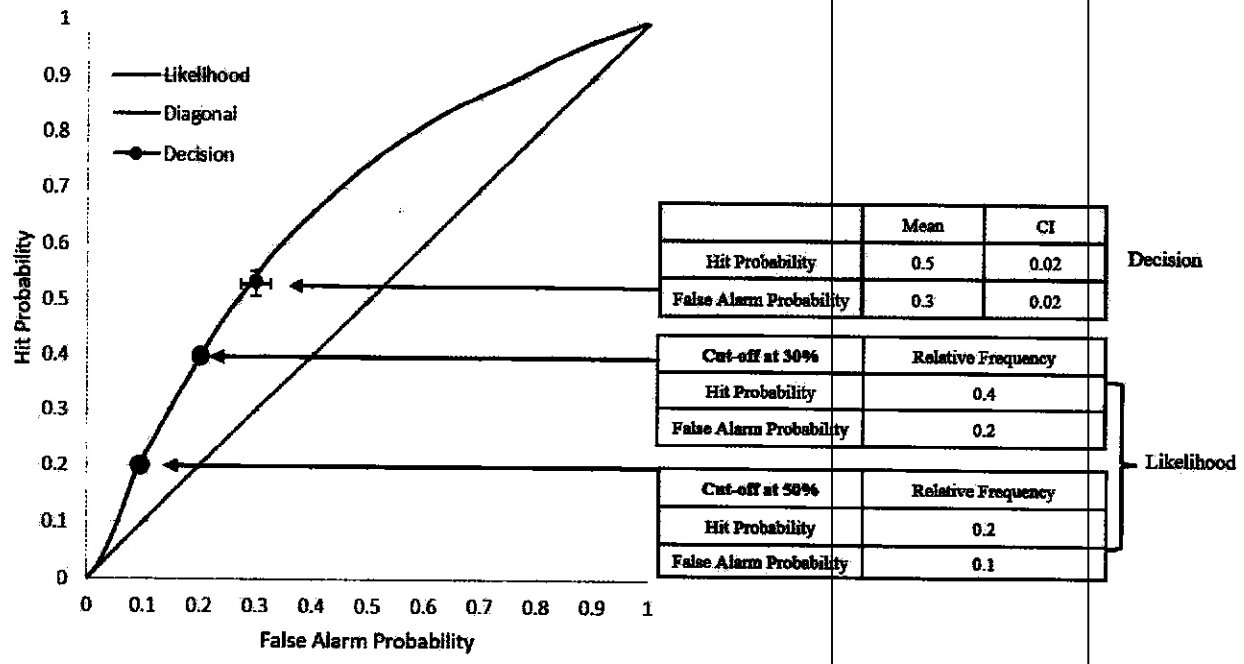


Figure 4
Calculation of the ROC plots with Hypothetical Data
For each of 20 hypothetical cut-off point, a pair of hit and false alarm probabilities was obtained from the relative frequency of likelihood ratings above the cut-off. For example, if the cut-off was 30% and 40% of the trials that resulted in tornados had a rating above the cut-off the hit probability would be .4. If 20% of the no tornado trials had a rating above the cut-off the probability of a false alarm would be .2 . There were 20 possible pairs of hit and false alarm probabilities. These pairs were plotted to form the ROC curve. The orange point was located on the curve based on the mean probabilities calculated from participants' actual decisions.

cut?

JP comments 3/12/25

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 and, ~~as mentioned above, that the subjective likelihood was variable while the subjective criterion had no variability~~. 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 (See Figure 5). This number, say 40% likelihood rating, was regarded as the calculated subjective criterion for this participant. With this method, a calculated subjective criterion was obtained for each participant, and a mean was calculated for each experiment.

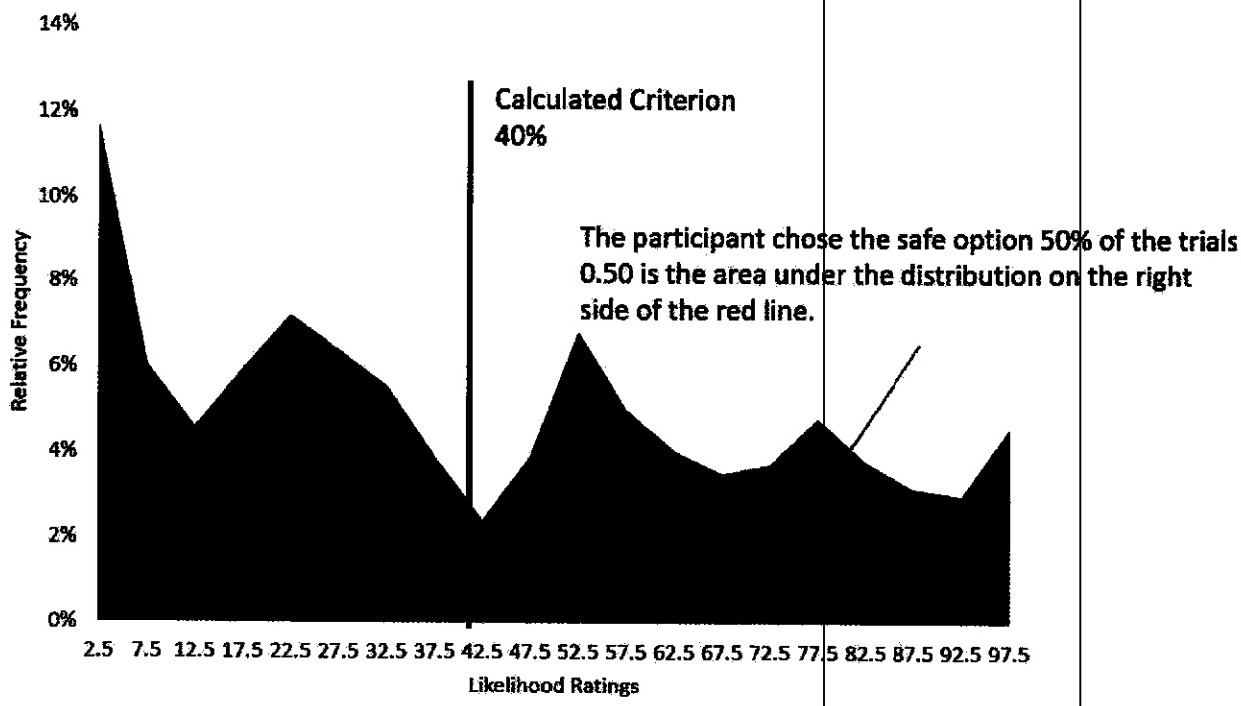


Figure 5
Example of Calculation of Subjective Criterion
The blue area is the likelihood rating distribution of a hypothetical participant. A calculated criterion was then found at 40% because the participant chose the safe option on 50% of the trials and 50% of the trials were on the right side of the 40% line.

is this real distribution?
If so, say so.

This is consistent with an effect of the different economically rational criteria.

In Figure 6, the calculated criterion is shown for tornado experiments 1 and 2. An independent t-test revealed that the mean calculated criterion in tornado experiment 1 (M = 31.0%, SD = 18.6%) was significantly lower than in tornado experiment 2 (M = 41.2%, SD = 13.8%) with a difference of -10.2% ($t(148.6) = 4.07$, $p < .001$).

In addition, the mean calculated criterion in each tornado experiment was compared to the respective economically rational criterion in two one-sample t-tests. In tornado experiment 1, the calculated criterion was significantly higher than the economically rational criterion of 9% with a difference of 22% ($t(82) = 10.77$, $p < .001$). In tornado experiment 2, the calculated criterion was significantly higher than the economically rational criterion 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 7 shows the likelihood ratings for tornado experiments 1 and 2 as a function of objective probabilities. The mean likelihood rating 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 7 plotting mean likelihood

Collapse over all objective probabilities

Need to compare the two expts to argue there was no effect of the different economically rational criteria

rating by each objective probability, shows that the overestimation was observed at most objective probability levels except for the extreme high end.

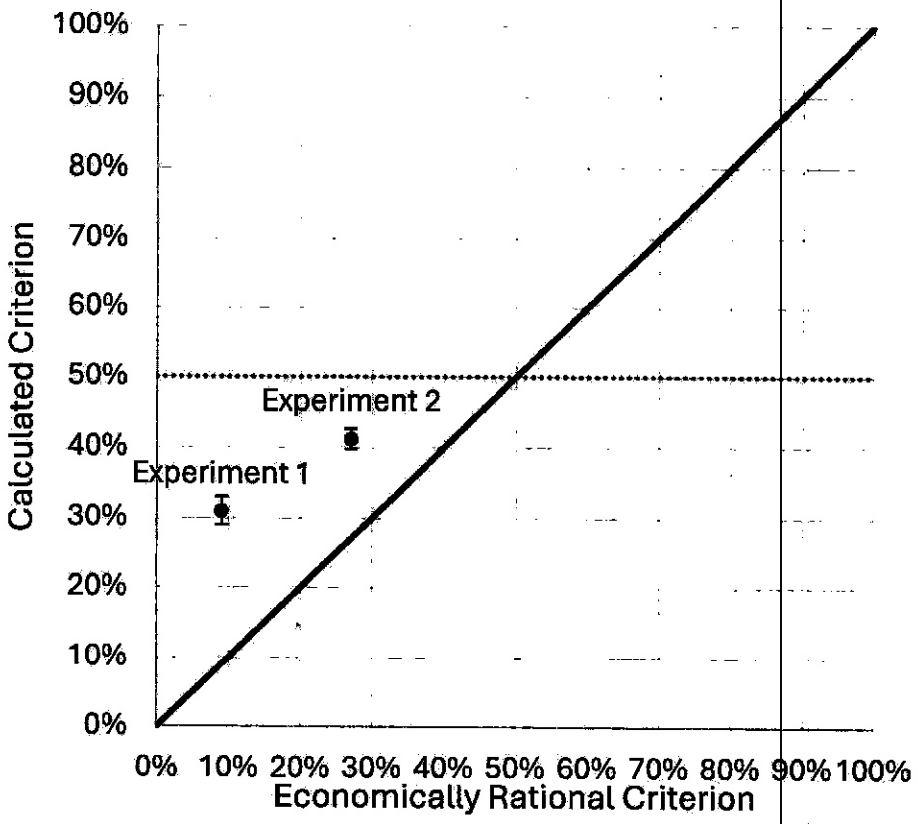


Figure 6
Calculated Subjective Criterion in the Two Tornado Experiments
The y axis represents the calculated criterion while the x axis represents the economically rational criterion. Diagonal line represents when the calculated criterion is the same as the economically rational criterion. The dashed line represents 50%.

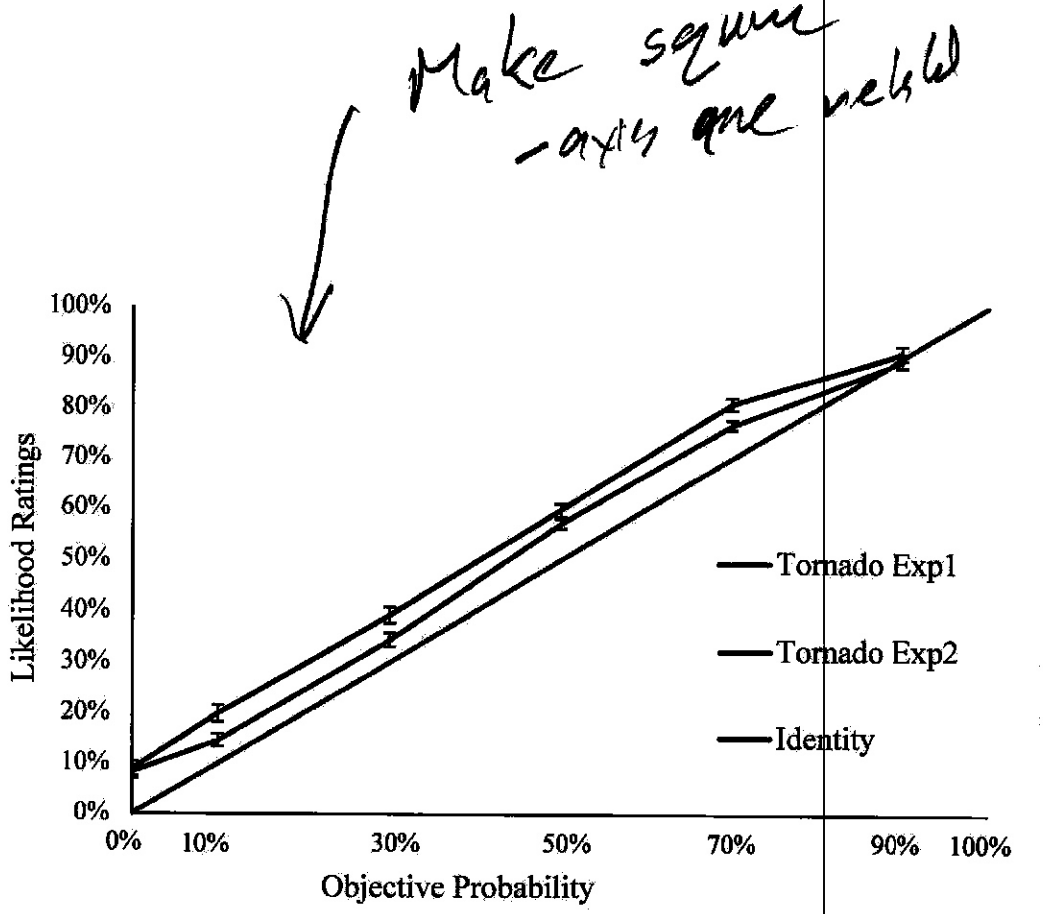


Figure 7
Likelihood Ratings as a Function of Objective Probabilities in Tornado Experiment 1 and 2
The y axis represents the likelihood ratings while the x axis represents the objective probabilities. The blue line represents the likelihood ratings of tornado experiment 1. The orange line represents the likelihood ratings of tornado experiment 2.

1) Testing a prediction of the Random Likelihood Model

Receiver Operating Characteristic Plots

The ROC analysis had two aims: 1) ~~Illustrating the random likelihood model method of calculating the subjective criterion~~, 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; See Figure 4 for a hypothetical ROC plot with both the curve (blue) and the point (orange)). 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 relative frequency 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 relative frequency 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 (See Figure 8 for an example). The 20 pairs were plotted as the ROC curve (the blue curve in Figure 4).

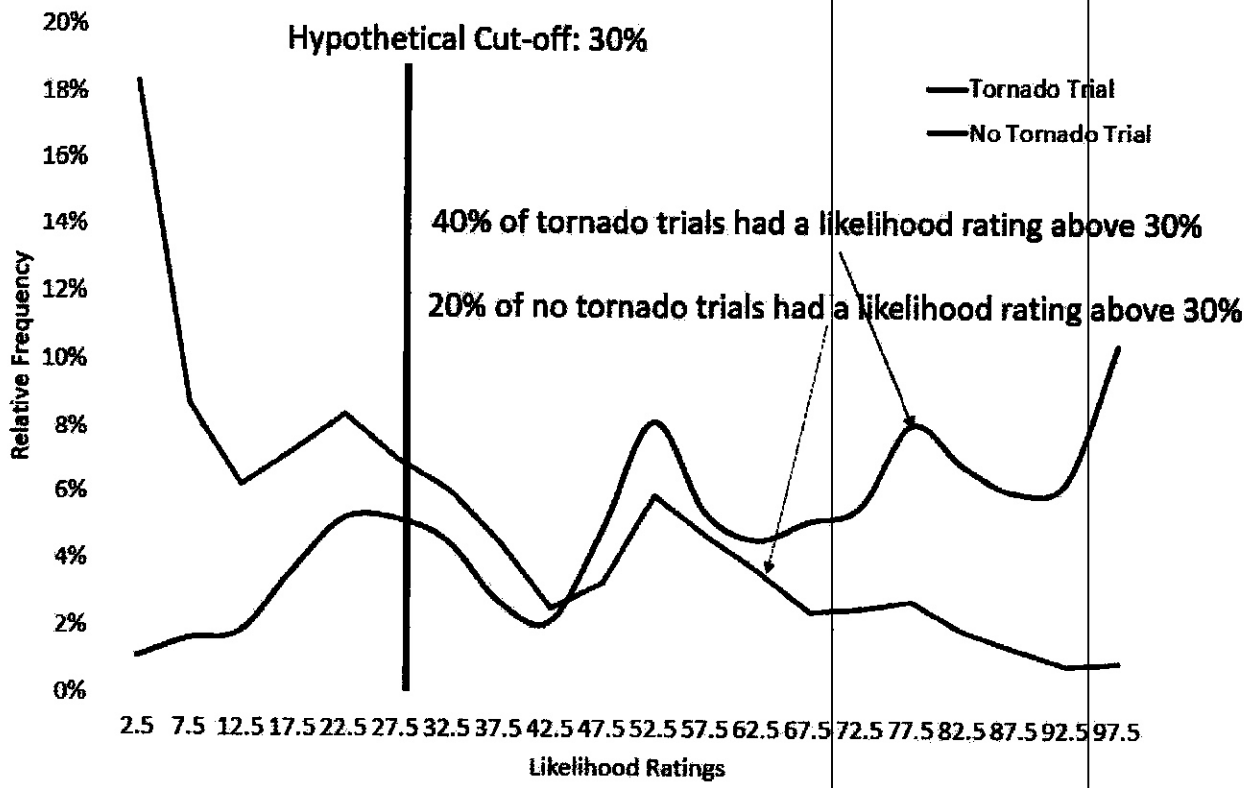


Figure 8
Example of Hit and False Alarm Probability Calculation with Hypothetical Data
The hypothetical cut-off is placed at 30%. Using this cut-off, 40% of tornado trials (hits) had a likelihood rating above 30%. The hit probability is therefore 40%. 20% of no tornado trials (false alarms) had a likelihood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

cut!

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

The first goal of the ROC analysis was to test the random likelihood model method of calculating the subjective criterion. ~~If participants always chose the safe option whenever their likelihood rating was above their subjective criterion, the hit and false alarm probabilities of the ROC curve from their likelihood ratings should align with the mean hit and false alarm probability of their decisions.~~ Figures 9 & 10 show the ROC plots for tornado experiment 1 and 2. In both experiments, the ROC curve and the 95% CI of the decision point overlapped. This indicates that the decision point was consistent with the ROC curve.

To address the second goal of the ROC analysis

~~In the next ROC plot analysis, sensitivity (ability to predict the tornado)~~, the mean percent area under the ROC curve was measured. 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.

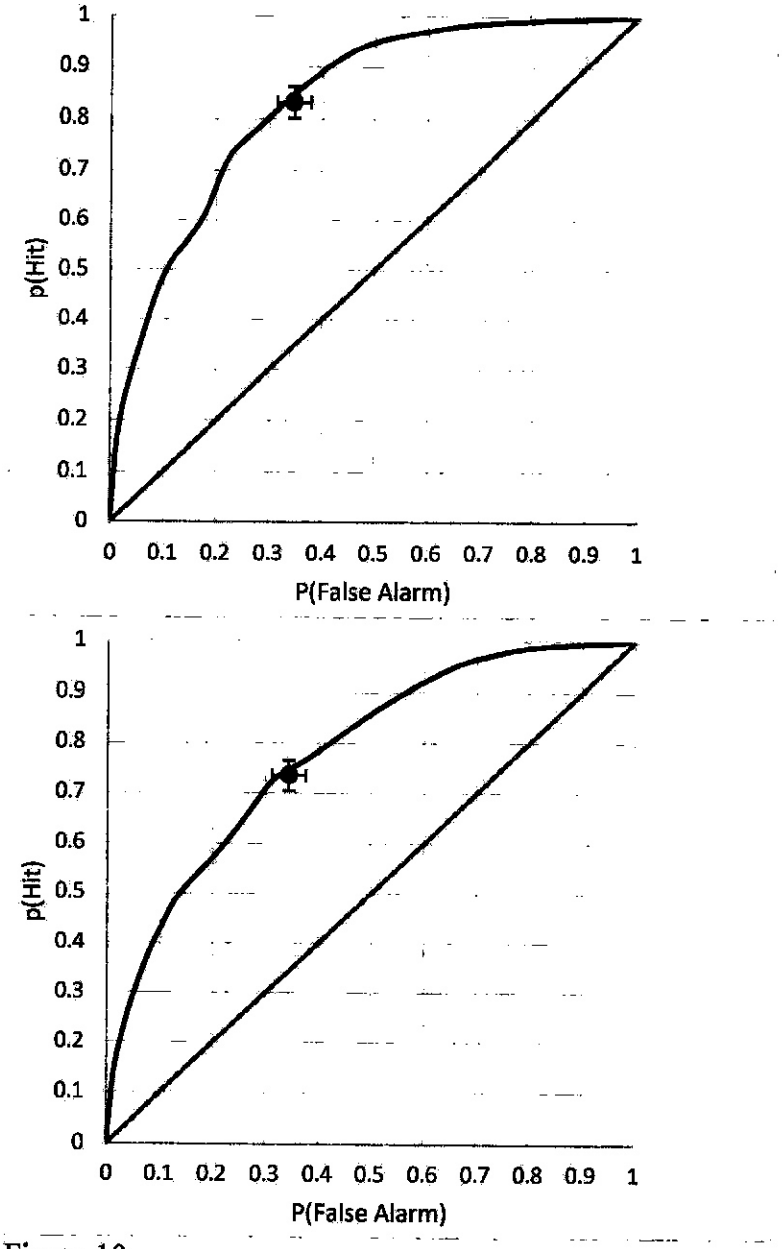


Figure 10
ROC Plot for Tornado Experiment 1(top) and 2 (bottom)
The y axis represents the probability of hits. The x axis represents the probability of false alarms. The orange dot indicates the proportion of hits and false alarms for each experiment based binary decisions.

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 condition nor a condition where the economically rational criterion was higher than 50%, 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.

However, 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., proportion of trials with a tornado) in the two experiments was different the mean likelihood ratings were not comparable with each other.

The difference in trial composition also prevented us from making a direct comparison in terms of mean sensitivity. This is because the two experiments differed in the proportion of extreme probability trials (10%, 90%) in which it was easier to predict a tornado. Therefore, although sensitivity differed between experiments, it was not possible to determine whether this

Say the result first, then talk about the reason

difference stemmed from the change of the economically rational criterion between the two experiments or the difference in trial composition.

In addition, the difference in trial structure of the two experiments prevented us from examining and comparing the relative frequency of choosing the safe option. This analysis was often used in previous naturalistic weather decision tasks to examine decision bias (Morss et al., 2010; Joslyn & LeClerc, 2013; Grounds & Joslyn, 2017; Demnitz & Joslyn, 2020; Klockow-McClain et al., 2020; Gulacsik et al., 2022; Burgeno & Joslyn, 2023; Qin et al., 2024). It could serve as an alternative verification of the signal detection analysis of decision bias. However, the relative frequency can be affected by trial composition and sensitivity, in addition to the decision bias. Therefore, a conclusion about decision bias based on a difference in relative frequency of choosing the safe option was not possible when the other two variables differed. In the new experiments reported in the next section, the trial composition was held constant among conditions to allow for these comparisons.

More importantly, in the ROC plots of both experiments, the 95% CI of the binary decision point overlapped with the ROC curve. This indicates that the proportion of hits and false alarms of participants' actual binary decisions was consistent with the ROC curve 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.

Experiment 1

The two experiments reported here employed a drought preparation task based on Demnitz & Joslyn (2020), using both a gain and a loss frame[2]. The goal was to distinguish a bias in subjective criterion from a bias in subjective likelihood and/or a difference in sensitivity. Experiment 1 focused on the centering effect on subjective criterion by manipulating the economically rational criterion. The idea was that the shift of the subjective criterion towards 50% due to centering can be observed by varying the economically rational criterion below and above 50%. Experiment 1 used a drought task in a loss frame. The economically rational criterion was manipulated to be higher, the same as, or lower than 50% to expose a centering effect such that participants' subjective criterion would shift towards 50%. This manipulation was expected to affect only the subjective criterion and should not affect the subjective likelihood or sensitivity.

Method

Participants

A total of 160 participants from the US were recruited from Prolific Academic in January 2024, a crowdsourcing platform for online research. After an elimination process, 157 participants were used in the analysis. Three were eliminated for failing the comprehension check (see the procedure section below). Each participant was paid $4 for participation plus a performance based monetary bonus. The mean age was 40 (SD = 13.91, range 20 to 80 years).

[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 due to low power.

probabilities were reliable (M = 36.5%), the mean objective probability and the proportion of drought trials were considered the same (36%) in the analyses.

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

Results

Analysis Overview

The hypothesis of experimental was that the mean subjective criterion would 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%. The manipulation of the economically rational criterion should not affect subjective likelihood or sensitivity. ~~In addition, the centering effect in likelihood ratings regardless of the manipulation of economically rational criterion was also examined. It was expected that, based on previous studies (Qin et al., 2024) and the reanalyses, there would be no centering effect in likelihood ratings.~~

The same set of dependent measures as the reanalyses were used in experiment 1. To measure the decision bias, the calculated criterion was analyzed. The consistency between the calculated and self-reported criterion was examined. ~~In addition, the relative frequency of choosing the safe option was analyzed as an alternative way to examine the decision bias.~~ If a centering effect was present, participants would choose the safe option less often (consistent with

a subjective criterion shifted from 25% to 50%) in the 25erc condition than the economically rational decisions. They would also choose the safe option more often (consistent with a subjective criterion shifted from 75% to 50%) in the 75erc condition than they should. They would be comparatively risk-neutral (consistent with a subjective criterion near 50%) in the 50erc condition. Likelihood ratings were analyzed to detect any bias in subjective likelihood. Finally, to measure the sensitivity, the area under the ROC curve was analyzed. The analysis of sensitivity tested whether it was possible to rule out the manipulation of economically rational criterion affecting it. A series of ANOVAs and t-tests were conducted. Holm-Bonferroni Method was used for planned and post hoc t-tests as well as planned pairwise comparisons under omnibus ANOVAs. Tukey method was used for post hoc pairwise comparisons under omnibus ANOVAs.

Decision Bias - Calculated Subjective Criterion

The first set of analyses examined the degree to which participants biased their decisions towards the safe option in the two conditions. The calculated subjective criterion was examined here along with its consistency with the self-reported subjective criterion and the relative frequency of choosing the safe option. The self-reported criterion was another operationalizations of the subjective criterion of the random likelihood model and signal detection theory. The relative frequency of choosing the safe option was an alternative way to examine the decision bias not under the purview of the model and the theory.

In Figure 22, the calculated criterion is shown as blue dots for 25erc, 50erc, and 75erc conditions. The mean calculated criterion was 35.0% (SD = 13.3%) in the 25erc condition,

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

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

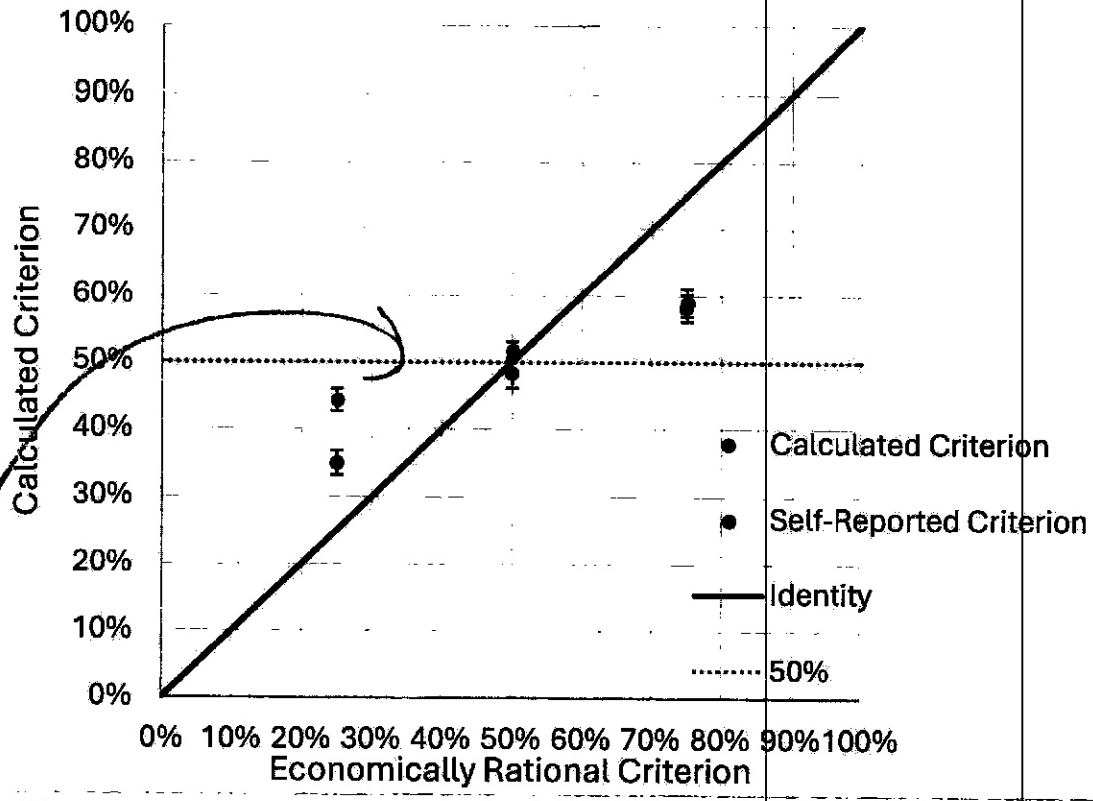


Figure 22
Calculated and Self-Reported Subjective Criterion in the Three Conditions of Experiment 1
In the 25% economically rational criterion condition, the calculated criterion was significantly higher than 25%. In the 50% economically rational criterion condition, the calculated criterion was not significantly different from 50%. In the 75% economically rational criterion condition, the calculated criterion was significantly lower than 75%.
The self-reported criterion shared a similar pattern to the calculated criterion.

Add lines for clarity

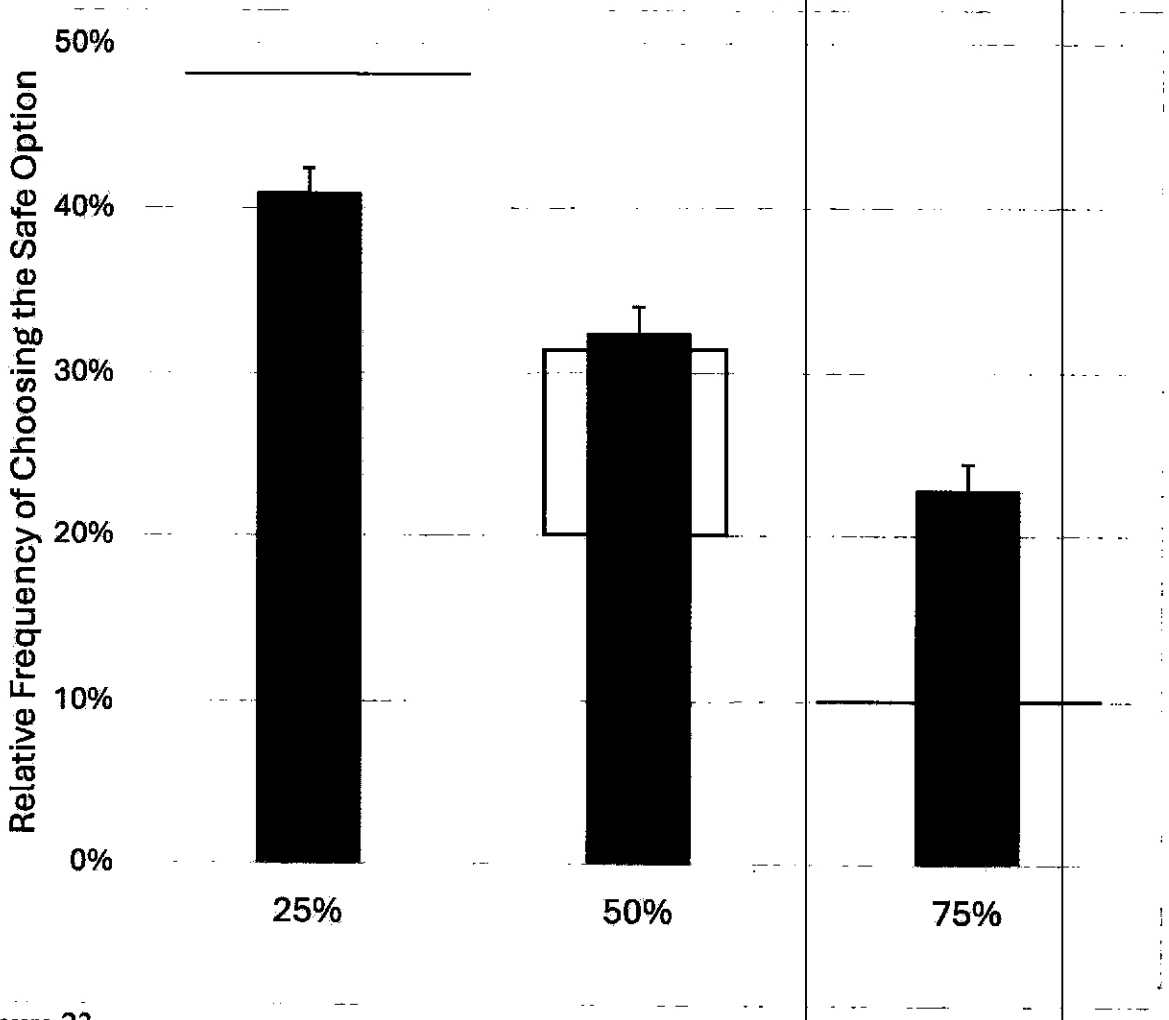


Figure 23
Relative Frequency of Choosing the Safe Option in the Three Conditions of Experiment 1
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red line in the 75% economically rational criterion condition shows the relative frequency of 10%. The red box in the 50% economically rational criterion condition shows the relative frequency between 20% to 32%. These lines are the relative frequency the participants would get if they always chose the safe option when the objective probability was above the economically rational criterion in their respective conditions.

CUT!

In addition

Finally, the pattern of self-reported criterion was consistent with that of the calculated criterion (Figure 23). Self-reported criterion was pulled towards 50% to a greater extent than the calculated criterion in the 25erc condition. The relative frequency of choosing the safe option also showed lower relative frequency than economically rational (risk-seeking) in the 25erc condition, no difference from economically rational (risk-neutral) in the 50erc condition, and higher than economically rational (risk-averse) in the 75erc condition (Figure 23). This result corroborated with the calculated criterion analysis. See supplemental material for test statistics.

So the full results for the self-report risk and other analyses can be found in the

Likelihood Ratings

Figure 24 shows the likelihood ratings as a function of objective probabilities in the three conditions. Collapsing over objective probability

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

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

Sensitivity

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

Note also that the curves
don't fall on the ROC
curve for all 3 conditions.

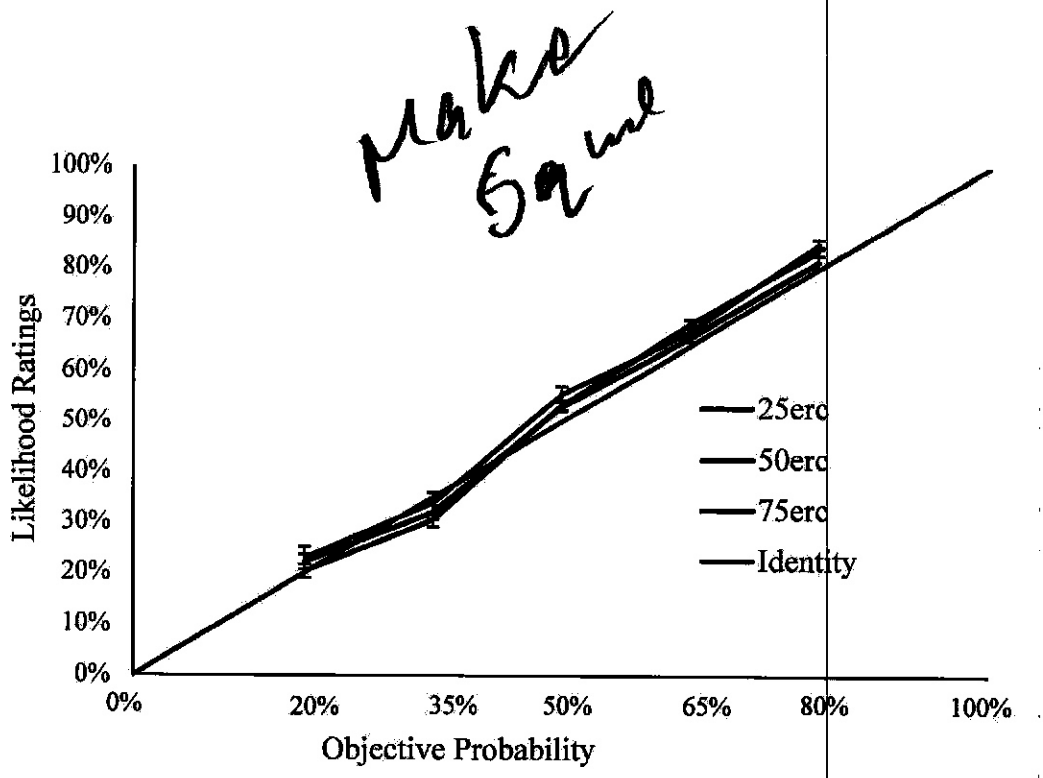
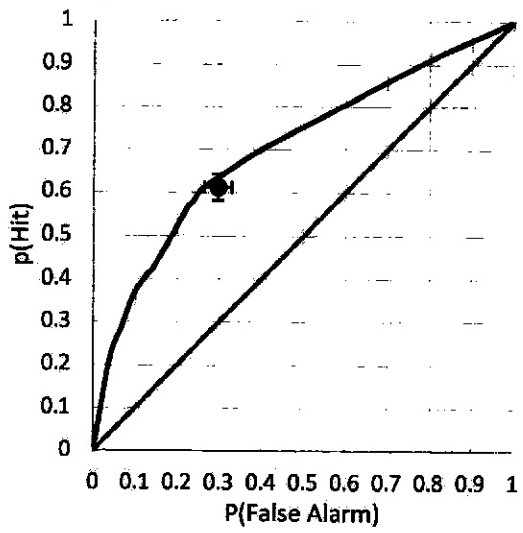
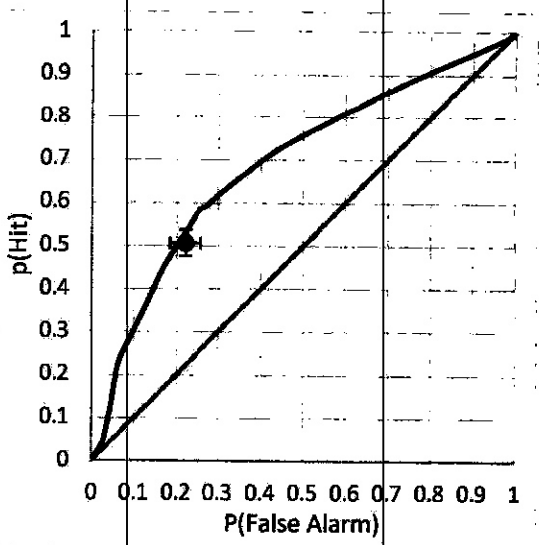


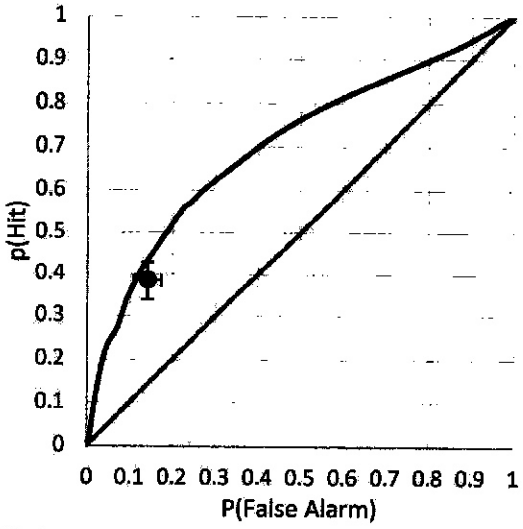
Figure 24
Likelihood Ratings in the 25%, 50%, and 75% Economically Rational Criterion Conditions as a Function of Objective Probability in Experiment 1
The blue line represents the 25 economically rational criterion condition. The orange line represents the 50 economically rational criterion condition.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



A) 25erc Condition



B) 50erc Condition



C) 75erc Condition

Figure 25
ROC Plot for the 25% Economically Rational Criterion Condition in Experiment 1
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 71.2% for the 25erc condition, 70.4% for the 50erc condition, and 70.6% for the 75erc condition.

Discussion

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

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

The results also yielded additional findings. In the ROC plots (Figure 25), the binary decisions (the decision dot indicating proportion of hits and false alarm) were consistent with the respective ROC curves, providing support for the calculation method of the calculated criterion.

Finally, as seen in Figure 24, there was a slight overestimation but no observed centering effect in likelihood ratings as indicated by comparing them to the objective probability levels.

Overall, experiment 1 yielded support for centering affecting the subjective criterion and in turn binary decisions, consistent with the subjective criterion hypothesis. Likelihood ratings and the sensitivity were not affected by the manipulation of economically rational criterion. Likelihood ratings were also not affected by the centering effect.

Experiment 2

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

Method

Participants

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 according prospect theory, the prediction was that there should be both a centering effect and a gain-loss framing effect on the subjective

criterion. As with experiment 1, the centering effect should shift the subjective criterion in the 25erc condition towards 50% while having no effect on the subjective criterion in the 50erc condition. Based on the utility function of prospect theory (Tversky & Kahneman, 1979), the gain-loss framing effect should lead to a higher subjective criterion in the loss condition than in the gain condition. Moreover, no interaction between these two effects was expected as their mechanisms should be theoretically independent. Finally, the manipulation of economically rational criterion and gain-loss framing were not expected to affect likelihood ratings or sensitivity.

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

Decision Bias - Calculated Subjective Criterion.

The first set of analyses examined the degree to which participants biased their decisions towards the safe option through calculated criterion. In Figure 26, the calculated criterion is shown for the four conditions. The calculated criterion in the loss frame condition was higher than in the gain frame condition. It was higher than the economically rational criterion in the 25erc condition and lower than the economically rational criterion in the 50erc condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion

(25erc, 50erc) manipulations as the independent variables were conducted on the calculated criterion. There was a main effect of the gain-loss framing manipulation such that in the loss frame (M = 40.6%, SD = 15.9%) the calculated criterion was 2.7% higher than in the gain frame (M = 37.9%, SD = 16.9%; F(1, 554) = 4.38, p = .037). There was a main effect of the economically rational criterion manipulation such that the calculated criterion in the 50erc condition (M = 43.7%, SD = 16.7%) was 9.2% higher than in the 25erc condition (M = 34.5%, SD = 15.0%; F(1, 554) = 46.76, p < .001). There was no significant interaction between the gain-loss framing and the economically rational criterion manipulation (F(1, 554) = 1.03, p = .31). The two manipulations both had an effect on the calculated criterion.

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

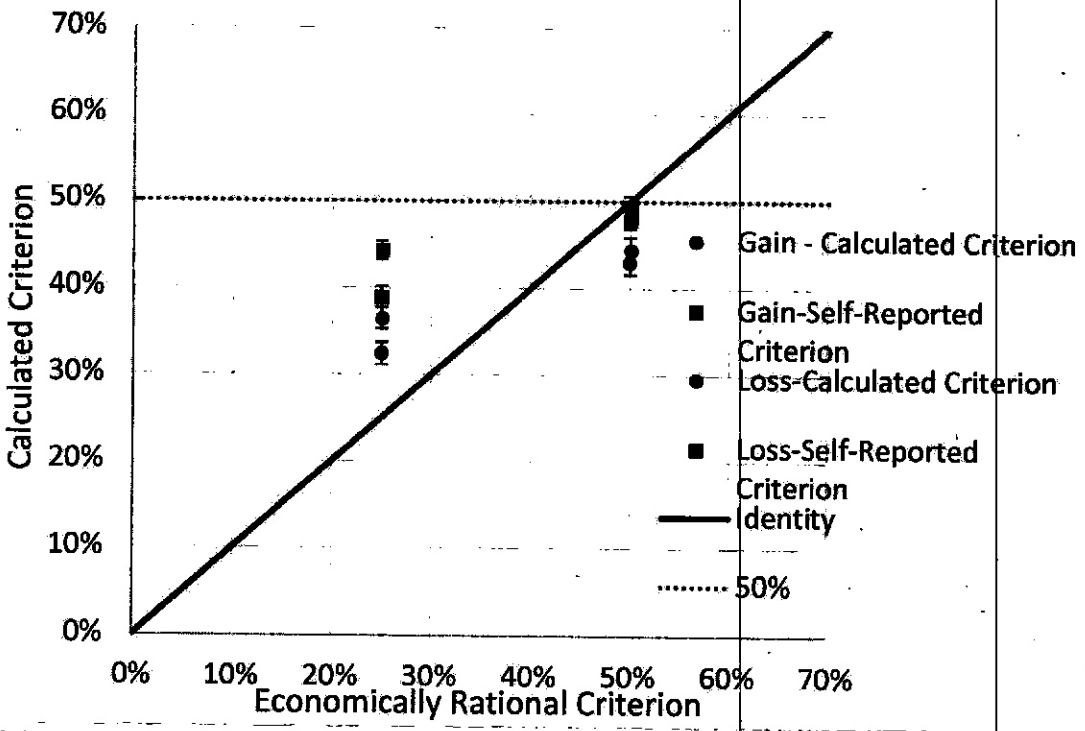


Figure 26
Calculated and Self-Reported Subjective Criterion in the Four Conditions of Experiment 2
In the 25erc gain frame condition, the calculated criterion was 7.5% higher than the economically rational criterion of 25%. In the 25erc loss frame condition, the calculated criterion was 11.6% higher than 25%. In the 50erc gain frame condition, the calculated criterion was 7.0% lower than the economically rational criterion of 50%. In the 50erc loss frame condition, the calculated criterion was 5.6% lower than 50%.

Focus on descriptive statistics not regression

Split into 2 panels and add titles

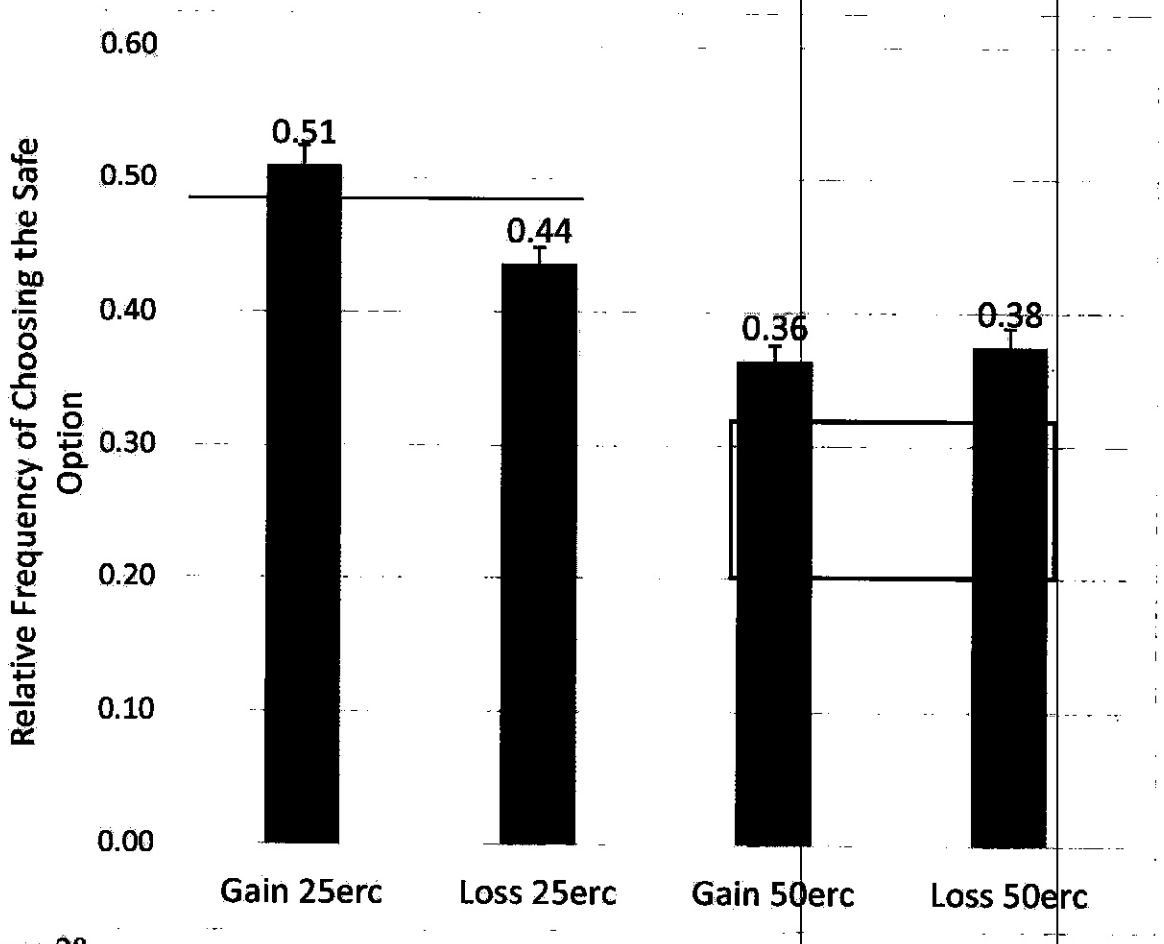


Figure 28
Relative Frequency of Choosing the Safe Option in the Four Conditions of Experiment 2
The red line in the 25% economically rational criterion condition shows the relative frequency of 48%. The red box in the 50% economically rational criterion condition shows the relative frequency between 20% to 32%. These lines are the relative frequency the participants would get if they always chose the safe option when the objective probability was above the economically rational criterion in their respective conditions.

CUT

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

Finally, like in experiment 1, the pattern of self-reported criterion was consistent with that of the calculated criterion (Figure 26). Self-reported criterion was pulled towards 50% to a greater extent than the calculated criterion all conditions. The relative frequency of choosing the safe option showed lower relative frequency than economically rational (risk-seeking) in the 25erc loss frame condition and higher relative frequency (risk-averse) in the 25erc gain condition and both 50erc conditions (Figure 28). This result corroborated with the calculated criterion analysis. See supplemental material for the statistics.

Likelihood Ratings

Figure 29 shows the likelihood ratings as a function of objective probabilities in the four conditions. The likelihood rating patterns were similar among the conditions. An ANOVA with the gain-loss framing and economically rational criterion manipulation (25erc, 50erc, and 75erc) as the independent variables, on the mean likelihood ratings was conducted. The 0.1% difference between the mean likelihood ratings in the gain frame condition (M = 39.9%, SD = 10.0%) and in the loss frame condition (M = 40.0%, SD = 10.3%) was not significant ($F(1, 554) = .017$, P = .68). The 0.3% difference between the mean likelihood ratings in the 25erc condition (M = 40.1%, SD = 9.2%) and in the 50erc condition (M = 39.8%, SD = 10.9%) was not significant ($F(1, 554) = 0.23$, P = .64). There was no significant interaction between the gain-loss framing and the economically rational criterion ($F(1, 554) = 0.36$, P = .55). This result suggests that likelihood ratings were not affected by either manipulation.

Make square

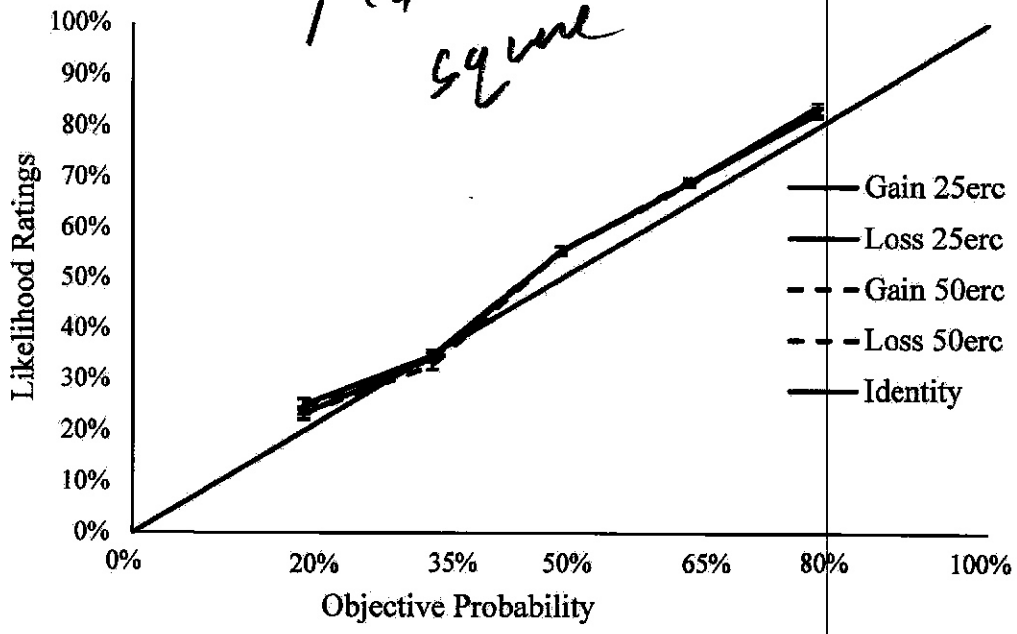


Figure 29

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

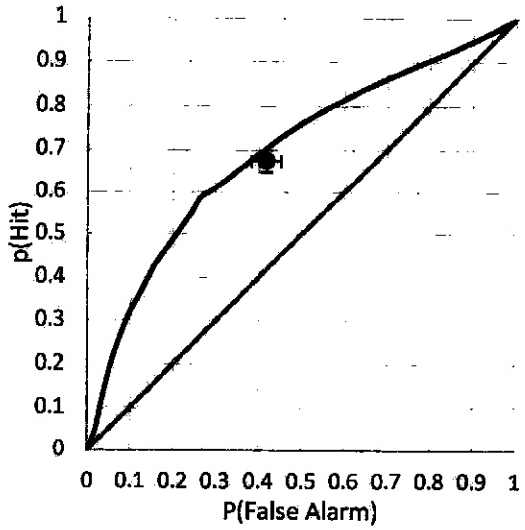
The green solid line represents the 25erc gain frame condition. The green dashed line represents the 50erc gain condition. The red solid line represents the 25erc loss frame condition. The red dashed line represents the 50erc loss condition. There was no observed [redacted] effects of the main conditions.

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

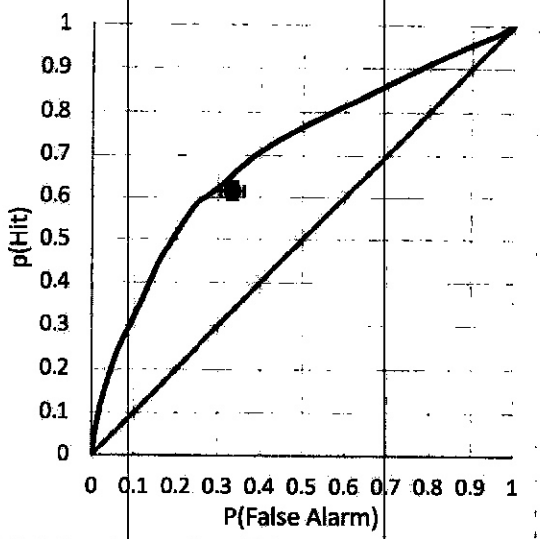
Sensitivity

The next analysis examined the sensitivity of participants' ability to predict the drought as measured by the area under the ROC curve. In Figure 30, ROC plots are shown for the four conditions respectively. The ROC curves of the four conditions were similar, indicating that both conditions had similar sensitivity or the ability to predict drought based on provided drought forecasts. All but 50erc gain frame conditions had the 95% CI of binary decisions overlapping

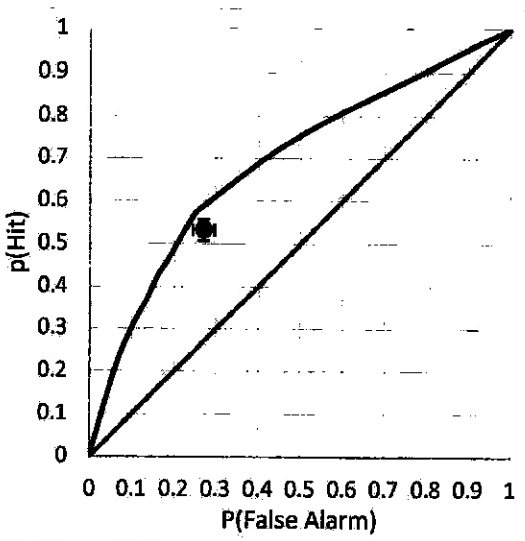
with the ROC curves. The mean percent Area under ROC curve was 70.0% (SD = 8.1%) in the 25erc gain frame condition, 71.1% (SD = 5.2%) in the 25erc loss condition, 69.8% (SD = 7.6%) in the 50erc gain condition, and 69.6% (SD = 6.9%) in the 50erc loss condition. An ANOVA with the gain-loss framing (gain frame, loss frame) and economically rational criterion (25erc, 50erc, and 75erc) manipulations as the independent variables on the mean percent area under the curve showed no significant main effect of the gain-loss framing ($F(1, 554) = 0.51$, $p = .47$) or the economically rational criterion manipulations ($F(1, 554) = 1.93$, $p = .17$). There also was not an interaction ($F(1, 554) = 0.94$, $p = .33$). This suggests that the sensitivity did not differ among the conditions and was not affected by the manipulation of gain-loss framing and economically rational criterion.



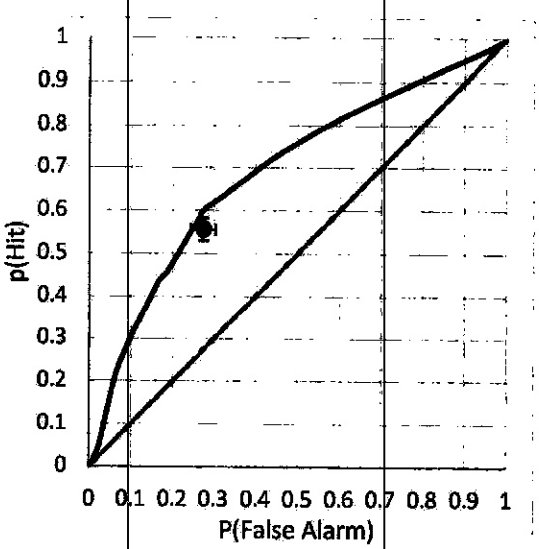
A) 25erc Gain Condition



B) 25erc Loss Condition



C) 50erc Gain Condition



D) 50erc Loss Condition

Figure 30
ROC Plot for the Four Conditions in Experiment 2
The blue curve is ROC curves created from likelihood ratings. The orange dot is created from binary decisions. The percent area under curve was 70.0% in the 25erc gain frame condition, 71.1% in the 25erc loss frame condition, 69.8% in the 50erc gain frame condition, 69.6%. in the 50erc loss condition.

59A

biased

The ~~subjective~~ criterion hypothesis made 4 predictions in Expmt 2. 1) Varying the gain-loss framing should affect the the subjective criterion.

2) Varying the economic context should affect the subjective criterion (as in Expmt 1).

3) Neither manipulation should affect likelihood ratings or sensitivity.

4) A centering effect would be present for the subjective criterion but not the other measures.

Discussion

Experiment 2 tested the prediction that both the manipulation of the economically rational criterion and gain-loss framing should affect the subjective criterion. These two manipulations should not interact in the subjective criterion as the gain-loss framing effect and the centering effect were considered theoretically independent. In addition, the manipulation of gain-loss framing and economically rational criterion should not affect subjective likelihood or sensitivity.

There was evidence for a centering effect in the subjective criterion exposed by the economically rational criterion manipulation, similar to the findings of experiment 1. The calculated subjective criterion were 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 higher in the 25erc condition than in the 50erc condition. Interestingly, the calculated criterion in the 50erc condition was slightly but significantly lower than 50%, a result not seen in experiment 1 where the calculated criterion did not differ from 50% in the 50erc condition. This difference was not seen in the self-reported criterion which held an overall similar pattern to the calculated criterion. One possible explanation is that participants considered somewhere slightly lower than 50% as the center of the range in their mind. Finally, similar to the gain-loss framing manipulation, the manipulation of the economically rational criterion did not affect likelihood ratings or sensitivity.

In terms of the gain-loss framing manipulation, the results showed that as predicted participants in the loss frame condition had a higher calculated subjective criterion than those in the gain frame condition. The participants in the loss frame condition also chose the safe option less often than those in the gain frame condition, showing a risk seeking decision bias. All these

results are consistent with a gain-loss framing effect where people are more risk seeking in a loss frame than in a gain frame. In addition, as the likelihood ratings and the sensitivity were not affected by manipulating the framing, it can be inferred that the gain-loss framing effect operates by altering the subjective criterion. Finally, it should be noted that participants in the loss condition reported this experiment to be slightly more difficult than participants in the gain condition. Therefore, the difficulty level might have been a confound here, if for some reason, participants became more risk-seeking with increased difficulty.

Next, as expected, the interaction between the gain-loss framing and the economically rational criterion manipulations failed to reach significance in the analyses on the calculated criterion. This implies that these two subjective criterion measurements' shift due to the gain-loss framing was not affected by centering. The relative frequency of choosing the safe option analysis found that participants were slightly risk-averse in both the gain frame and the loss frame in the 50erc condition, but they were, as predicted, only risk-averse in the gain frame while risk-seeking in the loss frame in the 25erc condition. This implies some interaction between the two manipulations might still exist.

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

The results also yielded additional findings. First, as with the previous experiments, the likelihood ratings were not affected by manipulating either gain-loss framing or economically rational criterion. However, they were overestimated compared to the proportion of drought trials

(practically the same as the mean objective probabilities). The overestimation but no centering effect was observed in likelihood ratings as a function of objective probability levels as seen in Figure 29. This is also consistent with experiment 1.

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

Overall, experiment 2 yielded support for both the gain-loss framing effect and the centering effect on subjective criterion and in turn the decisions, consistent with the subjective criterion hypothesis. The centering effect appeared to have a larger effect on the subjective criterion than the gain-loss framing effect when the economically rational criterion was at 25%. These two effects did not interact with other in most of the analyses reported here. Again, likelihood ratings and sensitivities were not affected by the manipulation of gain-loss framing or economically rational criterion.

General Discussion

Summary of Results