Sources of Bias in Naturalistic Decision Making Under Risk from A Signal Detection Perspective

Chao Qin

University of Washington

Abstract

Severe weather events such as tornadoes require people to take protective action even when the probability is low because their high severity. 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 study examined this phenomenon from a signal detection theory perspective. A random likelihood model was introduced to estimate the subjective criterion, the likelihood above which one takes protection actions. This model separates the subjective criterion from subjective likelihood, participants' perception of the probability of the weather. Two experiments manipulated the economically rational criterion (the criterion based on expected value theory) and gain-loss framing to examine their effect on the subjective criterion. When the economically rational criterion was manipulated, the subjective criterion was between the economically rational criterion and the center of the range (50%). When the gain-loss framing was manipulated, the subjective criterion was higher in a loss frame than a gain frame. In addition, participants showed an overestimation in subjective likelihood that was unaffected by the manipulations. The shifted subjective criterion overcame this overestimation and resulted

in risk-seeking decisions in some conditions. Thus, the shift of the subjective criterion is a source of risk-seeking decisions in naturalistic decision tasks. Potential interventions are discussed with the aim to improve the placement of the subjective criterion.

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 US National Weather Service when the probability of a tornado is 10% or higher (Qin et al., 2024). Due to the 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).

One reason contributing to this low uptake of protective action is that people do not receive probabilistic information (e.g., there is a 20% chance of tornado) of the severe weather events from the forecast to form an accurate risk perception. Indeed, many forecasts do not provide people with an estimated probability of the event, although such information is increasingly available (Gallo et al., 2016; Karstens et al., 2015; Joslyn & Savelli, 2021; Gulacsik et al., 2022; Qin et al., 2024). There is experimental 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, receiving the probability of the severe weather event is only part of the puzzle. Additional biases might have contributed to their risk-seeking decisions.

Naturalistic Weather Tasks Requiring Decision Making under Risk

There had been a plethora of experiments examining people's response to forecasts in naturalistic weather tasks. Many of them measured participants' perception of the risk given the forecast but did not include a decision task where participants made decisions with stimulated economical consequences (Baker, 1995; Morss et al., 2008; Ash et al., 2014; Lindell et al., 2016; Ash et al., 2014). While examining perception of risk is important, decision patterns could not be directly inferred from these experiments. On the other hand, there were several experiments that examined behaviors using naturalistic weather tasks requiring decision making under risk (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 decision tasks, participants were given numeric probabilistic forecasts for a possible severe weather event over a number of trials and made decisions under a point system. 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 the weather event. In the usual risky option, participants did not take any protective action and the weather event incurred a greater point penalty [redacted]. 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) Subjective likelihood: Participants' self-reported perception of the likelihood of the weather event, usually measured using a continuous scale from 0% to 100% ([redacted] 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 [redacted] rational standard [redacted] (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). Subjective likelihood was compared to objective probabilities, which were the probabilities of the weather event on each trial and usually 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). There is a probability at which the fixed cost of the safe option and the expected value of the risky option broke even. This probability is considered the economically rational probability threshold above which one should choose the safe option as it has the better expected value ([redacted]). Decisions made with this probability threshold therefore were an economically rational standard with which participants' binary decisions could be compared.

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

The findings of these studies were twofold. First, participants' 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 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~~ can 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).

The results about the binary decisions ~~could~~ can be explained by prospect theory (Kahneman & Tversky, 1979). 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~~ is 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

cut Fig 1

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 diminishing return 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 (less gain) than gaining $100 for sure (safe option) while a 50% chance of losing $200 (risky option) has a better expected utility (less loss) than losing $100 for sure (safe option). Therefore, people tend to be risk-averse in a gain frame while risk-seeking in a loss frame.

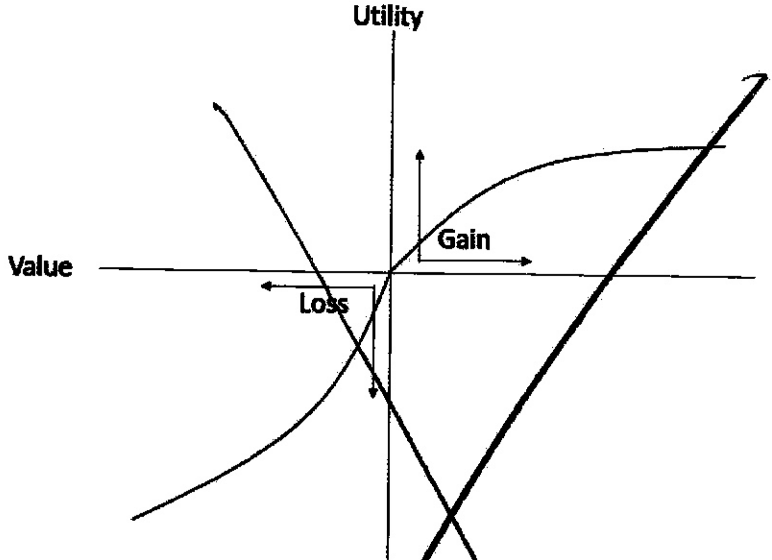


Figure 1
The Typical Gain/loss Utility Function
The x axis indicates the objective values of gains and losses. The y axis indicates the utility of these gains and losses to a decision-maker. The center is the reference point. The increase/decrease in values in relation to the reference point have diminishing 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.

cut

On the other hand, the result of overestimated self-reported subjective likelihood and risk-seeking decisions in a loss frame (Gulacsik et al., 2022; Qin et al., 2024) presents a curiosity if considered together. Subjective likelihood is thought to be among the components leading to people's decisions (Ferrell & McGoey, 1980; Krizan & Windschitl, 2007). 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, the studies reviewed here found that they were risk-seeking in their decisions (Gulacsik et al., 2022; Qin et al., 2024). This curiosity can not be explained by prospect theory, which does not tackle subjective likelihood as measured by self-reporting (Kahneman & Tversky, 1979).

This curiosity was also observed in a non-naturalistic, games of chance experiment, suggesting that this phenomenon is not limited to naturalistic weather tasks (Barron & Ursino, 2013). In this loss-frame study, participants provided subjective likelihood and chose between a safe option (sure loss) and a risky option (a gamble) with no provided probabilistic information. Both the probability of a large loss on each trial and the economically rational probability threshold in the risky lottery was low, at 15%, similar to low probability high severity weather events in real life. Participants' decisions showed risk-seeking, but their subjective likelihood showed overestimation. Therefore, this disconnect between the subjective likelihood and decisions was not because of the naturalistic weather settings.

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.

The study reported here 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. The study reported here 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~~ This 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. Therefore, a biased subjective criterion can potentially explain the disconnect between overestimated subjective likelihood and risk-seeking decisions.

Signal Detection Theory and Decision Making

Originated from perceptual experiments, signal detection theory concerns two separate psychological mechanisms: *Internal representation of the stimulus* and *subjective criterion* (See Figure 2, Macmillan & Creelman, 2005). ~~Separate~~ An internal representations ~~are~~ is generated, ~~one~~ representing ~~that~~ whether the stimulus is present ~~and the other that it is~~ or absent. ~~They are~~ It is 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 the 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 or the less noisy they are, the less the two distributions 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. It could come in many different forms, like a probability threshold if the internal representation was about the probability of an event.

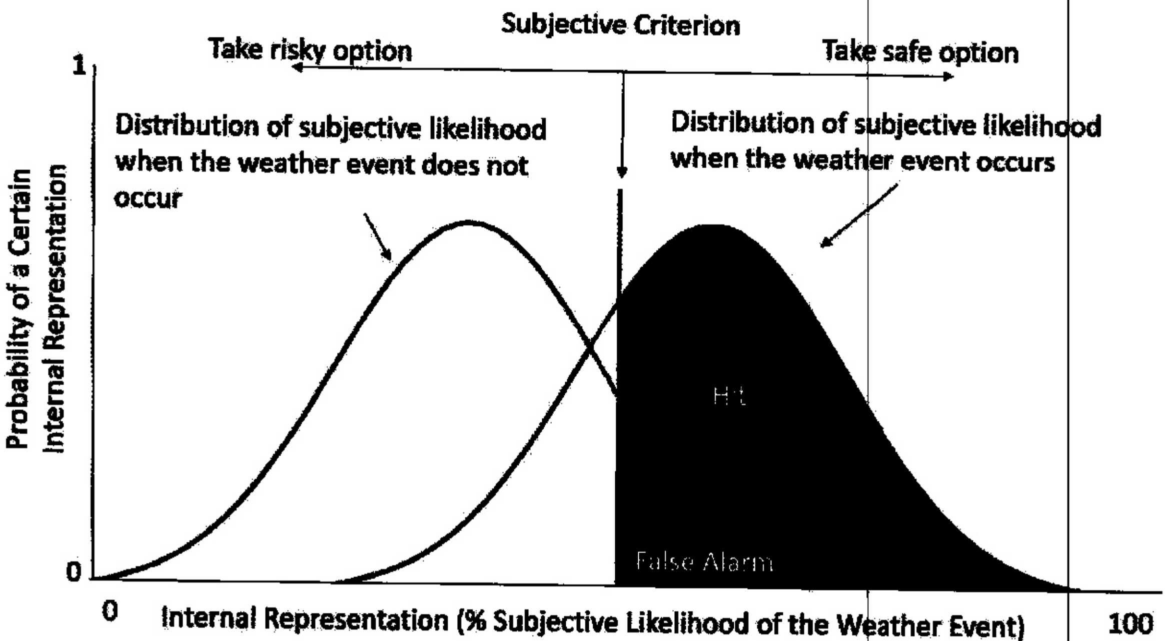


Figure 2
Random Likelihood Model Example Based on Signal Detection Theory
Note. The random likelihood model does not assume the shape of the internal representation distribution. Normal distributions are used in the graph as an example.

Move into Theory section

Move into theory section + integrated p16

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.

this needs a reference such as Ferrel & McGoey 1980?

[Move into theory section]

	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

Move to
Theory
section

Signal detection theory [redacted]e, stochastic internal representations often studied in decision literature (Busemeyer & Townsend, 1993; Thurstone 1994; Wallin et al 2018). Such stochastic representations mean that under the same external circumstances, one [redacted]t not make the same decision every time, just a higher probability towards a certain decision. This could cause [redacted]s probability of choosing the safe or risky option following an S-shaped function as opposed to a step function (See Figure 3). This S-shaped function was observed in previous studies (Busemeyer & Townsend, 1993; Qin et al., 2024). Signal detection theory that tackles variable internal representations could be integrated with behavioral economics theories such as prospect theory that considers the uncertainty in the outcomes to better understand decision-making processes (Lynn et al., 2018).

The application of signal detection theory in real life decision making under risk was often studied in the context of diagnostic decisions (Swets et al., 2000). For example, in medical diagnostic problems, a physician needs to diagnose whether a patient has cancer based on an X-ray. The physician's ability to interpret the X-ray is their sensitivity and the amount of evidence they need in order to diagnose cancer is their subjective criterion. Like in many diagnostic decisions, the utility of a hit is not the same as the utility of a false alarm. In this case, perhaps correctly diagnosing cancer when the patient actually has cancer is more important than incorrectly diagnosing cancer when the patient does not have cancer. Therefore, the physician should have a lower, more liberal criterion to ensure higher hits at the expense of higher false alarms. An application of signal detection theory in these diagnostic is to find the (economically) rational criterion and could increase diagnose accuracy and utility in domains like medical/psychiatric diagnoses, violent risk assessment, weather forecasting, and school admissions (Swets et al., 2000).

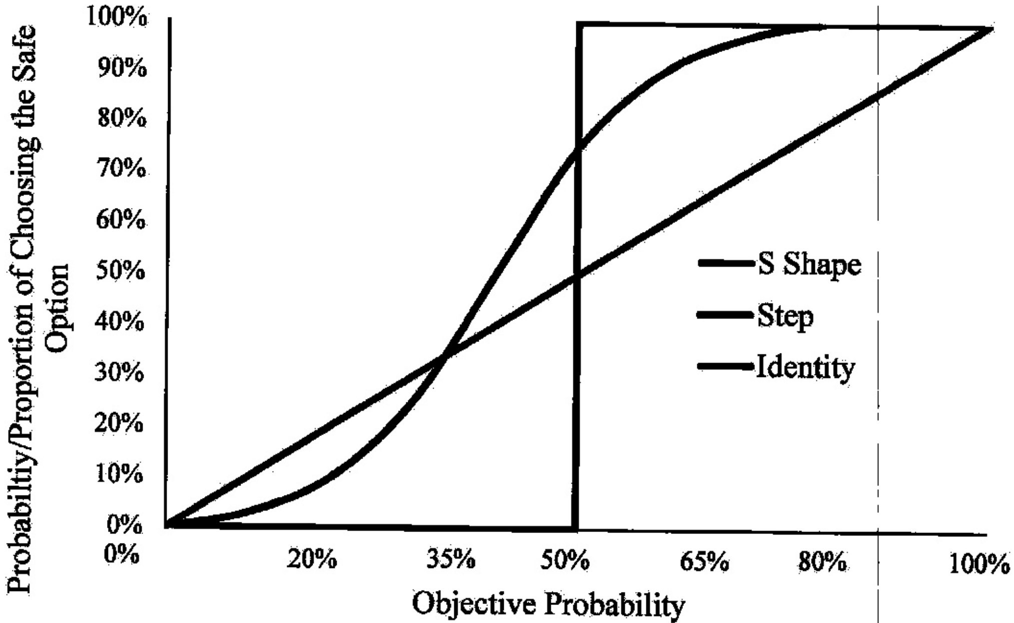


Figure 3
Hypothetical Probability/Proportion of Choosing the Safe Option in a Loss Frame Assuming that the economically rational criterion is 50%. In the S-shaped function, the probability/proportion of choosing the safe option increased in an S-shape as a function of objective probability. In the step function, participants never choose the safe option when below the criterion and always choose when above the criterion.

cut

Random Likelihood a[redacted]on Model

The four outcomes in the naturalistic weather decision tasks requiring decision-making under risk 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 2), 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 forecast information (~~objective probability in this case~~). The model does not assume a specific shape of distribution for this variability. This is because in this study subjective likelihood ~~was~~ elicited ~~directly~~ and ~~therefore its actual~~

used an estimate of the internal distribution,

distribution was known. For example, in Figure 2, 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. Such biases can ~~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 2). 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 study focused on the subjective criterion while the sensitivity and subjective likelihood were also examined.~~

In naturalistic weather tasks, prior studies have found that there is a small but consistent trend of overestimated subjective likelihood from the objective probability (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). 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~~. Therefore, the study reported here used the random likelihood model to examine a possible systematic bias in the subjective criterion ~~in these tasks~~. This study proposes that biases in the subjective criterion can lead to biased decisions. This is the subjective criterion hypothesis. The alternative is a bias in the likelihood distributions which is the subjective likelihood hypothesis.

Possible Sources of Bias on the Subjective Criterion

Two sources that might shift the subjective criterion from the economically rational criterion were the centering effect and the gain-loss framing effect. The centering effect suggests that when judging a quantity people tend to bias their judgement towards the center of the quantity's range (also called central tendency bias; Poulton, 1979; Olkkonen et al., 2014). 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. The study reported here 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.

On the other hand, it is also possible that the decision bias 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 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 subjective likelihood. These two sources are not mutually exclusive and might both contribute to a biased subjective criterion.

Overview of Analyses and Experiments

The study reported here 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.

T[redacted]ect then [redacted]d the subjective criterion hypothesis and possible sources of bias leading to risk-seeking decisions and overestimated subjective likelihood in a naturalistic weather decision task [redacted]. A drought task based on Demnitz & Joslyn (2020) [redacted]d. 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

---

[2] A pilot experiment examining the gain-loss framing effect yielded a trend in the expected direction but failed to reach significance due to low power. It was omitted here for brevity.

in subjective likelihood and/or a difference in sensitivity. Two manipulations were examined: 1) 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%. 2) 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.

This project had two experiments. Experiment 1 manipulated the economically rational criterion to examine centering on the subjective criterion and the manipulation's effect on subjective likelihood and sensitivity. Experiment 2 combined both the economically rational criterion and gain-loss framing manipulations to allow a comparison between the two effects within a single experiment.

Reanalysis of Previous Experiments

Method of Tornado Experiments

The two tornado experiments were from Qin et al., 2024. 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. 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 the study reported here, 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. On each trial, participants saw a tornado warning, rated how likely they thought the tornado would be (likelihood ratings, operationalization of subjective likelihood), 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. The outcomes of their decisions were represented with a point structure (see the section below). The experiment was framed in a loss frame. Participants' goal was to lose as few points as possible by the end of the experiment.

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 broke even with the cost of the safe option was 9% in tornado experiment 1 and 27% in tornado experiment 2. The 9% and 27% were 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.

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 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 the sensitivity (see Figure 4). The ROC curve also indicated 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.

Experiment 1	Safe Option	Risky Option
Tornado Occurred	Hit: -90 points	Miss: -1000 points
Tornado Did Not Occur	False alarm: -90 points	Correct rejection: 0 points

Experiment 2	Safe Option	Risky Option
Tornado Occurred	Hit: -270 points	Miss: -1000 points
Tornado Did Not Occur	False alarm: -270 points	Correct Rejection: 0 points

Table 2
Point Structure of the Two Tornado Experiments

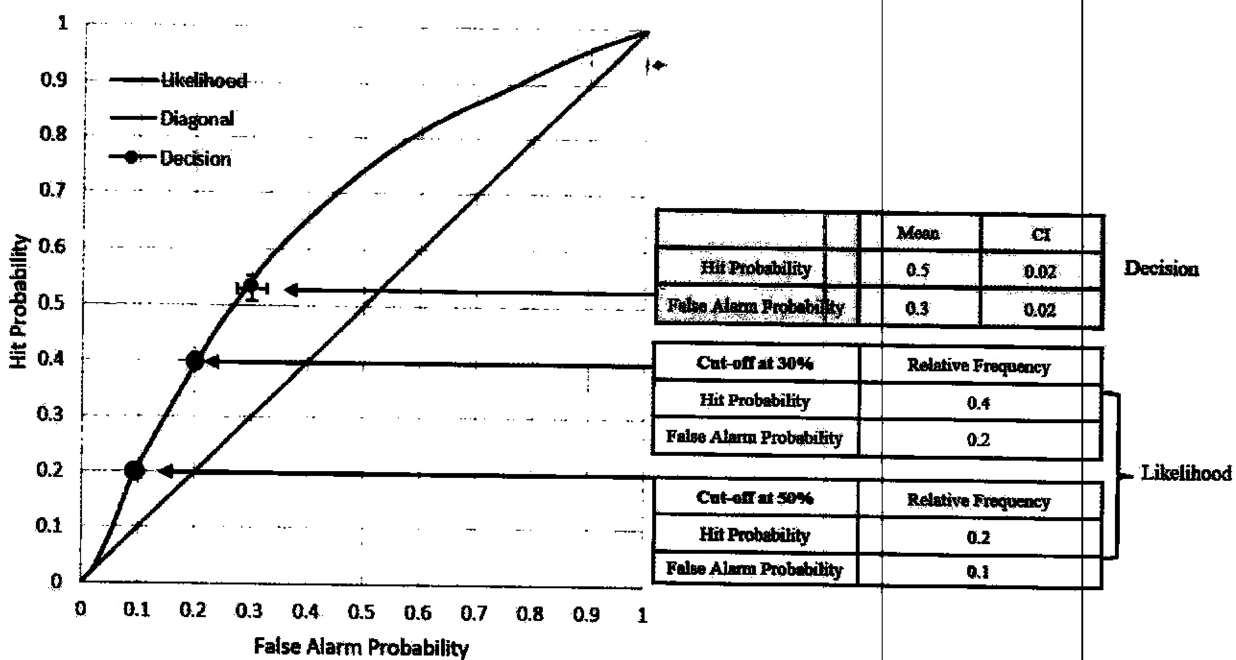


Figure 4

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 likelihood ratings and frequency of choosing the safe option 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 was higher than their subjective criterion and that the subjective likelihood followed a distribution while the subjective criterion had no variability.

~~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%, 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.

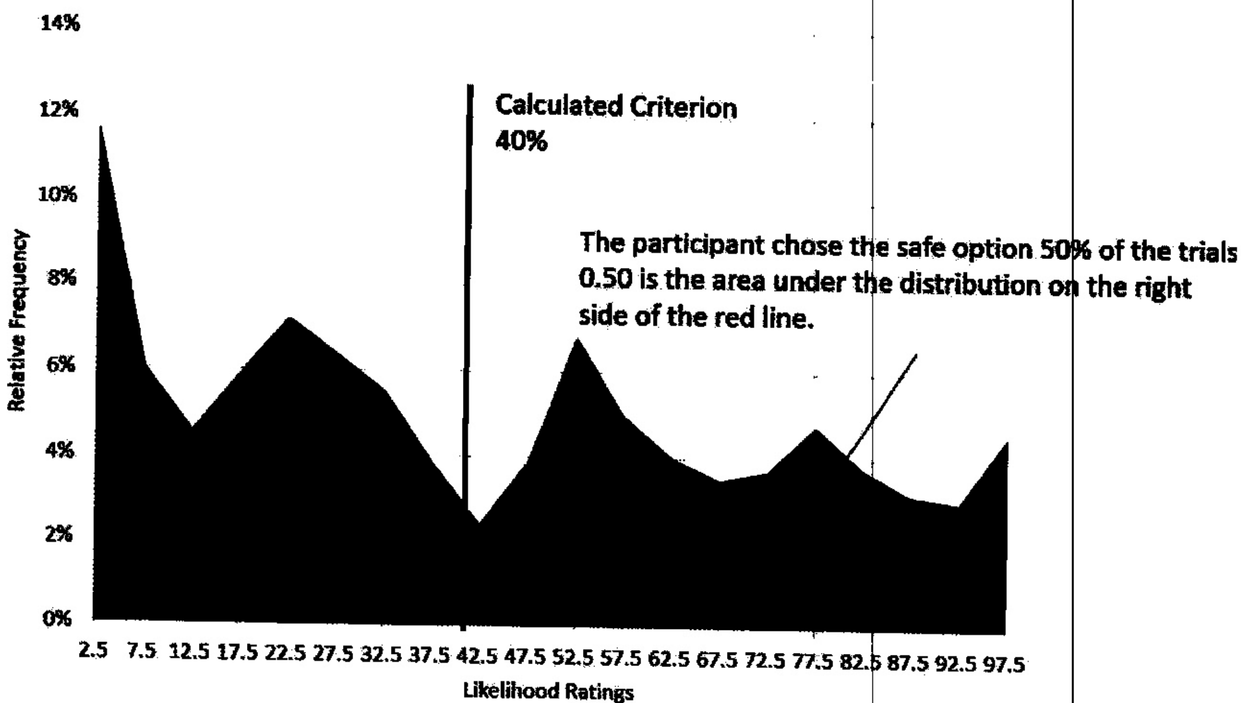


Figure 5

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 [redacted] 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.

found at

Explain Figure, axis, etc

In Figure 6, the calculated criterion is shown for tornado experiments 1 and 2. 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.

Likelihood Ratings

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% ($t(84) = 5.48$, $p < .001$). Therefore, the mean likelihood ratings were higher than the objective probabilities in both experiments. In addition, Figure 7 shows that the overestimation was observed at most objective probability levels except for the extreme high end.

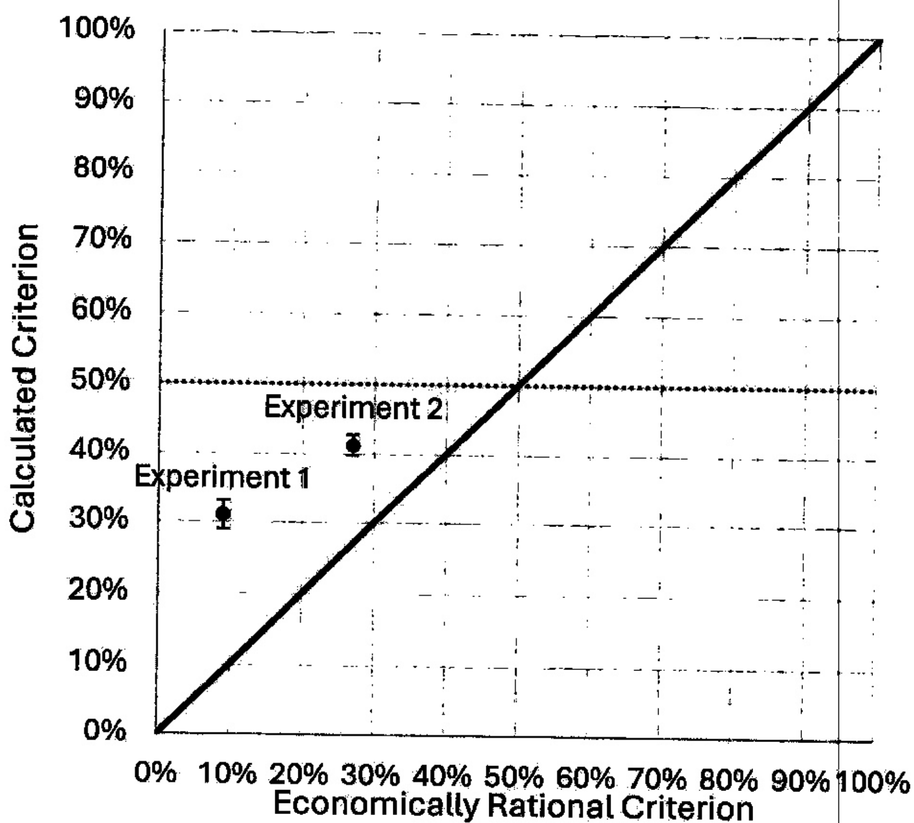


Figure 6
Calculated Subjective Criterion in the Two Tornado Experiments
The difference of 10.2% between the two experiments was significant

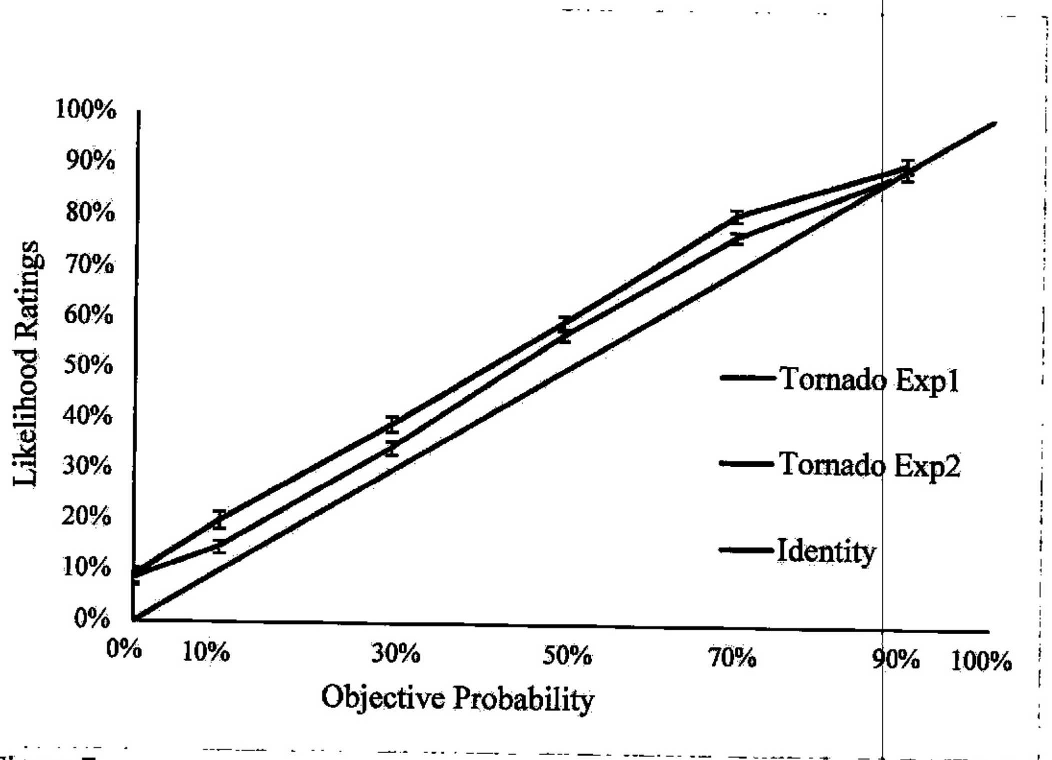


Figure 7
Likelihood Ratings as a Function of Objective Probabilities in Tornado Experiment 1 and 2
The blue line represents the likelihood ratings of tornado experiment 1. The orange line represents the likelihood ratings of tornado experiment 2. ~~There was no observed centering effect.~~

(put in text or ignore)

Receiver Operating Characteristic Plots

The ROC plots analysis had two aims: 1) Illustrating the random likelihood model 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 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)).

The ROC curves were created by estimating hypothetical 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). The assumption was that participants should have high likelihood ratings on tornado trials and low ratings on no tornado trials. The varying hypothetical cut-off on the likelihood rating distribution determined what was considered "high" and "low." 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 of likelihood ratings 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 distribution, a pair of hit 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 of likelihood ratings 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.

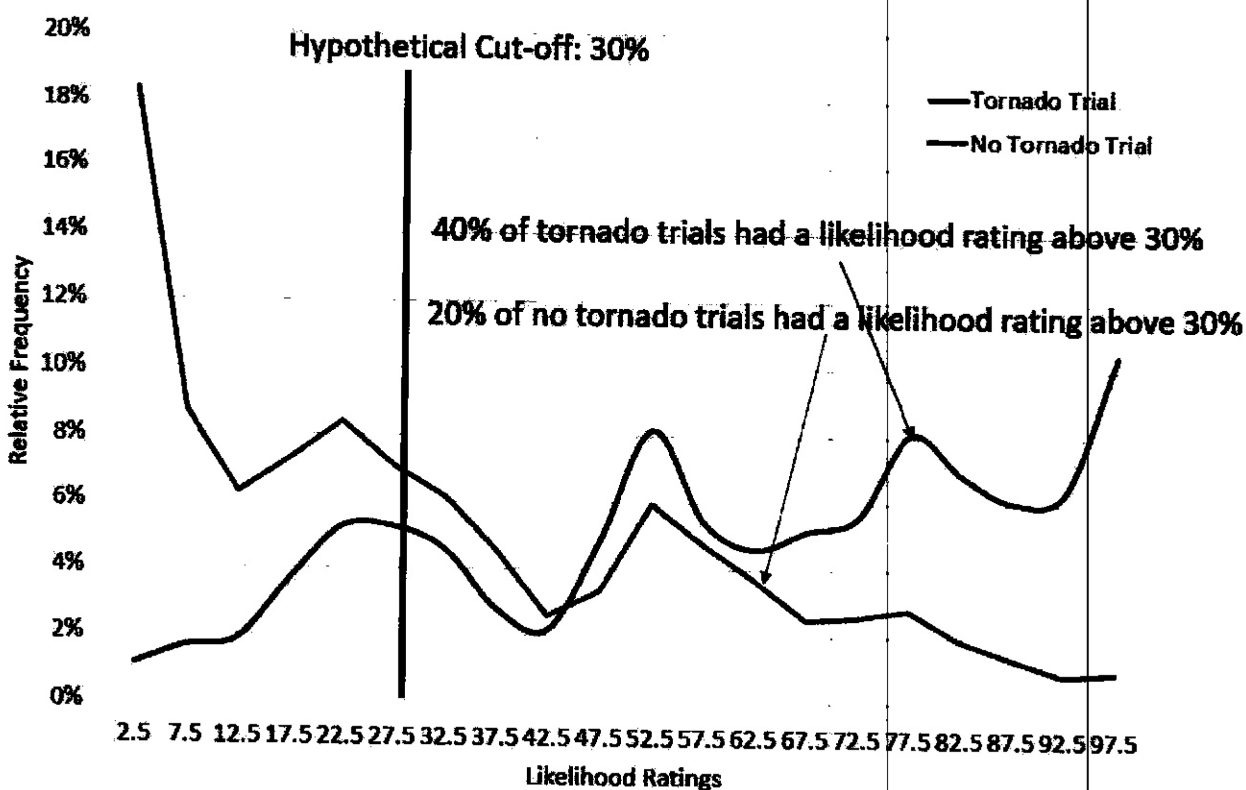


Figure 8

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 model method of calculating the subjective criterion. If participants always chose the safe option whenever their likelihood rating was above their subjective criterion, the hit and false alarm probabilities of their likelihood ratings should align with the hit and false alarm probabilities of decisions with a subjective criterion of the same number as the cut-off. 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. 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.

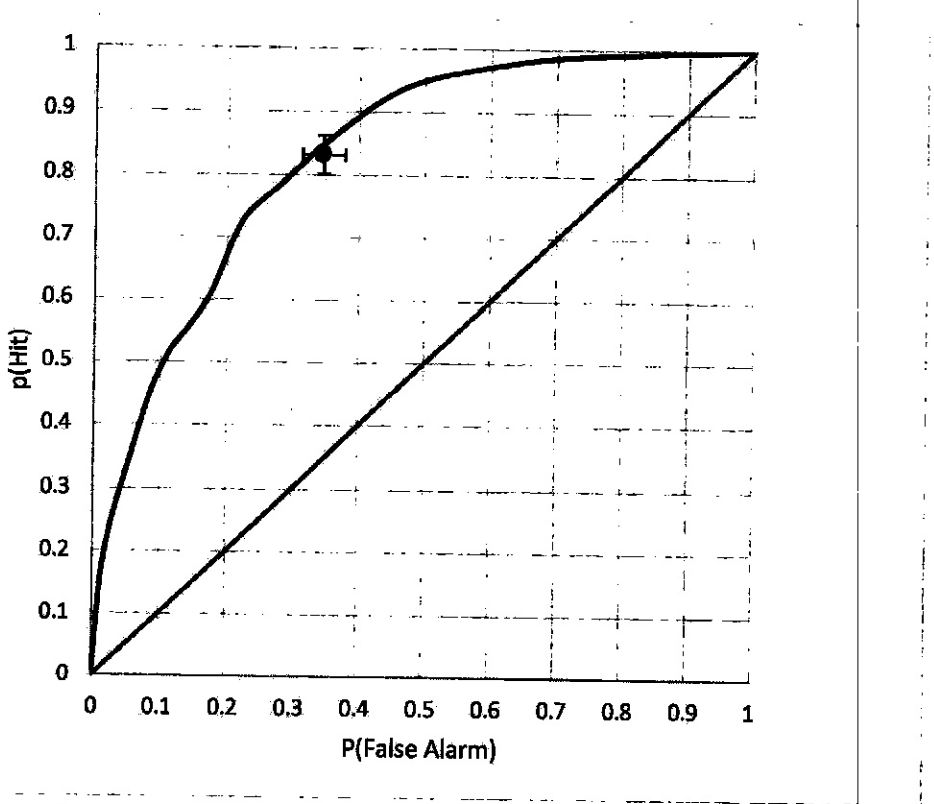


Figure 9
ROC Plot for Tornado Experiment 1
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The mean percent area under curve was 64.0%.

Combine Fig 9 + 10 using 2 panels

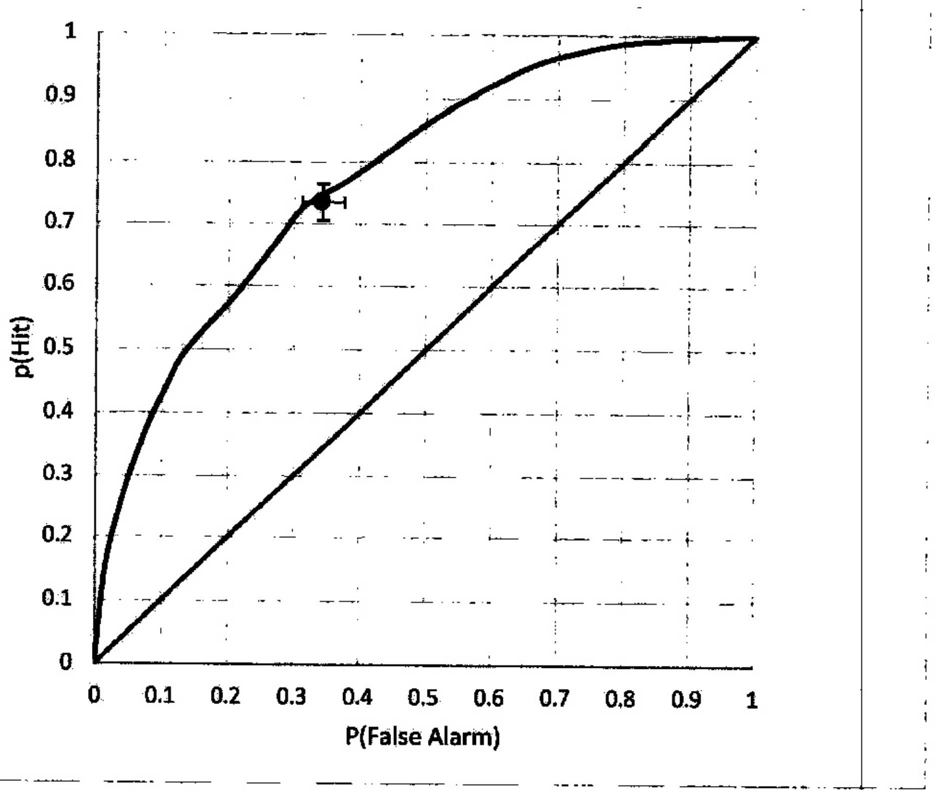


Figure 10
ROC Plot for Tornado Experiment 2
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The mean percent area under curve was 78.1%

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 frame 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 reported in this study, these conditions were added to distinguish the centering effect and the gain-loss framing effect.

The mean likelihood rating analysis showed that the likelihood ratings were significantly overestimated in both experiments. There was also 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 reported in this study, 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. Again, 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 whether 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 with mid-range (e.g., 50%) objective probabilities. The two experiments differed in the proportion of these trials. Hence in the new experiments reported here, the trial composition was held constant across the conditions to remove it as a potential confound.

More importantly, in the ROC plots of both 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 model 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 and sensitivity, in addition to the decision bias. Therefore, a conclusion about decision bias based on a difference in relative frequency of choosing the safe option was not possible when the other two variables differed. This was another reason why the trial composition was kept constant among conditions in the new experiments reported here. 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 centering effect and a gain-loss framing effect.

Experiment 1

Experiment 1 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%. Experiment 1 used a drought task in a loss frame. 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 (see the procedure section below). 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 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*. ~~See Appendix A for illustrations 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). 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. 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.

Point Structure. In order to simulate real life decisions with consequences, and to encourage participants to put forth their best effort, [redacted] point structure, [redacted], was implemented in which participants lost points based on their decision (See Table 4). [redacted] economically rational criterion was manipulated between groups. There were three economically rational criterion conditions (erc): 25% (25erc), 50% (50erc), and 75% (75erc). In all conditions 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 25erc, 50erc, and 75erc condition, the drought resistant crop (safe option) provided a sure loss of 100, 200 and 300 points respectively ([redacted]). All conditions had the same starting balance of 20,000 points. The bonus payment structure was set up 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. This payment threshold was set up to prevent participants from taking the simplistic approach of choosing the safe option in every single trial. For example, in the 25erc condition, if they chose the safe option in all 50 trials, they would end up with 15,000 points (20,000 – 50 * 100).

Trial Structure. After reading through the background information, point structure, and going through a practice trial, and two attention checks, 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). This percent chance will be referred to as the objective probability. It was calibrated to be roughly reliable. 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" (likelihood rating). On the second screen, participants saw the same forecast and chose which crop they wished to plant by pressing one of the two buttons (binary decision). 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" (25erc example) 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.

The 50 trials varied in their objective probabilities and occurrence of drought ([redacted]). 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.

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 4
Point Structure of the Three Conditions in Experiment 1

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. Participants were first asked a comprehension check question. They then indicated their (self-reported) criterion of their decision. This was an alternative way ~~to the calculated criterion~~ 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 and an open-ended question ~~asking them which part of the experiment~~ they found difficult to understand. 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 ~~reported~~ any problem in the experimental program. No ~~bugs or glitches~~ were reported. 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 [redacted] 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 [redacted] and the proportion of drought was 36% across 50 trials. 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 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.

Results

Analysis Overview

The hypothesis of experiment 1 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) and the reanalyses, there would be no centering effect in likelihood ratings.

The same set of dependent measures as the reanalyses were used in experiment 1. 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. 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.

Cut?

Move into methods

Decision Bias

The first set of analyses examined the degree to which participants biased their decisions towards the safe option in the two conditions. ~~Three~~ Two dependent variables were examined here: Calculated subjective criterion, and the 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~~ Varying the economically rational criterion manipulation showed a significant

(25erc, 50erc, and 75erc) ~~as the independent variable on the calculated subjective criterion~~ s[redacted]n 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, [redacted]). 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, c[redacted]). 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, [redacted]). This result was consistent with the centering effect such that the calculated criterion shifted towards 50% from the economically rational criterion.

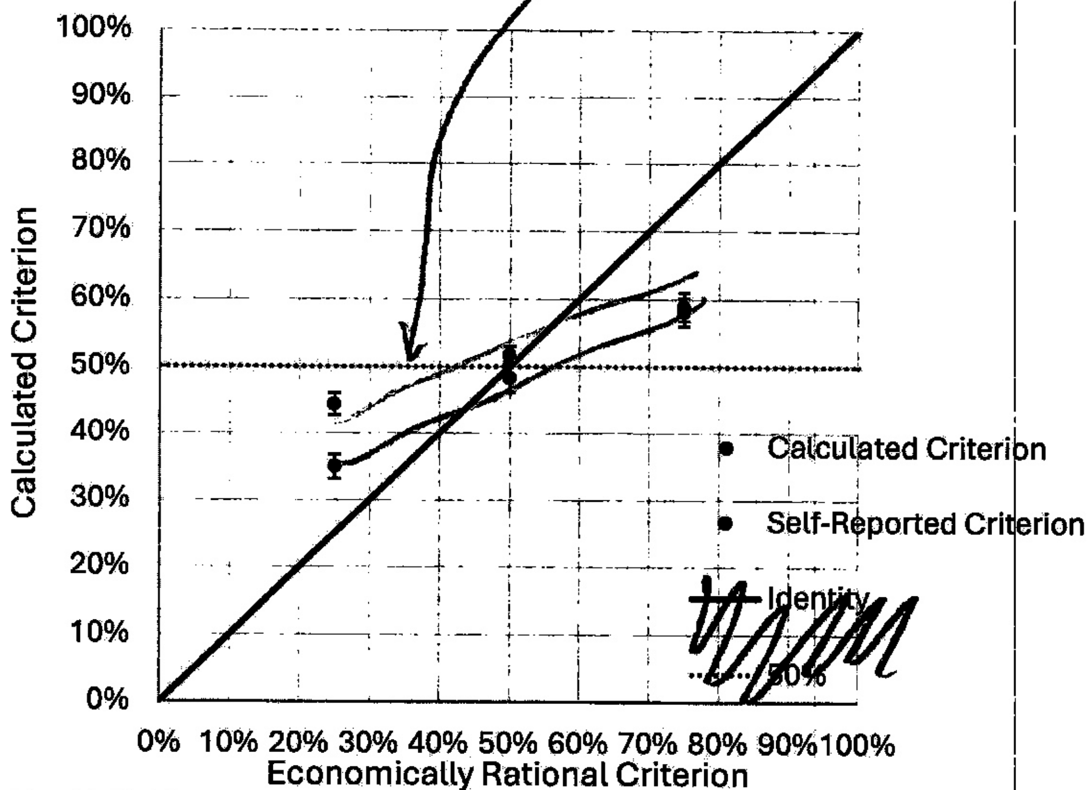


Figure 22

Calculated and Self-Reported Subjective Criterion in the Three Conditions of Experiment 1

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 [redacted] (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, [redacted]. 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, c[redacted]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, c[redacted]). 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, co[redacted] 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, $p < .001$, c[redacted]). 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. 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.

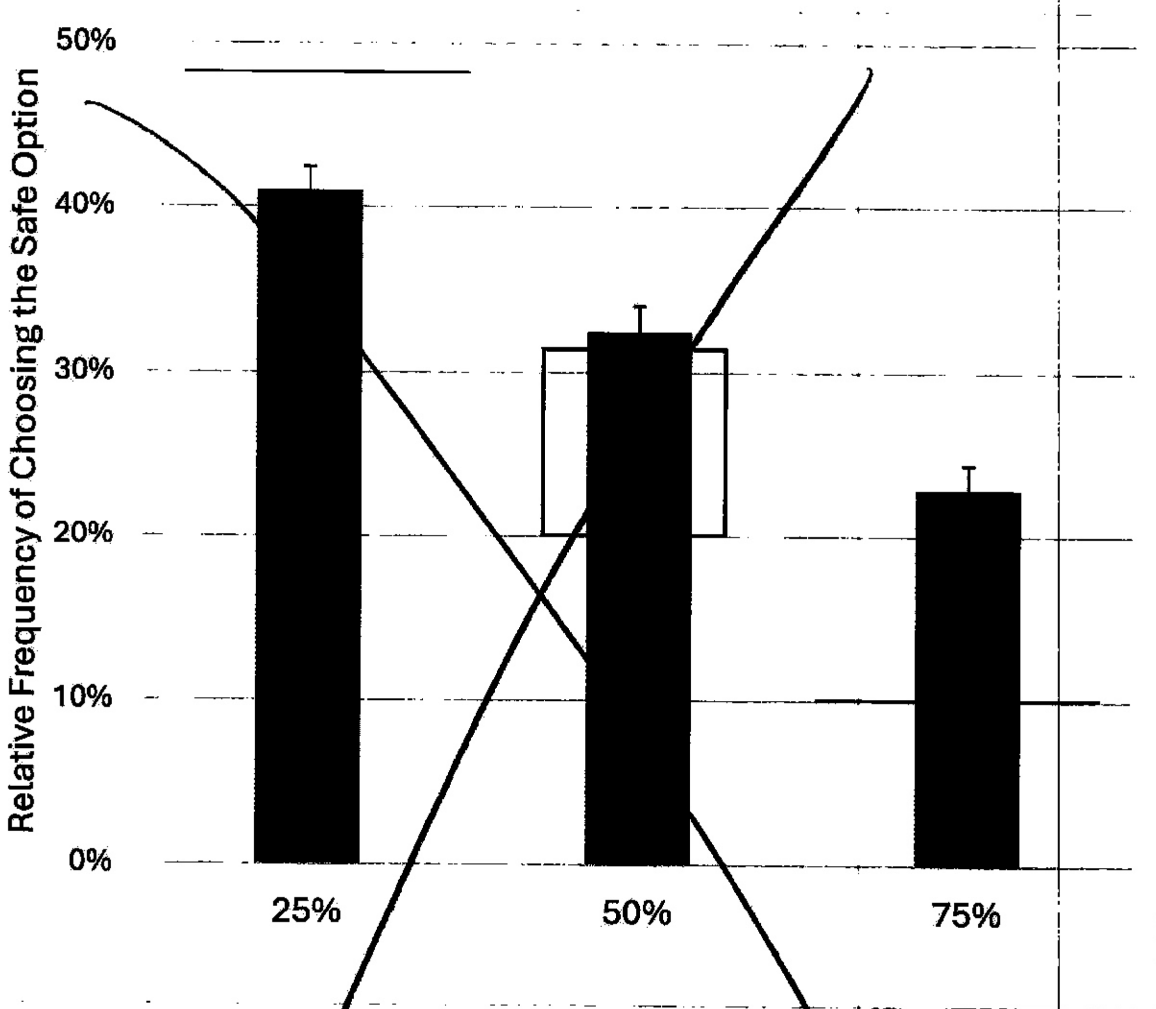


Figure 23
Relative Frequency of Choosing the Safe Option in the Three Conditions of Experiment 1
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red line in the 75% economically rational criterion condition shows the relative frequency of 10%. 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) had no significant effect 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 but one objective probability levels ~~except for 35%~~

(collapsing over objective probability)

Confusing. Cut.

The all closely followed the objective probability

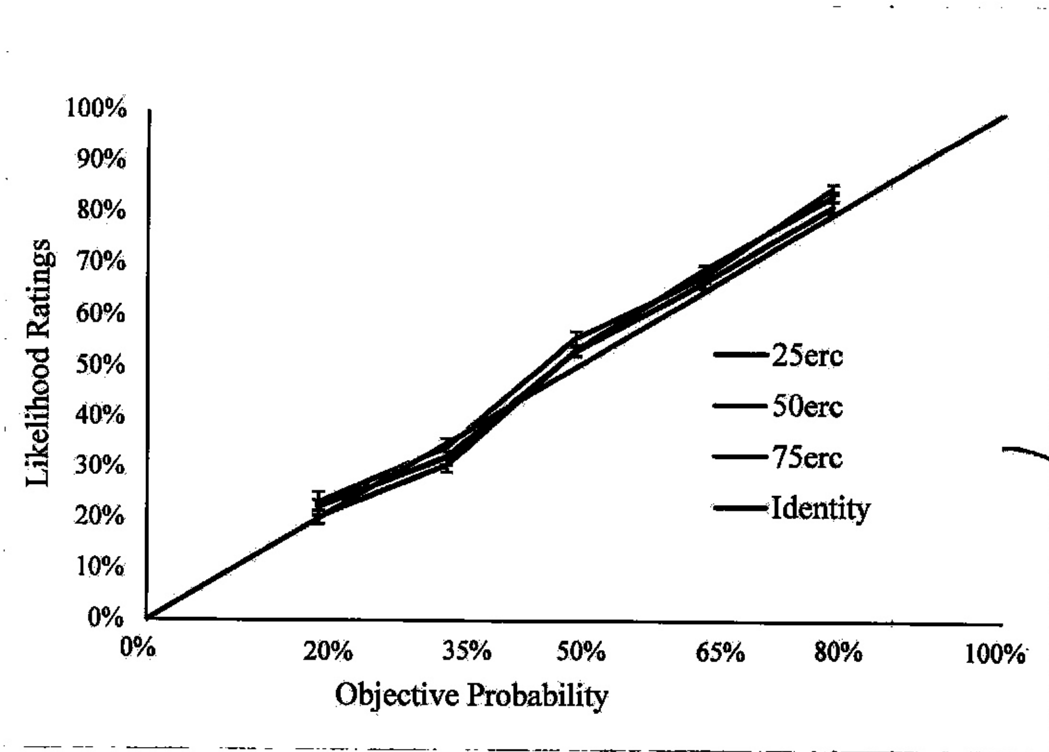


Figure 24
Likelihood Ratings in the 25%, 50%, and 75% Economically Rational Criterion Conditions as a Function of Objective Probability in Experiment 1
The blue line represents the 25 economically rational criterion condition. The orange line represents the 50 economically rational criterion condition. The purple line represents the 75 economically rational criterion condition. ~~There was no observed centering effect.~~

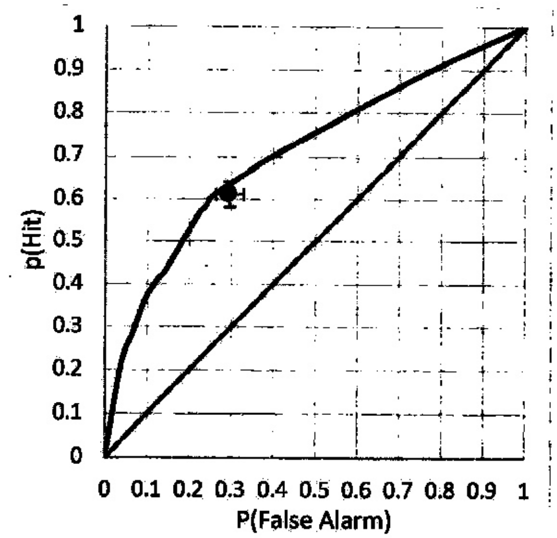
The likelihood ratings closely followed the objective probability with a small overestimation.

Make bigger esp the y axis

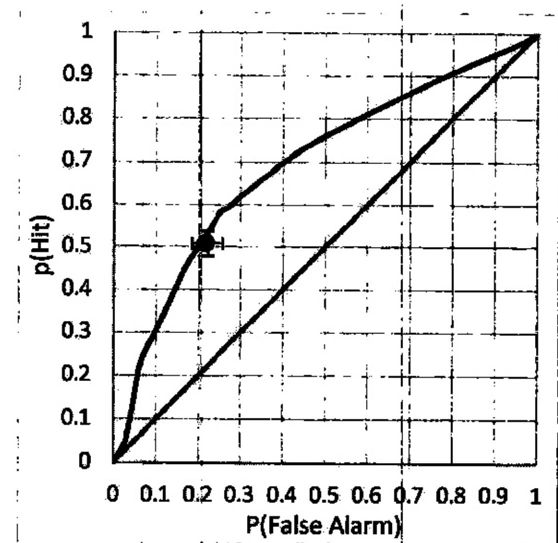
Sensitivity

The next analysis examined the sensitivity of 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.

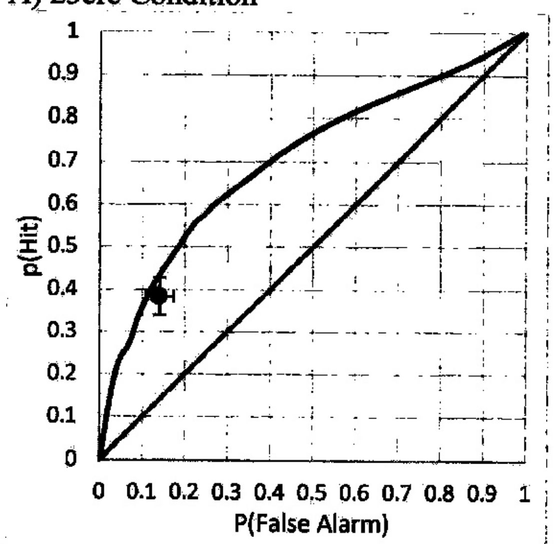
Note also that the bias decision fell on the ROC curve for all three conditions.



A) 25erc Condition



B) 50erc Condition



C) 75erc Condition

Figure 25
ROC Plot for the 25% Economically Rational Criterion 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% for the 25erc condition, 70.4% for the 50erc condition, and 70.6% for the 75erc condition.

Discussion

Experiment 1 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. 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

58 A

The subjective criterion hypothesis made three predictions in Experiment 1: 1) Varying the economically valid criterion would affect the subjects' centering; 2) This manipulation would not affect likelihood rating or sensitivity. 3) A centering effect would be present for the subjective criterion but not the other measures.

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, 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 1 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 2

Experiment 2 examines the gain-loss framing effect in addition to the centering effect with a larger sample size than experiment 1. It had two goals: 1) Examine the gain-loss framing effect; 2) Examine whether there is an interaction between the gain-loss framing effect and the centering effect. Experiment 2 used the same task as experiment 1 with both the economically rational criterion and gain-loss framing manipulated. The economically rational criterion used 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 a gain frame and a loss frame were used 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. 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 identical to experiment 1. Participants performed a drought decision task as described in experiment 1 with additional conditions and corresponding point structures. The gain frame used was equivalent to the loss frame with the drought resistant crop (safe option) yielding a sure gain while the regular crop (risky option) having the potential to yield a higher gain (see Appendix). 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 2, both a gain frame and a loss frame were used. There were two economically rational criterion levels (erc): 25% (25erc), and 50% (50erc). There were two frame levels: 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 1 while the point structure in the gain frame conditions was equivalent to their loss frame counterparts (See table 8). The beginning point balance and payment structure of the 25erc and 50erc conditions were the same as of experiment 1 were used.

25erc Condition Gain Frame		
	Safe Option	Risky Option
Drought Occurred	Hit: 300 points	Miss: 0 points
Drought Did Not Occur	False Alarm: 300 points	Correct Rejection: 400 points
25erc Condition Loss Frame		
	Safe Option	Risky Option
Drought Occurred	Hit: -100 points	Miss: -400 points
Drought Did Not Occur	False Alarm: -100 points	Correct Rejection: 0 points
50erc Condition Gain Frame		
	Safe Option	Risky Option
Drought Occurred	Hit: 200 points	Miss: 0 points
Drought Did Not Occur	False Alarm: 200 points	Correct Rejection: 400 points
50erc Condition Loss Frame		
	Safe Option	Risky Option
Drought Occurred	Hit: -200 points	Miss: -400 points
Drought Did Not Occur	False Alarm: -200 points	Correct Rejection: 0 points

Table 8
Point Structure of the Four Conditions in Experiment 2

Trial Structure. The trial structure was the same as experiment 1. 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. The post-survey questions were identical to experiment 1.

Design

Experiment 1 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. Like in experiment 1, 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, participants made likelihood ratings using a VAS and binary crop decisions [redacted]. 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. Unlike experiment 1 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 2 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 according to prospect theory, the prediction was that there should be both a centering effect and a gain-loss framing effect on the subjective criterion. As with experiment 1, 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. Based on the utility function of prospect theory (Tversky & Kahneman, 1979), the gain-loss framing effect should lead to a higher subjective criterion in the loss condition than in the gain condition. Moreover, no interaction between these two effects was expected as their mechanisms should be theoretically independent. Finally, the manipulation of economically rational criterion and gain-loss framing were not expected to affect likelihood ratings or sensitivity.

The same set of dependent measures were used as were used in reanalyses and experiment 1. To measure decision bias, the calculated criterion, the self-reported criterion, and the relative frequency of choosing the safe option were analyzed. 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%) [redacted] 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). The two manipulations both had an effect on the calculated criterion.

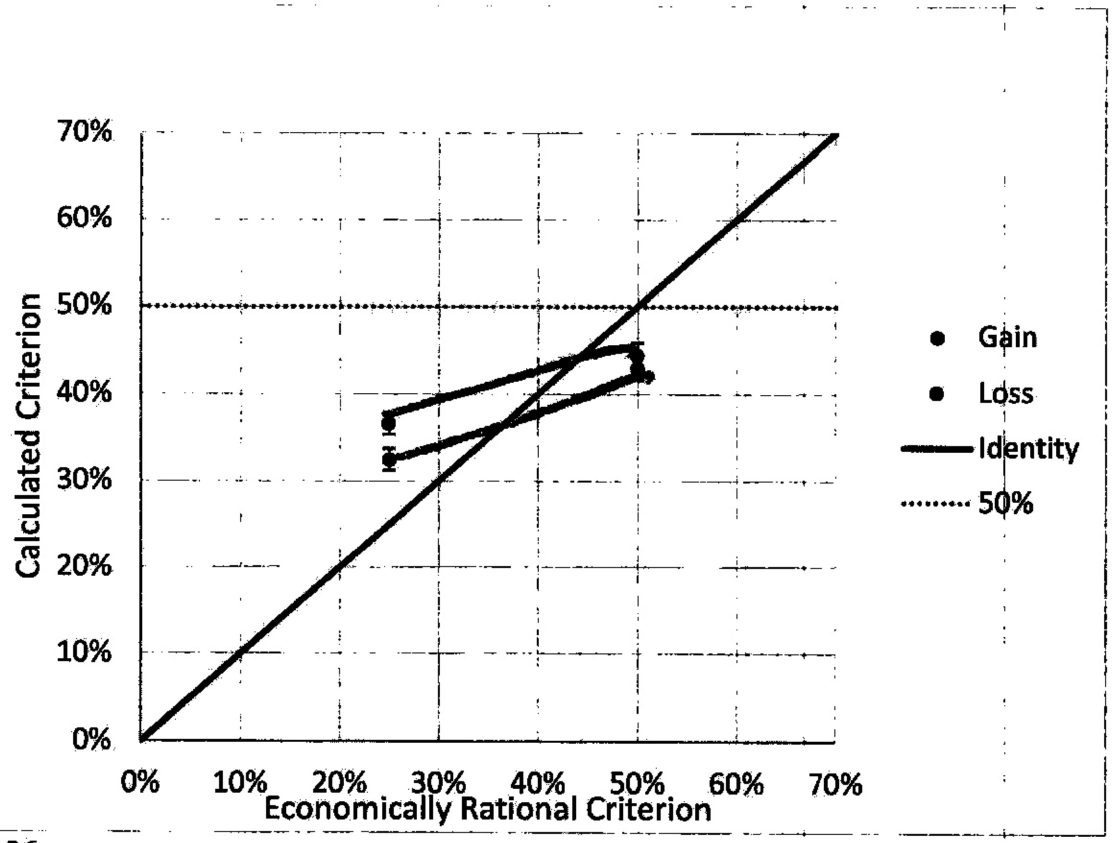


Figure 26
Calculated Subjective Criterion in the Four Conditions of Experiment 2
In the 25erc gain frame condition, the calculated criterion was 7.5% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the calculated criterion was 11.6% higher than 25%. In the 50erc 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 [redacted] framing effect. A higher signed deviation in the 25% condition than in the 50% condition indicates [redacted] 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 [redacted] and the economically rational criterion manipulation [redacted].

[redacted], to better understand the [redacted] 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 = .02[illegible]). 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 = [illegible]).

In summary, both manipulations affected the calculated criterion 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 ($F(1, 459) = 2.76$, $p = .097$. The two manipulations both had an effect on the self-reported criterion.

[Give difference of differences.]

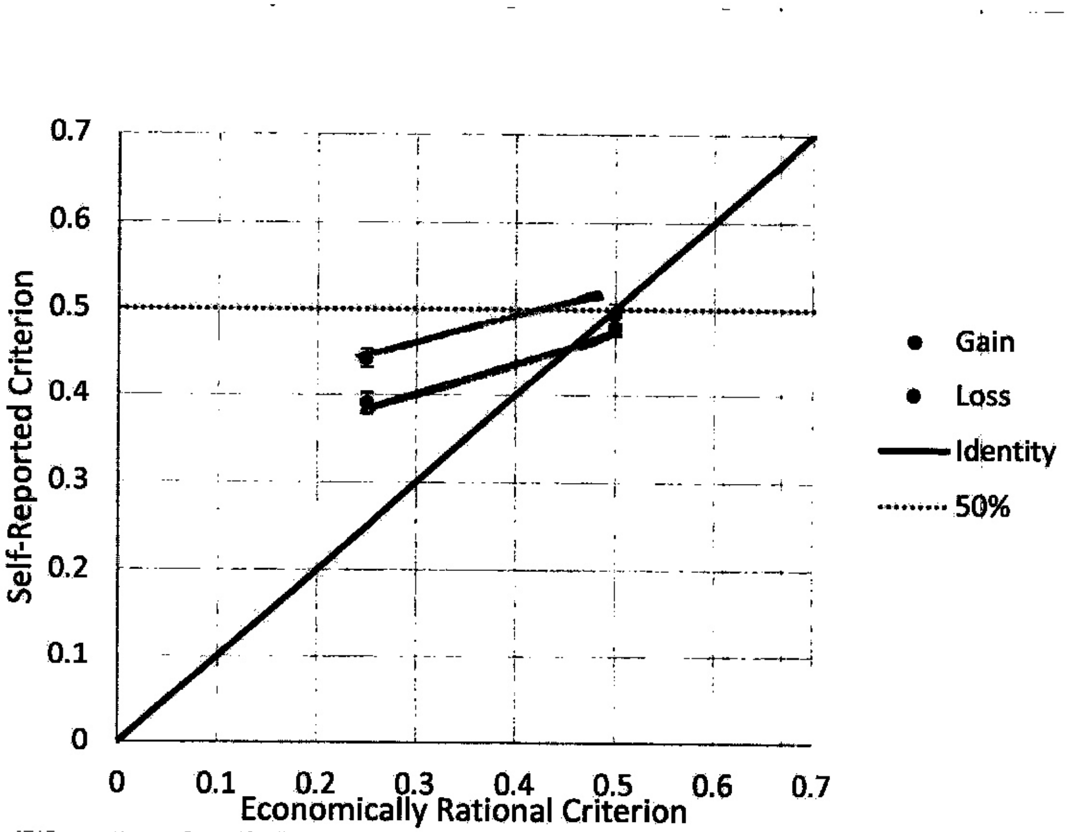


Figure 27
Self-Reported Subjective Criterion in the Four Conditions of Experiment 2
In the 25erc gain frame condition, the self-reported criterion was 14.0% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the self-reported criterion was 19.3% higher than 25%. 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.

measure the decision bias

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~~ changes in decision bias of both the gain-loss framing and the economically rational criterion manipulations ~~but no interaction~~.

[redacted], to better understand the ~~deviation~~ decision bias 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, [redacted]). 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, [redacted]). 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%

Give size of difference

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

In summary, both manipulations affected the ~~self-reported criterion~~ 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.

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, 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 ($F(1,554) = 10.21$, $p = .001$). The two manipulations both had an effect on the relative frequency. However, an interaction was detected.

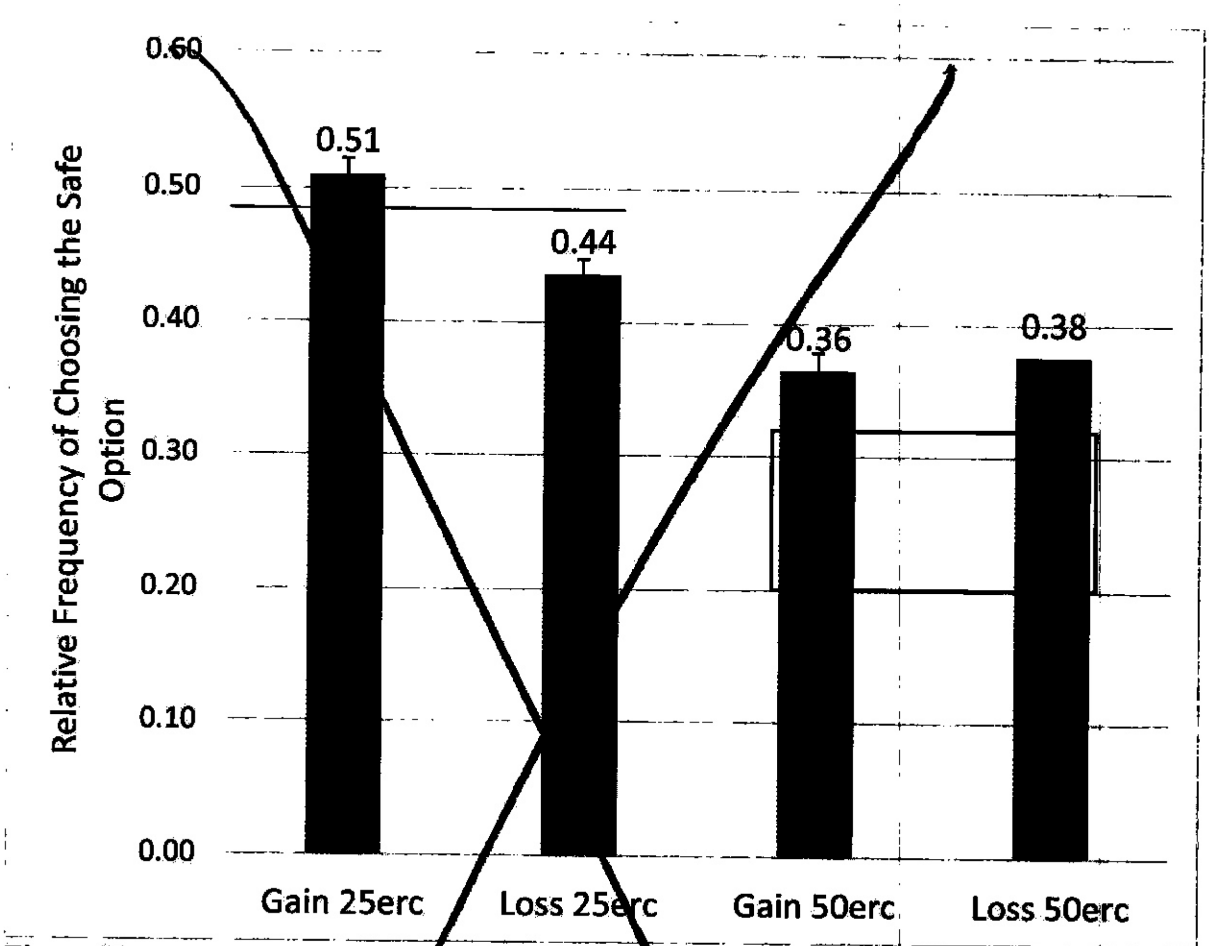


Figure 28
Relative Frequency of Choosing the Safe Option in the Four Conditions of Experiment 2
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red box in the 50% economically rational criterion condition shows the relative frequency between 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 [redacted] likelihood ratings were [redacted] affected by either manipulation.

differ by only 0.3%, and

Thus there is no sign that the

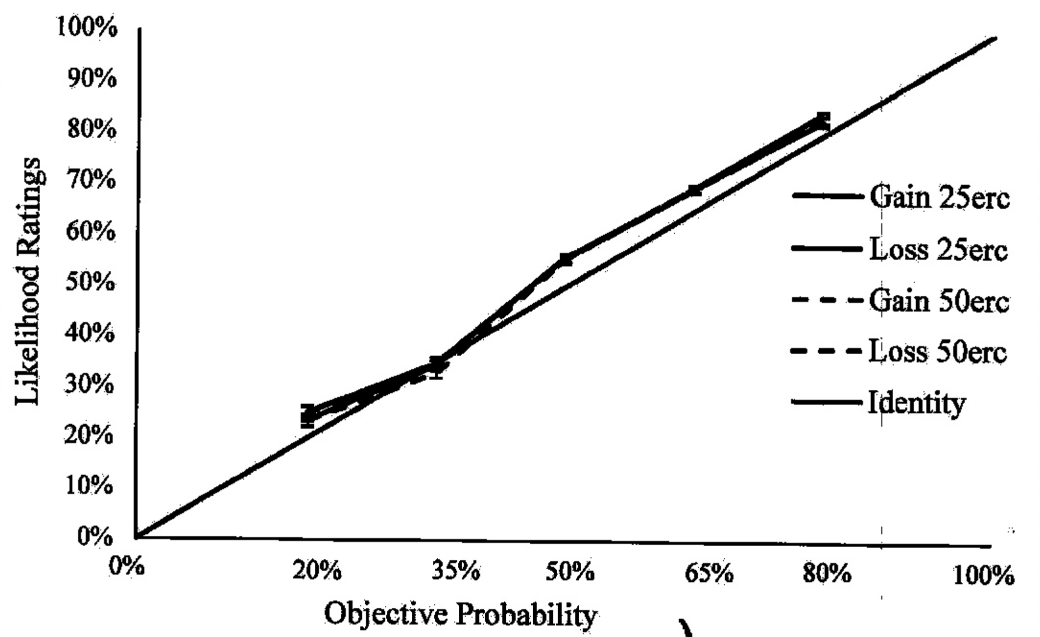


Figure 29
Likelihood Ratings in the Four Conditions as a Function of Objective Probability in Experiment 2

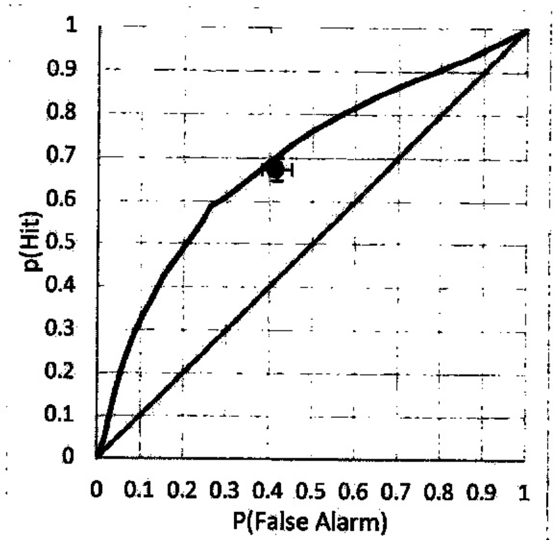
The green solid line represents the 25erc gain frame condition. The green dashed line represents the 50erc gain condition. The red solid line represents the 25erc loss frame condition. The red dashed line represents the 50erc loss condition. There was no observed centering effect.

Make Bigger especially Y axis

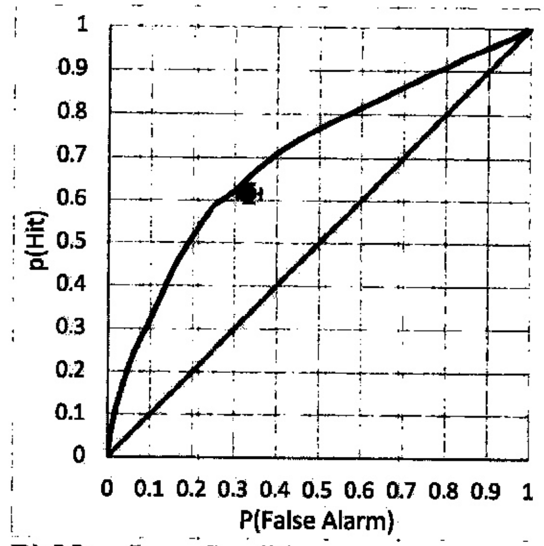
In addition, four post hoc one-sample t-tests compared the mean likelihood rating in each condition to the proportion of drought trials of 36%. A significant deviation suggests a bias in likelihood ratings. In the 25erc gain frame condition, the mean likelihood rating (M = 40.5%, SD = 9.8%) was significantly 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.

Sensitivity

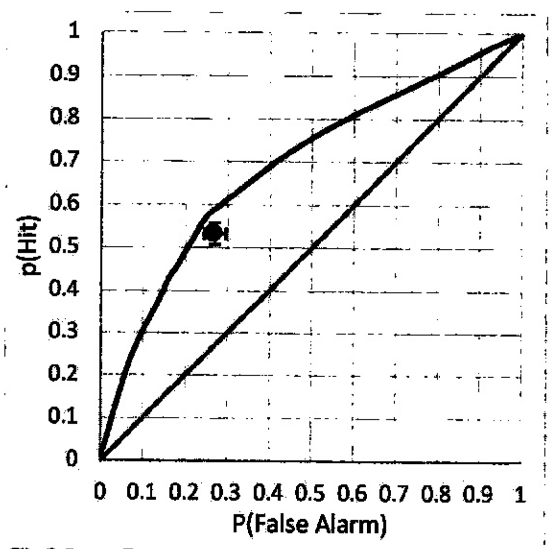
The next analysis examined the sensitivity of participants' ability to predict the drought, as measured by the area under the ROC curve. In Figure 30, ROC plots are shown for the four conditions respectively. The ROC curves of the four conditions were similar, indicating that both all conditions had 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.



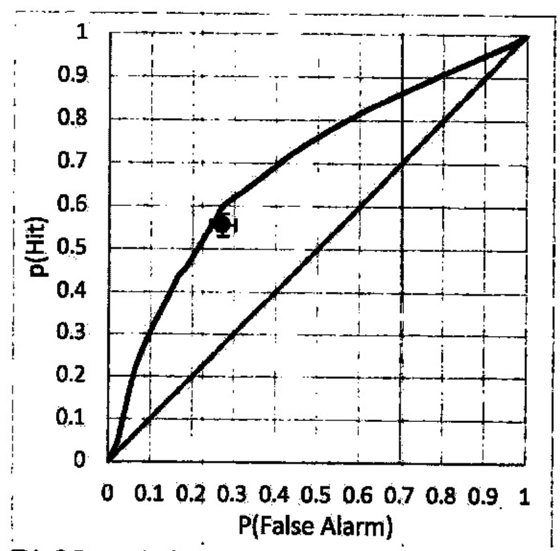
A) 25erc Gain Condition



B) 25erc Loss Condition



C) 25erc Loss Condition



D) 25erc Gain Condition

Figure 30
ROC Plot for the Four Conditions in Experiment 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 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

← Insert 82A

Experiment 2 tested the prediction that both the manipulation of the economically rational criterion and gain-loss framing 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.

There was evidence for a centering effect in the subjective criterion exposed by the economically rational criterion manipulation, similar to the findings of experiment 1. 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 and needs to be interpreted with caution. 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.

There was evidence for a framing effect,

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 where higher in the loss frame than those in the gain frame condition. 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

82A

The subjective criterion hypothesis made 4 predictions in Experiment 2. 1) Varying the gain-loss framing would affect the the subjective criterion.
2) Varying the economically rational criterion would affect the subjective criterion (as in Experiment 1).
3) Neither manipulation would affect likelihood ratings or sensitivity.
4) A centering effect would be present for the subjective criterion but not the other measures.

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 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 [redacted]re, if for some reason, participants became more risk-seeking with increased difficulty.

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 [redacted] size of the two effects, the gain-loss framing effect shifted the calculated criterion by 4.2% in the 25erc condition [redacted]. 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 caused a larger decision bias ~~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.

~~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.

[redacted], experiment 2 yielded support for both the gain-loss framing [redacted] and [redacted] centering [redacted] on subjective criterion and in [redacted], 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. A[redacted] 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 two new experiments used a signal detection theory perspective to examine the role centering and gain-loss framing on the subjective criterion in naturalistic tasks requiring decision making under risk. The gain-loss framing and the economically rational were manipulated. A centering effect, tested in experiments 1 and 2, was observed in both subjective criterion measures 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%. both subjective criterion measures were 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. Finally, the results of analysis of the ROC plots were consistent with the random likelihood model. In 6 out of 7 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. Neither subjective likelihood nor the sensitivity was not affected by manipulating either the economically rational criterion or frame 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

86 A

In summary, the effects of centering and framing were both consistent with the subjective criteria hypothesis and not the subjective likelihood hypothesis.

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 with probabilistic information than without, 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 two experiments here. The similar risk-seeking tendency was also observed in the 25erc loss condition of both 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 was 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 study did not take the participant level of variance into account. This means the ROC curves did not have confidence intervals. Therefore, 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 6 out of 7 conditions, suggesting that the subjective criterion and subjective likelihood jointly determine the decisions.

First, 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 as well. 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 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 Centering Effect

An[redacted] important finding was that the subjective criterion was affected by the centering effect (Poulton, 1979, Olkkonen et al., 2014). Interestingly, in experiment 1 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 2 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 1, 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 2, 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 the high end (e.g., 90%) but encompasses the low end (e.g., 10%). As a result, the center of the range was slightly lower than 50%.

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 similar tasks if 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.

One potential account ~~As for the mechanism~~ of the centering effect on the subjective criterion is anchoring and adjustment [redacted] (Kahneman, 2003). By Anchoring and adjustment [redacted] 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 [redacted] (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. An alternative [redacted]er [redacted]e 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

New Para

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. In summary, anchoring & adjustment and extreme-avoiding tendency are two accounts that might explain the centering effect.

The Gain-Loss Framing Effect and Prospect Theory

The findings of experiment 2 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. This result also meant that predictions of prospect theory are compatible with the random likelihood model and can be integrated to better explain decisions.

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: Centering and framing. 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 study), 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 [redacted] 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 [redacted] 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 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

The study reported here used the random likelihood model based on signal detection theory to better understand naturalistic weather decisions in loss conditions. 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. They 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

Appendix A

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 A1
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.

Cut all Appendix

Move into OSF?