Sources of Bias in Naturalistic Decision Making under Risk from a Signal Detection Perspective

Chao Qin, Susan Joslyn, John Palmer

University of Washington

Keywords: Probabilistic Forecast, Severe Weather Warning, Signal Detection Theory, Risk Perception, Decision Criterion

7/5/25

JP's Comments 7/9/25

Abstract

Severe weather requires protective action even at low probabilities. However, as revealed in naturalistic decision experiments, people often fail to protect when it is rational to do so, resulting in predominantly risk-seeking decisions. This study applied ~~a~~ signal detection theory perspective to better understand this phenomenon. Two hypotheses were tested: People are risk-seeking due to 1) too high criterion for taking protective action or 2) too low subjective likelihood of the event. In two experiments conducted in 2024, the economically rational criterion and gain-loss framing were manipulated. Results revealed that participants' subjective criterion was between the economically rational criterion and the center of the range, suggesting a centering effect. In addition, the subjective criterion was higher in a loss than a gain frame, suggesting a framing effect. In contrast, participants subjective likelihood ratings were unaffected by either manipulation, but tended to be inflated. ~~However, t~~ The shifted subjective criterion overcame the inflated subjective likelihood, resulting in risk-seeking decisions. Thus, we conclude that a shift of the subjective criterion ~~can lead to~~ is the source of the risk-seeking decisions in the naturalistic decision tasks studied here. The sample was reasonably representative in age, gender, and race, suggesting generalizability to the US population. Potential interventions to improve subjective criterion placement are discussed.

Public Significance Statement

In situations in which the probability of the severe weather is low but the severity of the potential outcome is high, people tend to take more risk than warranted. This research, using a novel signal detection approach, provides new insights into that behavior. It ~~suggests~~ shows that even when reliable probabilities are provided, people are risk seeking. It is not because people misperceive the likelihood of the event, but rather, because they have a higher than optimal subjective criterion for taking action. ~~The~~ These findings ~~reported here~~ provide new insights for designing behavioral interventions to increase the uptake of protective actions against severe weather events.

Acknowledgements

The reanalysis data used in this research was funded by NOAA VORTEX-SE grant ~~numbers~~ NA17OAR4590196 and NA17OAR4590197. ~~The National Center for Atmospheric Research is sponsored by the National Science Foundation~~. The new experiments ~~in this research~~ were partially funded by the Bolles Graduate Fellowship Fund of the University of Washington.

In risky ~~[illegible]~~ situations such as potential severe weather, people often need to make decisions for an uncertain future. For example, when facing a possible tornado, people must decide whether or not to seek shelter before they know if it will be necessary. Taking such action can cost time and resources but provide needed protection if a tornado occurs ~~in some cases~~. On the other hand, not taking protective action can save time and resources, but might expose decision-makers to potential harm if a tornado occurs. Due to the prospect of serious harm, people are often advised to take protective action when the probability of severe weather is low, for example, a 10% chance (Qin et al., 2024). Under these circumstances, severe weather often fails to occur at the residents' location, making protective action seem like a waste of time, potentially making them reluctant to follow future recommendations. Indeed, research on residents responses to warnings for events such as floods, tornadoes, and hurricanes, suggests that people often fail to comply with the recommendations (e.g., evacuation), a risk-seeking tendency (Baker, 1991; McKinley & Urbina, 2008; Parker et al., 2009; Smith & McCarty, 2009; Morss & Hayden, 2010; Nagele & Trainor, 2012; Gibbs & Holloway, 2013; Martin et al., 2017; Rashid et al., 2025).

One contribution to low compliance might be that residents do not routinely receive probabilistic information, although such information is increasingly available with its dissemination encouraged by scholars (National Research Council et al., 2006; Morss et al., 2008; Nagele & Trainor, 2012, Gallo et al., 2016; Karstens et al., 2015). Lacking probabilistic information, residents may misunderstand the likelihood or worse, they may regard the forecast as deterministic which may increase the perception that the warning was a false alarm when the severe weather fails to occur. This may, in turn, reduce trust in future warnings.

Naturalistic Weather Tasks Requiring Decision Making under Risk

Indeed, there is experimental evidence that providing probabilistic information helps people better understand the likelihood of the weather event, increases trust in the forecast, and allows them to make better decisions compared to those using a single value or deterministic forecasts (Ash et al., 2014; Joslyn & LeClerc, 2013; Klockow-McClain et al., 2020; Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). Decision quality in these experiments was determined by comparing participants' decisions to an economical rational standard. According to expected value theory, to maximize gains or minimize losses, one should choose the option with the best expected value (sum of the option's outcomes multiplied by the probability of occurrence; Tversky & Fox, 1995). The probability at which the expected value of the risky option (e.g., a chance of losing $10) and fixed cost of the safe option (e.g., always costing $5) break-even is the economically rational criterion: One should select the safe option when the probability of the adverse event is greater than the economically rational criterion.

In the experiments cited above, participants' decisions were closer to this rational standard when probabilistic forecasts were provided to them. The only exception was when the economically rational criterion was at very low probability levels (e.g., 10%; Morss et al., 2010, Qin et al., 2024) where no advantage was observed for probabilistic forecasts over deterministic forecasts. Moreover, although probabilistic forecasts generally improved decisions, participants still showed a risk-seeking tendency (not taking protective action as often as was warranted). Interestingly, this risk-seeking tendency was observed despite the fact that participants' self-reported perception of the likelihood of the weather event was fairly accurate or even slightly overestimated (Grounds & Joslyn, 2018; Burgeno & Joslyn, 2023; Gulacsik et al., 2022; Qin et al., 2024). This suggests that perceived probability of the severe weather event was only part of the equation. Additional biases must have contributed to participants' decisions.

Most of the decision tasks in the experiments cited above used a loss frame: Participants could only lose points with the goal to minimize point loss. However, some of the tasks involved a mixed gamble in which it was possible to both gain and lose points from the options. In mixed gamble tasks participants were overall risk-averse: They chose the safe option more often than was economically rational (Demnitz & Joslyn, 2020). This pattern, taking more risk than is economically rational for losses and less risk than is rational for gains or mixed gambles, is consistent with a gain-loss framing effect described by Prospect Theory (Kahneman & Tversky, 1979).

However, the result of inflated subjective likelihood and risk-seeking decisions in a loss frame (Gulacsik et al., 2022; Qin et al., 2024) presents a curiosity if considered together. When all else is equal and unbiased, overestimating likelihood should lead 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 the opposite tendency (Gulacsik et al., 2022; Qin et al., 2024). This combination of biases was also observed in a game of chance experiment, in which only losses were possible but no probabilistic information was provided to participants (Barron & Ursino, 2013). Participants' decisions showed a risk-seeking bias, but they overestimated likelihood. This suggests that this curiosity is not limited to naturalistic weather tasks or to tasks in which the probability is provided.

This curiosity is important because it suggests that there is another component in the decision-making process that leads to the risk-seeking bias described above. In the study reported here, we ask whether the observed risk-seeking tendency could be explained by a higher than rational subjective criterion, the likelihood above which one decides to take protective action.

A higher than rational subjective criterion can explain a risk-seeking decision bias when perceived likelihood is well calibrated to objective probabilities. Theoretically a high enough subjective criterion could also counteract an overestimated subjective likelihood and lead to risk-seeking decisions. For example, i~~f the probability of the adverse event is 30% and~~ suppose the economically rational criterion for choosing the safe option is 20%~~, one should do so. However,~~ Then they should choose the safe option when the likelihood of an adverse event is greater than 20% ~~a person might choose the risky option in this case even if their perceive likelihood is 40%,~~ But, if ~~because~~ their subjective criterion is ~~much~~ higher than the rational criterion at, for example, 50%. Then, even an ~~In this case, despite the~~ overestimated subjective likelihood~~, the end result~~ can result in a decision ~~was~~ biased toward the risky option. Therefore, a biased subjective criterion can explain the disconnect between overestimated subjective likelihood and risk-seeking decisions. A similar explanation has been used for biases in the motivated reasoning literature. For instance, people require more evidence (higher criterion) to be convinced that their preferred conclusion is false compared to the nonpreferred conclusion (Kunda, 1990; Windschitl et al., 2010).

Signal Detection Theory and Decision Making

In the work presented here, a signal detection approach was used to isolate the impact of participants' subjective criterion. Signal detection theory, typically used to understand perceptual processes, concerns two separate psychological mechanisms: Internal representations ~~referred to as the signal of the stimulus and that of the background, as well as the~~ that represent the strength of a stimulus and a subjective criterion ~~in signal~~ that represents the strength required by the decision maker to identify it as present (See Figure 1; Macmillan & Creelman, 2005). Separate internal representations are generated, ~~one~~ by the presence of the stimulus ~~and the other by background that would be present even when the~~ or by the absence of the stimulus ~~is absent~~ (two distributions in Figure 1). Both are assumed to be noisy and follow some distributions

~~typically normal~~. Because of this noise, the same ~~external signal~~ stimulus might be internally represented differently each time. The ability to differentiate internal representations of the stimulus present versus absent is the person's *sensitivity*. Sensitivity is jointly determined by the distance between the distributions relating to the two internal representation distributions ~~and~~ their noise. The further apart or the less noise, ~~the less the two distributions overlap, resulting in~~ the greater the sensitivity.

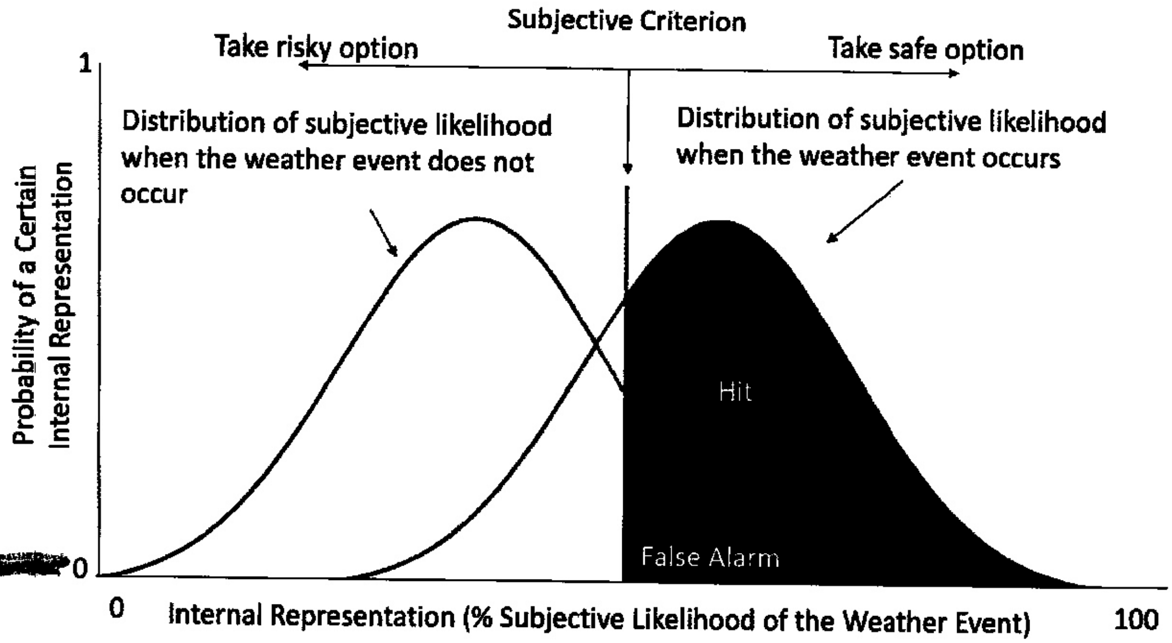
The subjective criterion is the ~~strength of the signal~~ value of the internal representation above which the presence of the stimulus is reported (red line in Figure 1). The placement of the subjective criterion indicates whether there is a decision bias towards reporting a presence or an absence of the stimulus. Under the same internal representation, a lower subjective criterion leads to an increase in both hits (reporting stimulus is present when it is) and false alarms (reporting stimulus is present when it is not)~~, while a high~~. In contrast, a higher subjective criterion 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.

Assuming stochastic internal representations such as those of signal detection theory is not new to decision theory (Busemeyer & Townsend, 1993; Thurstone, 1994; Wallin, et al., 2018). For stochastic representations, one does not necessarily make the same decision based on the same external information every time. This means that, in the context of weather decision tasks, one does not take protective action every time the objective probability is above the economically rational criterion. Signal detection theory with variable internal representations can be integrated with behavioral economics theories such as prospect theory that consider the uncertainty in the outcomes to better understand decision-making processes (Lynn et al., 2015).

Figure 1

Random Likelihood Model Example Based on Signal Detection Theory

page break



Note. The random likelihood model need not assume the shape of the internal representation distribution. Normal distributions are used in the graph as an example.

~~In addition,~~ signal detection theory has been applied to the study of diagnostic decisions in real life decision making under risk (Swets et al., 2000). For example, 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 a combination of their sensitivity and their subjective criterion the amount of evidence required to diagnose the image as indicating cancer ~~(subjective criterion)~~. As with many diagnostic decisions, the utility of a miss (reporting no stimulus when the stimulus is present) is not the same as the utility of a false alarm. In this example, perhaps not diagnosing cancer when the patient actually has cancer is less acceptable than incorrectly diagnosing cancer when the patient does not have cancer. Therefore, the physician should have a lower, more liberal criterion (requiring less evidence) to increase hits at the expense of higher false alarms. Signal detection theory has been applied in these situations is to identify the economically rational criterion to increase outcome utility in many domains ~~like~~ including medical/psychiatric diagnoses, violent risk assessment, weather forecasting, and school admissions (Swets et al., 2000). In previous work, some studies aimed to provide a prescriptive rational criterion for decision makers (Swets et al., 2000). Others aimed to distinguish effects on "sensitivity" and "bias" ~~like~~ similar to the research reported here (Harvey et al., 2012).

Table 1

Perceptual and

Stimulus-Response Table Used in ~~Signal Detection Theory and Adaptation in~~ Weather Tasks

	~~Stimulus-Response Table~~ Perceptual Tasks	
	Stimulus Present	Stimulus Absence
Reporting a presence	Hit	False Alarm
Reporting an absence	Miss	Correct Rejection
	Weather Tasks	
	Severe Weather Present	Severe Weather Absence
Taking protective actions	Hit	False Alarm
Not taking actions	Miss	Correct Rejection

Double space lines

Random Likelihood Model

Signal detection theory is applied in this project by assuming a random likelihood model. It is an analog to a random utility model in which a given value is translated to a noisy utility (Bockenholt, 2006). As illustrated in Table 1, the four outcomes in the naturalistic weather decision tasks can be aligned with the four outcomes in signal detection theory (hit, false alarm, miss, correct rejection), ~~allowing the application of the theory to participants' behavior in these tasks~~ (Ferrell & McGoey, 1980, Harvey et al., 2012). Here The internal representation ~~of interest here~~ is the participants subjective likelihood ~~[illegible]~~ that the weather event will occur, using a similar method to a previous application of signal detection theory (Ferrell & McGoey, 1980). The subjective criterion is the subjective likelihood above which the participants decide to take protective action. We assume that the subjective likelihood of the weather event has variability while the subjective criterion has no variability.

This model assumes that subjective likelihood varies over ~~the~~ trial~~s~~ even when given the same forecast information which includes probabilities. This assumption is consistent with previous studies with naturalistic decision-making tasks showing that subjective likelihood ratings differed in the same participant in different trials with the same objective probability, even when ~~which were~~ the probabilities of the weather event on each trial that were calibrated to be roughly reliable[1] ~~and provided to participants in experimental conditions~~ (Demnitz & Joslyn, 2020; Gulacsik et al., 2022; Qin et al., 2024). The model does not assume a specific shape of distribution for this variability. This is because subjective likelihood ratings were elicited directly ~~over several trials~~ to provide an estimate of the distribution. ~~In this project,~~ The random variability of subjective likelihood is ~~considered~~ likely the result of an amalgamation of many factors.

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

For example, experience with weather events in preceding trials has been found to affect the behavior in the subsequent 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).

The random likelihood model assumes that the subjective criterion has no variability, consistent with the typical application of signal detection theory (Macmillan & Creelman, 2005). In other words, we assumed that people chose the safe option whenever their subjective likelihood was above their subjective criterion, which may or may not be the same as the economically rational criterion. T~~his is a simplifying assumption. In reality, people might change their subjective criterion between decisions depending on the outcome of previous trials (Demnitz & Joslyn, 2020).~~

This model allows for bias in subjective criterion as well as subjective likelihood. This combination of biases can shift one's decision away from the economically rational decision in three separate ways. 1) A higher subjective criterion can lead to risk-seeking bias while a lower one can lead to a risk-averse bias. 2) Higher subjective likelihood shifts the internal representations upward, potentially ~~above the subjective criterion, effectively~~ potentially leading to a risk adverse bias ~~decision bias towards the safe option~~. Lower subjective likelihood shifts the internal representations downward, potentially ~~below the subjective criterion~~ leading to a risk seeking bias ~~towards the risky option~~. 3) A greater sensitivity either resulting from increased distance between the distributions or reduced variability can lead to more hits and fewer false alarms regardless of the decision bias. The first two ways are biases where participants make more either risk-seeking or risk-averse errors. The last way ~~was not a bias as it affects both errors equally.~~ might also introduce biases depending on how the distributions shift.

The study reported here used the random likelihood model to examine two possible explanations for the tendency toward risk-seeking decisions: A bias in the subjective criterion (biased criterion hypothesis) versus a bias in the subjective likelihood distribution (biased likelihood hypothesis).

Possible Explanations for Biases in Subjective Criterion

~~Furthermore~~, we ~~investigated~~ focused on two explanations possible for ~~a possible~~ shift in the subjective criterion away from the economically rational criterion. The first is the centering effect, the tendency to bias ~~quantity~~ judgements towards the center of the response range (also called central tendency bias; Poulton, 1979; Olkkonen et al., 2014). Originally observed in perceptual judgements (noise volume, distance, color perception), this effect has been observed in judgement and decision making ~~like~~ such as the 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 present in the subjective criterion in decision-making tasks. For instance, an economically rational criterion at 20% might result in a subjective criterion above 20% closer to the midpoint of 50%, while an economically rational criterion at 80% might lead to a subjective criterion below 80%, again, closer 50%.

A second ~~source of~~ explanation for decision bias is gain-loss framing, ~~as explained by prospect theory~~ (Tversky & Kahneman, 1979). Risk-aversion in the gain frame might be due to a subjective criterion lower than what is economically rational, while a risk-seeking tendency in the loss frame might be due to a higher than rational subjective criterion. These two sources of bias, centering and framing, are not ~~necessarily~~ mutually exclusive and might both contribute to a biased subjective criterion.

Overview of Analyses and Experiments

To begin, two previously published experiments from Qin et al., 2024 were reanalyzed using the random likelihood model to determine the subjective likelihood, subjective criterion, and the sensitivity in the context of both framing and centering, ~~to the degree possible~~. Then two new naturalistic weather decision experiments were conducted to more systematically examine the biased criterion and ~~hypothesis and its potential causes as well as the~~ biased likelihood hypothes~~i~~es.

Transparency and Openness

For data availability, hypothesis, and analysis plan of the reanalyses (Qin et al., 2024), please see the transparency and openness sections of the original paper. Analytic codes and additional materials for reanalyses are available upon request.

For the two new experiments reported below, hypotheses, experimental design, procedure, method, elimination criteria and data analysis plans of were preregistered at the Open Science Framework at https://doi.org/10.17605/OSF.IO/RUFD8 on Jan. 21th, 2024 (experiment 1) and https://doi.org/10.17605/OSF.IO/V368N on March. 29th, 2024 (experiment 2). Registration occurred after data collection but before data analyses. Data is available at the Open Science Framework ~~at~~ https://osf.io/bspqn/?view_only=249fe2542ed747829c42dd35e3f6e609 (experiment 1) and https://osf.io/sv57z/?view_only=116867f6874341cd85e2cc18b6586664 (experiment 2). Analytic codes, and additional materials for the two new experiments are available upon request.

use simpler URLs

Reanalysis of Previous Experiments

Method of Tornado Experiments

In ~~the~~ two previously published experiments (Qin et al., 2024), participants made decisions, based on a tornado warning, about whether to take shelter from possible tornadoes (safe option) or not (risky option). The tornado warnings were presented either with or without the probability of a tornado, manipulated between groups. The reanalysis presented here included only the probabilistic conditions. In these conditions, participants were given reliable objective probabilities of 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 differences between these conditions were not the concern of the study reported here, they were combined. There were 83 participants from experiment 1 and 85 from experiment 2 in the reanalyses.

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 a tornado would be (likelihood ratings), and decided between the safe option and the risky option (binary decision). At the end of each trial, they were told whether the tornado occurred. As mentioned above, there was a point cost to shelter and a larger point penalty if participants failed to shelter and a tornado hit their location. Therefore only losses were possible. 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 a tornado hit. Participants could mitigate this loss completely by choosing the safe option for a fixed cost. The cost was 90 points in tornado experiment 1 and 270 points in tornado experiment 2. Therefore, the economically rational criterion was 9% in tornado experiment 1 and 27% in tornado experiment 2. Another difference was that in tornado experiment 1 the proportion of trials with a tornado was 23.5% while in tornado experiment 2 it was 38.2%.

Results

Three sets of analyses were conducted for each experiment to examine ~~the~~ three possible ways the decisions could be affected. Ananlyses were conducted on 1) mean subjective criterion, 2) mean likelihood ratings, and 3) sensitivity, operationalized as Receiver Operating Characteristic (ROC) plots. ROC plots were constructed from participants likelihood ratings and decisions and show how different subjective criteria can impact participants' hit and false alarm probability given the same sensitivity. The greater the area under the curve the greater the sensitivity (see Figure 2). Inferential statistics of t-tests were conducted with an alpha of .05.

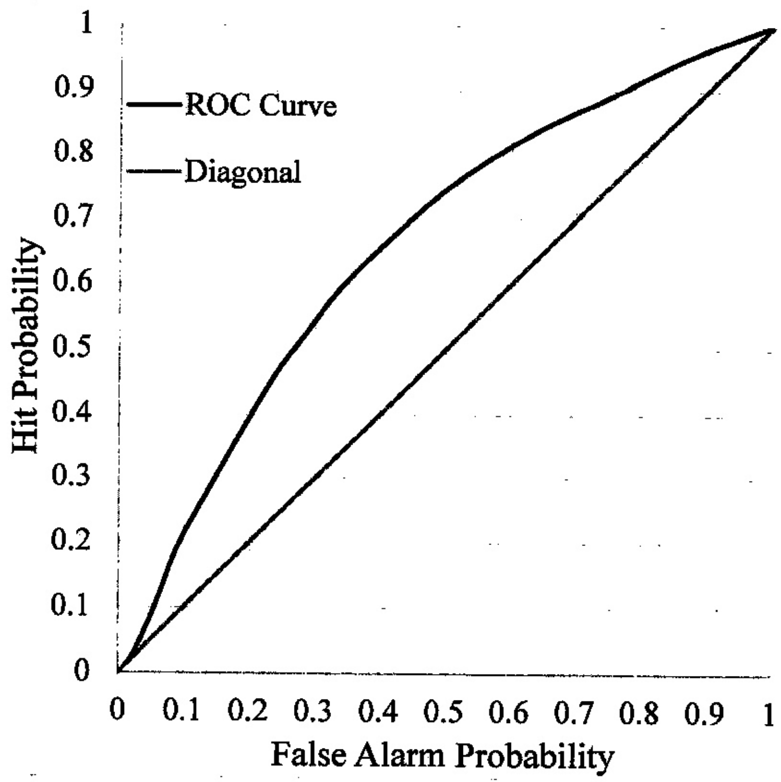
Table 2

Point Structure of the Two Tornado Experiments

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

Figure 2

A Typical Curve



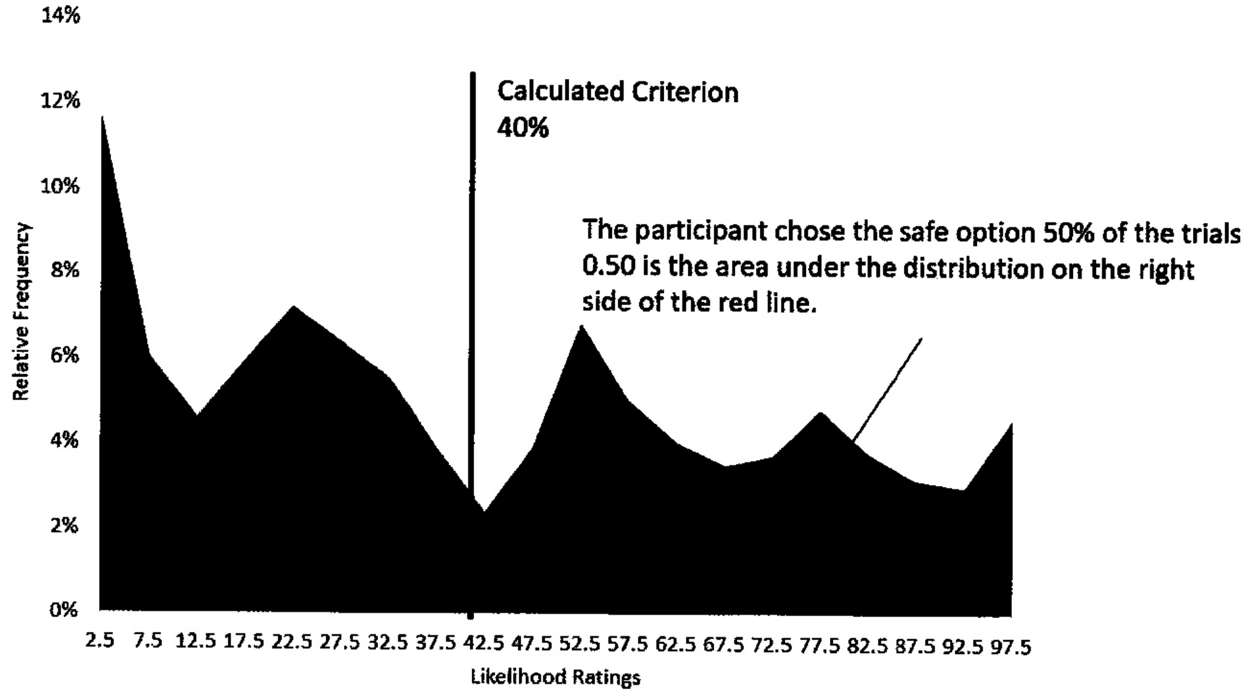
Note. The y axis is the probability of a hit when the probability of a false alarm is the value on the x axis. A lower criterion is indicated by an increase in the probability of both a hit and a false alarm, following the blue curve from bottom left to upper right.

Calculated Subjective Criterion

The subjective criterion was estimated using each participant's likelihood ratings and proportion of trials on which they chose the safe option. The assumption was that participants chose the safe option whenever their subjective likelihood was higher than their subjective criterion. Thus, for each participant, the cumulative proportion of likelihood ratings falling between X% chance and 100% was determined such that it matched the proportion of trials on which the participant chose the safe option. The likelihood at the end point (X%) is referred to as the *calculated subjective criterion*. For example, if a participant chose the safe option on 50% of the trials, a number was located on their likelihood rating distribution such that 50% of the trials had a likelihood rating higher than that number (Figure 3). This number, say 40% likelihood (the red line in Figure 3), was the calculated subjective criterion for this participant because 50% of trials fell above it. With this method, a calculated subjective criterion was obtained for each participant, and a mean was calculated for each experiment.

Figure 3

Illustration of the Calculation of Subjective Criterion



Note. The blue area is an example of a likelihood rating distribution.

In Figure 4, the mean calculated criterion for each experiment is shown as a blue dot in relation to the economically rational criterion. 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% ($t(148.6) = 4.07$, $p < .001$). This is consistent with an effect of the economically rational criterion, ~~because it~~ that differed between ~~was different in each~~ experiment. LS

In addition, the mean calculated criterion in each experiment was directly compared to the economically rational criterion in that experiment using one-sample t-tests. In both cases the mean calculated criterion was higher than the economically rational criterion. In tornado experiment 1, the calculated criterion (M = 31.0%) 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 ((M = 41.2%) was again, significantly higher than the economically rational criterion of 27% with a difference of 14.2% ($t(84) = 9.81$, $p < .001$). Thus, the calculated criterion was higher than the economically rational criterion in both experiments, suggesting a biased subjective criterion. Moreover, the bias was larger in the experiment with the lower economically rational criterion.

Likelihood Ratings

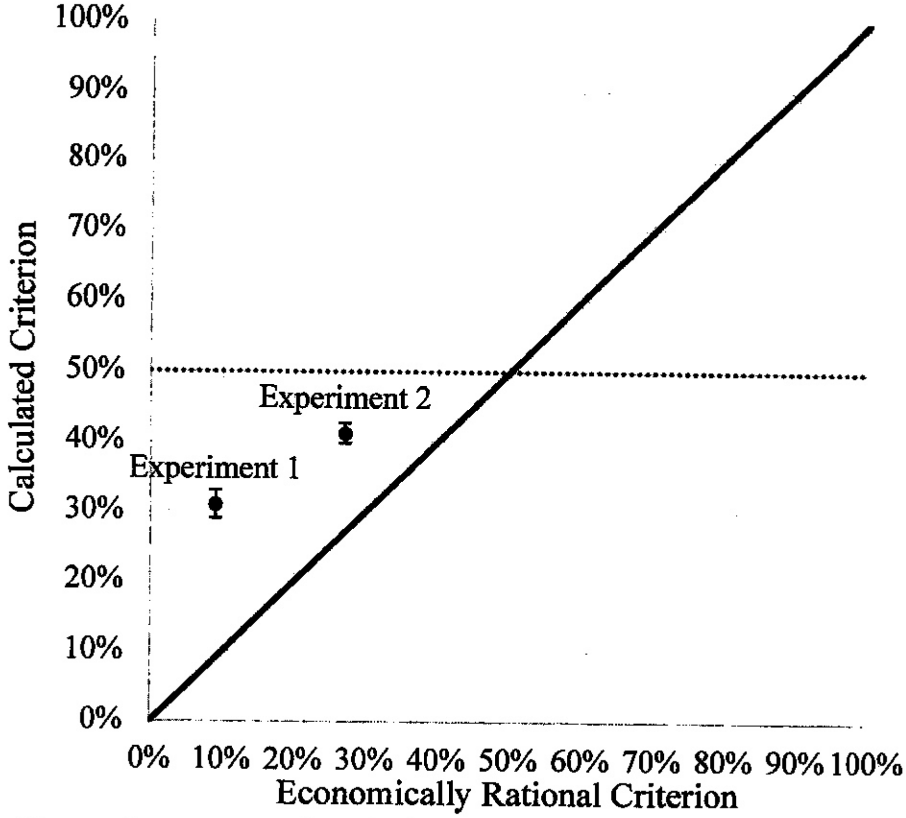
In each experiment, one-sample t-tests were used to compare the mean likelihood rating over all trials with differing probability levels to an operationalization of mean objective probability, the proportion of trials in which a tornado occurred in that experiment[2]. In both experiments mean likelihood ratings were higher. In tornado experiment 1, the mean likelihood

[2] This is equivalent to the mean objective probability which was reliable.

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% ($t(84) = 5.48$, $p < .001$). Figure 5 shows the likelihood ratings for tornado experiments 1 and 2 as a function of the objective probabilities. It confirms that the likelihood was overestimated for all objective probability levels except for the extreme high end. Because the mean objective probability was different between the two experiments (due to different trial composition), it was not possible to compare them to determine whether there was an effect of different economically rational criteria on subjective likelihood.

Figure 4

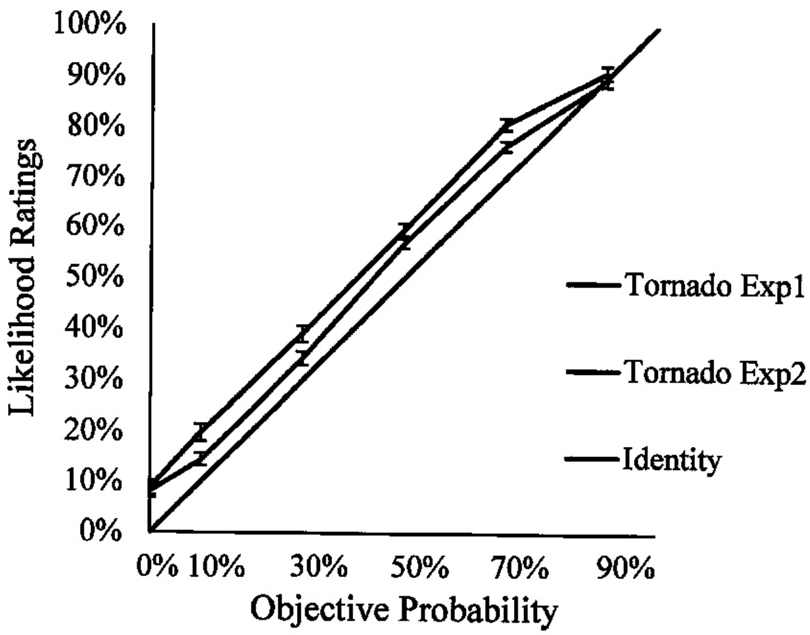
Calculated Subjective Criterion in the Two Tornado Experiments



Note. The y axis represents the calculated criterion while the x axis represents the economically rational criterion. Diagonal line represents when the calculated criterion is the same as the economically rational criterion. The dashed line represents a calculated criterion of 50%. Error bars show standard errors of the mean.

Figure 5

Likelihood Ratings as a Function of Objective Probabilities in the Two Tornado Experiments



Note. The y axis represents the likelihood ratings while the x axis represents the objective probabilities. The blue line represents the likelihood ratings of tornado experiment 1. The orange line represents the likelihood ratings of tornado experiment 2. Error bars show standard errors of the mean.

Receiver Operating Characteristic Plots

The ROC analysis was conducted to test the assumption of the random likelihood model that subjective likelihood and the subjective criterion jointly determine the decision. It also allowed us to examine whether the sensitivity (participants' ability to predict the tornado as revealed by their shelter decisions) differed between experiments. The ROC plots were composed of two parts: 1) ROC curves 2) Points indicating the hit and false alarm rates of decisions in each experiment. For a similar approach see Harvey et al. (2012). The calculation of each is explained below.

The ROC curves show the relation between the estimated hit and false alarm probabilities based on the likelihood rating distribution conditionalized by whether there was a drought or not and a varying hypothetical cut-off at 5% intervals from 0% to 100% on the distribution, using a method similar to Ferrel & McGoey (1980). A hit was defined as a trial with a likelihood rating above the cut-off and a tornado occurred. The hit probability was the proportion of tornado trials above that cut-off. A false alarm was a trial with a likelihood rating above the hypothetical cut-off and a tornado did not occur. The false alarm probability was the proportion of no tornado trials above the cut-off. By varying the cut-off by 5% steps from 0% to 100%, a pair of hit and the false alarm probabilities at each step (20 in total) were calculated and plotted as the ROC curve in Figure 6. As 68 trials per participant were too few data points to populate 20 steps to create participant level ROC curves, trials from all participants were lumped to create a singular ROC curve for each experiment.

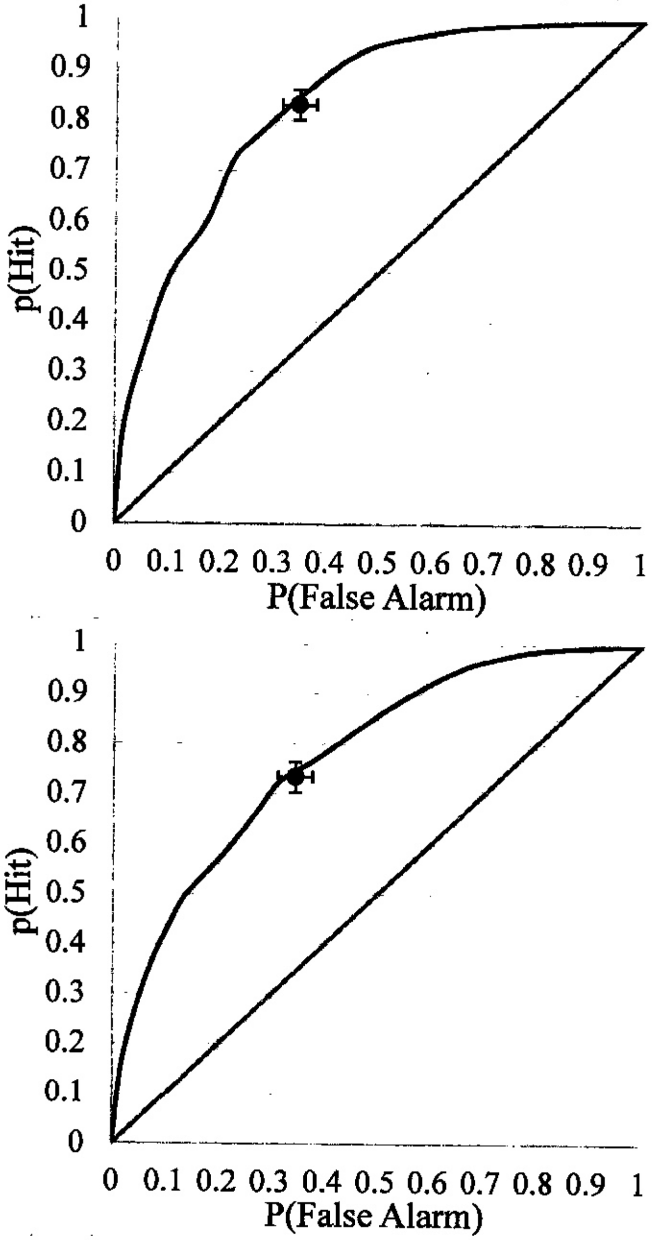
Next, a point representing the mean proportion of hits and false alarms of participants' binary decisions in each experiment was added to the appropriate ROC plot. A hit was when the participant chose the safe option and the tornado occurred. A false alarm was when the participant chose the safe option and no tornado occurred.

The first goal of the ROC analysis was to test the random likelihood model method of calculating the subjective criterion. If participants chose the safe option whenever their likelihood rating was above their subjective criterion, the hit and false alarm probabilities of the ROC curve from their likelihood ratings should align with the mean hit and false alarm probability of their decisions. Figure 6 shows the ROC plots for tornado experiment 1 and 2. In both experiments, the ROC curve and the 95% confidence interval of the decision point overlapped. This indicates that the decision point was consistent with the ROC curve.

To address the second goal of the ROC plot analysis, the mean percent area under the ROC curve was calculated. The greater the area under the curve the greater the sensitivity. 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. However, as different trial compositions can affect sensitivity, a conclusion about the effect of altering the economically rational criterion cannot be drawn.

Figure 6

ROC Plots for Tornado Experiment 1(Top) and 2 (Bottom)



Note. The y axis represents the probability of hits. The x axis represents the probability of false alarms. The orange dot indicates the proportion of hits and false alarms for each experiment based binary decisions. Error bars show standard errors of the mean.

Discussion

The reanalysis of the tornado experiments indicated that the participants' subjective criterion was higher than the economically rational criterion in both experiments, indicating 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, consistent with the centering effect. Because the tornado experiments did not test a gain frame known to lead to risk averse behavior (subjective criterion expected to be higher for gain than loss), nor a condition where the economically rational criterion was higher than 50% (subjective criterion expected to be lower than rational), it is not possible to distinguish between these two explanations. Hence in the new experiments reported in the next section, these conditions were added to distinguish the centering effect and the gain-loss framing effect.

The reanalysis presented above also suggests that the risk-seeking tendency observed in these two experiments was not due to biased subjective likelihood. The mean likelihood rating analysis showed that likelihood was significantly over rather than underestimated in both experiments. Subjective likelihood ratings were also not affected by centering as there were similar amounts of overestimation across all objective probability levels except for the highest level. The higher-than-rational subjective criterion counteracted the overestimated likelihood ratings and led to risk-seeking decisions. Unfortunately, as the trial composition (e.g., mean objective probability of a tornado) in the two experiments was different, the mean likelihood ratings were not comparable with each other to examine whether changing the economically rational criterion had an effect on subjective likelihood.

Next, the mean sensitivity differed between the two experiments. Again, the difference in trial composition also prevented us from making a direct comparison in terms of mean sensitivity here. This is because the two experiments differed in the proportion of extreme probability trials (10%, 90%) in which it was easier to predict whether a tornado would hit, translating to a higher sensitivity. Therefore, it was not possible to determine whether this difference stemmed from the change of the economically rational criterion between the two experiments or the difference in trial composition.

Finally, in the ROC plots of both experiments, the 95% confidence interval of the binary decision point overlapped with the ROC curve. This indicates that the proportion of hits and false alarms based on participants' binary decisions was consistent with the ROC curve of hits and false alarms based on likelihood ratings. This in turn suggests that the subjective criterion and subjective likelihood were the sole determinants of the decisions, validating the random likelihood model approach to calculating the subjective criterion.

Overall, the reanalyses indicated a bias in the subjective criterion that overcame the bias in subjective likelihood and led to risk-seeking decisions. It left the door open to both a centering effect and/or a gain-loss framing effect on the subjective criterion. However, the experimental design and the difference in trial composition between experiments that led to confounds in several analyses prevent stronger conclusions.

no para

The goal of the two new experiments reported below was to distinguish the cause of the bias in subjective criterion and rule out the effect of subjective likelihood. They employed a drought preparation task based on Demnitz & Joslyn (2020), using both a gain and a loss frame[3].

Experiment 1

Experiment 1 investigated the effect of centering on subjective criterion by manipulating the economically rational criterion at, below and above 50%. If centering is present the subjective

[3] Two pilot experiments, not reported here, examined the gain-loss framing effect. One yielded a significant effect and the other yielded a trend in the expected direction but failed to reach significance.

criterion should not shift in the first case, shift up in the second and down in the third. By the biased criterion hypothesis, this manipulation was expected to affect only the subjective criterion and not 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. Demographic data was provided by Prolific. 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 (<1%) for whom Prolific provided no data. There were 12 Asians (8%), 13 African Americans (8%), 112 Whites (71%), 12 mixed (8%), 6 other (4%) and 2 (1%) where Prolific had no data. This experiment was approved by the University of Washington Human Subject Division.

Procedure and Stimuli

The experiment was hosted on Qualtrics. The link to the Qualtrics survey was posted on Prolific inviting those who were residents of the US to participate. The electronic informed-consent form was displayed on the first page and participants were instructed to click next and continue if they consented. See Supplementary Materials S1 for illustrations of the task, including instructions and survey questions.

In the task, henceforth called the *drought 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) based on a drought forecast for the upcoming season. Participants were told that a potential drought might lead to a loss for farmers reflected in a point loss for participants, compared to regular, non-drought seasons. They were then introduced to the point structure that tracks the outcomes of their decisions. They were paid a monetary bonus commensurate with their point balance at the end of the experiment. Participants' goal was to lose as few points as possible.

Point Structure. In order to simulate real life decisions with consequences, and to encourage participants to put forth their best effort, a point structure mentioned above was implemented (See Table 3). Point loss values determined the economically ration criterion, 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 (See table 3). In all conditions participants were endowed with the same starting balance of 20,000 points. The bonus payment structure was set up to make the bonus 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 the instructions and going through a practice trial, and answering two attention check questions (See Supplementary Materials), participants began the 50 experimental trials. Each trial comprised three screens. On the first screen, participants saw a forecast which described the probability that drought would occur on that trial (e.g., "The latest climate forecast indicates a 35% chance of drought in the upcoming season for farmer-client 1"). We refer to this as the *objective probability* because it was calibrated to be roughly reliable. The mean objective probability (M = 36.5%) matched the proportion of droughts across the 50 trials (36%). Objective probability had five within-subject levels: 20%, 35%, 50%, 65%, and 80% (see Supplementary Materials S2). The trial order was randomized for each participant.

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) labeled "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 in terms of point deductions.

Table 3
Point Structure of the Three Conditions in Experiment 1

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

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

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) decision criterion for choosing the safe option. This was an alternative to the calculated criterion for estimating the subjective criterion for choosing the safe option. The self-reported criterion was asked at the end of the experiment and thus was based on experience from the 50 trials, similar to a previous study (Joslyn & Grounds, 2015). However, the self-reported criterion was a conscious post-hoc assessment based on memory. As such, it might not match participants' actual decision-making process which might also be affected by unconscious trial-specific processes. Next, participants rated how difficult the task was and answered an open-ended question asking them which part of the experiment they found difficult to understand. These two questions were meant to check the challengingness of the goals in the loss frame task Finally, participants reported any problems 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 used a 3 x 6 mixed design. There was one between-group independent variable: Economically rational criterion with three levels: 25erc, 50erc, and 75erc. There was one within-group independent variable: Objective probability of a drought with five levels: 20%, 35%, 50%, 65% and 80%.

Assuming the biased criterion hypothesis and if

Results

Analysis Overview

~~According to~~ the biased criterion is ~~due to~~ centering ~~hypothesis~~, the mean subjective criterion should shift away from the economically rational criterion towards 50%. This means 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%. ~~However~~ In addition, the manipulation of the economically rational criterion should not affect subjective likelihood or sensitivity. Additional analyses on self-reported criterion and the proportion of trials choosing the safe option yielded results consistent with the calculated criterion analyses and are reported in the Supplementary Materials S3.

The same set of dependent measures as those used in the reanalyses were analyzed in experiment 1. To test the biased criterion hypothesis, the calculated criterion was analyzed. And to test the subjective likelihood hypothesis, Likelihood ratings and sensitivity (area under the ROC curve) were also analyzed ~~to rule out effects on these variables due to the manipulation of economically rational criterion~~.

For all dependent variables, a series of ANOVAs and t-tests were conducted. Holm-Bonferroni Method was used for planned and post hoc t-tests as well as planned pairwise comparisons under omnibus ANOVAs. Tukey method was used for post hoc pairwise comparisons under omnibus ANOVAs.

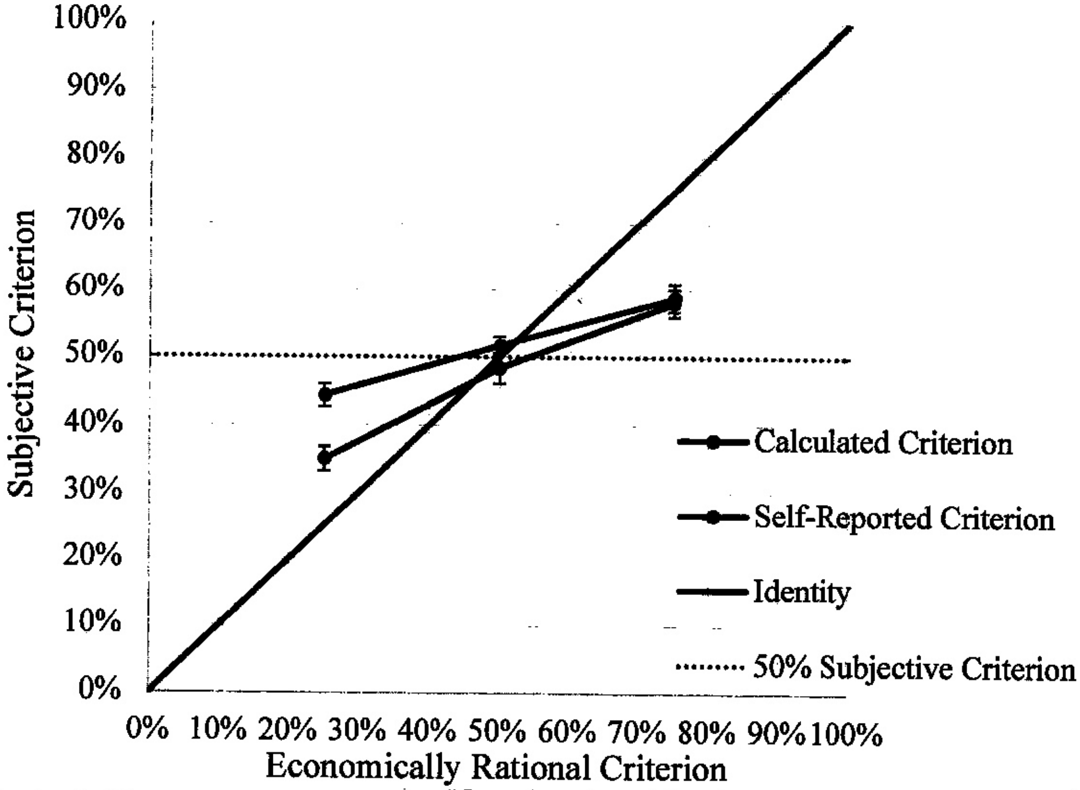
Calculated ~~Subjective~~ Criterion

In Figure 7, 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 on the calculated criterion with economically rational criterion (25erc, 50erc, and 75erc) as the independent variable showed a main effect ($F(2,154) = 32.53$, $p < .001$). In planned pairwise comparisons the 50erc condition had a significantly higher calculated criterion than the 25erc condition with a difference of 13.4% ($t(154) = 4.54$, $p < .001$, corrected alpha = .025). The 75erc condition had a significantly higher calculated criterion than the 50erc condition with a difference of 9.7% ($t(154) = 3.30$, $p = .001$, corrected alpha = .05).

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

Figure 7

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



Note. 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%. ~~The self-reported criterion~~ shared a similar pattern to the ~~calculated criterion.~~ Error bars show standard errors of the mean.

In addition, the pattern of self-reported criterion was consistent with that of the calculated criterion, although the self-reported criterion was pulled towards 50% to a greater extent than the calculated criterion in the 25erc condition. Similarly, the pattern of the proportion of trials choosing the safe option was also consistent with that of the calculated criterion. The proportion was lower than economically rational proportion (risk-seeking) in the 25erc condition, not different from economically rational (risk-neutral) in the 50erc condition, and higher than economically rational (risk-averse) in the 75erc condition. These results corroborate the calculated criterion analysis. The full results of the self-reported criterion and the proportion of trials choosing the safe option ~~is~~ are reported in the Supplementary Material S3.

Likelihood Ratings

Figure 8 shows the likelihood ratings as a function of objective probabilities in the three conditions. An ANOVA on the mean likelihood ratings collapsed over all objective probability levels with the economically rational criterion manipulation as the independent variable (25erc, 50erc, and 75erc) showed no significant differences ($F(2,154) = 0.81$, $p = .45$). 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. In addition, to assess the accuracy of participants likelihood ratings, post hoc one-sample t-tests compared the mean likelihood rating in each condition with the proportion of drought trials (36%). None of the three differences was significant (25erc: 1.1% ($t(53) = 1.16$, $p = .25$); 50erc: 4.3% ($t(48) = 1.94$, $p = .06$, corrected alpha = .017; 75erc 1.8% ($t(53) = 1.88$, $p = .07$, corrected alpha = .025)). This suggests that the likelihood of drought ratings was close to the proportion of drought trials and not affected by the manipulation of economically rational criterion. In addition, as seen in Figure

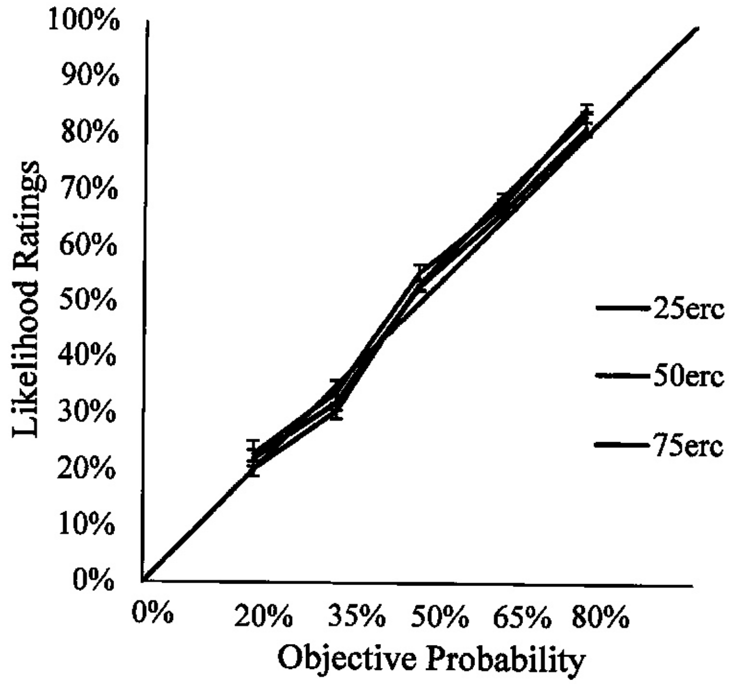
8, the three conditions had similar likelihood rating patterns. Likelihood ratings were slightly overestimated at all objective probability levels except for 35%.

Sensitivity

The next analysis examined the sensitivity or participants' ability to predict the drought, as measured by the area under the ROC curve. In Figure 9, ROC plots are shown for the 25erc, 50erc, and 75erc conditions. 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 on the mean percent area under curve, with the economically rational criterion as the independent variable (25erc, 50erc, and 75erc) failed to reach 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 dot representing the mean binary decisions fell on the ROC curve in all three conditions suggesting again that the decisions were consistent with participants making decisions based on subjective likelihood and subjective criterion.

Figure 8

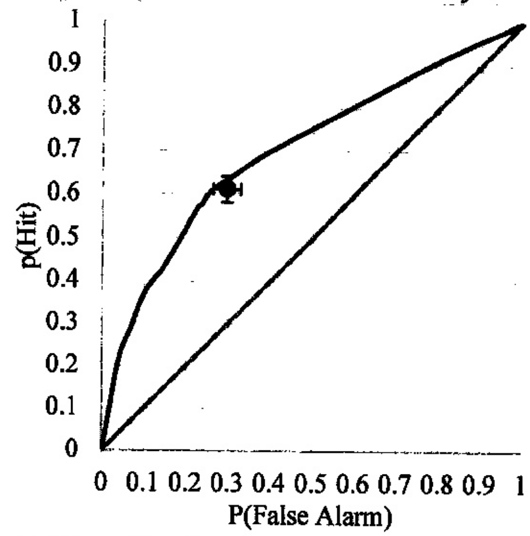
Likelihood Ratings as a Function of Objective Probability in Experiment 1



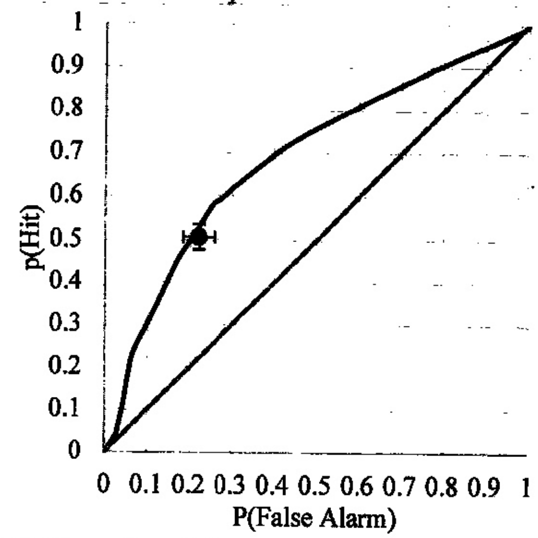
Note. The blue line represents the 25 economically rational criterion condition. The orange line represents the 50 economically rational criterion condition. The purple line represents the 75 economically rational criterion condition. There was no observed centering effect.

Figure 9

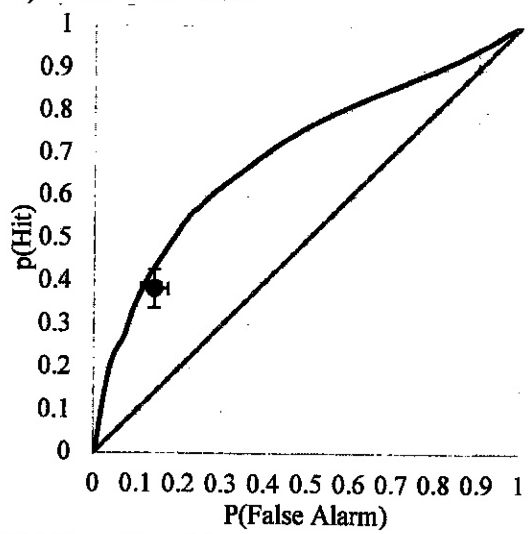
ROC Plot for the 25% Economically Rational Criterion Condition in Experiment 1



A) 25erc Condition



B) 50erc Condition



C) 75erc Condition

Note. 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. Error bars show standard errors of the mean.

Discussion

Experiment 1 tested two predictions based on the biased criterion hypothesis: 1) manipulating the economically rational criterion ~~would~~ should reveal a centering effect such that calculated criterion would shift towards 50%; 2) The likelihood ratings and sensitivity would not be affected. Supporting this hypothesis, the calculated criterion was between 25% and 50% in the 25erc condition (a shift up towards 50%), and between 50% and 75% in the 75erc condition (a shift down towards 50%). Meanwhile there was little shift in the 50erc condition. These results were corroborated by analyses of self-reported criterion and the proportion of trials choosing the safe option in which similar patterns were observed. Thus, in this loss framed task, participants were risk-seeking in the 25erc condition, risk-neutral in the 50erc condition, and risk-averse in the 75erc condition.

In addition ~~However~~, as predicted by the biased criterion hypothesis, manipulating the economically rational criterion did not affect the likelihood ratings ruling it out as an explanation for participants biased choices. As seen in Figure 8, there was a slight overestimation but no observed centering effect in likelihood ratings as indicated by comparing them to the objective probability levels. In addition, sensitivity was also not affected by manipulating economically rational criterion. This suggests that changing the economically rational criterion does not affect people's perception of the probability of drought or their ability to predict the drought.

Next, in the ROC plots (Figure 9), the binary decisions (the decision dot indicating proportion of hits and false alarm) were consistent with the respective ROC curves, suggesting a common basis for the likelihood ratings and the binary decision, ~~that the subjective criterion and the subjective likelihood were the sole determinant of the binary decisions~~. This provides support for the calculation method of the calculated criterion.

In summary, experiment 1 yielded support for centering affecting the subjective criterion but not subjective likelihood or the sensitivity, consistent with the biased criterion hypothesis. An interesting finding is that participants showed risk-aversion when the economically rational criterion was high even in a loss frame. This hints that the gain-loss framing effect might not be powerful enough to ensure risk-seeking behaviors in a loss frame at least in this scenario. ~~Finally, the result also supports the Random Likelihood Model method of calculating the criterion.~~

Experiment 2

In Experiment 2, the primary goal was to examine the gain-loss framing effect in addition to the centering effect. Therefore, both a gain frame and a loss frame were included. Experiment 2 used the same drought-related decision task as experiment 1.

Method

Participants

A total of 591 participants from the US were recruited from Prolific Academic in March 2024. 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[4]. Twenty-two were eliminated for failing the comprehension check (same question as experiment 1). As with Experiment 1, each participant was paid $4 plus a performance based monetary bonus. Demographic data was provided by Prolific. 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 (1<%) for whom Prolific provided no data. There were 62 Asians (11%), 73 African Americans (13%), 359 Whites (64%), 37 mixed race (7%), 24 other

[4] Same procedure was used in experiment 1 with no participant eliminated.

(4%) and 3 (1%) for whom Prolific provided no data. This experiment was approved by the University of Washington Human Subject Division.

Procedure and Stimuli

The procedure, trial structure and stimuli were similar to experiment 1. Only exceptions are noted here. In particular, a gain frame condition was added that was similar to the loss frame (See Table 4), with the drought resistant crop (safe option) yielding a sure gain while the regular crop (risky option) had the potential to yield a higher gain (see Supplementary Materials S2). In addition, the 75erc condition was omitted in order to maximize the number of participants in the each condition. The 75erc was excluded as it was the least realistic among the three levels as severe weather events usually require people to take protective action at a lower probabilities.

In terms of dependent variables, the same questions were asked as with experiment 1 (See Supplementary Materials S2). In Experiment 2, an ANOVA on difficulty ratings with gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) as the independent variables revealed that participants 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 ($M = 10.5$, $SD = 16.6$, $F(1, 554) = 4.94$, $p = .027$).

Point Structure. The point structure in the 25erc and 50erc loss frame conditions was identical to that of experiment 1 while the point structure in the gain frame conditions had the equivalent risk (See table 4). The beginning point balance and payment structure were the same as the 25erc and 50erc conditions of experiment 1.

Table 4

Point Structure of the Four Conditions in Experiment 2

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

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 five levels: 20%, 35%, 50%, 65% and 80%.

Results

Analysis Overview

As with Experiment 1 we expected the results to indicate a biased subjective criterion. In Experiment 2 we tested two mechanisms for such a bias, gain-loss framing and centering. Based on the utility function of prospect theory (Tversky & Kahneman, 1979), the loss frame should lead to a higher subjective criterion than the gain frame. 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. No interaction between these two effects was expected as their mechanisms are theoretically independent. Neither manipulation was expected to affect likelihood ratings or sensitivity.

Calculated Subjective Criterion

Figure 10 shows the calculated criterion for the four conditions. As predicted by prospect theory, the calculated criterion in the loss frame (M = 40.6%, SD = 15.9%) was higher than in the gain frame (M = 37.9%, SD = 16.9%). As predicted by the centering effect, it was higher than the economically rational criterion in the 25erc condition (M = 34.5%, SD = 15.0%) and lower than the economically rational criterion in the 50erc condition (M = 43.7%, SD = 16.7%). An ANOVA on the calculated criterion with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) revealed a main effect of the gain-loss framing manipulation such that in the loss frame the calculated criterion was 2.7% higher than in the gain frame (; $F(1, 554) = 4.38$, $p = .037$). There was a main effect of the economically rational criterion such that the calculated criterion in the 50erc condition was 9.2% higher than in the 25erc condition (; $F(1, 554) = 46.76$, $p < .001$). There was no significant interaction ($F(1, 554) = 1.03$, $p = .31$).

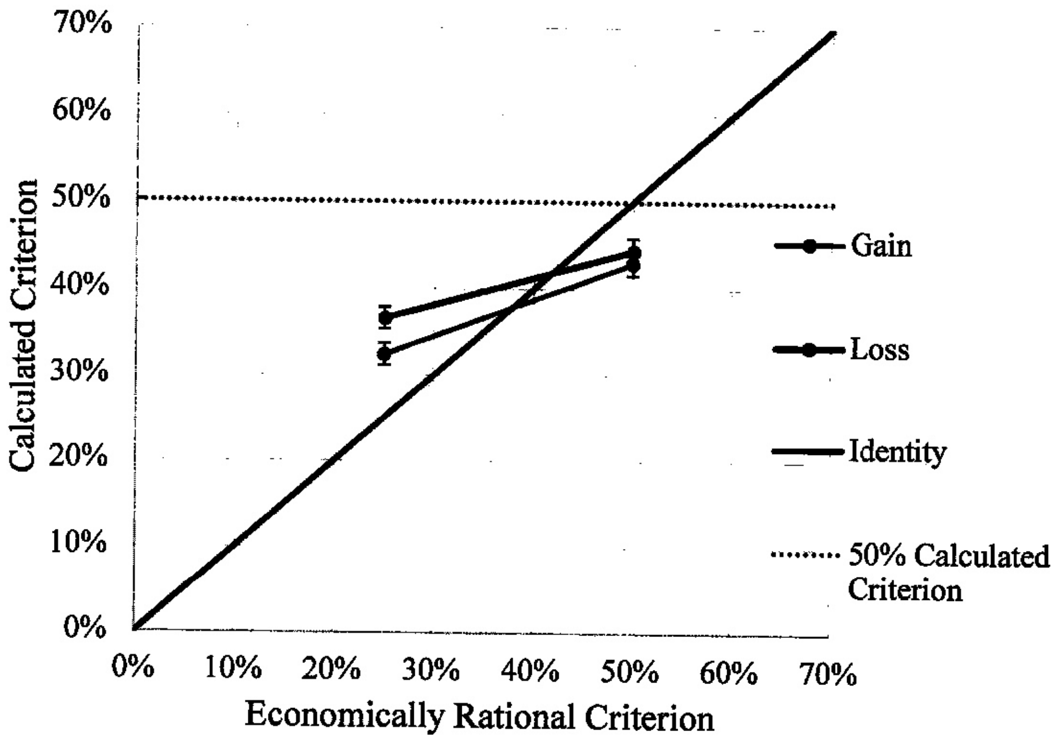
To measure the deviation from economically rational criterion in each condition, four planned one-sample t-tests compared the calculated criterion with the economically rational criterion in each condition. In the 25erc gain frame condition, the calculated criterion (M = 32.5%, SD = 15.4%) was significantly higher than the economically rational criterion of 25% with a difference of 7.5% ($t(140) = 5.76$, $p < .001$, corrected alpha = .017). In the 25erc loss frame condition, the calculated criterion (M = 36.6%, SD = 14.3%) was significantly higher than 25% with a difference of 11.6% ($t(129) = 9.27$, $p < .001$, corrected alpha = .013). In the 50erc gain frame condition, the calculated criterion (M = 43.0%, SD = 16.8%) was significantly lower than the economically rational criterion of 50% with a difference of -7.0% ($t(148) = 5.10$, $p < .001$, corrected alpha = .025). In the 50erc loss frame condition, the calculated criterion (M = 44.4%, SD = 16.5%) was also significantly lower than 50% with a difference of -5.6% ($t(137) = 137$, $p < .001$, corrected alpha = .05).

In summary, both manipulations affected the calculated criterion and they did not interact with one another. While the calculated criterion in the 25erc condition was higher than the economically rational criterion as predicted, the calculated criterion in the 50erc condition was unexpectedly lower than the economically rational criterion.

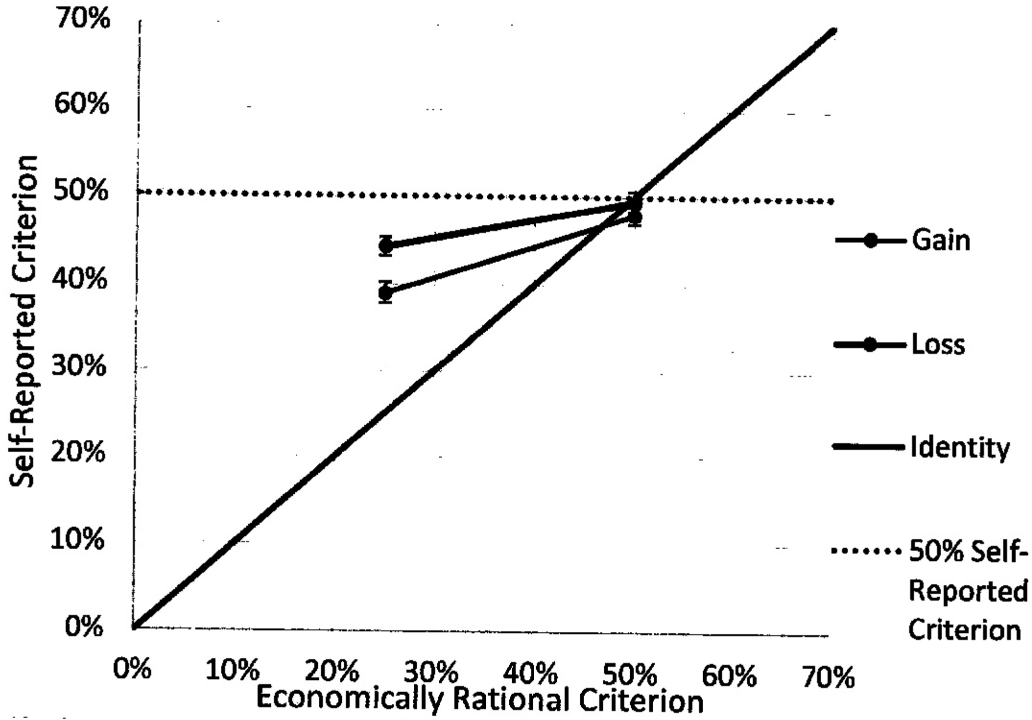
In addition, like in experiment 1, the pattern of self-reported criterion was consistent with that of the calculated criterion. The proportion of trials choosing the safe option showed lower proportion than economically rational (risk-seeking) in the 25erc loss frame condition and higher proportion (risk-averse) in the 25erc gain condition and both 50erc conditions . This result also corroborated the calculated criterion analysis. The full results are reported in the Supplementary Material S4.

Figure 10

Calculated and Self-Reported Subjective Criterion in the Four Conditions of Experiment 2



Note. Error bars show standard errors of the mean.



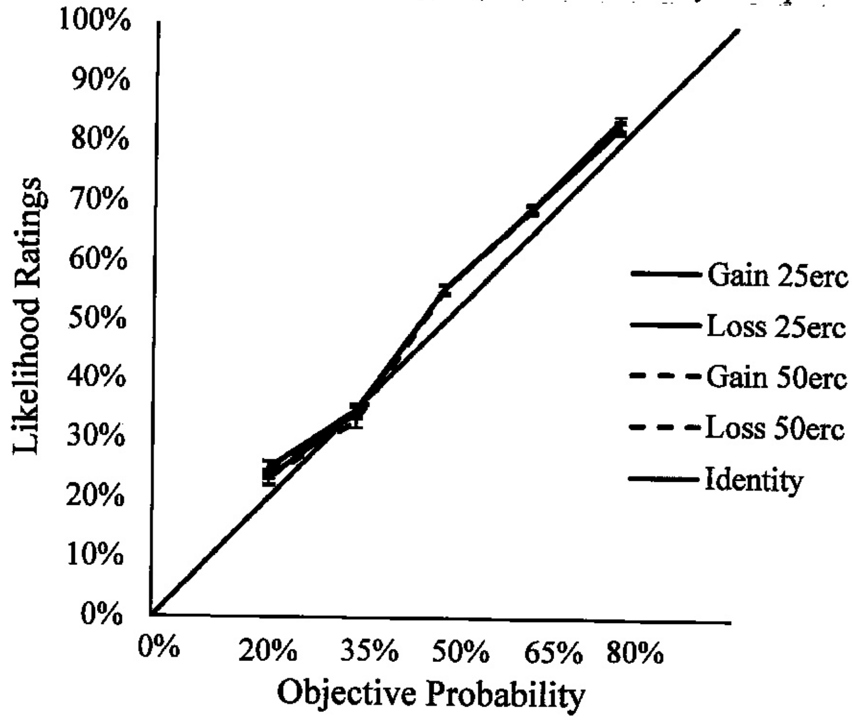
Note. In the 25erc gain frame condition, the calculated criterion was 7.5% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the calculated criterion was 11.6% higher than 25%. In the 50erc gain frame condition, the calculated criterion was 7.0% lower than the economically rational criterion of 50%. In the 50erc loss frame condition, the calculated criterion was 5.6% lower than 50%. The self-reported criterion followed a similar pattern.

Likelihood Ratings

Figure 11 shows mean likelihood ratings as a function of objective probabilities in the four 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 collapsed over all objective probability levels revealed that it was similar across conditions. There was a 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%) that 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 also, not significant ($F(1, 554) = 0.23$, $P = .64$). There was no significant interaction ($F(1, 554) = 0.36$, $P = .55$).

Figure 11

Likelihood Ratings a Function of Objective Probability in Experiment 2



Note. 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 effect of the manipulations.

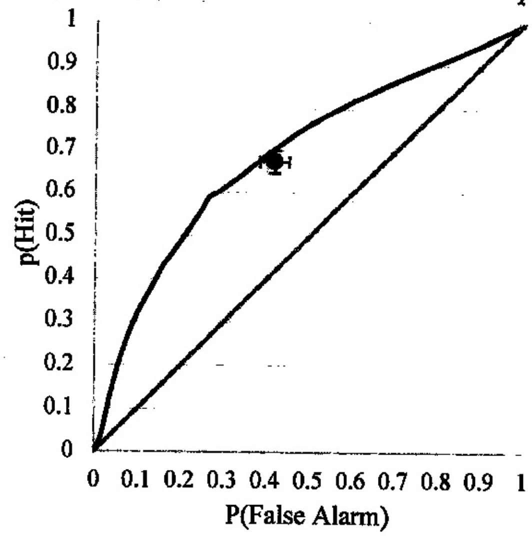
In addition, four post hoc one-sample t-tests compared the mean likelihood rating in each condition to the proportion of drought trials (36%) showing that in each case, likelihood ratings were higher. In the 25erc gain frame condition, the mean likelihood rating (M = 40.5%, SD = 9.8%) was significantly 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). Thus, likelihood ratings were overestimated compared to the proportion of drought trials regardless of the gain-loss framing or the economically rational criterion manipulation.

Sensitivity

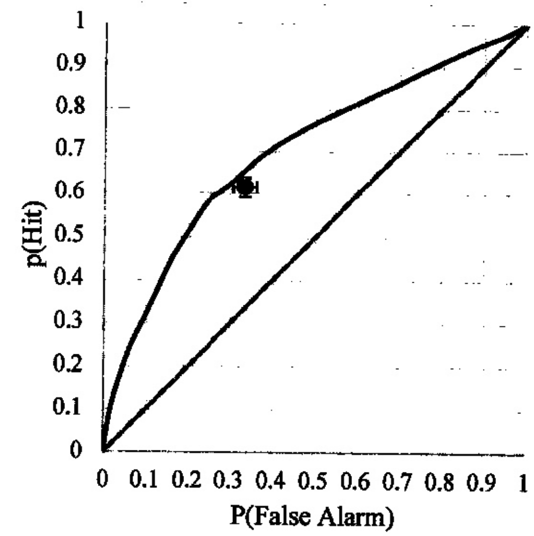
The next analysis examined sensitivity (participants' ability to predict the drought), as measured by the area under the ROC curve. Figure 12 shows ROC plots for the four conditions showing similar curves, indicating similar sensitivity in all four conditions. In addition, the 95% confidence interval of binary decisions overlapped with the ROC curves in all but the 50erc gain frame condition. 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 on the mean percent area under the curve with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc, and 75erc) as the independent variables showed no significant main effect of gain-loss framing ($F(1, 554) = 0.51$, $p = .47$) or economically rational criterion ($F(1, 554) = 1.93$, $p = .17$). There was no significant interaction ($F(1, 554) = 0.94$, $p = .33$). This suggests that the sensitivity was not affected by the manipulation of gain-loss framing and economically rational criterion and that the decisions were determined by subjective likelihood and subjective criterion.

Figure 12

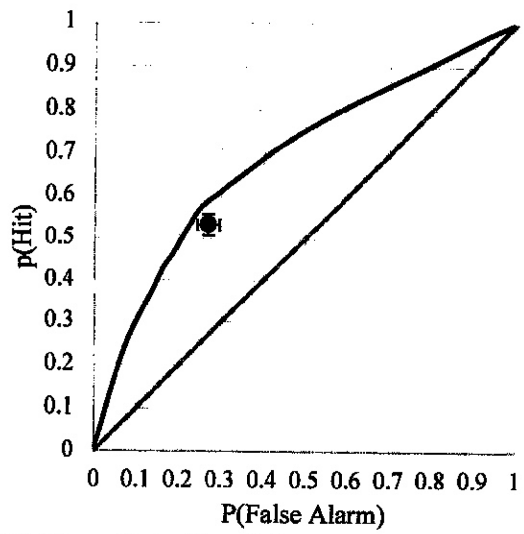
ROC Plot for the Four Conditions in Experiment 2



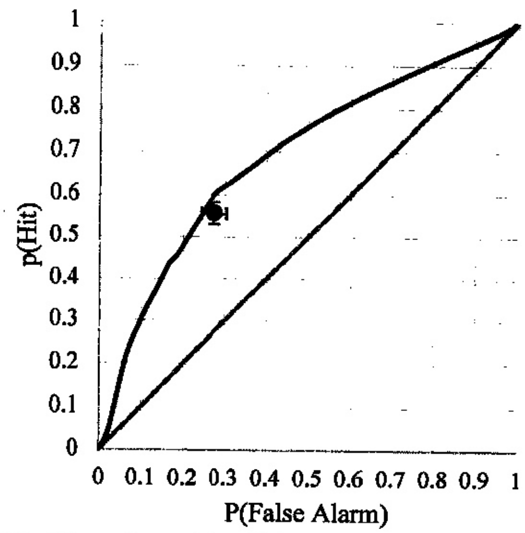
A) 25erc Gain Condition



B) 25erc Loss Condition



C) 50erc Gain Condition



D) 50erc Loss Condition

Note. 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. Error bars show standard errors of the mean.

for centering, with 9

Discussion

Experiment 2 tested predictions about the mechanisms that account for a biased criterion: for 1) gain-loss framing ~~in which~~ the subjective criterion should be ~~is~~ higher in the loss than in the gain frame; 2) economically rational criterion below 50% ~~which causes~~ the subjective criterion should ~~to~~ migrate toward the center of the range (50%). Both prediction were observed and did not interact with one another. Nor did ~~they~~ the manipulations impact likelihood ratings.

The mean calculated subjective criterion was higher than 25% in the 25erc condition suggesting movement toward the center (50%). The difference between the calculated subjective criterion and the economically rational criterion was greater in the 25erc condition than in the 50erc condition. Interestingly, the calculated criterion in the 50erc condition was slightly but significantly lower than 50%. This last result was not seen in Experiment 1 where the calculated criterion did not significantly differ from 50% in the 50erc condition as expected, despite a ~~slight~~ trend in the same direction. In addition ~~However~~, the self-reported criterion ~~confirmed the full pattern observed in Experiment~~ was also not significantly different than 50% in Experiment 1. Thus, the deviation in calculated criterion in the 50erc condition observed in Experiment 2 should be interpreted with caution. Finally, as expected and observed in Experiment 1, manipulating the economically rational criterion did not affect likelihood ratings or sensitivity.

There was also evidence of a framing effect. The calculated subjective criterion in the loss frame condition was higher than that in the gain frame condition. This results in ~~leading~~ participants ~~to~~ choosing the safe option less often and showing a risk-seeking decision bias. In addition, ~~as~~ the likelihood ratings and sensitivity were not affected by manipulating the framing. Thus, 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 the task to be slightly more difficult than those in the gain condition. Increased difficulty might constitute a confound, if for some reason, participants became more risk-seeking with increased difficulty.

It is also interesting to note that the effect sizes for the two manipulations differed in the 25erc condition which is similar to real world low-probability high-impact severe weather events. Gain-loss framing shifted the calculated criterion by 4.2% in the 25erc condition while the centering effect shifted the calculated criterion 9.5% above 25%. Therefore, at the 25% economically rational criterion, the centering effect had a larger effect on the subjective criterion than the gain-loss framing effect.

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 objective probabilities ~~See~~ as shown in Figure 11. This is also consistent with experiment 1 and suggests that biased subjective likelihood ~~can in no way have~~ did not ~~contributed~~ contribute to the risk-seeking behavior observed in the loss frame.

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 conditionalized by whether there was a drought or not.)~~ in all conditions except for the 50erc gain frame condition ~~suggesting that decisions were jointly determined by subjective likelihood and subjective criterion~~. However, in the 50erc gain frame condition alone, ~~which was not tested in previous experiments,~~ the binary decisions were below the ROC curve, ~~indicating that the decisions were not solely based on subjective likelihood and subjective criterion~~. This is the only inconsistency in binary decisions observed across ~~all~~ the two 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.

This consistency suggests that the likelihood ratings and binary decisions have a common source as proposed in the rating likelihood model.

Overall, experiment 2 found effects of both the gain-loss framing effect and the centering effect on subjective criterion ~~that explain the biased criterion~~. Centering appeared to have a larger effect on the subjective criterion than the gain-loss framing when the economically rational criterion was at 25%.

General Discussion

People often make decisions that incur more risk than is normatively optimal. The research reported here used a novel signal detection theory approach to examine some of the ~~psychological~~ reasons for this phenomenon. ~~Here~~, specifically we applied ~~the~~ a random likelihood model ~~from the signal detection theory perspective~~ to this context. The ~~work reported here~~ results support~~s~~ the biased criterion hypothesis, ~~proposing~~ a shift in the subjective criterion, ~~operationalized as the likelihood of the adverse event at which a person decided to take a safe over a risky option.~~ According to this hypothesis, risk seeking is caused by ~~waiting until~~ requiring that the likelihood of the adverse event is higher than what is optimal according to the normative standard. The results were not consistent with the alternative biased subjective likelihood hypothesis, ~~that biased choices are the product of participant's biased prediction of the severe weather events, operationalized as~~ a shift in the subjective likelihood ~~was rejected~~.

The decision biases observed here are consistent with previous research using similar naturalistic decision tasks in which risk-seeking tendency was observed for loss frames and low economically rational criteria (Gulacsik et al., 2022; Qin et al., 2024). The risk-averse decisions in the gain ~~than in the loss~~ frame ~~observed here~~ were also observed in previous studies using either mixed-gamble tasks with a greater overall chance to gain than lose (Demnitz & Joslyn, 2020) or pure gain frame tasks (Tversky & Kahneman, 1979).

The research reported here is also consistent with previous research on subjective likelihood showing that people have accurate, albeit slightly overestimated, perception of the likelihood of the severe weather event when probabilistic information is provided (Gulacsik et al., 2022; Qin et al., 2024). Despite the overestimation of perceived likelihood, a similar risk-seeking tendency was also observed in the 25erc loss condition of both experiments reported here. This is curious, because all else being equal, overestimated likelihood ~~would~~ leads to risk-averse rather than risk-seeking decisions. However, the ~~novel~~ signal detection approach used here revealed that the overestimation of likelihood was counteracted by a biased subjective criterion. Participants had a ~~much~~ higher-than-rational subjective criterion ~~in loss frames~~ when the economically rational criterion was low (<50%). The higher-than-rational subjective criterion counteracted the overestimated subjective likelihood, leading to risk-seeking decisions. In summary, the findings reported here are consistent with previous studies while providing important insight into why people can have relatively accurate perception of the likelihood of the weather event while still making suboptimal decisions.

Random Likelihood Model

Signal detection theory was applied using a random likelihood model ~~with the assumption that~~ that can account for both likelihood ratings and binary decisions al~~one determined participants' decisions~~. To test ~~this assumption~~ the common account of both measures, ROC plots were calculated from likelihood ratings conditionalized by whether there was a drought or not. In 6 out of 7 cases the participants' binary decisions (proportion of hits and false alarms in decisions) were consistent with the ROC curve, validating this approach.

The analysis based on the random likelihood model makes important contributions to our understanding of risky decision making. Previous research examined only the proportion of trials in which participants chose the safe option, which documents the bias but does not explain it. This new approach allowed us to identify the subjective criterion as ~~a~~ the source of the decision bias by isolating it from ~~other factors such as~~ effect on subjective likelihood ~~and the sensitivity~~. In addition, the analysis based on the random likelihood model is independent of trial stricture allowing for comparison between studies with different trial compositions, ~~problematic~~ This is not possible when the dependent variable is the proportion of safe choices. That said, it is important to note that analyses of proportion of trials choosing the safe option can be conducted with a smaller number of trials and without measuring likelihood ratings, an advantage over the analysis of calculated criterion. Overall, the results of the experiments support the application of the random likelihood model to examine decisions in similar naturalistic tasks requiring decision making under risk.

The Centering Effect

Another important finding was that the subjective criterion was affected by the centering effect (Poulton, 1979, Olkkonen et al., 2014). In the new experiments reported here, the economically rational criterion was manipulated at, above, and below 50% in Experiments 1 and at or below 50% in experiment 2. A centering effect was observed in both, showing a shift in the subjective criterion toward 50% when the economically rational criterion was below (25%) or above (75%) the center of the range. In addition, the subjective criterion stayed comparatively close to 50% when the economically rational criterion was 50%. ~~However,~~ The subjective likelihood and sensitivity were unaffected by this manipulation.

It ~~appears~~ is possible that the centering effect did not shift the subjective criterion to the precise center of the range (50%). Instead, two of three measurements suggest that the shift was towards someplace lower than 50%. In Experiment 1, the calculated criterion in the 75erc condition shifted downward to a greater extent than the calculated criterion shifted upward in the 25erc condition. In Experiment 2, the calculated criteria in both the gain and loss frame 50erc conditions were significantly lower than 50%, at 43% and 44% respectively. One possible explanation is that participants' preconceived notion of severe weather events was that the potential harm ~~should be~~ is much greater than the cost of protection to take action. Therefore, they adopted a criterion that was slightly lower than 50%.

A potential mechanism for the centering effect is anchoring and adjustment (Kahneman, 2003). When people estimate an unknown quantity, they sometimes start with a number recently encountered and then adjust away from that anchor to make their estimate. The adjustment is often insufficient. In the case of the centering effect observed in the experiments reported here, it is possible that the center of the range of possible probabilities (50%) was ~~used to estimate subjective criterion and became~~ an anchor from which adjustment was insufficient. Alternatively, it ~~may~~ might be the participants intentionally shifted the subjective criterion toward the center of the likelihood range to avoid extreme decisions. This is similar to the compromise effect observed in decision making (Simonson, 1989).

~~However,~~ Centering did not appear to affect subjective likelihood in the reanalyses or new experiments reported here. Perceptual studies tend to suggest that when the information regarding the correct response in question is available, the centering effect is reduced (Radvansky et al., 1995). In the current experiments, the available probabilistic information ~~may~~ might explain the absence of a centering effect in subjective likelihood. Indeed, in the deterministic conditions of some previous studies ~~in which objective probabilities were not provided on each trial~~, effects similar to centering were observed such that the mean subjective likelihood estimate was close to 50% (Demnitz & Joslyn, 2020; Qin et al., 2024). However, when probabilities were provided ~~here and in the probabilistic conditions of the aforementioned studies,~~ the centering effect on subjective likelihood disappeared.

The Gain-Loss Framing Effect

In Experiment 2, framing was manipulated. In line with the predictions of prospect theory, the subjective criterion was higher, that is, more risk-seeking, in a loss compared to a gain frame. It should be noted that participants were also risk-seeking in the gain frame with 25% economically rational criterion, just less so than in the loss frame. A similar set of patterns was observed in the shift in self-reported criterion and the proportion of trials choosing the safe option confirming this analysis. Again, subjective likelihood and sensitivity were unaffected. 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 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.

However, the effect size of the gain-loss framing in this study was small (only 3% difference at 25% erc) in comparison to classic experiments in prospect theory (e.g., Tversky & Kahneman, 1979). Two potential accounts might explain this. First, the multi-trial nature of the experiments reported here might have led participants to consider the big picture of all trials as opposed to one trial at a time (narrow framing) when making decisions. When considering a series of risky decisions together as opposed to each trial independently, people may have been less susceptible to the effects of gain-loss framing and better able to consider the long term strategies that underly an appreciation of expected value (Thaler. 1999). An alternative explanation is that in classic prospect theory experiments, the expected values of the safe and risky options were the same or similar (e.g., the disease problem in Tversky & Kahneman, 1981). In the experiments reported here, the expected value of the risky option varied trial by trial and might differ from that of the safe option drastically. ~~Therefore,~~ I ~~it~~ was possible that gain-loss framing became less salient when the expected values differ which ~~between the safe and risky option~~ by ~~a large extent~~. This might be the case if participants have some appreciation of expected value. If so, they would ~~may~~ be less swayed by framing when the expected value of the two options differs greatly. ~~Using the disease problem (Tversky & Kahneman, 1981) as an example, participants might overwhelmingly choose the safe option regardless of the frame if the safe option is changed to saying 550 people/losing 50 people while the risky option remains unchanged, as the former now has much superior expected value.~~

Potential Behavioral Interventions

The current experiments revealed that in situations in which the economically rational criterion is low, people tend to be risk-seeking due to a higher-than-rational subjective criterion regardless of the frame. This aligns with an abundance of previous research (Baker, 1995; Joslyn & LeClerc, 2013; Atreya et al., 2015; LeClerc & Joslyn, 2015; Qin et al., 2024). In real life situations, a response is often required for adverse weather events that have a low probability of occurrence and a high potential for casualties and severe damage. This translates into a low objective probability and a low economically rational criterion. Therefore, one important application of the findings of the current experiments is to design behavioral interventions to reduce the risk-seeking decision bias based on the mechanisms revealed here, a shift of the subjective criterion due to centering and framing. Changing the gain-loss framing is a conceptually straight forward intervention, even though switching to a gain frame might not completely eliminate the risk-seeking tendency. Framing the consequences of the event in a gain frame should shift the subjective criterion closer to the economically rational criterion ~~when the stake is low~~ as shown in the current experiments. However, while some events can be relatively easily framed as a gain (e.g., the 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).

~~On the other hand,~~ Interventions based on centering ~~could~~ might also be implemented. One such potential intervention is to provide the economically rational criterion (e.g., "you should choose the safe option when the probability was higher than 25%"). Unfortunately, previous tests of this hypothesis were not promising (Joslyn & Leclerc, 2012). However, in ~~this~~ that case no probabilistic information was provided. 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 would make better decisions with both a subjective criterion close to the economically rational criterion and accurate subjective likelihood. Alternatively, it might help to simply bring attention to the criterion by asking participants to estimate the economically rational probability in situations where such estimation is possible, as has been seen in similar biases (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 number as the circumstance is different from person to person. In short, direct knowledge of the economically rational criterion might be a way to reduce the centering effect on the subjective criterion, ~~although there are a number of issues that would have to be addressed.~~

Another possible intervention is to communicate the precise numeric probabilistic information on a smaller and lower range of probabilities, perhaps through forecast visualizations. For example, the forecast could provide a visualized ruler showing a range of probabilities from 0% to 50% and indicating a specific point as the likelihood of a given weather event, e.g. 15%. This may cause users to constrict their perceived range of possible likelihoods to 0%-50%, making 25% the center toward which estimates would shift.

Limitations and Future Directions

One limitation of the current project is the limited examination of the connection between the subjective criterion and expected utility which is the cornerstone of research in decision making under risk (Von Neumann & Morgenstern, 1947; Tversky & Kahneman, 1979; Weber, 1994). According to prospect theory the economically rational criterion is determined by the expected [redacted] utility of options. Similarly, subjective criterion ~~might~~ can be determined in part by the utility of options. While the difference in subjective criterion due to gain-loss framing found here is consistent with the influence of utility, it is not conclusive as this study was not designed to evaluate how [redacted] expected utility differed among conditions. ~~Therefore, expected utility might be one determinant of the subjective criterion.~~ This random likelihood model application of signal detection theory in conjunction with prospect theory to the study of decision making may be a profitable new direction for future research.

Next, while the current experiments provided a case for the usefulness of ~~applying signal~~ [redacted] the random likelihood model ~~in naturalistic tasks requiring decision~~

~~making under risk~~, the model operates on the assumption that the subjective criterion has no variability and is fixed across decisions. This is unlikely to be the case in real world situations. Indeed, a previous study found that after recent encounters with adverse weather events, decisions became more risk-averse while the subjective likelihood remain unchanged (Demnitz & Joslyn, 2020). This suggests that the subjective criterion can change between decisions based on recent trial experience. Recent outcomes might not be the only reason for changes in the subjective criterion. The objective probability might serve as an anchor that shifts the subjective criterion from one trial to the next. The current experiments could not examine these possible effects because there were not enough trials to conduct within-group analysis at different objective probability levels. A similar experiment with a higher number of trials or a different experimental paradigm could be used to explore this issue.

Finally, another future direction is to incorporate the perceived efficacy of protective actions into the model as another variable that can have stochastic distributions. In experiments ~~works~~ like the study reported here, protective actions had perfect efficacy. ~~however~~ This is rarely the case in real world situations. According to protection motivation theory, in addition to one's risk perception (probability and severity of an adverse event), the efficacy of protective actions is also critical to their uptake (Rogers, 1975). In hurricane and flood situations, for instance, perceived efficacy was linked to compliance with evacuation orders (Demuth et al., 2016; Liu et al., 2024). When the efficacy of protective actions is not perfect, people's perceived efficacy might vary, potentially following a stochastic distribution. Perceived efficacy of the safe option might in turn affect the placement of the subjective criterion and as a result, people's decisions. Therefore, it is important to take perceived efficacy into account for a more comprehensive understanding of how people make decisions under risk in anticipation of severe weather events.

Constrains on Generality

The new experiments and the reanalyses used an online US nationwide sample that was fairly representative in terms of age, gender, race. The new experiments did not measure education level. However, since the sample was recruited from the same pool (Prolific), on a similar topic and structure, and with other demographic characteristics similar to the studies used in the reanalyses, the education level can be assumed to be ~~similar, which~~ was on average more educated than a representative sample (see Qin et al., 2024). For that reason, the results might not extend to those with ~~very~~ low levels of education. nevertheless, As the sample was fairly representative overall, we consider it suitable for generalizing the findings to the US population.

The stimuli were chosen to represent a fairly common severe weather/~~climate~~ event, ~~across the world with the primary goal to examine the underlying cognitive processes~~. While personal experience with droughts might affect participants' overall risk tolerance and risk perception, we do not believe it would systematically confound our independent variables. Therefore we do not believe the conclusions reported here depend on any other participant, material, or context factors.

Conclusion

This study ~~reported here~~ used a random likelihood model based on a ~~novel~~ signal detection theory ~~approach~~ to better understand the risk-seeking tendency in naturalistic weather decisions. This ~~method~~ approach distinguishes between the influence of subjective criterion and subjective likelihood on decision bias, ~~suggesting that~~ In these experiments, the risk-seeking bias is due primarily to the subjective criterion.

centered heading