The Advantage of Numeric Uncertainty Information For Decision-Making With Two Thresholds

Jee Hoon Han

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Reading Committee:

Susan Joslyn

Andrea Stocco

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

The Advantage of Numeric Uncertainty Information For Decision-Making With Two Thresholds

Jee Hoon Han

Chair of the Supervisory Committee:
Susan Joslyn
Department of Psychology

Evidence suggests that people make better decisions based on forecasts with numeric uncertainty estimates (e.g., 30%) compared to single value (deterministic) forecasts. Much of this work has focused on binary tasks. However, most real-world situations include more array of options. Advantages for uncertainty information may not be seen with more options because of the increase in cognitive load. The present study aims to answer whether the advantage for probabilistic compared to deterministic forecasts persists in situations with three options. To investigate, a school closure paradigm with three-option (close, delay, open) was compared to two-option condition (close, open). The results suggest that the advantages for probabilistic forecasts hold across task complexity. Moreover, participants in two-option condition showed greater proportion of risk-seeking errors than risk-averse errors, whereas the tendency was

reversed in three-option condition. The study suggests that numeric uncertainty information can be beneficial in complex situation.

The Advantage of Numeric Uncertainty Information For Decision-Making With Two Thresholds

A variety of important decisions are made under uncertainty, including those about medical treatment, retirement investments, and those related to severe weather. In the weather domain, our focus here, it is now possible to quantify the uncertainty in many cases, such as 30% chance of snow accumulating one inch or more. Moreover, several laboratory-based experiments have shown that people make better decisions and have greater trust in weather forecasts when numeric uncertainty information, such as probability, is included (Joslyn & Grounds, 2015; Joslyn & LeClerc, 2012; Nadav-Greenberg & Joslyn, 2009; Savelli & Joslyn, 2013). Somewhat surprisingly, advanced education is not required to reap the benefits of numeric uncertainty information. The advantage is seen in groups with a high school education or less as well as in groups with an advanced degree (Grounds & Joslyn, 2018, Grounds, Joslyn & Otsuka, 2017).

Although the results from previous studies provide strong evidence for the benefits of probabilistic forecasts, the degree to which these benefits will extend to more complex real-world situations remains unknown. Unlike the binary decision tasks employed in previously mentioned studies (Joslyn & LeClerc, 2012; Nadav-Greenberg & Joslyn, 2009), many of the decisions made in natural settings involve multiple options. For example, when administrators must decide whether to close schools in the face of severe weather, such as snow, in addition to considering closing or staying open, they may consider the intermediate option of delaying school start time a few hours when conditions are likely to be less extreme. In other words, in the real world there may be multiple different appropriate actions depending on the situation.

As the number of options increases, however, the amount of information to be retained and evaluated by the decision maker increases. Take for example, the simplified school closure decision with two possible options: staying open or closing. If the threshold for closing is six or more inches, whenever snow accumulation is expected to exceed six inches, the best choice is to close school. However, remaining open is the best option when snow accumulation is expected to be less than six inches. Thus, there are four possible decision-outcome pairs to consider: 1) remain open and snow accumulation is less than six inches, 2) remain open and accumulation is more than six inches, 3) close and accumulation is less than six inches, and 4) close and accumulation is more than six inches. Adding an intermediate option, delaying school, appropriate when there is a moderate amount of snow, e.g., between one and six inches, increases the relevant decision-outcome pairs to nine. Three are optimal: 1) stay open and it snows less than one inch, 2) delay and it snows one or more but less than six inches, and 3) close and it snows more than six inches. There are also six sub-optimal outcomes: 4) stay open and it snows more than one inch but less than six inches, 5) stay open and it snows more than six inches, 6) delay and it snows less than one inch, 7) delay and it snows more than six inches, 8) close and it snows less than one inch, and 9) close and it snows more than one inch but less than six inches. Consequently, adding one more option may make the situation much more complex for decision makers.

Indeed, as the situation becomes more complex, the cognitive load also increases, increasing the chance for reasoning errors. Cognitive load refers to the overall amount of mental resource that is required for the task (Chandler & Sweller, 1991), in this case, making a decision. Cognitive load can be measured by various methods such as participant ratings (e.g., task challengingness), a simultaneous secondary task to gauge the decrement in primary task performance (e.g., Whitney, Rinehart & Hinson, 2008), response time (increasing implies greater load) or, physiological measures (e.g., Haapalainen et al., 2010; DeLeeuw & Mayer, 2008; Paas et al., 2003).

Evidence suggests increasing cognitive load stresses what psychologists refer to as “working memory,” roughly synonymous with consciousness (Baddeley, 1983), long known to have severe limitations (Miller, 1956). Thus, increasing cognitive load tends to result in underperformance in various types of cognitive tasks including decision making (Hinson, Jameson & Whitney, 2003; Hitch, 1978; Kane & Engle, 2000; Kane, Bleckley, Conway & Engle, 2001; Whitney et al., 2008). For instance, increasing the number of options from two to three leads to an increase in impulsive behavior resulting in greater discounting for delayed monetary rewards (Hinson et al., 2003). Additionally, participants are more likely to rely on heuristics to make decisions when asked to simultaneously retain a list of unrelated memory items (Whitney et al., 2008).

Therefore, when combined with additional decision options and outcomes, the need to evaluate uncertainty information may further increase cognitive load. Unlike a single value forecast, referred to here as a deterministic forecast because it implies a single outcome, a forecast that includes uncertainty requires the decision maker to consider the likelihoods of several possible outcomes throughout the decision process. Thus, numeric uncertainty forecasts that improve decision quality with binary decisions, may not be fully evaluated when more options are involved, and hence may not lead to the same advantages seen in previous research. Nevertheless, there is some preliminary evidence suggesting otherwise (Joslyn & Grounds, 2015), although the additional option involved circumstances that are so rare, it may have been considered irrelevant on most trials. To summarize, previous work has shown that people can benefit from numeric uncertainty information in decision-making tasks. Moreover, there is some preliminary evidence that the advantage may extend to more complex situations. However, additional research is required to determine whether the advantage for uncertainty forecasts persists in tasks with multiple options that are relevant for all trials.

Therefore, to test that question, the study reported here employed a school closure decision task similar to that described above, in which half of participants made decisions with two options and half with three options. Half of each group had only a single value forecast, while the other half of participants were also provided with probabilistic forecasts. If participants in both the simple (two-option) and complex (three-option) task, show improved performance with probabilistic forecasts, it suggests that numeric uncertainty information can be helpful for non-experts in more complex and realistic situations. In addition, in this task only losses are possible, because the choice is between the cost of protection and a larger probabilistic loss. Therefore, we expect the decision errors to be risk seeking rather than risk averse (Kahneman & Tversky, 1979). In other words, we expect participants to take protective action, closing or delaying schools, less often than would be optimal from the perspective of expected value (Bernoulli, 1954) although there is some evidence that the reverse tendency (risk averse) is observed under cognitive load (Deck & Jahedi, 2015).

Methods

Participants

A total of 191 (38% female) participants were recruited from Amazon Mechanical Turk and received a flat rate monetary compensation based on the federal minimum wage at the time, which was $7.25 per hour, as well as a bonus payment commensurate with the performance, which averaged to $3.15 per person. Mean age was 37.14 years (range 18 – 73). The study was reviewed by the University of Washington Human Subject Division and qualified for exemption with no greater than minimal risk.

Procedure

After participants provided informed consent and indicated their gender and age, they read instructions, which explained the task scenario and the cost-loss structure. Participants were to assume the role of decision consultant giving advice to school districts about how to respond to possible snowstorms. The advice depended on the predicted snow accumulation provided in weather forecasts.

In the three-option condition participants were to advise 1) closing if they expected the snow to accumulate six inches or more the next day to prevent accidents and injuries, 2) delaying one hour if they expected the snow to accumulate one inch or more, but less than six inches, and 3) staying open if they expected the snow to accumulate less than one inch. In the two-option condition, participants were to advise 1) closing if they expected the snow to accumulate six inches or more or 2) staying open otherwise. There were 61 trials representing four hypothetical winter seasons from mid-November to mid-March. Each forecast was issued on Monday, and the decision was made for Tuesday. Participants were told that each school was in a different school district to encourage them to consider each trial to be independent.

In order to motivate participants, they were provided with a point balance and told that their goal was to retain as many points as possible. Giving advice to close the school cost two points to represent the cost of makeup days. Giving advice to delay the school for one hour (only in the three-option condition) cost one point to represent time lost. There was no charge for giving advice to stay open. However, if the participant advised staying open and six inches or more of snow were observed the next day, the participant was penalized eight points to reflect possible accidents and injuries. Similarly, if the participant advised delaying one hour in the three-option condition, and six inches or more of snow were observed the next day, the participant was penalized four points. The penalty was reduced to represent the fact that some damage could be mitigated by delaying one hour[1]. If the participant advised staying open and one inch or more but less than six inches of snow were observed the next day, the participant was penalized two points. See table 1. Thus, due to this cost-loss structure and based expected value theory (Bernoulli, 1954), it was optimal to close schools whenever the probability of six or more inches of snow was 25% chance or greater for both the two and three-option conditions (see equation 1 & 2). It was optimal to delay school in the three-option condition whenever the probability of one or more inches was greater than 50% and six or more inches was less that 25% (see equation 3).

In two and three-option condition,

25% (Percent Chance of 6+") * -8 (Penalty for staying open) = -2 (Closing Cost) (1)

In three-option condition,

25% (Percent Chance of 6+") * -4 (Penalty for delaying) = -1 (Delaying Cost) (2)

50% (Percent Chance of 1+") * -2 (Penalty for staying open) = -1 (Delaying Cost) (3)

Thus, on every trial, there was an economically optimal decision (See results section). The starting budget for the two conditions differed (228 points in three- and 200 points in two-option condition), to compensate for the fact that participants in the three-option condition would

---

[1] Notice (see equations 1-3) that the reducing the penalty when the participant chooses to delay (cost = 1) and snow accumulation exceeds six inches, allows us to maintain the optimal probability threshold for closing at 25% in both conditions.

spend 28 more points when following the optimal strategy than would those in the two-option condition.

Table 1

Cost-Loss Structure

	Decision		
	Stay Open	Delay 1 Hour [a]	Close
Cost	0 points	1 point	2 points
Snow accumulation		Penalty	
Snow < 1”	none	none	none
1” ≤ Snow < 6” [a]	2 points	none	none
Snow > 6”	8 points	4 points	none

[a] Delay 1 hour option and 1” ≤ Snow < 6” threshold were only available in the three-option condition

To further motivate participants, a cash bonus was given commensurate with ending point balance, one dollar for every five points above 106 points. This threshold was chosen so that participants would not use the simple but unrealistic strategy of closing schools on all trials, which would result in spending 122 points in three option condition (228-122 = 106). The same threshold was used for both three-option and two-option conditions.

In order to familiarize participants with the cost-loss structure, they were guided through example trials illustrating the possible choices and outcomes. Then, they answered two questions about total point deduction in hypothetical situations to check whether they understood the point structure correctly (e.g., *How many points would be deducted if you chose to close, and 7 inches of snow fell?).* The same two questions were repeated until the participant got both of them right. Then, participants began the 61-trial sequence. On each trial, participants were shown the date and the day with the forecast. The forecast was visible while participants provided their own estimates for the number of inches of snow for the next day as well as the least and the greatest number of inches that would not surprise them. These questions were added to encourage participants to attend to the forecasts. Then, they rated how much they trusted the forecast using a 6-point drop-down menu ranging from “Not at all” to “Completely”.

After these questions, participants were shown the decision screen, where the weather forecast was shown again. Participants made their decision by clicking a button labeled either “Close”, “Delay 1 Hour” (in three-option condition only), or “Stay Open” button. Captions were added below each button to remind participants that “Close” meant “I think snow accumulation will be 6 inches or more”, “Delay 1 Hour” meant “I think snow accumulation will be 1 inch or more but less than 6 inches”, and “Stay Open” meant “I think snow accumulation will be less than 1 inch,” or “I think snow accumulation will be less than 6 inches” in the case of two-option condition. After making their decision, participants were shown the observed snow accumulation of the next day and were informed of any loss that occurred. Participants, then rated how much they trusted the forecast again and moved to the next trial. However, the trust measures were not properly stored and will not be mentioned again. The current balance of points was present on every screen throughout the trial.

After the first trial, 30th trial (halfway through the trial sequence), and 61st trial (at the end), participants were asked how challenging the task was and responded on a 6-point drop-down menu ranging from “Not at all” to “Completely”. This was measured to check whether the participants in the three-option conditions perceived the task to be more cognitively challenging than the participants in the two-option conditions did.

Stimuli

About half of the participants received a single value forecast for snow accumulation in inches (deterministic forecast, e.g., “3 inches of snow”). The other half received probability of exceedance forecasts (percent chance that snow accumulation will exceed a specified value, e.g., "43% chance of 1 inch or more," and/or "14% chance of 6 inches or more.") in addition to the single value forecast. In the three-option condition both the probability of exceeding one and six inches were given while in the two-option condition only the probability of exceeding six inches was shown.

The single value forecast, observed snow accumulation, and the probabilistic forecasts for exceeding six inches were based on historical data from the National Weather Service from the Eastern region of the United States. The single value forecasts ranged from zero to six inches ($M = 2.75$), and the observed snow accumulation ranged from zero to nine inches ($M = 2.02$). The probabilistic forecasts for six inches or more ranged from 0% to 43% ($M = 16.30\%$). The probabilistic forecasts were adjusted slightly so that they were reliable when forecasts were binned according to the percent chance specified (0-4%, 5-14%, 15-24%, 25-34%, 35-44%)[2]. The frequency of the observed snow accumulating six or more inches fell within the range of the bin that included the probabilistic forecast (e.g., the observed accumulation was six or more inches on two out of 10 trials where the forecasted probability was between 15 and 24% chance). Similarly, the range of the probabilistic forecast for one or more inches was from 17% to 93% ($M = 58.85\%$) and reliable when binned (see table 2). The probabilistic forecasts for one or more inches were not provided in the historical data set. They were generated so that they would correlate with the probabilistic forecasts for six or more inches, deterministic forecasts, and observed values. The average correlation between the forecasts (deterministic & probabilistic one or more = .84; deterministic & probabilistic six or more = .79; probabilistic one or more &

---

[2] The historical data upon which forecasts were based were collected on a single day, although they were presented as sequential. As a result, there was an unrealistic slight negative correlation ($r = -.4$) between the observed value and probabilistic forecasts within the 25-34% range category of the data. Therefore, three forecasts and observations were re-paired in that category to increase the correlation.

probabilistic six or more = .96) were higher than the average correlation between observed value and each of the forecast (observed & deterministic = .44; observed & probabilistic one or more = .59; observed & probabilistic six or more = .58.

Table 2

Summary of Probabilistic Forecasts

			6 + inches				1 + inches			
Optimal Decision	Bin	Total trials	Probability Range	Mean Probabilities (real data)	Observed frequencies (real data)	Percent of trials with 6+ inches	Probability Range	Mean Probabilities (artificial data)	Observed frequencies (real data)	Percent of trials with 1+ inches
Open	Bin 1	21	0-4%	0.70%	0	0.00%	0-49%	32.85%	10	47.61%
Delay	Bin 2	10	5-14%	8.70%	1	10.00%	50-55%	52.00%	5	50.00%
	Bin 3	10	15-24%	19.00%	2	20.00%	56-75%	65.10%	6	60.00%
Close	Bin 4	10	25-34%	30.70%	3	30.00%	76-85%	81.60%	8	80.00%
	Bin 5	10	35-44%	39.50%	4	40.00%	86-95%	90.50%	9	90.00%

Design

A 2(forecast format) by 2(complexity) between groups design was used. Forecast format had two levels, one that included the deterministic forecast alone and the other that also included probabilistic forecasts. Complexity had two levels with one level having two-options and the other level having three-options. Participants were randomly assigned to one of the four resulting conditions.

Results

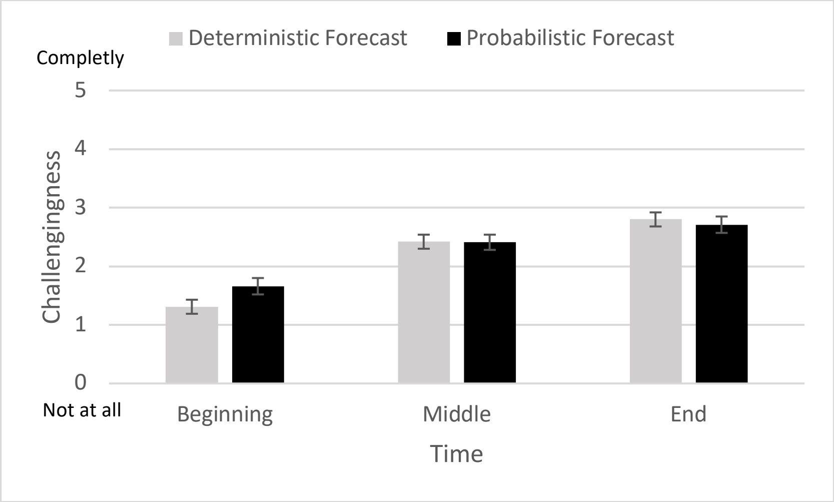
First, participants' snow accumulation estimates were analyzed to determine whether or not they paid attention to the task. Out of 191 participants across all conditions, 13 participants were excluded because the average upper bound estimate was lower than the average lower bound estimate. Then, the challengingness measure was analyzed to determine whether there were differences due to task complexity. In addition to the challengingness, the time participants took to make decisions were analyzed to see if participants in three-option condition took longer time to make decisions. The main hypotheses, however, concerned whether decision quality increased with addition of probabilistic forecasts or declined with task complexity (two- vs. three-option). Decision quality was operationalized by expected value. To further investigate the difference between the two- and three-option conditions, decision errors were analyzed exploratorily. The main hypotheses were tested by conducting a series of analysis of variance (ANOVA) with an alpha level of .05. Effect sizes were measured with generalized eta squared values for ANOVAs. All the post hoc comparisons were Bonferroni corrected for familywise error (α = .0167), and the effect sizes were measured with Cohen's d. The hypotheses tested were registered with the Open Science Framework (https://osf.io/jmvq6).

Challengingness

First, we examined challengingness ratings. Although we expected participants to rate the task more challenging in three-option condition and when provided with probabilistic forecasts, that was not the case. A mixed model three-way analysis of variance (ANOVA) was performed on challengingness rating with time (beginning, middle, end of experiment) as the within groups factor and complexity (two- vs. three-option) and forecast format (deterministic vs. probabilistic) as between groups factors. Neither the main effect of complexity nor the main effect for forecast format reached significance. However, the two-way interaction between time and forecast format was significant (F (2, 348) = 4.59, p = .03, η_G^2=.01). See figure 1. At the beginning, participants in the probabilistic condition rated the task as more challenging (M = 1.66, SD = 1.37) than did participants in the deterministic condition (M = 1.32, SD = 1.16). However, the pattern reversed at the end with participants in deterministic condition rating the task as slightly more challenging (*M* = 2.80, *SD* = 1.19) than participants in probabilistic condition did (*M* = 2.71, *SD* = 1.28). No other interactions reached significance. There was an unpredicted main effect for time (*F*(2, 348) = 112.95, *p* < .001, η_G^2=.16). The post-hoc pairwise comparison revealed that the mean challengingness rating was significantly lower at the beginning (*M* = 1.48, *SD* = 1.27) than in the middle (*M* = 2.43, *SD* = 1.20, *t*(177) = -10.32, *p* < .001, *d* = .78). The mean challengingness rating was significantly lower in the middle (*M* = 2.43, *SD* = 1.20) than at the end (*M* = 2.76, *SD* = 1.23, *t*(177) = -5.17, *p* < .001, *d* = .37). Thus, perceived challengingness increased with time. Moreover, although participants perceived probabilistic forecast to be more challenging in the beginning as predicted, that pattern disappeared by the end, regardless of the number of options.

Figure 1

The challengingness score based on forecast format over the course of time



Note. *The error bars represent +/- 1 standard error.*

Decision time

The challengingness rating is an explicit measure of cognitive load. However, cognitive load can also be measured implicitly. For instance, the time it takes to make a decision can reflect the amount of cognitive processing required or cognitive load (Rubinstein, 2007). Decision times were compared between two-option and three-option conditions. The average of the difference between the time the participant entered and left the decision page was calculated for each participant[3]. On average, participants in three-option condition took longer to make a decision (*M* = 4.64, *SD* = 2.23) than did participants in two-option condition (*M* = 3.53, *SD* = 1.25). An ANOVA was performed on mean decision time with forecast format (deterministic vs. probabilistic) and complexity (two- vs. three-option) as between groups factors revealed a main effect of decision complexity ($F(1, 174) = 16.90$, $p < .001$, $\boxed{\eta_G^2 = .09}$). Neither the main effect of forecast format nor the interaction effect was significant. Thus, the decision time analysis demonstrated that participants took longer to make a decision in three-option condition compared to two-option condition, suggesting that three-option condition increased the cognitive load.

Decision quality: expected value

Next, we examined decision quality to determine whether it was reduced as cognitive load increased in the three-option condition or increased when probabilistic forecasts were included. On each trial, there was an optimal decision based on comparing the expected value to the cost of closing or delay. Since we are using cost-loss scenario (no gains are possible), we will refer to expected value as “expected loss” henceforth. Calculating the expected loss of participants’ decisions allowed us to determine how it compared to the optimal decision. Expected loss was calculated for every trial. When participants advised school closure, the cost of closing school (-2 points) was assigned in both two- and three-option condition. In three-option condition, when the participants decided to delay one hour, the expected loss was the cost

---

[3] From the total of 10,858 trials from 178 participants, the time stamp data of 70 trials, which is 0.6 percent, were lost and not included in the analysis.

to delay (-1) added to the product of the possible penalty for not closing (-4) and the probability for snow accumulation of six inches or more (see equation 5). When participants decided to stay open, the expected loss was the sum of the expected loss of the two possible outcomes, snow accumulating six inches or more and snow accumulating one inch or more but less than six inches (See equation 4). In two-option condition, when the participant decided to stay open, the expected loss was the product of the probability of snow accumulating six inches or more and the penalty (-8). See equation 6.

Three-Option Condition Expected Value

When stayed open: (-8) * (prob 6+”) + (-2) * (prob 1+” – prob 6+”) (4)

When delayed 1 hour: (-4) * (prob 6+”) + (-1) (5)

Two-Option Condition Expected Value

When stayed open: (-8) * (prob 6+”) (6)

For each participant, expected loss (and/or cost) for every trial was summed. Then, the participant’s expected loss was subtracted from the sum of optimal expected loss. The optimal choice was the one on a given trial with the smallest loss. The sum of the optimal expected loss over all trials was -63.36 in the two- and -85.78 in the three-option conditions. If a participant followed the optimal strategy, the difference score would be zero, otherwise it would be negative. Thus, the smaller the difference, the closer performance is to the optimal strategy.

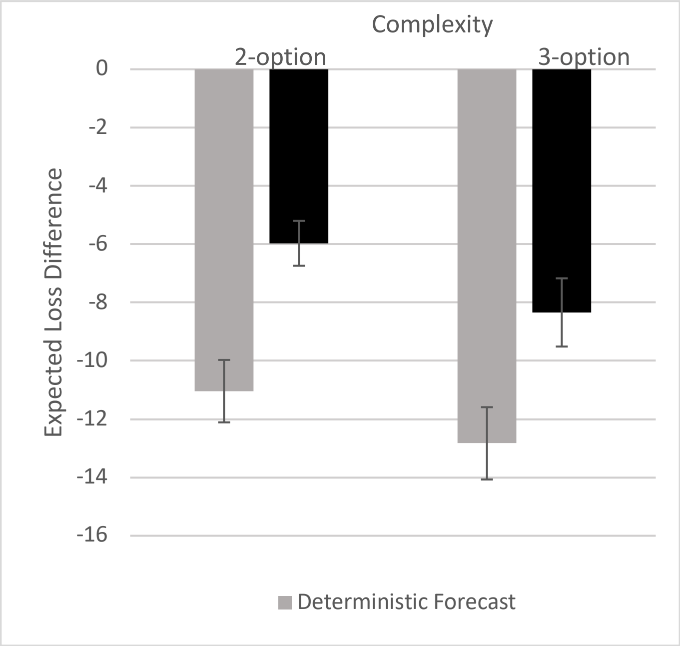
Supporting our hypotheses, participants with probabilistic forecasts had smaller difference in expected loss than participants with deterministic forecasts, indicating better decision quality. A two-way ANOVA was conducted on mean difference score with forecast format (deterministic vs. probabilistic) and complexity (two- vs. three-option) as the between groups factors. There was a significant main effect for forecast format ($F(1, 174) = 19.50$, $p <$

.001, $\boxed{\eta_G^2 = .10}$). The difference score of participants in probabilistic forecast condition (*M* = -7.18, *SD* = 6.43) was significantly smaller (better) than that of participants in deterministic forecast condition (*M* = -11.88, *SD* = 7.77). There was also a marginally significant main effect of complexity (*F*(1, 174) = 3.65, *p* = .058, $\boxed{\eta_G^2 = .02}$). The difference score of participants in two-option condition (*M* = -8.70, *SD* = 6.33) was slightly smaller (better) than that of participants in three-option condition (*M* = -10.56, *SD* = 7.94). The interaction did not reach significance, see figure 2.

Thus, as predicted participants made better quality decisions with probabilistic than with deterministic forecasts. In addition, participants in three-option condition performed worse in terms of the expected value than in the two-option condition.

Figure 2

Analysis of the expected loss



Note. *Difference score between the optimal expected loss and each participant's expected loss. The shorter bar indicates better performance. The error bars represent +/- 1 standard error.*

Error Analysis

In order to determine whether participants made more risk seeking errors as is predicted by prospect theory (Kahneman & Tversky, 1979), a series of exploratory analyses was conducted on errors. Any decision that was not optimal was counted as an error. Errors were categorized as either risk seeking or risk averse. Risk seeking errors were decisions for which the chosen option cost less than the optimal option but had a bigger potential penalty. On the other hand, risk averse errors were decisions for which the chosen option had a larger cost than the expected loss of the optimal option.

To determine which error was more prevalent, the proportion of each error type was calculated for each participant. The proportion of risk seeking (averse) errors was calculated by dividing the number of risk seeking (averse) errors by the total number of trials in which making a risk seeking (averse) error was possible. Therefore, the trials for which the optimal choice was to stay open were not included in risk seeking errors because that type of error was not possible (see table 3). Similarly risk averse errors only included trials for which the optimal strategy was to delay one hour and/or stay open.

Table 3

Categorization of decision errors in two- and three-option condition

		Participant Decision		
		Stay Open	Delay 1 Hour [a]	Close
Optimal Decision	Stay Open	Optimal Decision	Error (Risk Averse)	Error (Risk Averse)
	Delay 1 Hour [a]	Error (Risk Seeking)	Optimal Decision	Error (Risk Averse)
	Close	Error (Risk Seeking)	Error (Risk Seeking)	Optimal Decision

[a] *Delay 1 hour option was only available in the three-option condition.*

When comparing these errors, participants in two-option condition made a higher proportion of risk seeking errors than risk averse errors, which aligns with the predictions of prospect theory (Kahneman & Tversky, 1979). However, the tendency was reversed in three-option condition such that participants made a higher proportion of risk averse errors than risk seeking errors. A mixed model ANOVA conducted on the proportion of errors with forecast format (deterministic vs. probabilistic) and complexity (two- vs. three-option) as between group factors and risk type (risk seeking vs. risk averse) as within group factors revealed a significant main effect of forecast format, with a higher proportion of errors in the deterministic (*M* = .30, *SD* = .18) as compared to the probabilistic forecast condition (*M* = .23, *SD* = .18), *F*(1, 174) = 25.77, *p* < .001, η_G^2=.05. In addition, there was a significant main effect of complexity such that the proportion of errors was higher in the three- (*M* = .33, *SD* = .17) than in the two-option condition (*M* = .21, *SD* = .18, *F*(1, 174) = 65.41, *p* < .001, η_G^2=.12). The main effect of risk type failed to reach significance. However, there was a significant two-way interaction between risk type and complexity (*F*(1, 174) = 20.16, *p* < .001, η_G^2=.07). In two-option condition the proportion of risk seeking errors (*M* = .26, *SD* = .22) was greater than risk averse errors (*M* = .16, *SD* = .12). Whereas in three-option condition the proportion of risk averse errors (*M* = .37, *SD* = .17) was greater than risk seeking errors (*M* = .29, *SD* = .15).

Thus, the error analysis revealed that participants with probabilistic forecasts made fewer errors overall. In addition, as the complexity of the task increased, participant made more errors and a greater proportion of these were risk averse errors instead of risk seeking errors.

Discussion

This experiment represents strong evidence that providing numeric uncertainty information can help people make better quality decisions not only in simple binary decision situations but also in more complex situations with an intermediate option. Participants with probabilistic forecasts made decisions with better expected value and fewer errors compared to participants with deterministic forecasts. Moreover, the degree of benefit conferred by numeric uncertainty information did not differ significantly between the simpler binary task and more complex task with three options. Although participants' decisions were not optimal from the economic point of view, providing numeric uncertainty information helped them to make decisions that were closer to that standard.

Here we have shown that the advantage for numeric uncertainty information is maintained as the number of options increases, increasing cognitive load. This contradicts the evidence provided by Hinson et al. (2003) that showed adding an option reduced decision quality. One possible explanation for the difference in results is that in our study, participants learned the outcome of their decisions immediately after they were made, whereas in the Hinson et al. (2003) task they did not. Feedback may have allowed participants in our study to assess and adjust their strategy, eventually coping better with the additional options.

Although the results presented here are encouraging, whether they extend to situations with even more options is yet unknown. The addition of an intermediate option in the present experiment increased the number of outcomes to consider by five, to a total of nine. However, it is possible that this did not constitute sufficient cognitive load to stress working memory capacity. Indeed, in real-life, much more complicated situations are often considered. This will be an important line to pursue in future research.

Interestingly, participants did not explicitly rate the three-option condition to be significantly more challenging than the two-option condition despite the fact that it took them significantly longer to make the decision. Thus, although there may well have been an increase in cognitive load that stressed working memory capacity, participants may not have been sufficiently aware of the increase for it to be reflected in self-reported challengingness measure. Similar relationships have also been observed between other explicit judgements and implicit measures such a memory confidence and memory accuracy (Berger & Herringer, 1991; Penrod & Cutler, 1995; Tomes & Katz, 1997). Indeed, the decision time analysis, which is a behavioral measure rather than self-report measure, demonstrated that participants in three-option condition took more than a second longer on average than participants in two-option condition, an increase of 31%. In addition, errors increased and expected value was slightly lower in the three-option condition suggesting that decision quality also declined. Taken together, this suggests that as expected, the increase in options in the task tested here stressed cognitive capacity. Despite this fact, there was no detectable decrease of the advantage for probabilistic forecasts. In both the two- and three-option conditions, people made better decisions when they had uncertainty forecasts than when they did not. Nonetheless, future research should test more complex situations, as well as other decision domains and populations, to determine whether there is a point at which providing uncertainty information does not benefit decision making.

Another interesting and unexpected result was the shift from a predominance of risk seeking errors in the two-option to risk averse errors in the three-option condition. According to prospect theory (Kahneman & Tversky, 1979), in many situations, people tend to prefer to take a chance than to incur a cost up front. Indeed, the error analysis revealed the expected risk seeking tendency among participants in two-option condition. However, when another option was added to make the task more cognitively challenging, participants made more risk averse errors compared to risk seeking errors. The finding is also in line with previous studies showing that increasing cognitive load, increases risk averse choices in both gain and loss scenarios (Deck & Jahedi, 2015) and across load producing techniques (Deck, Jahedi & Sheremeta, 2021). The current study contributes to this line of work by providing evidence that adding a third option, which increases the complexity of the primary task, also leads to risk averse tendency.

In sum, our findings suggest that providing numeric uncertainty information in the form of a probability can be beneficial in a more complex situation with three-options, as it is in a binary decision situation. Moreover, adding an intermediate option led participants to make a smaller proportion of risk seeking as compared to risk averse errors in this loss scenario, suggesting that an intermediate protection option may be beneficial where safety is a priority.

References

Bernoulli, D. (1954). Exposition of a new theory of risk evaluation. *Econometrica*, *22*, 23-36.

Baddeley, A. D. (1983). Working memory. *Philosohical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1110), 311-324.

Berger, J. D., & Herringer, L. G. (1991). Individual differences in eyewitness recall accuracy. *The Journal of Social Psychology*, *131*(6), 807-813. https://doi.org/10.1080/00224545.1991.9924667.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293-332.

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97-119. https://doi.org/10.1016/j.euroecorev.2015.05.004.

Deck, C., Jahedi, S., & Sheremeta, R. (2021). On the consistency of cognitive load. *European Economic Review*, *134*, 103695. https://doi.org/10.1016/j.euroecorev.2021.103695.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of educational psychology*, *100*(1), 223-234. https://doi.org/10.1037/0022-0663.100.1.223.

Grounds, M. A., & Joslyn, S. L. (2018). Communicating weather forecast uncertainty: Do individual differences matter? *Journal of experimental psychology: applied*, *24*(1), 18-33.

Grounds, M. A., Joslyn, S., & Otsuka, K. (2017). Probabilistic interval forecasts: An individual differences approach to understanding forecast communication. *Advances in Meteorology*, *2017*, 1-18.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010, September). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Copenhagen, Denmark. https://doi.org/10.1145/1864349.1864395.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 298-306.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*(3), 302-323.

Joslyn, S. L., & Grounds, M. A. (2015). The use of uncertainty forecasts in complex decision tasks and various weather conditions. *Journal of Experimental Psychology: Applied*, *21*(4), 407-417.

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of experimental psychology: applied*, *18*(1), 126-140.

Kahneman, D. Tversky. A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336-358.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of experimental psychology: General*, *130*(2), 169-183.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81-97.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209-227.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, *38*(1), 63-71. https://doi.org/10.1207/S15326985EP3801_8.

Payne, J. W. & Bettman, J. R. (2004). Walking with the Scarecrow: The information-processing Approach to Decision Research. D. J. Koehler & N. Harvey (Eds.). *Blackwell Handbook of Judgment and Decision Making* (pp. 110-132). Oxford: Blackwell Publishing Ltd.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, *1*(4), 817-845. https://doi.org/10.1037/1076-8971.1.4.817.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal*, *117*(523), 1243-1259. https://doi.org/10.1111/j.1468-0297.2007.02081.x.

Savelli, S., & Joslyn, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, *27*(4), 527-541.

Tomes, J. L., & Katz, A. N. (1997). Habitual susceptibility to misinformation and individual differences in eyewitness memory. *Applied Cognitive Psychology: The Official Journal* *of the Society for Applied Research in Memory and Cognition*, *11*(3), 233-251. https://doi.org/10.1002/(SICI)1099-0720(199706)11:3<233::AID-ACP447>3.0.CO;2-V.

Whitney, P., Rinehart, C. A., & Hinson, J. M. (2008). Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic bulletin & review*, *15*(6), 1179-1184