JP Comments 5/21/25

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

Chao Qin, Susan Joslyn, John Palmer

University of Washington

Create Separate Title Page

Abstract

Severe weathers require protective action even at low probability. In naturalistic decision experiments based on these situations, people often ~~did~~ not take protective action when it is economically rational to do so (risk-seeking). This study applied a signal detection theory perspective to this setting: A person might make risk-seeking decisions due to an overly high criterion for taking protective action or overly low perceived likelihood of the event. In two experiments in 2024, the economically rational criterion and gain-loss framing were manipulated. According to an analysis based on signal detection theory, when the economically rational criterion was manipulated, the subjective criterion was between the economically rational criterion and the center of the range. When gain-loss framing was manipulated, the subjective criterion was higher in a loss than a gain frame. Participants also overestimated subjective likelihood ~~that was unaffected by the manipulations~~. However, the shifted subjective criterion overcame this overestimation, resulting in risk-seeking decisions. Thus, we conclude that the shift of the subjective criterion can lead to risk-seeking decisions in naturalistic decision tasks. The sample was representative in age, gender, and race, suggesting generalizability to the US population. Potential interventions to improve the placement of the subjective criterion were discussed.

What about acknowledgments?
What about funding?

page #s?

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

put on title page

Public Significance Statement

This research demonstrates that in addition to people's perceived likelihood of a severe weather event, their subjective criterion, the perceived likelihood above which they will take protective action, is another factor that can be biased. In the situation where the probability of the severe weather is low but the severity is high, people tend to take more risk than warranted, not because they perceive an inaccurate likelihood of the event, but because they have a high subjective criterion. The findings here provide new avenues for designing behavioral interventions to increase the uptake of protective actions against severe weather events.

This needs revision to make more clear. — Susan?

Should all headings be bold?

Start page #s from beginning

Introduction

In risky decision making about real world events such as severe weather events, people often need to make decisions for an uncertain future. For example, when facing a possible tornado, people must decide whether to take protective actions such as taking shelter. Taking protective action costs time and resources but can protect them from harm. On the other hand, not taking protective actions might expose decision-makers to potential harm but can save time and resources if a tornado does not materialize. Due to the potential serious harm of these events, people are advised to take protective actions even when the probability is low. For example, tornado warnings are issued by the US National Weather Service when the probability of a tornado is as low as 10% (Qin et al., 2024). Due to the low probability, the severe weather event often fails to occur at the residents' location, making protective actions seem like a waste of time. Indeed, research on people's response to real forecasts and warnings in severe weather events, such as floods, tornadoes, and hurricanes, showed that people often failed to comply with the recommended protective actions (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). For example, a study found that in 2004, the Charlotte County in the Southwestern Florida, hit by three hurricanes and suffering heavy damage from Hurricane Charley, had a low proportion of residents evaluating at least once (36%; Smith & McCarty, 2009).

One reason contributing to not take protective actions in severe weather events might be that people do not receive probabilistic information (e.g., there is a 40% chance of tornado) of the event from the forecast which may reduce trust. Indeed, many forecasts do not provide people with an estimated probability of the event (deterministic forecasts), although such experiments were conducted to examine the biased criterion hypothesis and the biased likelihood hypothesis.

Reanalysis of Previous Experiments

Transparency and Openness

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

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 tornados warning 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 was 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 the tornado would be (likelihood ratings), and decided between the safe option and the risky option (binary decision). At the end of each trial, they were told whether the tornado occurred. As mentioned above, there was a point cost to shelter and a larger point penalty if participant 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. This cost was 90 points in experiment 1 and 270 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 experiment 1 the overall 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 to examine the three possible ways the decisions could be affected: Mean calculated subjective criterion for each experiment, mean likelihood ratings, and Receiver Operating Characteristic (ROC) plots constructed from likelihood ratings and decisions. ROC plots are used in signal detection theory to indicate the sensitivity of participants for predicting a tornado. In this case, the greater the area under the curve the greater the sensitivity (see Figure 2). The ROC curve also indicated how participants' hit and false alarm probability changed given different subjective criterion. Inferential statistics in the form of t-tests were conducted with an alpha of .05.

Calculated Subjective Criterion

The mean calculated subjective criterion was estimated using each participant's likelihood ratings and frequency of choosing the safe option in each experiment. Its calculation was based on the assumption that participants chose the safe option whenever their subjective likelihood was higher than their subjective criterion in accordance with the random likelihood model. For each participant, the cumulative proportion of likelihood ratings falling between X% chance and 100% was calculated such that the proportion of trials in between matched the proportion of trials on which the participant chose the safe option. This point was called the *calculated subjective criterion* because if the participant always chose the safe option when their likelihood rating was above this point, the proportion of trials with a likelihood rating higher than that would be the same as the observed proportion of trials in which they chose the safe option. 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 on 50% of the trials had a likelihood rating higher than this number (50% of the blue area from the right Figure 3). This number, say 40% likelihood rating (the red line in Figure 3), was regarded as the calculated subjective criterion for this participant. With this method, a calculated subjective criterion was obtained for each participant, and a mean was calculated for each experiment.

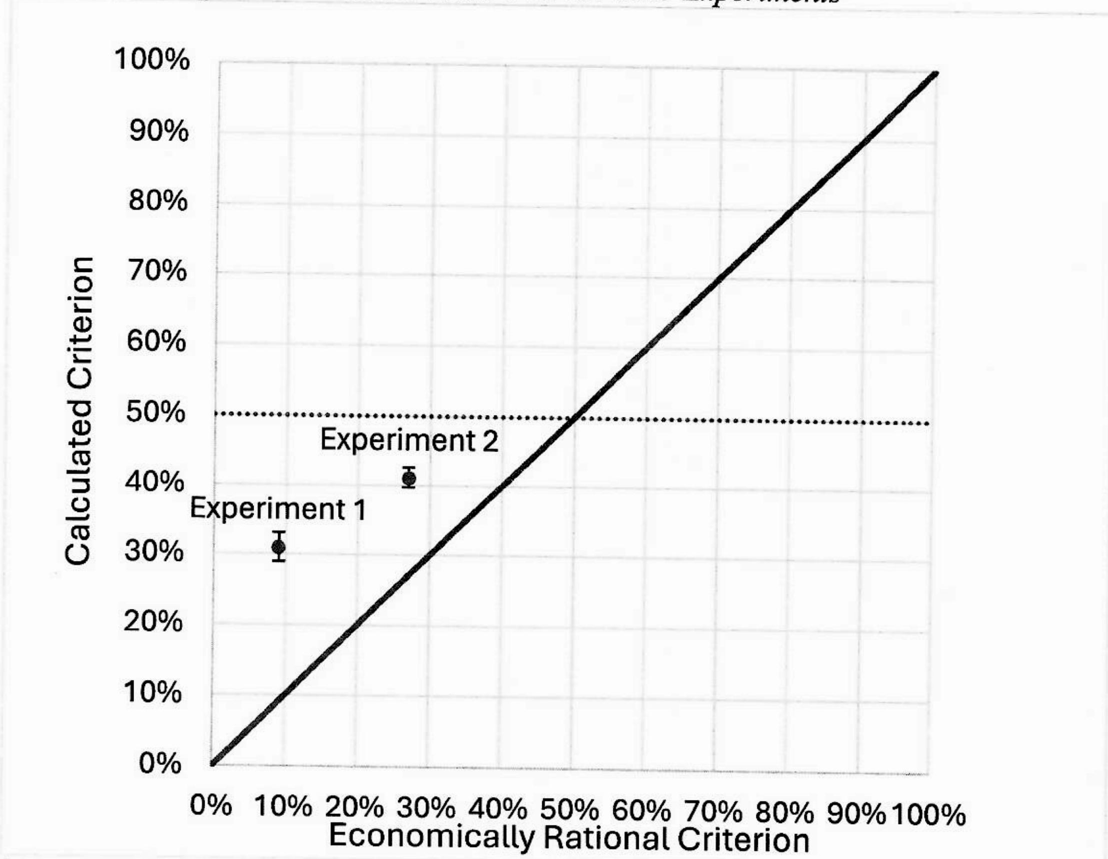
In Figure 4, the calculated criterion is shown for tornado experiments 1 and 2 as the blue dots in relation to the economically rational criterion in these two experiments. An independent t-test revealed that the mean calculated criterion in tornado experiment 1 (M = 31.0%, SD = 18.6%) was significantly 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 different economically rational criterion. In addition, the mean calculated criterion in each tornado experiment was compared to the respective economically rational criterion in two one-sample t-tests. In tornado experiment 1, the calculated criterion was significantly higher than the economically rational criterion of 9% with a difference of 22% (t(82) = 10.77, p < .001). In tornado experiment 2, the calculated criterion was significantly higher than the economically rational criterion of 27% with a difference of 14.2% (t(84) = 9.81, p < .001). The calculated criterion was higher than the economically rational criterion in both experiments, suggesting a biased subjective criterion, and more so in the experiment with the lower economically rational criterion.

Likelihood Ratings

Figure 5 shows the likelihood ratings for tornado experiments 1 and 2 as a function of objective probabilities. The mean likelihood rating collapsed over all objective probability levels in each experiment was compared to the proportion of trials in which a tornado occurred in that experiment (equivalent to the mean objective probability) with two one-sample t-tests. In tornado experiment 1, the mean likelihood rating (M = 33.7%, SD = 10.0%) was significantly higher than the proportion of tornado trials of 23.5% with a difference of 10.2% (t(82) = 9.33, p < .001). In tornado experiment 2, the mean likelihood rating (M = 42.7%, SD = 7.5%) was 4.5% higher than the proportion of tornado trials of 38.2% (t(84) = 5.48, p < .001). Therefore, the mean likelihood ratings were higher than the proportion of tornado trials in both experiments. In addition, Figure 5 plotting mean likelihood rating by each objective probability, shows that the overestimation was observed at most objective probability levels except for the extreme high end. Finally, as the mean objective probability was different between the two experiments (due to different trial compositions), the comparison of likelihood ratings between the two experiments cannot reveal whether there was an effect of economically rational criterion on subjective likelihood.

Figure 4

Calculated Subjective Criterion in the Two Tornado Experiments

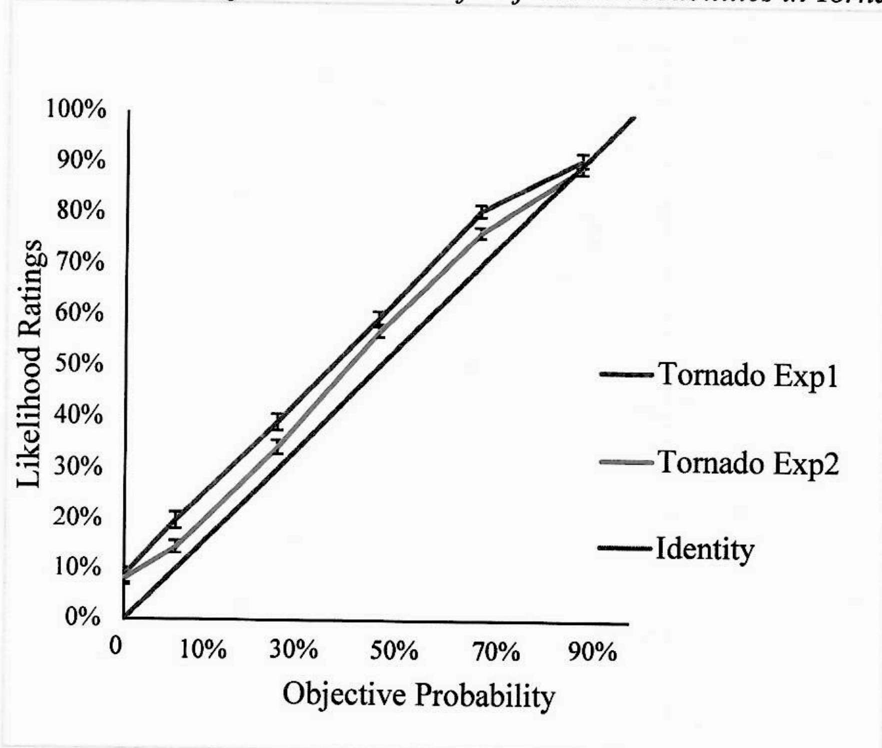


Note. The y axis represents the calculated criterion while the x axis represents the economically rational criterion. Diagonal line represents when the calculated criterion is the same as the economically rational criterion. The dashed line represents 50%.

a calculated criterion of

Figure 5

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



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.

Receiver Operating Characteristic Plots

The ROC analysis had two aims: 1) Testing the assumption of the random likelihood model that subjective likelihood and the subjective criterion jointly determine the decision, 2) Examining whether the sensitivity (participants' ability to predict the tornado as revealed by their binary decisions) was different between experiments. The ROC plots were composed of two parts: 1) Receiver Operating Characteristic curves created from participants' likelihood ratings and actual tornado occurrence; 2) Points on the plots representing the outcomes of participants' binary decisions (safe option/risky option). For a similar approach see Harvey et al. (2012)

The ROC curves were created by estimating hit and false alarm rates based on the likelihood rating distribution, the actual tornado occurrence, and a varying hypothetical cut-off at 5% intervals from 0% to 100% on the distribution, using a method similar to Ferrel & McGoey (1980). At each point, 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 at 5% steps from 0% to 100% on the likelihood rating distribution, a pair of hit and the false alarm probabilities at each step (20 in total) was calculated. The 20 pairs were plotted as the ROC curve (the curve in Figure 6).

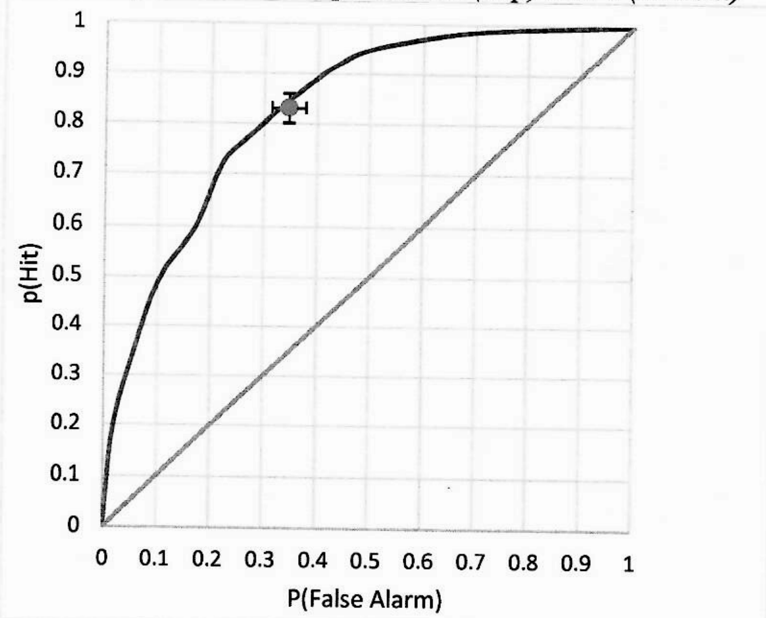
Next, a point representing the mean proportion of hits and false alarms of participants' actual binary decisions was added to the ROC plot. A hit was when the participant chose the safe option and the tornado occurred. A false alarm was when the participant chose the safe option and no tornado occurred.

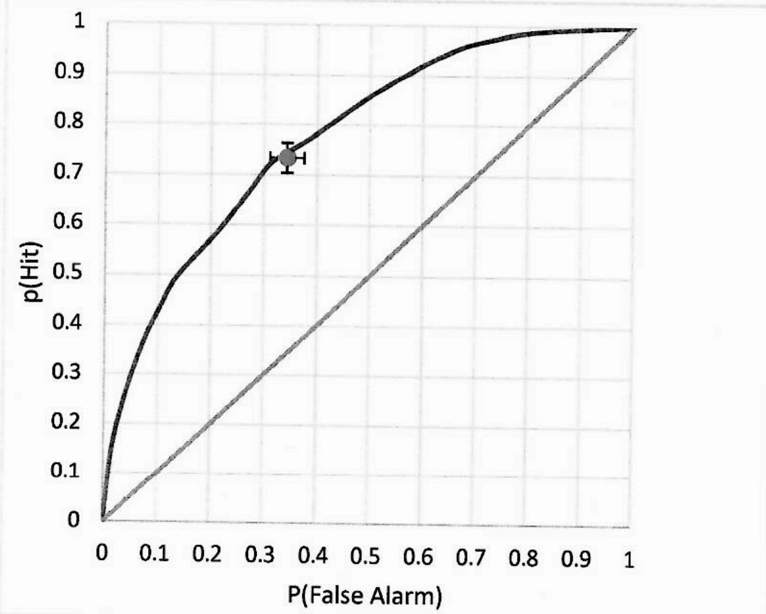
The first goal of the ROC analysis was to test the random likelihood model method of calculating the subjective criterion. If participants chose the safe option whenever their likelihood rating was above their subjective criterion, the hit and false alarm probabilities of the ROC curve from their likelihood 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% ~~CI~~ confidence interval of the decision point overlapped. This indicates that the decision point was consistent with the ROC curve.

To address the second~~ary~~ goal of the ROC plot analysis, the mean percent area under the ROC curve was ~~measured~~ calculated. The greater the area under the curve the greater the sensitivity. An ~~independent~~ one sample t-test revealed that the mean percent area under the ROC curve was significantly lower in tornado experiment 1 (M = 64.0%, SD = 6.8%) than in tornado experiment 2 (M = 78.1%, SD = 6.8%) with a difference of -14.1% ($t(165.86) = 13.41$, $p < .001$). This indicates a worse sensitivity in experiment 1 than experiment 2.

Figure 6

ROC Plot 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.

Discussion

The reanalysis of the tornado experiments indicates that 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. This result is also consistent with the centering effect. Because the tornado experiments did not have a gain frame (subjective criterion expected to be higher for gain than loss) condition 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 the likelihood was significantly over rather than under estimated in both experiments. Subjective likelihood ratings were also not consistent with a centering effect as there were similar amounts of overestimation across all objective probability levels except for 90%, which was less. The higher-than-rational subjective criterion counteracted the overestimated likelihood ratings and led to risk-seeking decisions instead. 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.

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

[Point of this sentence not clear.]

(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% CI of the binary decision point overlapped with the ROC curve. This indicates that the proportion of hits and false alarms of participants' actual binary decisions was consistent with the ROC curve based on likelihood ratings and actual tornado occurrence. This in turn suggests that the subjective criterion and subjective likelihood (likelihood ratings) were the sole determinants of the decisions, 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 in the two tornado experiments. It left the door open to a possible centering effect and a gain-loss framing effect on the subjective criterion. The difference in trial composition also led to confounds in several analyses.

The two new experiments reported below employed a drought preparation task based on Demnitz & Joslyn (2020), using both a gain and a loss frame[2]. The goal was to distinguish the cause of the bias in subjective criterion and rule out the effect of subjective likelihood and/or a difference in sensitivity with experiments with the same trial compositions.

Experiment 1

Experiment 1 focused on the centering effect on subjective criterion by manipulating the economically rational criterion at, below and above 50%. If centering is present subjective criterion should not shift in the first case, shift up in the second and down in the third. By the

[2] 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 [redacted]r.

biased criterion hypothesis, this manipulation was expected to affect only the subjective criterion and not the subjective likelihood or sensitivity.

Transparency and Openness

Hypotheses, experimental design, procedure, method, elimination criteria and data analysis plans were preregistered at the Open Science Framework at https://doi.org/10.17605/OSF.IO/RUFD8 on Jan. 21th, 2024. Registration was after data collection but before data analyses. Data is available at the Open Science Framework at https://osf.io/bspqn/[redacted]. Analytic codes, and additional materials for experiment 1 are available upon request.

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

After providing informed consent, participants were provided instructions to a task in which they decided which crop to plant based on climate projections concerning possible droughts. This task is henceforth called the *drought task*. See Supplementary Materials S1 for illustrations of the drought task, including the instructions shown to the participants and the questions they answered. In the task, participants played the role of an agricultural consultant who advised farmers on whether to plant a drought resistant crop (safe option) or a regular crop (risky option). The outcomes of their decisions were tracked with a point structure (see the section below). Their goal was to have as many points as possible by the end of the experiment. They were paid a monetary bonus commensurate with their point balance at the end of the experiment.

At the beginning of the task, participants read background information on the threat of drought to farmers' crops and their own role in the task. Participants were told that the potential drought might incur a loss, compared to regular, non-drought seasons. They achieved their goal by losing as few points as possible. They were then introduced to the point structure of the task and then performed a practice trial.

Point Structure. In order to simulate real life decisions with consequences, and to encourage participants to put forth their best effort, the point structure, mentioned above, was implemented in which participants lost points based on their decision (See Table 3). The economically rational criterion was manipulated between groups. There were three economically rational criterion conditions (erc): 25% (25erc), 50% (50erc), and 75% (75erc). In all conditions the regular crop (risky option) provided a loss of 0 points if there was no drought and a loss of 400 points if there was a drought. In the 25erc, 50erc, and 75erc condition, the drought resistant crop (safe option) provided a sure loss of 100, 200 and 300 points respectively (See table 3). All conditions had the same starting balance of 20,000 points. The bonus payment structure was set up to make it roughly equivalent in the three conditions. In the 25erc condition, participants were paid $1 for every 1,000 points in their balance above a payment threshold of 15,000 points. In the 50erc condition, participants were paid $1 for every 2,000 points in their balance above 10,000 points. In the 75erc condition, participants were paid $1 for every 3,000 points in their balance above 5000 points. This payment threshold was set up to prevent participants from taking the simplistic approach of choosing the safe option in every single trial. For example, in the 25erc condition, if they chose the safe option in all 50 trials, they would end up with 15,000 points (20,000 – 50 * 100).

Trial Structure. After reading through the background information, point structure, and going through a practice trial, and two attention checks, participants began the 50 experimental trials. In each trial, participants saw three screens. On the first screen, participants saw a forecast which described the probability that drought would occur (e.g., The latest climate forecast indicates a 35% chance of drought in the upcoming season for farmer-client 1). This percent chance will be referred to as the objective probability. It was calibrated to be roughly reliable. Participants then moved the slider on a visual analog scale (VAS) with anchor points *impossible* and *certain* to answer the question "Move the marker to indicate what you think the likelihood of a drought is" (likelihood rating). On the second screen, participants saw the same forecast and

Post-Task Questions. After completing all 50 trials, a summary of participants' decisions and their outcomes across the trials was shown along with several questions. Participants were first asked a comprehension check question. They then indicated their (self-reported) criterion of their decision. This was an alternative way to the calculated criterion of getting the subjective criterion. This self-reported criterion was asked at the end of the experiment and thus reflected the decision criterion based on experience from the 50 trials, similar to a previous study (Joslyn & Grounds, 2015). This self-reported criterion was also conscious. As such, it might not match participants' actual decision-making process which might also be affected by unconscious processes. Next, participants rated how difficult the task and an open-ended question asking them which part of the experiment they found difficult to understand. These two questions were meant to check whether the descriptions of the goal in the gain or loss frame conditions had a different difficulty which would have been a confound. Participants reported no difference in the difficulty of understanding the two conditions. Finally, participants reported any problem in the experimental program. No bugs or glitches were reported. After completion of all questions, participants were thanked and provided with a unique completion code to enter into Prolific to verify their participation and receive payment.

Design

Experiment 1 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 six levels: 20%, 35%, 50%, 65% and 80%. Other parameters were fixed: The frame was the loss frame from experiment 1 and the proportion of drought was 36% across 50 trials. As the objective probabilities were reliable (M = 36.5%), the mean objective probability and the proportion of drought trials were considered the same (36%) in the analyses.

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

Results

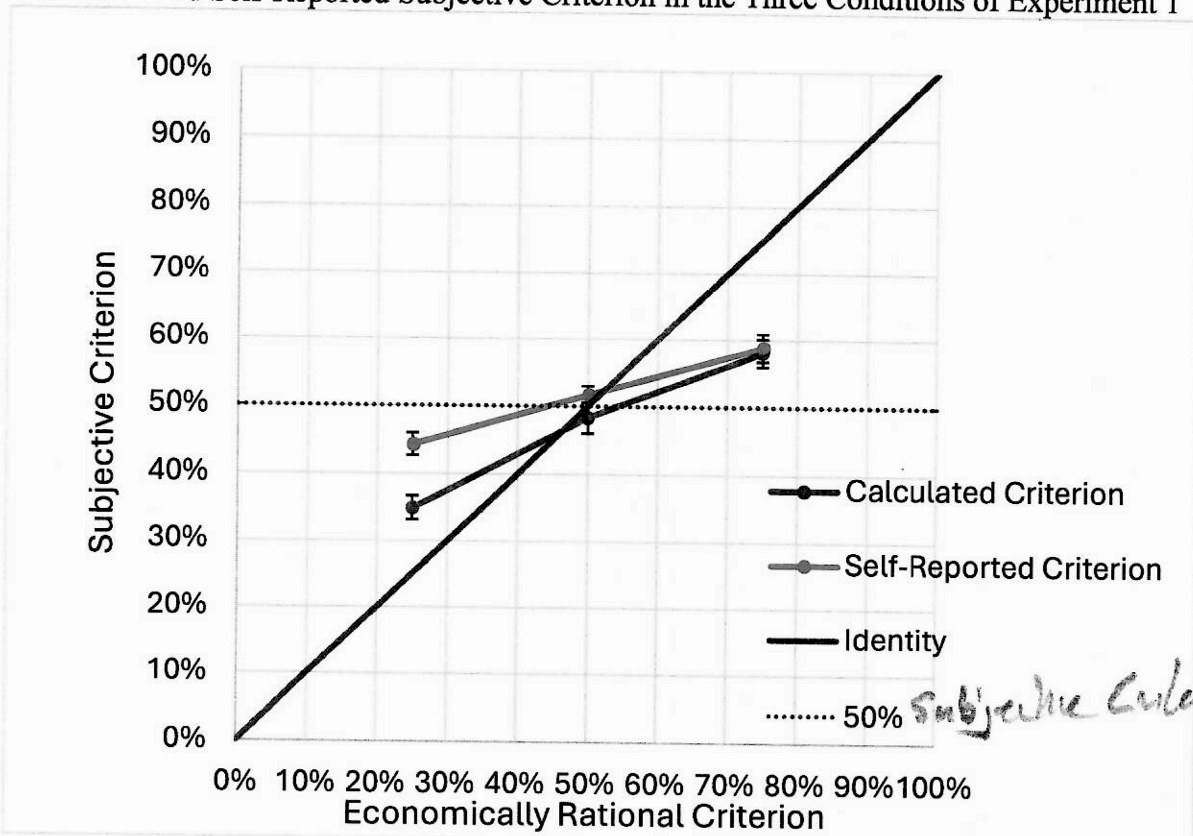
Analysis Overview

If there was a centering effect, the mean subjective criterion should shift 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%. By the biased criterion hypothesis, the manipulation of the economically rational criterion should not affect subjective likelihood or sensitivity.

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

For all dependent variables, 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.

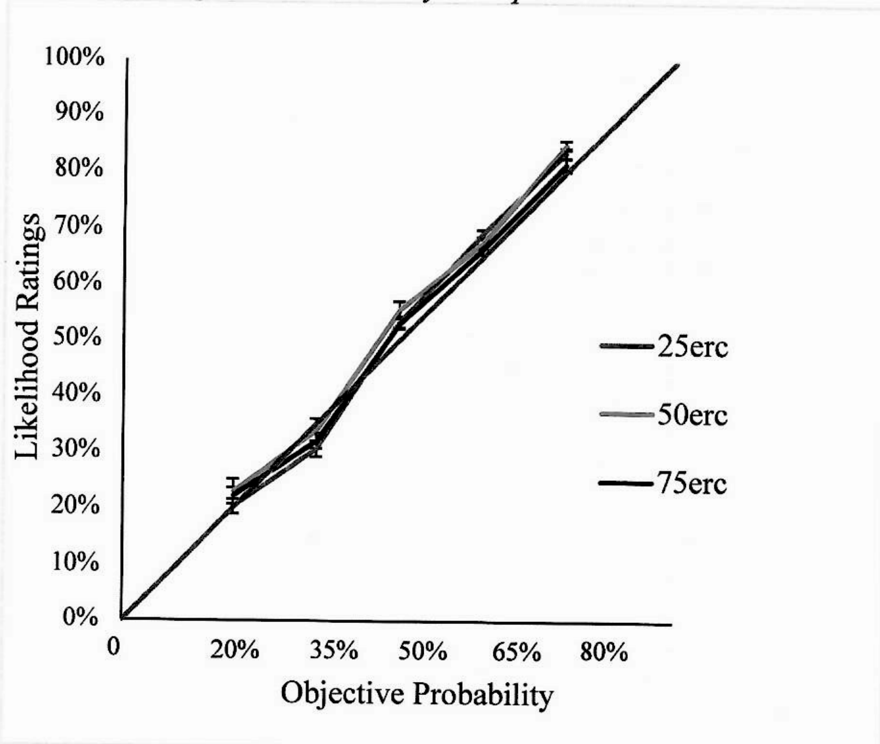
Figure 7
Calculated and Self-Reported Subjective Criterion in the Three Conditions of Experiment 1



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.

Figure 8

Likelihood Ratings ~~in the 25%, 50%, and 75% Economically Rational Criterion Conditio~~ns 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.

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 with a larger sample size than experiment 1, with a secondary goal to examine whether there was an interaction between two. Experiment 2 used the same task as experiment 1. The economically rational criterion was manipulated between groups using the 50erc and the 25erc but not the 75erc condition. Excluding the less realistic 75erc condition maximized the number of participants in each condition. Severe weather events usually require people to take protective action at a low probability which corresponds to a low economically rational criterion. Both a gain frame and a loss frame were used to examine the gain-loss framing effect.

Transparency and Openness

Hypotheses, experimental design, procedure, method, elimination criteria and data analysis plans were preregistered at the Open Science Framework at https://doi.org/10.17605/OSF.IO/V368N on March. 29th, 2024. Registration was after data collection but before data analyses. Data is available at the Open Science Framework at https://osf.io/sv57z/?view_only=116867f6874341cd85e2cc18b6586664. Analytic codes, and additional materials for experiment 2 are available upon request.

This is only needed once

Method

Participants

A total of 591 participants from the US were recruited from Prolific Academic in March 2024, a crowdsourcing platform for online research. After the elimination process, 558 participants were used in the analysis. Eleven were eliminated for having a lower than 0.7 ReCAPTCHA score, a bot detection system used by Qualtrics survey platform. Twenty-two were eliminated for failing the comprehension check (same question as experiment 1). 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<%) where Prolific did not provide data. There were 62 Asians (11%), 73 African Americans (13%), 359 Whites (64%), 37 mixed (7%), 24 other (4%) and 3 (1%) where Prolific had no data. This experiment was approved by the University of Washington Human Subject Division.

Procedure and Stimuli

The procedure was identical to experiment 1with a drought decision task, although there were additional conditions and corresponding point structures. A gain frame condition was added that was equivalent to the loss frame, 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). Participants goal was to have as many points as possible by the end of the experiment. They were paid a monetary bonus commensurate with their ending point balance.

Point Structure. ~~In Experiment 2, wit~~h both a gain frame and a loss frame, ~~there were also~~ two economically rational criterion levels, 25% (25erc), and 50% (50erc). These two v~~ariables~~

were fully crossed creating four between groups conditions. The point structure in the 25erc and 50erc loss frame conditions were identical to those of experiment 1 while the point structure in the gain frame conditions had the equivalent risk to their loss frame counterparts (See table 4). The same beginning point balance and payment structure of the 25erc and 50erc conditions of experiment 1 were used.

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

Design

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

In terms of dependent variables, as with experiment 1, participants made likelihood ratings using a VAS and binary crop decisions on each trial. At the end of the experiment, participants answered a comprehension check question, reported their criterion, difficulty of understanding the task, and any glitches they encountered as with experiment 1. Unlike experiment 1 where no difference in the difficulty between conditions was reported, an ANOVA with gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc) manipulations as the independent variables, on the difficulty ratings revealed that participants in experiment 2 reported the loss frame condition (M = 14.2, SD = 22.4, range 0 to 100) to be slightly more difficult to understand than the gain frame condition (M = 10.5, SD = 16.6, $F(1, 554) = 4.94$, $p = .027$).

Results

Analysis Overview

Based on results from experiment 1 and the biased criterion hypothesis, the prediction was that there should be both a centering effect and in line with the tenants of prospect theory, a gain-loss framing effect on the subjective criterion. As with experiment 1, the centering effect should shift the subjective criterion in the 25erc condition towards 50% while having no effect on the subjective criterion in the 50erc condition. Based on the utility function of prospect theory (Tversky & Kahneman, 1979), the loss frame should lead to a higher subjective criterion than the gain frame. Moreover, no interaction between these two effects was expected as their mechanisms ~~should be~~ are theoretically independent. Neither manipulation was expected to affect likelihood ratings or sensitivity.

The same set of dependent measures was used. To test the biased criterion hypothesis, the calculated criterion was analyzed. Next, likelihood ratings were analyzed to detect biases in subjective likelihood. Finally, to measure the sensitivity, the area under the ROC curve was analyzed.

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 ($F(1, 554) = 1.03$, $p = .31$).

Figure 10

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

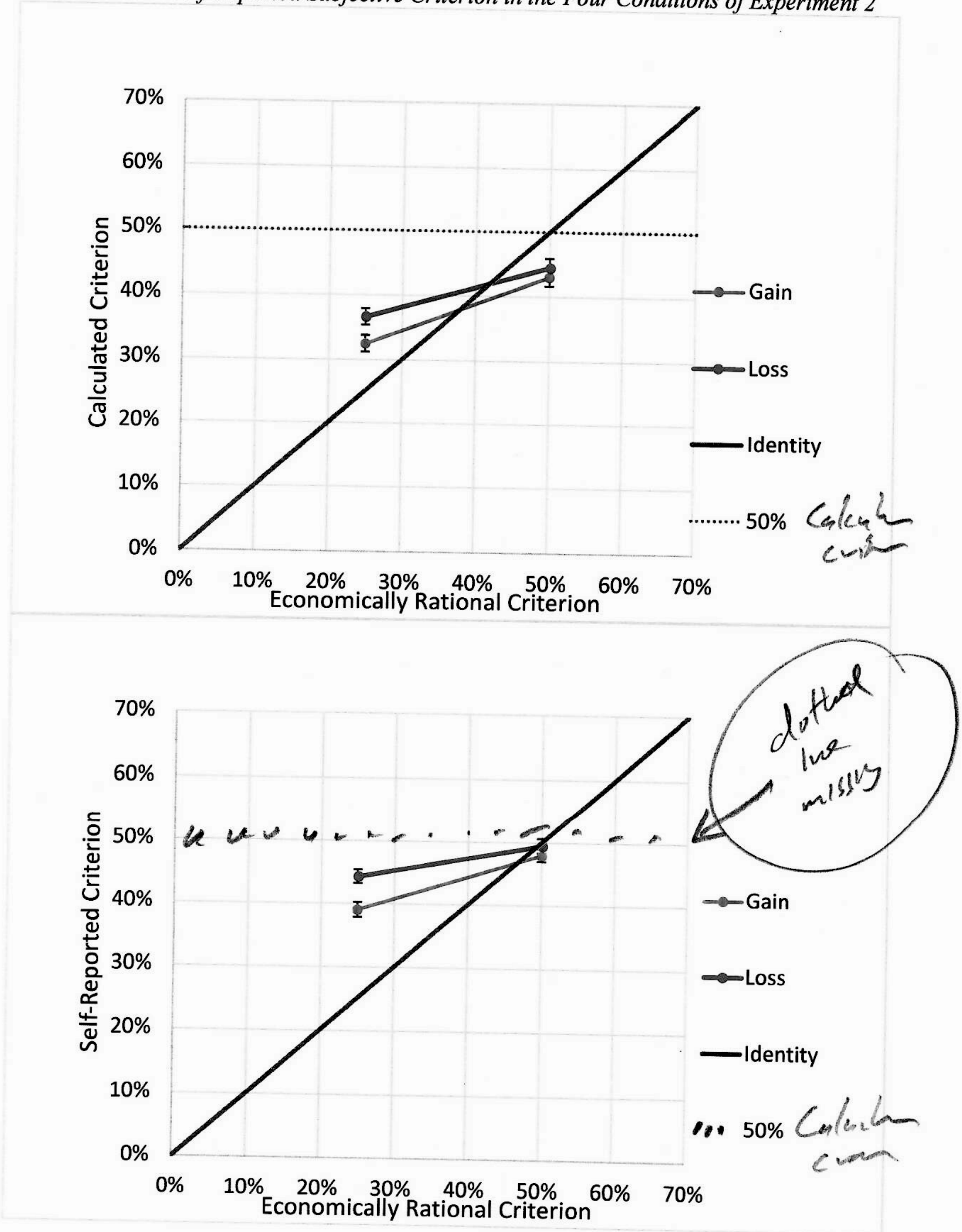
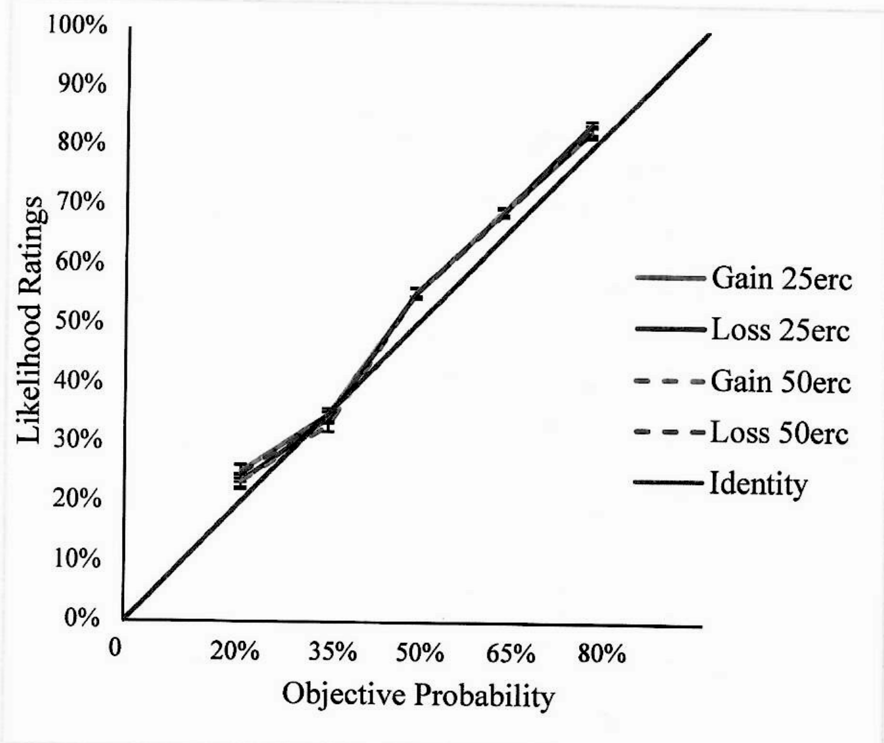


Figure 11

Likelihood Ratings ~~in the Four Conditions~~ as 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.

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 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 former~~.

the subjective criterion.

References

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

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

Fix Font

Baker, E. J. (1991). Hurricane evacuation behavior. International Journal of Mass *Emergencies and Disasters, 9*, 287–310.

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

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

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

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

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432–459.

Check Fonts throughout references.

25	20	0	50	80	0

Cut this page