criterion than the economically rational criterion of 25% in the 25erc condition, a subjective criterion close to the economically rational criterion of 50% in the 50rc condition, and a lower subjective criterion than the economically rational criterion of 75erc condition. Thus, the result of this measure was consistent with the two measures of the subjective criterion. Therefore, all three measures showed the centering effect in subjective criterion.

Next, as expected, the manipulation of economically rational criterion did not affect the likelihood ratings or the sensitivity. This suggests that changing the economically rational criterion itself does not affect people's perception of the probability of drought or their ability to predict the drought.

The results also yielded additional findings. First, the self-reported criterion was significantly higher than the calculated criterion in the 25erc condition. This was consistent with experiment 1. On the other hand, the self-reported criterion was similar to the calculated criterion in the 50erc and 75erc conditions. This suggests that participants were better at estimating their own criterion in these two conditions. Next, in the ROC plots (Figure 25), the binary decisions (the decision dot indicating proportion of hits and false alarm) were consistent with the respective ROC curves, providing support for the calculation method of the calculated criterion. Finally, as seen in Figure 24 and consistent with experiment 1, there was a slight overestimation but no observed centering effect in likelihood ratings as indicated by comparing them to the objective probability levels.

Overall, experiment 2 yielded support for centering affecting the subjective criterion and in turn binary decisions, consistent with the subjective criterion hypothesis. Likelihood ratings and the sensitivity were not affected by the manipulation of economically rational criterion. Likelihood ratings were also not affected by the centering effect.

Experiment 3

Experiment 3 examined the gain-loss framing effect and the centering effect combined with a larger sample size than experiment 1 and 2 combined. It had two goals: 1) Examine whether there is an interaction between the gain-loss framing effect and the centering effect; 2) Reexamine the gain-loss framing effect with higher statistical power compared to experiment 1. Experiment 3 used the same task as experiment 1 and 2 with both the economically rational criterion and gain-loss framing manipulated. The economically rational criterion was manipulated in the same way as experiment 2, using the 50erc or the 25erc but not the 75erc condition. The reason for excluding the 75erc condition was to maximize the number of participants in each condition. Real life severe weather events usually require people to take protective action at a low probability which corresponds to a low economically rational criterion. Therefore, the 75erc condition was not as realistic as the 50erc or the 25erc conditions and therefore the most disposable among the three. It was hence excluded to maximize participants into other conditions. Both the gain frame and loss frame from experiment 1 were used in experiment 3 to examine the gain-loss framing effect.

Method

Participants

A total of 591 participants from the US were recruited from Prolific Academic in March 2024, a crowdsourcing platform for online research. After the elimination process, 558 participants were used in the analysis. Eleven were eliminated for having a lower than 0.7 ReCAPTCHA score, a bot detection system used by Qualtrics survey platform. Twenty-two were eliminated for failing the comprehension check (same question as experiment 1). Each participant was paid $4 for participation plus a performance based monetary bonus, same as experiment 1 and 2. The mean age was 39 (SD = 11.91, range 18 to 81 years). There were 229 (41%) females, 327 (59%) males, 1 (<1%) who preferred not to say, and 1 where Prolific did not provide data.

Procedure and Stimuli

The procedure was mostly identical to experiment 1 and 2. Participants performed a drought decision task as described in experiment 1 and 2. The outcomes of their decisions were tracked with a point structure mostly identical to the ones used in experiment 1 and 2. Their goal was to have as many points as possible by the end of the experiment. They were paid a monetary bonus commensurate with their ending point balance.

Point Structure. In Experiment 3, both the gain frame and the loss frame were used. The point structure used in experiment 3 was mostly unchanged from experiment 1 and 2. There were two economically rational criterion conditions (erc): 25% (25erc), and 50% (50erc). There were two frame conditions: gain frame and loss frame. These two variables were fully crossed so there were four conditions in total. The point structure in the 25erc and 50erc loss frame conditions were identical to those of experiment 2 while the point structure in the 25erc gain frame condition was identical to the gain condition in experiment 1 (See table 8). The only new condition was the 50erc gain frame condition, where participants chose between a safe option of gaining 200 points and a risky option of gaining 400 points if a drought did not occur and 0 points if a drought occurred (Table 8). The same beginning point balance and payment structure from experiment 2, 25erc and 50erc conditions were used.

25erc Condition Gain Frame		
	Safe Option	Risky Option
Drought Occurred	300 points	0 points
Drought Did Not Occur	300 points	400 points
25erc Condition Loss Frame		
	Safe Option	Risky Option
Drought Occurred	-100 points	-400 points
Drought Did Not Occur	-100 points	0 points
50erc Condition Gain Frame		
	Safe Option	Risky Option
Drought Occurred	200 points	0 points
Drought Did Not Occur	200 points	400 points
50erc Condition Loss Frame		
	Safe Option	Risky Option
Drought Occurred	-200 points	-400 points
Drought Did Not Occur	-200 points	0 points

Table 8
Point Structure of the Four Conditions in Experiment 3

Trial Structure. The trial structure was the same as experiment 2. After reading through the background information, point structure, and going through a practice trial, participants saw two attention check questions. Participants then began the 50 experimental trials whose composition was identical to experiment 1 and 2. The post-survey questions were identical to experiment 1 and 2.

Design

Experiment 3 used a 2 x 2 x 6 mixed design. There were two between-group independent variables: Economically rational criterion with two levels: 25erc, and 50erc and gain-loss framing with two levels: Gain frame and loss frame. There was one within-group independent variable: objective probability of a drought with levels: 20%, 35%, 50%, 65% and 80%. The proportion of drought was 36% across 50 trials, same as experiment 1 and 2. Like in experiment 1 and 2, as the objective probabilities (shown to participants) were reliable (M = 36.5%), the mean objective probability and the proportion of drought trials were considered the same (36%) in the analyses.

In terms of dependent variables, as with experiment 1 and 2, participants made likelihood ratings using a VAS and binary crop decisions on each trial. At the end of the experiment, participants answered a comprehension check question, reported their criterion, difficulty of understanding the task, and any glitches they encountered as with experiment 1 and 2. Unlike experiment 1 and 2 where no difference in the difficulty between conditions was reported, an ANOVA with gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables, on the difficulty ratings revealed that participants in experiment 3 reported the loss frame condition (M = 14.2, SD = 22.4, range 0 to

100) to be slightly more difficult to understand than the gain frame condition (M = 10.5, SD = 16.6, $F(1, 554) = 4.94$, $p = .027$).

Results

Analysis Overview

Based on results from experiment 1 and 2, the prediction was that there should be both the gain-loss framing effect and the centering effect on the subjective criterion. As with the prediction for experiment 1, the gain-loss framing effect should lead to a higher subjective criterion in the loss condition than in the gain condition. As with experiment 2, the centering effect should shift the subjective criterion in the 25erc condition towards 50% while having no effect on the subjective criterion in the 50erc condition. Moreover, no interaction between these two effects was expected as their mechanisms should be theoretically independent. The gain-loss framing effect stems from the utility function of prospect theory (Tversky & Kahneman, 1979) while the centering describes people's tendency to shift their judgement towards the center of the range (Poulton, 1979). Finally, the manipulation of gain-loss framing and economically rational criterion were not expected to affect likelihood ratings or sensitivity.

The same set of dependent measures were used as were used in experiment 1 and 2. To measure decision bias, the calculated criterion and the self-reported criterion were analyzed. In addition, the relative frequency of choosing the safe option was analyzed as an alternative way to examine the decision bias. Next, likelihood ratings were analyzed to detect any bias in subjective likelihood. Finally, to measure the sensitivity, the area under the ROC curve was analyzed. A series of ANOVAs and t-tests were conducted. Holm-Bonferroni Method was used for planned and post hoc t-tests.

Decision Bias

The first set of analyses examined the degree to which participants biased their decisions towards the safe option.

Calculated Subjective Criterion. In Figure 26, the calculated criterion is shown for the four conditions. The calculated criterion in the loss frame condition was higher than in the gain frame condition. It was higher than the economically rational criterion in the 25erc condition and lower than the economically rational criterion in the 50erc condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables were conducted on the calculated criterion. There was a main effect of the gain-loss framing manipulation such that in the loss frame (M = 40.6%, SD = 15.9%) the calculated criterion was 2.7% higher than in the gain frame (M = 37.9%, SD = 16.9%; $F(1, 554) = 4.38$, $p = .037$). There was a main effect of the economically rational criterion manipulation such that the calculated criterion in the 50erc condition (M = 43.7%, SD = 16.7%) was 9.2% higher than in the 25erc condition (M = 34.5%, SD = 15.0%; $F(1, 554) = 46.76$, $p < .001$). There was no significant interaction between the gain-loss framing and the economically rational criterion manipulation ($F(1, 554) = 1.03$, $p = .31$). Similar to experiment 1 and 2, the two manipulations both had an effect on the calculated criterion.

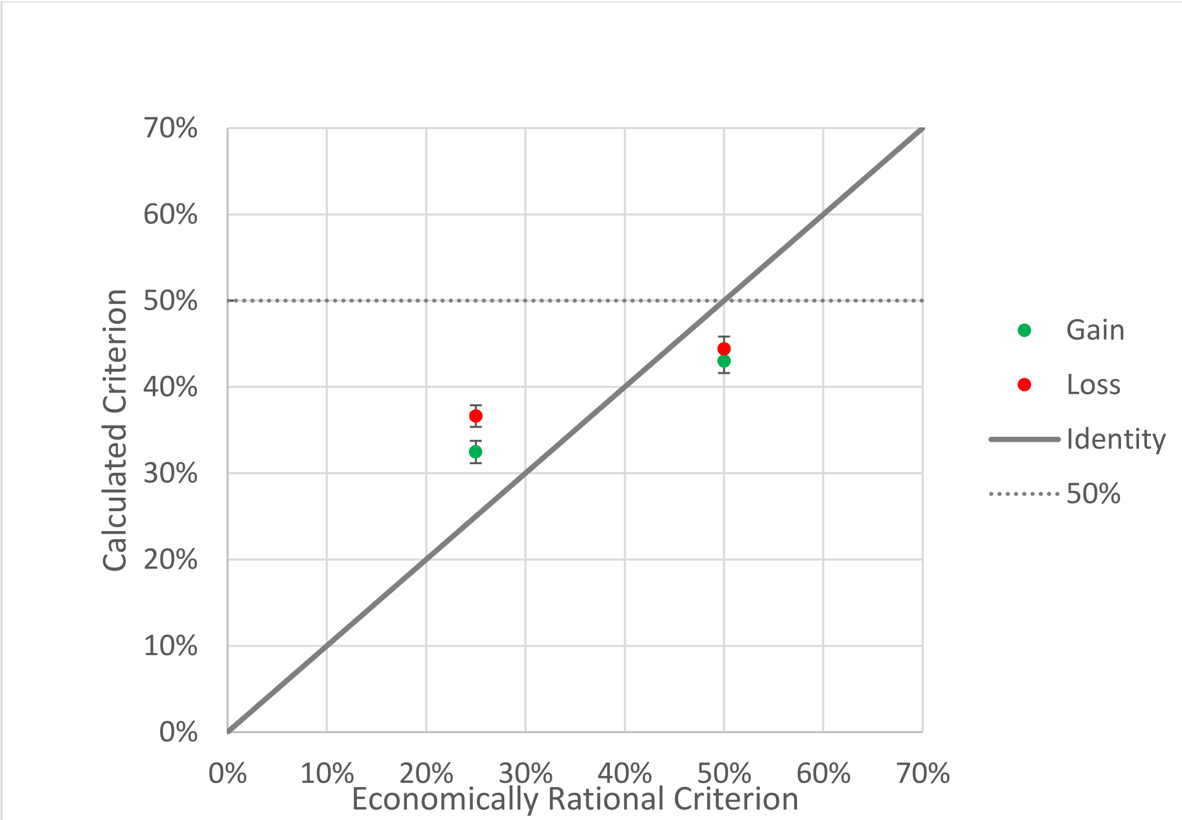


Figure 26
Calculated Subjective Criterion in the Four Conditions of Experiment 3
In the 25erc gain frame condition, the calculated criterion was 7.5% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the calculated criterion was 11.6% higher than 25%. In the 50erc gain frame condition, the calculated criterion was 7.0% lower than the economically rational criterion of 50%. In the 50erc loss frame condition, the calculated criterion was 5.6% lower than 50%.

To better detect the interaction between gain-loss framing and centering, the signed deviation of the calculated criterion from the economically rational criterion was analyzed. A higher (more positive) signed deviation in the loss frame condition than in the gain frame condition indicates the gain-loss framing effect. A higher signed deviation in the 25% condition than in the 50% condition indicates the centering effect. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables were conducted on the signed deviation. The signed deviation in the loss frame condition ($M = 2.8\%$, $SD = 17.7\%$) was significantly higher than in the gain frame condition ($M = 0.2\%$, $SD = 17.7\%$) with a difference of 2.6% ($F(1, 554) = 4.38$, $p = .037$). The signed deviation of calculated criterion in the 25erc condition ($M = 9.5\%$, $SD = 15.0\%$) was significantly higher than in the 50erc condition ($M = -6.3\%$, $SD = 16.5\%$) with a difference of 15.8% ($F(1, 554) = 139.51$, $p < .001$). The interaction between the gain-loss framing and the economically rational criterion manipulations failed to reach significance ($F(1, 554) = 1.03$, $p = .31$). In summary, the result showed the expected main effects of both the gain-loss framing and the economically rational criterion manipulations but no interaction.

In addition, to better understand the deviation in each condition, four planned one-sample t-tests compared the calculated criterion with the economically rational criterion in each condition (25erc gain frame, 25erc loss frame, 50erc gain frame, and 50erc loss frame). In the 25erc gain frame condition, the calculated criterion ($M = 32.5\%$, $SD = 15.4\%$) was significantly higher than the economically rational criterion of 25% with a difference of 7.5% ($t(140) = 5.76$, $p < .001$, corrected alpha = .017). In the 25erc loss frame condition, the calculated criterion ($M = 36.6\%$, $SD = 14.3\%$) was significantly higher than 25% with a difference of 11.6% ($t(129) = 9.27$, $p < .001$, corrected alpha = .013). In the 50erc gain frame condition, the calculated criterion

(M = 43.0%, SD = 16.8%) was significantly lower than the economically rational criterion of 50% with a difference of -7.0% ($t(148) = 5.10$, $p < .001$, corrected alpha = .025). In the 50erc loss frame condition, the calculated criterion (M = 44.4%, SD = 16.5%) was also significantly lower than 50% with a difference of -5.6% ($t(137) = 137$, $p < .001$, corrected alpha = .05).

Taken together, these results offer support for the prediction that both manipulations affected the calculated criterion. Both a gain-loss framing effect and a centering effect were observed, but they did not interact with one another. While the calculated criterion in the 25erc condition was as predicted by centering higher than the economically rational criterion, the calculated criterion in the 50erc condition was unexpectedly significantly lower than the economically rational criterion.

Self-Reported Criterion. In Figure 27, the self-reported criterion is shown for shown for the four conditions. The self-reported criterion in the loss frame condition was higher than in the gain frame condition. It was higher than the economically rational criterion in the 25erc condition and smaller than the economically rational criterion in the 50erc condition. Among the 558 participants, 464 (83%) indicated that they had used a criterion for their decision and 94 (17%) indicated that they did not. One participant among the 464 was removed from the analysis as they provided a self-reported criterion of 505, which was out of the range from 0 to 100. The following analyses included only the remaining 463 participants. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables were conducted on the self-reported criterion. There was a main effect of the gain-loss framing manipulation such that in the loss frame (M = 46.8%, SD = 12.2%) the self-reported criterion was 3.2% higher than in the gain frame (M = 43.6%, SD

= 13.0%; $F(1, 459) = 9.06$, $p = .003$). There was a main effect of the economically rational criterion manipulation such that the self-reported criterion in the 50erc condition ($M = 48.7\%$, $SD = 11.8\%$) was 7% higher than in the 25erc condition ($M = 41.7\%$, $SD = 12.7\%$; $F(1, 459) = 39.64$, $p < .001$). There was no significant interaction between the gain-loss framing and the economically rational criterion manipulation ($F(1, 459) = 2.76$, $p = .097$). Similar to experiment 1 and 2, the two manipulations both had an effect on the self-reported criterion.

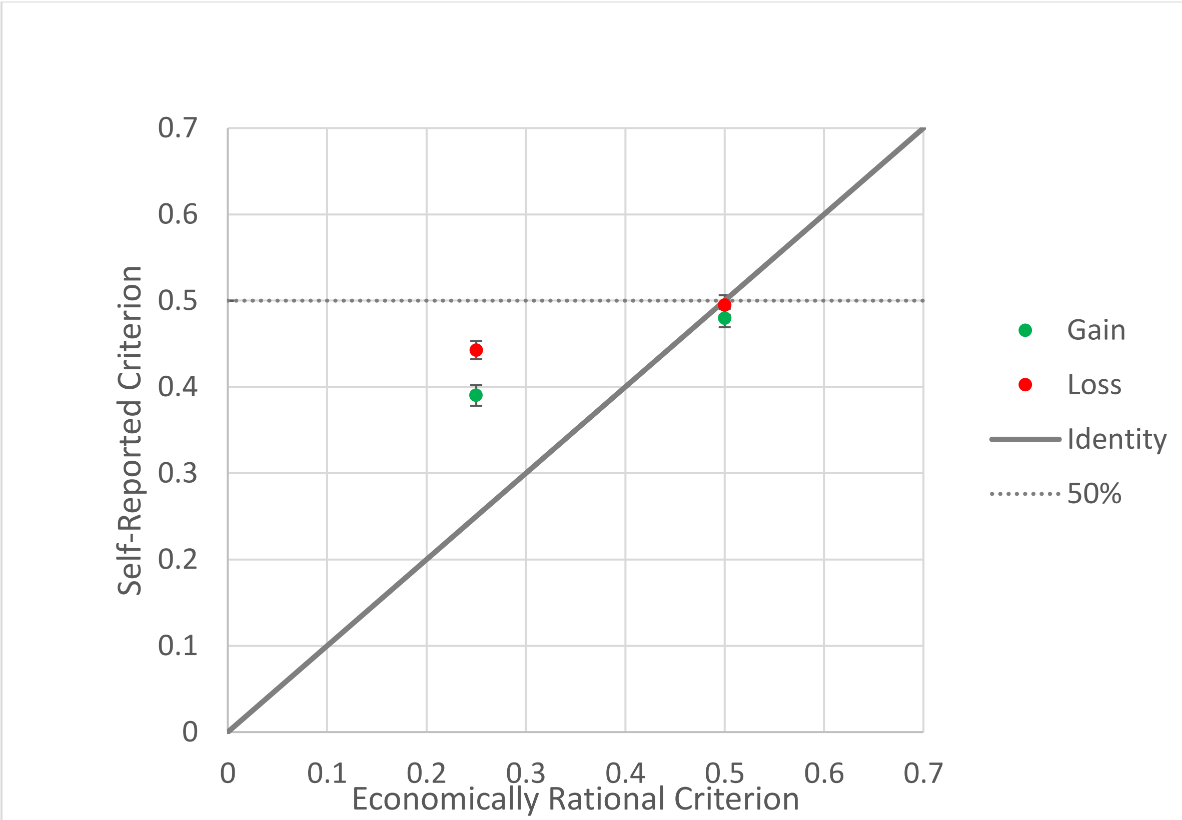


Figure 27
Self-Reported Subjective Criterion in the Four Conditions of Experiment 3
In the 25erc gain frame condition, the self-reported criterion was 14.0% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the self-reported criterion was 19.3% higher than 25%. In the 50erc gain frame condition, the self-reported criterion was 2.0% lower than the economically rational criterion of 50%. The difference was insignificant. In the 50erc loss frame condition, the self-reported criterion was 0.5% lower than 50%, but the difference was insignificant.

To better detect the interaction between gain-loss framing and centering, the signed deviation of the self-reported criterion from the economically rational criterion was analyzed with the same set of analyses and predictions for the calculated criterion. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables was conducted on the signed deviation among the remaining participants. The mean signed deviation in the loss frame condition ($M = 9.6\%$, $SD = 15.5\%$) was significantly higher than in the gain frame condition ($M = 5.8\%$, $SD = 14.6\%$) with a difference of 3.8% ($F(1, 459) = 9.06$, $p = .003$). The signed deviation of self-reported criterion in the 25erc condition ($M = 16.7\%$, $SD = 12.7\%$) was significantly higher than in the 50erc condition ($M = -1.3\%$, $SD = 11.8\%$) with a difference of 15.4% ($F(1, 459) = 253.32$, $p < .001$. The interaction between framing and centering failed to reach significance ($F(1, 459) = 2.76$, $p = .097$). In summary, the result showed the expected main effects of both the gain-loss framing and the economically rational criterion manipulations but no interaction.

In addition, to better understand the deviation in each condition, four post hoc one-sample t-tests compared the self-reported criterion with the economically rational criterion in each condition (25erc gain frame, 25erc loss frame, 50erc gain frame, and 50erc loss frame). In the 25erc gain frame condition, the self-reported criterion ($M = 39.0\%$, $SD = 12.8\%$) was significantly higher than the economically rational criterion of 25% with a difference of 14.0% ($t(114) = 11.73$, $p < .001$, corrected alpha = .013). In the 25erc loss frame condition, the self-reported criterion ($M = 44.3\%$, $SD = 12.0\%$) was also significantly higher than 25% with a difference of 19.3% ($t(115) = 17.31$, $p < .001$, corrected alpha = .017). In the 50erc gain frame condition, the -2% difference between the self-reported criterion ($M = 48.0\%$, $SD = 11.6\%$) and 50% was not significant ($t(121) = 1.92$, $p = .057$). In the 50erc loss frame condition, the -0.5%

difference between the self-reported criterion (M = 49.5%, SD = 12.0%) and 50% was not significant (t(109) = 0.45, p = .66).

Taken together, these results suggest, as predicted and as with the calculated criterion, both manipulations affected the self-reported criterion. Both a gain-loss framing effect and a centering effect were observed, but they did not interact. The self-reported criterion was as predicted higher than the economically rational criterion in the 25erc condition and not different from the economically rational criterion in the 50erc condition. The significant difference between the calculated criterion and 50% in the 50erc condition was not seen in the self-reported criterion.

Finally, four planned paired t-tests compared the self-reported criterion and the calculated criterion in each condition (25erc gain frame, 25erc loss frame, 50erc gain frame, and 50erc loss frame) to explore their differences. In the 25erc gain frame condition, the self-reported criterion was significantly higher than the calculated criterion (M = 31.6%, SD = 13.3%) with a difference of 7.4% (t(114) = 4.30, p < .001, corrected alpha = .013). In the 25erc loss frame condition, the self-reported criterion was significantly higher than the calculated criterion (M = 36.8%, SD = 14.0%) with a difference of 7.5% (t(115) = 2.77, p < .001 corrected alpha = .017). In the 50erc gain frame condition, the self-reported criterion was significantly higher than the calculated criterion (M = 42.3%, SD = 15.0%) with a difference of 5.7% (t(121) = 3.29, p = .001, corrected alpha = .05). In the 50erc loss frame condition, the self-reported criterion was significantly higher than the calculated criterion (M = 41.9%, SD = 12.8%) with a difference of 7.6% (t(109) = 5.02, p <.001, corrected alpha = .025). This suggests that the self-reported criterion was significantly higher than the calculated criterion in all four conditions among those who reported a self-reported criterion.

Relative Frequency of Choosing the Safe Option. As with experiment 1 and 2, the participants' relative frequency of choosing the safe option was also analyzed. In Figure 28, the relative frequency is shown for the four conditions. Participants had higher relative frequency than the economically rational based relative frequency, the frequency based on objective probabilities and the economically rational criterion, in all but 25erc loss frame condition, where the relative frequency was lower. The economically rational based relative frequency was 48% in the 25erc condition, and 20% to 32% in the 50erc condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables were conducted on the relative frequency. There was a main effect of the gain-loss framing manipulation such that in the gain frame ($M = 43\%$, $SD = 18\%$) the relative frequency was 3% higher than in the loss frame ($M = 40\%$, $SD = 16\%$; $F(1,554) = 5.65$, $p = .018$). There was a main effect of the economically rational criterion manipulation such that the relative frequency in the 25erc condition ($M = 47\%$, $SD = 16\%$) was 10% higher than in the 50erc condition ($M = 37\%$, $SD = 15\%$; $F(1,554) = 59.11$, $p < .001$). There was a significant interaction between the gain-loss framing and the economically rational criterion manipulation such that the gain-loss framing manipulation had a greater effect at 25erc compared to 50erc ($F(1,554) = 10.21$, $p = .001$). Similar to experiment 1 and 2, the two manipulations both had an effect on the relative frequency. However, an interaction was detected.

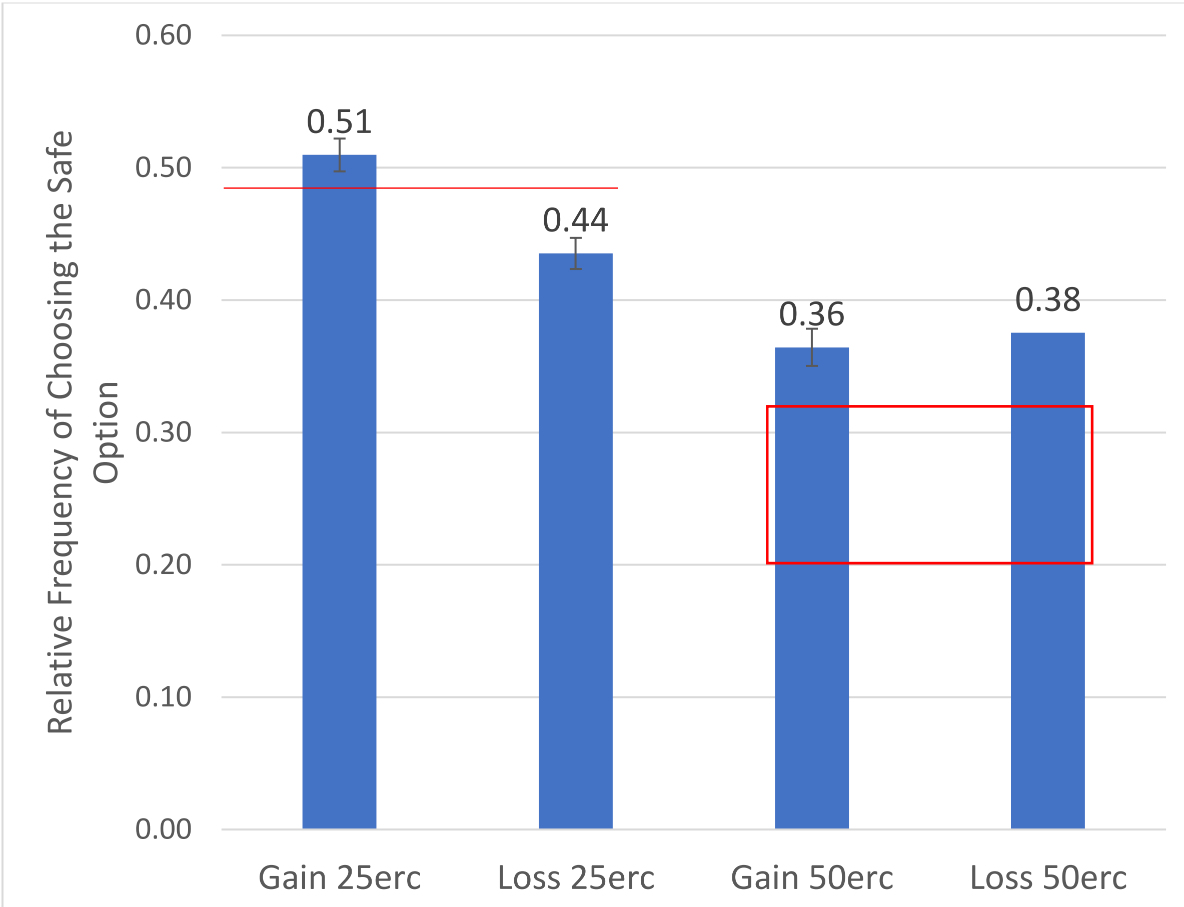


Figure 28
Relative Frequency of Choosing the Safe Option in the Four Conditions of Experiment 3
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red box in the 50% economically rational criterion condition shows the relative frequency between 20% to 32%. These lines are the relative frequency the participants would get if they always chose the safe option when the objective probability was above the economically rational criterion in their respective conditions.

To better detect the interaction between gain-loss framing and centering, the signed deviation of the relative frequency from the economically rational based relative frequency was analyzed with the same set of analyses for the calculated criterion and the self-reported criterion. A higher (more positive) signed deviation in the loss frame condition than in the gain frame condition suggests a gain-loss framing effect. A higher (more positive) signed deviation in the 50erc condition compared to the 25erc condition can be the result of the centering effect shifting the subjective criterion in the 25erc condition upward (creating a negative deviation of relative frequency) but not the 50erc condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables were conducted on the signed deviation. The mean signed deviation in the gain frame condition ($M = 3.7\%$, $SD = 16.1\%$) was significantly higher than in the loss frame condition ($M = 0.7\%$, $SD = 16.2\%$) with a difference of 3.0% ($F(1,554) = 5.65$, $p = .018$). The mean signed deviation in the 50erc condition ($M = 5.0\%$, $SD = 15.4\%$) was significantly higher than in the 25erc condition ($M = -0.6\%$, $SD = 16.5\%$) with a difference of 5.6% ($F(1,554) = 18.42$, $p < .001$). There was an interaction between the two manipulations such that the difference of the relative frequency between the gain frame and the loss frame was higher at the 25erc condition than in the 50erc condition ($F(1,554) = 10.21$, $p = .001$). As expected, these results suggest that increasing the economically rational criterion and using a loss frame made people choose the safe option less often. However, there is an unexpected interaction suggesting that the framing manipulation had little effect at the 50erc condition while having a bigger effect at the 25erc condition.

In addition, four planned one-sample t-tests compared the relative frequency in each condition to the economically rational based relative frequency to further examine participants' risk-taking tendencies and to explore the unexpected interaction. A relative frequency significantly higher than the economically rational based relative frequency suggests a risk-averse tendency. A relative frequency significantly lower than the economically rational based relative frequency suggests a risk-seeking tendency. In the 25erc gain frame condition, the relative frequency was significantly higher than 48% with a difference of 3% ($t(140) = 1.99$, $p = .049$, corrected alpha = .050). In the 25erc loss frame condition, the relative frequency was significantly lower than 48% with a difference of -4% ($t(129) = 3.61$, $p < .001$, corrected alpha = .025). In the 50erc gain frame condition, the relative frequency was significantly higher than 32% with a difference of 4% ($t(148) = 3.77$, $p < .001$, corrected alpha = .017). In the 50erc loss frame condition, the relative frequency was also significantly higher than 32% with a difference of 6% ($t(137) = 3.93$, $p < .001$, corrected alpha = .013). This result suggests that participants were overall risk averse in the 25erc gain, 50erc gain and 50erc loss conditions. They were risk seeking in the 25erc loss condition. The interaction was that participants changed their risk-taking tendency between the gain frame and the loss frame in the 25erc condition but not in the 50erc condition.

Likelihood Ratings

Figure 29 shows the likelihood ratings as a function of objective probabilities in the four conditions. The likelihood rating patterns were similar among the conditions. An ANOVA with the gain-loss framing and economically rational criterion manipulation (25erc, 50erc, and 75erc) as the independent variables, on the mean likelihood ratings was conducted. The 0.1% difference between the mean likelihood ratings in the gain frame condition ($M = 39.9\%$, $SD = 10.0\%$) and in the loss frame condition ($M = 40.0\%$, $SD = 10.3\%$) was not significant ($F(1, 554) = .017$, P

= .68). The 0.3% difference between the mean likelihood ratings in the 25erc condition (M = 40.1%, SD = 9.2%) and in the 50erc condition (M = 39.8%, SD = 10.9%) was not significant ($F(1, 554) = 0.23$, $P = .64$). There was no significant interaction between the gain-loss framing and the economically rational criterion ($F(1, 554) = 0.36$, $P = .55$). This result suggests that likelihood ratings were not affected by either manipulation.

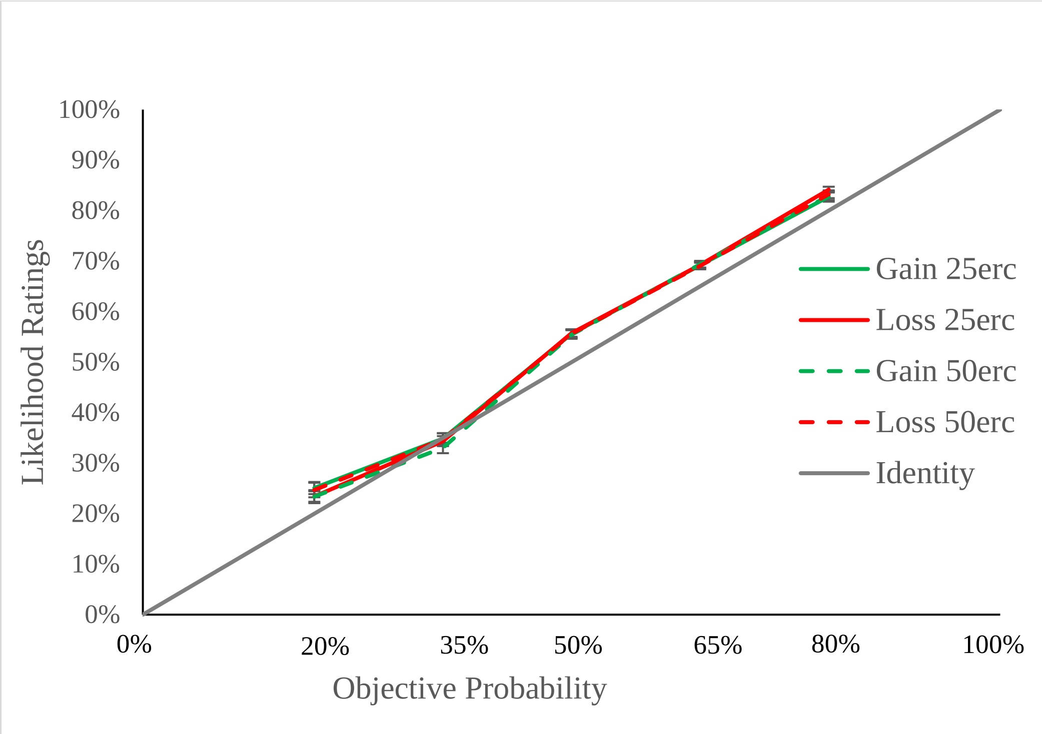
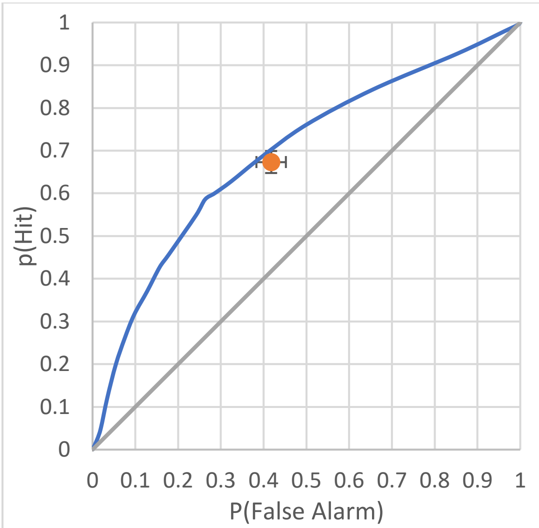


Figure 29
Likelihood Ratings in the Four Conditions as a Function of Objective Probability in Experiment 3
The green solid line represents the 25erc gain frame condition. The green dashed line represents the 50erc gain condition. The red solid line represents the 25erc loss frame condition. The red dashed line represents the 50erc loss condition. There was no observed centering effect.

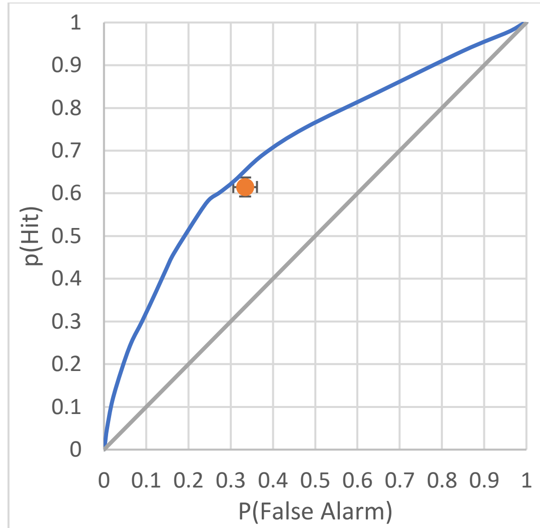
In addition, four post hoc one-sample t-tests compared the mean likelihood rating in each condition to the proportion of drought trials of 36%. A significant deviation suggests a bias in likelihood ratings. In the 25erc gain frame condition, the mean likelihood rating ($M = 40.5\%$, $SD = 9.8\%$) was significantly higher than the proportion of drought trials with a difference of 4.5% ($t(140) = 5.45$, $p < .001$, corrected alpha = .013). In the 25erc loss frame condition, the mean likelihood rating ($M = 39.7\%$, $SD = 8.5\%$) was significantly higher than the proportion of drought trials with a difference of 3.7% ($t(129) = 4.93$, $p < .001$, corrected alpha = .017). In the 50erc gain frame condition, the mean likelihood rating ($M = 39.3\%$, $SD = 10.1\%$) was significantly higher than the proportion of drought trials with a difference of 3.3% ($t(148) = 4.01$, $p < .001$, corrected alpha = .025). In the 50erc loss condition, the mean likelihood rating ($M = 40.3\%$, $SD = 11.8\%$) was also significantly higher than the proportion of drought trials with a difference of 4.3% ($t(137) = 4.28$, $p < .001$, corrected alpha = .05). This suggests that, as with the previous two experiments, the likelihood ratings were overestimated compared to the proportion of drought trials regardless of the gain-loss framing or the economically rational criterion manipulation. Finally, as seen in Figure 29, the four conditions had similar likelihood rating patterns. Likelihood ratings were overestimated at all objective probability levels except for 35%, same as experiment 1 and 2.

Sensitivity

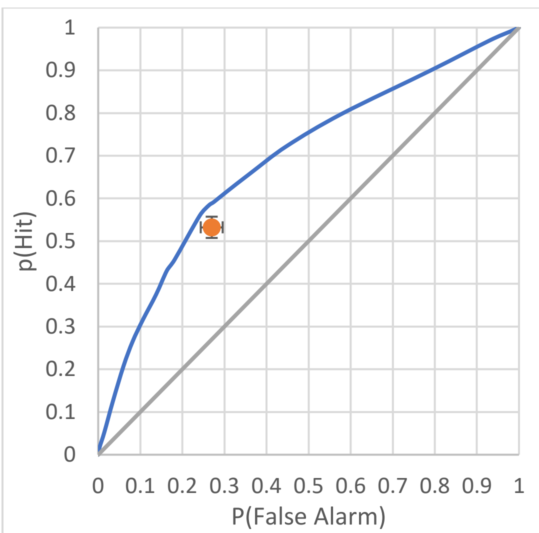
The next analysis examined the sensitivity or participants' ability to predict the drought, as measured by the area under the ROC curve. In Figure 30, ROC plots are shown for the four conditions respectively. The ROC curves of the four conditions were similar, indicating that both conditions had similar sensitivity or the ability to predict drought based on provided drought forecasts. All but 50erc gain frame conditions had the 95% CI of binary decisions overlapping with the ROC curves. The mean percent Area under ROC curve was 70.0% (SD = 8.1%) in the 25erc gain frame condition, 71.1% (SD = 5.2%) in the 25erc loss condition, 69.8% (SD = 7.6%) in the 50erc gain condition, and 69.6% (SD = 6.9%) in the 50erc loss condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc, and 75erc) manipulations as the independent variables on the mean percent area under the curve showed no significant main effect of the gain-loss framing ($F(1, 554) = 0.51$, $p = .47$) or the economically rational criterion manipulations ($F(1, 554) = 1.93$, $p = .17$). There also was not an interaction ($F(1, 554) = 0.94$, $p = .33$). This suggests that the sensitivity did not differ among the conditions and was not affected by the manipulation of gain-loss framing and economically rational criterion.



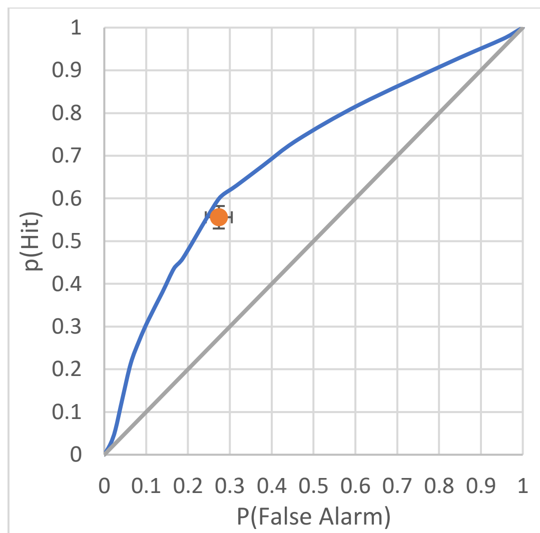
A) 25erc Gain Condition



B) 25erc Loss Condition



C) 25erc Loss Condition



D) 25erc Gain Condition

Figure 30
ROC Plot for the Four Conditions in Experiment 3
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 70.0% in the 25erc gain frame condition, 71.1% in the 25erc loss frame condition, 69.8% in the 50erc gain frame condition, 69.6%. in the 50erc loss condition.

Discussion

Experiment 3 tested the prediction that both the manipulation of gain-loss framing and the economically rational criterion (exposing a centering effect) should affect the subjective criterion. These two manipulations should not interact in the subjective criterion as the gain-loss framing effect and the centering effect were considered theoretically independent. In addition, the manipulation of gain-loss framing and economically rational criterion should not affect subjective likelihood or sensitivity.

In terms of the gain-loss framing manipulation, the results showed that as predicted participants in the loss frame condition had higher calculated and subjective criterion than those in the gain frame condition. However, the difference between the subjective criterion of the gain and loss frame conditions in experiment 3 was smaller than that in experiment 1 (calculated criterion: 2.9% in exp1, 2.6% in exp3; self-reported criterion: 5.6% in exp1, 3.8 in exp3). Thus, the significance in experiment 3 was likely due to the increased power. The participants in the loss frame condition also chose the safe option less often than those in the gain frame condition, showing a risk seeking decision bias. All these results are consistent with a gain-loss framing effect where people are more risk seeking in a loss frame than in a gain frame. In addition, as the likelihood ratings and the sensitivity were not affected by manipulating the framing, it can be inferred that the gain-loss framing effect operates by altering the subjective criterion. Finally, it should be noted that, unlike in experiment 1, participants in the loss condition reported this experiment to be slightly more difficult than participants in the gain condition. Therefore, the difficulty level might have been a confound here, if for some reason, participants became more risk-seeking with increased difficulty.

There was also evidence for a centering effect in the subjective criterion exposed by the economically rational criterion manipulation, similar to the findings of experiment 2. Both the calculated and self-reported criterion were higher than 25% in the 25erc condition suggesting movement toward the center (50%). The difference between the subjective criterion (calculated and self-reported) and the economically rational criterion was higher in the 25erc condition than in the 50erc condition. Interestingly, the calculated criterion in the 50erc condition was slightly but significantly lower than 50%, a result not seen in experiment 2 where the calculated criterion did not differ from 50% in the 50erc condition. This difference was not seen in the self-reported criterion. One possible explanation is that participants considered somewhere slightly lower than 50% as the center of the range in their mind. Finally, similar to the gain-loss framing manipulation, the manipulation of the economically rational criterion did not affect likelihood ratings or sensitivity.

Next, as expected, the interaction between the gain-loss framing and the economically rational criterion manipulations failed to reach significance in the analyses on both the calculated and subjective criterion. As the analyses were conducted on the signed deviation of the calculated and subjective from the economically rational criterion, it is fair to make the inference that the lack of interaction implies that these two subjective criterion measurements' shift due to the gain-loss framing was not affected by centering. However, there was an unexpected interaction between the two manipulations in the relative frequency of choosing the safe option. Participants were slightly risk-averse in both the gain frame and the loss frame in the 50erc condition, but they were, as predicted, only risk-averse in the gain frame while risk-seeking in the loss frame in the 25erc condition. With no interaction detected in the subjective criterion, likelihood ratings, or the sensitivity, it is unclear how this interaction came to be if subjective criterion and subjective likelihood were the sole determinant of the decisions.

In terms of effect sizes of the two effects, the gain-loss framing effect shifted the calculated criterion by 4.2% in the 25erc condition and 2.7% overall. The centering effect shifted the calculated criterion 9.5% (average between the gain and loss frame) above 25% in the 25erc condition. Therefore, it appears that at 25% economically rational criterion, the centering effect had a larger effect on the subjective criterion than the gain-loss framing effect.

The results also yielded additional findings. First, as with the previous experiments, the likelihood ratings were not affected by manipulating either gain-loss framing or economically rational criterion. However, they were overestimated compared to the proportion of drought trials (practically the same as the mean objective probabilities). The overestimation but no centering effect was observed in likelihood ratings as a function of objective probability levels as seen in Figure 29. This is also consistent with experiment 1 and 2.

Next, participants showed overall risk-averse tendency when operationalized as relative frequency to choose the safe option in all but 25erc loss frame condition where they were risk-seeking as predicted. This risk-seeking tendency in the 25erc loss frame condition is consistent with previous studies using a loss frame and an economically rational criterion lower than 50% (Gulacsik et al., 2022; Qin et al., 2024). The risk-averse relative frequency in both the 50erc gain and loss frame conditions was consistent with the observed lower-than-rational subjective criterion in these conditions. The risk-seeking frequency in the 25erc loss frame condition was consistent with the observed higher-than-rational subjective criterion. However, the risk-averse relative frequency in the 25erc gain frame condition was not consistent with the observed higher-than-rational subjective criterion. This inconsistency could be explained by overestimated likelihood ratings, which in this case perhaps counteracted the risk seeking tendency brought by the higher-than-rational subjective criterion and shifted the decisions towards risk aversion.

Next, in the ROC plots, the binary decisions (summarized by proportion of hits and false alarms in decisions) were consistent with the respective ROC curves (from likelihood ratings and actual tornado occurrence), suggesting that the subjective criterion and the subjective likelihood were the sole determinant of the binary decisions, in all conditions except for the 50erc gain frame condition. However, in the 50erc gain frame condition alone, which was not tested in previous experiments, the binary decisions were below the ROC curve, indicating a worse sensitivity in the binary decisions than in the likelihood ratings. This is the only inconsistency in binary decisions observed across all three experiments. It is unknown whether this inconsistency was a statistical false negative or whether there were some unknown effects on participants' behavior in this condition.

Overall, experiment 3 yielded support for both the gain-loss framing effect and the centering effect on subjective criterion and in turn the decisions, consistent with the subjective criterion hypothesis. The centering effect appeared to have a larger effect on the subjective criterion than the gain-loss framing effect when the economically rational criterion was at 25%. These two effects did not interact with other in most of the analyses reported here. Again, likelihood ratings and sensitivities were not affected by the manipulation of gain-loss framing or economically rational criterion.

General Discussion

Summary of Results

This research examined potential psychological mechanisms that might account for a shift in subjective criterion in naturalistic weather decision leading to a bias towards risky decision making. The subjective criterion (calculated and self-reported) was subject to two separate effects: a gain-loss framing effect and a centering effect. The reanalyses of previous experiments and three new experiments used a signal detection theory perspective to examine the role of gain-loss framing and centering on the subjective criterion in naturalistic tasks requiring decision making under risk. The gain-loss framing and the economically rational criterion (to expose a centering effect in the subjective criterion) were manipulated. The subjective criterion (calculated and self-reported) was found to be higher, more risk-seeking in a loss frame compared to a gain frame, consistent with the gain-loss framing effect as described by prospect theory. This effect was tested in Experiment 1 and 3. A trend in the expected direction was observed in experiment 1 but it failed to reach significance. The effect reached significance in experiment 3 with greater statistical power. A centering effect, tested in experiments 2 and 3, was observed in the subjective criterion (both calculated and self-reported) such that it shifted towards 50% when the economically rational criterion was 25% or 75% while stayed comparatively close to 50% when the economically rational criterion was 50%. Finally, the results of analysis of the ROC plots were consistent with the random likelihood model. In 8 out of 9 conditions the participants' binary decisions were consistent with the ROC curve based on their likelihood ratings and drought occurrence. This validates the estimation of the calculated criterion using likelihood ratings and binary decisions.

In addition to examining the subjective criterion, the current experiments also examined the effect of the manipulations on the subjective likelihood and sensitivity. Subjective likelihood was not affected by manipulating either the frame or economically rational criterion in any of the experiments. Similarly, the sensitivity was not affected by either manipulation in any of the experiments.

Naturalistic Tasks Requiring Decision Making under Risk

The findings here are consistent with the subjective criterion hypothesis and previous studies using similar naturalistic decision tasks. In terms of decisions shown by relative frequency of choosing the safe option, the risk-seeking decisions in the 25erc loss frame condition in all three experiments were consistent with previous studies using loss frames and a low economically rational criterion (Gulacsik et al., 2022; Qin et al., 2024). The risk-averse decisions in the gain frame were also consistent with previous naturalistic and artificial gamble studies using either mixed-gamble tasks with overall more chance to gain points than lose or pure gain frame tasks (Tversky & Kahneman, 1979; Demnitz & Joslyn, 2020).

Previous studies showed that while overall people can have relatively accurate albeit slightly overestimated perception of the likelihood of the severe weather event when provided with probabilistic information compared to when such information is absent, they somehow showed risk-seeking decisions in loss scenarios (Gulacsik et al., 2022; Qin et al., 2024). The same small overestimation was observed in the three experiments here. The similar risk-seeking tendency was also observed in the 25erc loss condition of all three experiments. This curiosity of risk-seeking decisions and overestimated subjective likelihood can be explained by people's biased subjective criterion as found in the current experiments. People had a higher-than-rational subjective criterion (calculated and self-reported) in loss frames when the economically rational criterion is low (<50%). The higher-than-rational subjective criterion might have counteracted the overestimated subjective likelihood, leading to risk-seeking decisions. On the other hand, in the gain conditions here or the mixed-gamble study (Demnitz & Joslyn, 2020), the subjective criterion, although still higher-than-rational, might not be high enough to fully counteract the overestimated subjective likelihood. This leads to risk-averse decisions. In summary, the findings here are consistent with previous studies with similar naturalistic tasks requiring decision making under risk and can provide insight into why people can have relatively accurate perception of the likelihood of the weather event while still making suboptimal decisions.

Random Likelihood Model

The results of the comparisons between the ROC curves and the binary decisions support the application of the random likelihood model in similar naturalistic tasks requiring decision-making under risk. The ROC curves created in this project did not take the participant level variance into account. This means the ROC curves did not have confidence intervals and the analysis of the consistency between the binary decisions (proportion of hits and false alarms in decisions) and the ROC curves (from likelihood ratings and actual drought occurrence) was conservative. Even so, the analysis found consistency in 8 out of 9 conditions, suggesting that the subjective criterion and subjective likelihood jointly determine the decisions.

With the binary decisions and the ROC curve consistent, the analysis based on the random likelihood model can supplement the analysis of the relative frequency of choosing the safe option to examine the risk-seeking/risk-averse tendency in decision-making tasks. This is an important advancement because the relative frequency can be affected by trial composition, which makes it problematic to compare between studies or conditions within a study with different trial compositions. The analysis based on the random likelihood model, on the other hand, isolated the bias in the subjective criterion from the bias in the subjective likelihood and difference in sensitivity. This isolation allows the comparison of decision bias between studies with different trial compositions. As seen in the reanalyses, while the relative frequency of choosing the safe option between the two tornado experiments could not be compared due to different trial compositions, the calculated subjective criterion could be compared. However, the analysis of relative frequency has advantages over the analysis based on the random likelihood model. The analysis based on the random likelihood model requires a large number of trials to acquire a comprehensive subjective likelihood distribution to calculate the subjective criterion and construct ROC curves. On the other hand, the analysis of the relative frequency of choosing the safe option can be conducted with a smaller number of trials and without measuring likelihood ratings. Overall, the result of the experiments supports the application of the random likelihood model to examine decisions in similar naturalistic tasks requiring decision making under risk.

The Gain-Loss Framing Effect and Prospect Theory

The findings of experiment 3 indicate that the subjective criterion, but not subjective likelihood, was affected by the gain-loss framing effect as described by prospect theory (Kahneman & Tversky, 1979). This indicates that, as revealed by the application of the random likelihood model, the risk-seeking/averse tendency predicted by the utility function of prospect theory can be partially explained by a shift in the subjective criterion. The subjective criterion in a loss frame is higher than that in a gain frame, leading to a greater risk-seeking tendency in the former. On the other hand, it is unclear how the other important component of prospect theory, the probability weighting function, could not be revealed using the random likelihood model.

This is because the current paradigm could not examine the decision weights as a function of objective probabilities.

The Centering Effect

Another important finding was that the subjective criterion was affected by the centering effect (Poulton, 1979, Olkkonen et al., 2014). Interestingly, in experiment 2 with a 75erc condition, when the economically rational criterion was higher than 50%, the subjective criterion was lower-than-rational, leading to risk-averse decisions even in a loss frame. This suggests that the centering effect is quite powerful, even able to overcome the upward shift of the subjective criterion brought by the loss frame. Indeed, in experiment 3 with both the gain-loss framing and the centering effect combined, centering seemed to have a greater effect size than the gain-loss framing.

In addition, it appears that the centering effect did not shift the subjective criterion to the precise center of the range (50%). Instead, it appears to have shifted the subjective criterion towards someplace lower than 50%. Two pieces of evidence support this: 1) In experiment 2, the calculated criterion in the 75erc condition shifted downward to a greater extent than the calculated criterion shifting upward in the 25erc condition; 2) In experiment 3, the calculated criterion in both the gain and loss frame 50erc conditions was significantly lower than 50%, at 43% and 44% respectively. One explanation is that perhaps participants' preconceived notion of severe weather events was that the potential harm should be much greater than the cost of protective actions. Therefore, they expect a narrower range of criterion that does not encompass high economically rational criteria (e.g., 90%) but encompasses low economically rational criteria (e.g., 10%). Therefore, the center of the range was slightly lower than 50%. If this is the case, possible interventions can be created from this explanation to reduce the risk seeking behavior that results from a higher-than-rational criterion (see the intervention section below).

In addition, based on Figure 7, 19, 24, and 29, no centering effect was observed in the subjective likelihood across objective probability levels, consistent with previous studies (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). In perceptual studies, it was found that when the information regarding the quantity in question was available, the centering effect was reduced (Radvansky et al., 1995). In the current experiments, the available probabilistic information can explain the absence of a centering effect in subjective likelihood. Indeed, effects similar to the centering effect were observed in subjective likelihood in naturalistic weather decision task conditions where no probabilistic information was given (Demnitz & Joslyn, 2020; Qin et al., 2024). In these studies, when participants were told that the weather event in question would happen instead of being provided with probabilistic information, their subjective likelihood was close to 50%. Therefore, providing direct information of the economically rational criterion (e.g., “you should choose the safe option when the probability is above 25%”) might be a way to counteract the centering effect that occurs with subjective criterion.

As for the mechanism of the centering effect on the subjective criterion, anchoring and adjustment is a potential account (Kahneman, 2003). Anchoring and adjustment is when people are estimating a quantity, they start their estimate on a number that recently comes to mind and then adjust their estimate. The adjustment is often insufficient. In the case of the centering effect, it is possible that the center of the range of the subjective criterion was used as a default estimate and an anchor due to a lack of direct information on the economically rational criterion (e.g., “you should choose the safe option when the probability is above 25%”). In this way, the subjective criterion starts at the center of the range and then shifts towards the economically rational criterion, albeit insufficiently. The anchoring account can also explain why subjective likelihood was not affected by a centering effect in the current experiments. Probabilistic information served as a greater anchor replacing the center of the range. Another possible account is that people might wish to avoid extreme decisions and thereby shift their subjective criterion towards the center of the range which they might consider less extreme. This is similar to the compromise effect observed in decision making (Simonson, 1989). This compromise effect suggests that when choosing between three or more choices with equal utility but different characteristics, people tend to choose the middle one and avoid the extreme ones. For example, say there are three cars of equal overall utility: 1) A cheap and uncomfortable one; 2) A moderately priced and moderately comfortable one; 3) An expensive and comfortable one. People are more likely to choose Car 2 over Cars 1 and 3 as it is a less extreme compromise between the latter two. In summary, anchoring & adjustment and extreme-avoiding tendency are two accounts that might explain the centering effect.

Potential Behavioral Interventions

The current experiments revealed that in situations where the economically rational criterion is low, people tend to be risk-seeking due to a higher-than-rational subjective criterion. There are two psychological effects that contributed to the shift of the subjective criterion away from the economically rational criterion: Framing and centering. In real life situations, adverse weather events have a low probability of happening but might cause high casualties and severe damage. This translates into a low objective probability and a low economically rational criterion. In naturalistic weather decision tasks simulating these situations, people tend to be risk-seeking, exposing themselves to unnecessary risks (Baker, 1995; Joslyn & LeClerc, 2013; Atreya et al., 2015; LeClerc & Joslyn, 2015; Qin et al., 2024). Therefore, one important application of the findings of the current experiments is to design behavioral interventions to reduce the risk-seeking decision bias and improve decisions in these situations. Changing the gain-loss framing is a conceptually straight forward intervention. Framing the consequences of the event in a gain frame should shift the subjective criterion closer to the economically rational criterion when the latter is low as shown in the current experiments. However, while some events can be relatively easily framed as a gain (e.g., disease problem (Tversky & Kahneman, 1981), droughts (this project), others are not easily framed as a gain (e.g., tornadoes, frozen roadways; Grounds & Joslyn, 2018; Gulacsik et al., 2022; Qin et al., 2024).

For the centering effect, a potential intervention is to directly provide the economically rational criterion (e.g., "you should choose the safe option when the probability was higher than 25%"). In naturalistic tasks requiring decision making under risk with no probabilistic information, this method was found to be no more effective than simply suggesting the economically rational decision to participants (e.g., "In this trial you should choose the safe option," Joslyn & Leclerc, 2012). However, as there was no probabilistic information in these conditions, it was possible that while the subjective criterion was close to the economically rational criterion, the subjective likelihood was not accurate, leading to suboptimal decisions. The effectiveness of directly supplying both economically rational criterion information and probabilistic information has not been tested. It is possible that when both are supplied, participants could make better decisions with both a subjective criterion close to the economically rational criterion and accurate subjective likelihood. A similar intervention is to ask participants to estimate the economically rational criterion in situations where such estimation is possible. In a study about the compromise effect where there was missing information regarding the options, the compromise effect was reduced when participants were asked to infer the missing information themselves (Cheng et al., 2012). Of course, interventions involving information about the economically rational criterion might not be feasible in many real-life scenarios involving members of the public. This is because in these scenarios, the potential loss is difficult to quantify and cannot be represented by one generalized number as the circumstance is different from person to person. On the other hand, such interventions might be feasible for policy makers who are concerned with economical loss or human casualty statistics that can be represented by numbers. In short, direct knowledge of the economically rational criterion might be a way to reduce the centering effect on the subjective criterion.

Another possible intervention is to constrict the range of the subjective criterion in order to shift the center of the range closer to the economically rational criterion. As the adverse weather and climate events usually have a small probability of happening, this intervention perhaps can be done by constricting people's perceived range of the objective probability. One possible way to constrict the perceived range is to show a narrower probability range in the forecast. For example, the forecaster can restrict the range of the probability shown to the audience and display any probability greater than 50% as "greater than 50%." As the probability in the forecast has a range of only 0% - 50%, the audience might fixate on this range and constrict their perceived range of likelihood to the same range, not unlike the anchoring and adjustment heuristic (Kahneman, 2003). This way, the center of the range becomes 25% instead of 50%. The centering effect should thereby shift the subjective criterion towards 25% instead of 50%. On the other hand, not presenting the actual probabilities in situations where the probability is over 50% might reduce trust (Joslyn & LeClerc, 2013), perhaps due to people's perception that the forecast is withholding information. Overall, constricting the range might be another way to counteract the centering effect.

Limitations and Future Directions

One limitation of the current project is the limited examination of the connection between the subjective criterion and utility which is the cornerstone of research in decision making under risk (Von Neumann & Morgenstern, 1947; Tversky & Kahneman, 1979; Weber, 1994). The finding that the gain-loss framing had an effect on the subjective criterion but not subjective likelihood hinted at some connection between the subjective criterion and utility. One simple hypothesis is that the placement of the subjective criterion is mostly determined by the expected utility of options, just like the economically rational criterion is determined by the expected value of options. If this is the case, many psychological mechanisms in utility might also be applicable to the subjective criterion. This interaction between signal detection theory and utility theories might open doors to both theoretical and applied possibilities. Therefore, a future direction is to examine the connection between the subjective criterion and the utility.

Next, while the current experiments provided a case for the usefulness of applying the random likelihood model in naturalistic tasks requiring decision making under risk, the model still operates on the assumption that the subjective criterion has no variability and fixed across decisions. However, a previous study found that after recent encounters with adverse weather events, one's decisions became more risk-averse while the subjective likelihood remain unchanged (Demnitz & Joslyn, 2020). This suggests that the subjective criterion could change between decisions based on their recent trial experience. This might not be the only reason one changes their subjective criterion. The objective probability might serve as an anchor that shifts the subjective criterion from one trial to the next. However, the paradigm in the current experiments could not examine these effects. This is because in the current experiments, there were not enough trials to conduct within-group analysis at different objective probability levels. For example, there were only five trials with 80% objective probability per participant. They were not enough to acquire a comprehensive likelihood rating distribution to calculate a participant level subjective criterion. A similar experiment with a higher number of trials can be used to conduct these analyses, but a much longer experiment can be logistically difficult to conduct. Therefore, one future direction is to develop a paradigm able to examine the within-group effects on the subjective criterion while still logistically easy to conduct.

Conclusion

This dissertation project used the random likelihood model based on signal detection theory to better understand naturalistic weather decisions in loss scenarios. In many previous experiments, participants failed to protect themselves compared to the economically rational standard, even though they overestimated the likelihood of the severe weather event. The current experiments manipulated gain-loss framing and economically rational criterion to observe the impact on these measures: The subjective criterion (likelihood above which one chooses the safe option), subjective likelihood (perception of the event's probability), and sensitivity (the ability to predict the event). The subjective criterion was affected by both manipulations. It was more risk-seeking in the loss frame than in the gain frame, consistent with prospect theory. It also shifted towards the center of the range for extreme economically rational criterions, consistent with a centering effect. Neither manipulation affected subjective likelihood or sensitivity. These two effects led to a higher-than-rational subjective criterion when the economically rational criterion was low and counteracted the overestimated subjective likelihood to result in risk-seeking decisions. With these two sources of bias identified, one can develop behavioral interventions based on these two effects in order to improve people's decisions. More generally, the random likelihood model was found to be useful and can be used in similar naturalistic settings.

References

Ash, K. D., Schumann, R. L., & Bowser, G. C. (2014). Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Weather, Climate, and Society, 6*(1), 104–118. https://doi.org/10.1175/wcas-d-13-00021.1

Atreya, A., Ferreira, S., & Michel-Kerjan, E. (2015). What drives households to buy flood insurance? New evidence from Georgia. *Ecological Economics, 117*, 153–161. https://doi.org/10.1016/j.ecolecon.2015.06.024

Baker, E. (1995). Public response to hurricane probability forecasts. *The Professional Geographer, 47*(2), 137-147. https://doi.org/10.1111/j.0033-0124.1995.00137.x

Barron, G., & Ursino, G. (2013). Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology, 39*, 278–286. https://doi.org/10.1016/j.joep.2013.09.002

Bockenholt, U. (2006). Thurstonian-Based Analyses: Past, Present, and Future Utilities. *Psychometrika*, *71*(4), 615–629. https://doi.org/10.1007/s11336-006-1598-5

Burgeno, J. N., & Joslyn, S. L. (2023). The Impact of Forecast Inconsistency and Probabilistic Forecasts on Users' Trust and Decision-Making. *Weather, Climate, and Society, 15*(3), 693–709. https://doi.org/10.1175/WCAS-D-22-0064.1

Cheng, Y.-H., Chou, C.-A., Kao, D. T., & Chuang, S.-C. (2012). The Effect of Incomplete Information on the Compromise Effect. *Judgment and Decision Making*, *7*(2), 196–204. https://doi.org/10.1017/S193029750000303X

Demnitz, R., & Joslyn, S. (2020). The Effects of Recency and Numerical Uncertainty Estimates on Overcautiousness. *Weather, Climate, and Society*, *12*(2), 309–322. https://doi.org/10.1175/WCAS-D-19-0115.1

Ferrell, W. R., & McGoey, P. J. (1980). A Model of Calibration for Subjective Probabilities. *Organizational Behavior and Human Decision Processes*, *26*(1), 32-53.

Gulacsik, G., Joslyn, S. L., Robinson, J., & Qin, C. (2022). Communicating Uncertainty Information in a Dynamic Decision Environment. *Weather, Climate, and Society*, *14*(4), 1201–1216. https://doi.org/10.1175/WCAS-D-21-0186.1

Grounds, M. A., & Joslyn, S. L. (2018). Communicating Weather Forecast Uncertainty: Do Individual Differences Matter? *Journal of Experimental Psychology. Applied*, *24*(1), 18–33. https://doi.org/10.1037/xap0000165

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Kahneman, D. (2003). A Perspective on Judgment and Choice: Mapping Bounded Rationality. *The American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

LeClerc, J., & Joslyn, S. (2015). The Cry Wolf Effect and Weather-Related Decision Making. *Risk Analysis*, *35*(3), 385–395. https://doi.org/10.1111/risa.12336

Lindell, M., Huang, S., Wei, H., & Samuelson, C. (2016). Perceptions and expected immediate reactions to tornado warning polygons. *Natural Hazards*, *80*(1), 683-707. https://doi.org/10.1007/s11069-015-1990-5

Joslyn, S. L., & Grounds, M. A. (2015). The Use of Uncertainty Forecasts in Complex Decision Tasks and Various Weather Conditions. *Journal of Experimental Psychology. Applied*, *21*(4), 407–417. https://doi.org/10.1037/xap0000064

Joslyn, S., & LeClerc, J. (2013). Decisions With Uncertainty: The Glass Half Full. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, *22*(4), 308–315. https://doi.org/10.1177/0963721413481473

Joslyn, S., & Savelli, S. (2021). Visualizing Uncertainty for Non-Expert End Users: The Challenge of the Deterministic Construal Error. *Frontiers in Computer Science (Lausanne)*, *2*. https://doi.org/10.3389/fcomp.2020.590232

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Klockow-McClain, K. E., McPherson, R. A., & Thomas, R. P. (2020). Cartographic Design for Improved Decision Making: Trade-Offs in Uncertainty Visualization for Tornado Threats. *Annals of the American Association of Geographers*, *110*(1), 314-333. https://doi.org/10.1080/24694452.2019.1602467.

Krizan, Z., & Windschitl, P. D. (2007). The Influence of Outcome Desirability on Optimism. *Psychological Bulletin*, *133*(1), 95–121. https://doi.org/10.1037/0033-2909.133.1.95

Lench, H. C., Smallman, R., Darbor, K. E., & Bench, S. W. (2014). Motivated perception of probabilistic information. *Cognition*, *133*(2), 429–442. https://doi.org/10.1016/j.cognition.2014.08.001

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

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

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: effects of internal and external noise. *Journal of Vision (Charlottesville, Va.)*, *14*(11), 5–5. https://doi.org/10.1167/14.11.5

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, *86*(4), 777–803. https://doi.org/10.1037/0033-2909.86.4.777

Qin, C., Joslyn, S., Savelli, S., Demuth, J., Morss, R., & Ash, K. (2024). The Impact of Probabilistic Tornado Warnings on Risk Perceptions and Responses. *Journal of Experimental Psychology. Applied*, *30*(1), 206–239. https://doi.org/10.1037/xap0000486

Radvansky, G. A., Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Uncertainty in estimating distances from memory. *Memory & Cognition*, *23*(5), 596–606. https://doi.org/10.3758/BF03197261

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *The Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Tversky, A., & Fox, C. R. (1995). Weighing Risk and Uncertainty. *Psychological Review*, *102*(2), 269–283. https://doi.org/10.1037/0033-295X.102.2.269

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science (American Association for the Advancement of Science)*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/bf00122574

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (Second edition.). Princeton University Press.

Weber, E. U. (1994). From Subjective Probabilities to Decision Weights: The Effect of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events. *Psychological Bulletin*, *115*(2), 228–242. https://doi.org/10.1037/0033-2909.115.2.228

Windschitl, P.D. & Weber, E.U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70%—right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*(6), 1514-1553. https://doi.org/10.1037//0278-7393.25.6.1514.

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes*, *111*(1), 33–47. https://doi.org/10.1016/j.obhdp.2009.08.003