Displaying Uncertainty: Investigating the Effects of Display Format and Specificity

Ann M. Bisantz, Stephanie Schinzing Marsiglio and Jessica Munch

Human Factors: The Journal of the Human Factors and Ergonomics Society 2005 47: 777
DOI: 10.1518/001872005775570916

Displaying Uncertainty: Investigating the Effects of Display Format and Specificity

Ann M. Bisantz, Stephanie Schinzing Marsiglio, and **Jessica Munch,** University at Buffalo, Buffalo, New York

We conducted four studies regarding the representation of probabilistic information. Experiments 1 through 3 compared performance on a simulated stock purchase task, in which information regarding stock profitability was probabilistic. Two variables were manipulated: display format for probabilistic information (blurred and colored icons, linguistic phrases, numeric expressions, and combinations) and specificity level (in which the number and size of discrete steps into which the probabilistic information was mapped differed). Results indicated few performance differences attributable to display format; however, performance did improve with greater specificity. Experiment 4, in which participants generated membership functions corresponding to three display formats, found a high degree of similarity in functions across formats and participants and a strong relationship between the shape of the membership function and the intended meaning of the representation. These results indicate that participants can successfully interpret nonnumeric representations of uncertainty and can use such representations in a manner similar to the way numeric expressions are used in a decision-making task. Actual or potential applications of this research include the use of graphical representations of uncertainty in systems such as command and control and situation displays.

INTRODUCTION

Despite extensive research on decision making under uncertainty, and on human difficulties in reasoning and making decisions using probabilistic information (Kahneman, Slovic, & Tversky, 1982; Nisbett, Krantz, Jepson, & Kunda, 1983), important questions remain regarding how uncertainty should be represented to decision makers. Answers to such questions have clear implications for the design of information displays that support decision making. Research in scientific visualization, geographic information systems, and linguistic description of probabilities has addressed the representation of uncertainty. However, the representations studied in that research have had either limited evaluation regarding their effectiveness or impact on decision making or limited extension to decision support applications. Human factors studies on the effect of graphical representations of uncertainty (Andre & Cutler, 1998; Kirschenbaum & Arruda, 1994) on decision making have provided preliminary information suggesting that such displays may be useful in decision making; however, these studies have been relatively limited in scope.

Techniques in Uncertainty Visualization

Research based in scientific visualization and geographic information systems (GISs) has explored graphical parameters that could be used to code the uncertainty in large data sets, including attributes of scene geometry (e.g., color, shading, and bumpiness; Pang, Wittenbrink, & Lodha, 1997) and traditional graphic variables used in cartography (e.g., texture, color, orientation, and shape; MacEachren, 1992). Pang et al. (1997) suggested the use of glyphs (graphical forms such as arrows or vertical lines), in part because they free other graphical dimensions (color, texture) for other purposes. Glyphs have

Address correspondence to Ann M. Bisantz, 438 Bell Hall, Department of Industrial Engineering, University at Buffalo, The State University of New York, Amherst, NY 14020; bisantz@eng.buffalo.edu. **HUMAN FACTORS,** Vol. 47, No. 4, Winter 2005, pp. 777–796.

been used to represent magnitude and direction of winds and ocean currents along with the uncertainties in these dimensions. In one example, the general shape of the glyph was an arrow, with the width of the arrowhead representing uncertainty in heading and multiple arrowheads representing uncertainty in magnitude (Lodha, Sheehan, Pang, & Wittenbrink, 1996). Lodha et al. (1996) claimed that the use of such visualizations results in an integrated graphic that users interpret holistically. In this way, these graphics can be seen as examples of object displays, which have been proposed for the integrated display of system information (Bennett, Toms, & Woods, 1993; Carswell & Wickens, 1996). Other suggested techniques include pairs of graphics showing the data value and corresponding uncertainty (either simultaneously or alternating; MacEachren, 1992), animation (through the degree of motion), and sound (Pang et al., 1997). However, such representations have not been systematically evaluated.

Linguistic Representations of Uncertainty

There has been extensive research on the interpretation and use of linguistic expressions as uncertainty representations (Budescu & Wallsten, 1995; Wallsten & Budescu, 1995). Linguistic representations include phrases such as “likely” or “seldom.” Such phrases have been described as vague representations of uncertainty, as compared with precise numeric probability values (Wallsten, 1990; Wallsten & Budescu, 1995). Researchers have compared participants’ use of such expressions with that of numeric representations in placing bets and selecting gambles. General findings from such studies have indicated little difference in judgment performance across linguistic and numeric representations of probability values (Budescu & Wallsten, 1990; Budescu, Weinberg, & Wallsten, 1988; Erev & Cohen, 1990).

Additionally, research that has captured individual understandings of linguistic expressions using fuzzy membership functions (Budescu et al., 1988; Wallsten, Budescu, Rapaport, Zwick, & Forsyth, 1986) has suggested large individual differences in the interpretation and generation of linguistic expressions. Fuzzy membership functions describe the degree to which an individual believes that a particular linguistic expression is representative of a probability value. A probability value may have a high degree of membership for one phrase, a low degree of membership for a second, and an intermediate degree of membership for a third. Likewise, for a particular phrase, some probability values will have higher degrees of membership and some will have lower degrees. For example, Figure 1 shows a hypothetical set of membership functions for three linguistic expressions of probabilities (“unlikely,” “even chance,” and “probable”). Notice that the membership values for “unlikely” are highest for lower values of probability and taper off for higher values of probability. Membership values for “even chance”

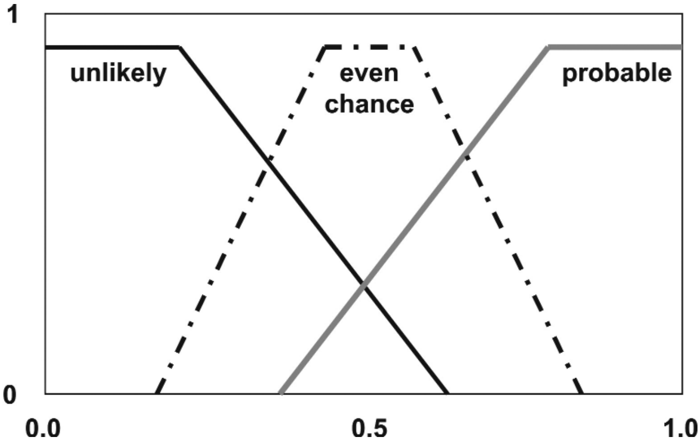


Figure 1. Hypothetical membership functions corresponding to three linguistic probability phrases.

are highest for the middle range of probabilities, and for "probable" they are highest for the higher values.

The use of membership functions to represent people's use of linguistic probability expressions has provided a theoretical basis from which to understand similarities in decision outcomes based on linguistic and numeric expressions (Wallsten et al., 1986). Researchers have theorized that when making decisions, people treat both numeric and linguistic representations of uncertainty in a vague way, which can be described using membership functions. They then sample from the probabilities most represented by the phrase or numeric expression (i.e., those with membership function values higher than some threshold), essentially converting the vague representations to a point estimate of uncertainty (Budescu & Wallsten, 1995). These point estimates are selected probabilistically (weighted according the value of the membership function) so that the expected value of the point estimate corresponds (for functions that are symmetric and unimodal) to the probability with the maximum membership value. Thus, in the hypothetical case shown in Figure 1, when faced with the term "even chance," people will reason using the probabilities that have the maximum membership value (here, values around $p = .5$). Similarities in decision outcomes between linguistic and numeric representations are thus attributable to similarities in understanding the vagueness underlying these representations (Wallsten & Budescu, 1995).

Graphical Uncertainty Representations and Decision Making

Finally, a limited number of studies have investigated the effect of graphical representations of uncertainty on decision making in environments with some elements of complexity. Andre and Cutler (1998) found that using rings proportional in size to the level of uncertainty regarding the location of an entity improved performance in a collision avoidance task. They also found that numeric, arc, and ring representations of uncertainty regarding an aircraft's heading improved decision-making performance in a simulated antiaircraft task, compared with a no-aid condition, under conditions of high uncertainty. Kirschenbaum and Arruda (1994) compared linguistic and graphical (ellipse-shaped) representations of uncertain submarine locations and found some advantage in estimating distance to target for the graphical representation, also for situations with greater uncertainty. These studies have provided a preliminary indication that graphical representations of uncertainty may provide value to decision makers and that the impact may be greater under conditions of higher uncertainty.

Finger and Bisantz (2002) tested degraded or blurred icons to convey uncertainty regarding the identity of a radar contact as hostile or friendly. An initial classification study found that participants could sort, order, and rank icons from five sets intended to represent different levels of uncertainty. A second experiment in which participants had to identify the status of radar contacts as either hostile or friendly represented the contacts and probabilistic estimates of their identities in one of three ways: with degraded icons and probabilities, with nondegraded icons and probabilities, and with degraded icons only. Results showed that task performance was similar across representations, with some advantages (for speed and correct identifications) for the condition with only degraded icons.

Similar to the categorization of linguistic uncertainty expressions, degraded graphical formats can be considered vague representations of uncertainty because they do not present explicit numeric information. The use of blurring or degradation is consistent with observations made in earlier research. For instance, MacEachren (1992) suggested using blurred lines to indicate uncertain boundaries, blurred images or fill to indicate uncertainty in a type of land cover, and lower levels of resolution or detail to represent less certain data. Recommendations from research in linguistic representations of probabilities have suggested that the presentation of uncertainty should be "as precise as warranted by the available information" (Wallsten, 1990, p. 28).

Research Goals

The current research had the following goals: (a) to replicate and extend the study reported in Finger and Bisantz (2002) by comparing degraded icon sets with other graphical representations (colored icons), linguistic representations, and combinations of graphical, linguistic, and numeric representations; (b) to investigate differences in the utility of the various representations when tasks require a graduated response, rather than a yes/no decision, and when the specificity level of the uncertainty (the degree to which probabilities are mapped to more or less precise ranges) varies; and (c) to collect and evaluate membership functions associated with the graphical representations to see if they are similar to those of linguistic expressions and, therefore, to see if theories regarding the use of vague, linguistic representations might be appropriately applied to vague, graphical representations. In particular, although research in linguistic representations of probabilities has noted strong individual differences in how the expressions are interpreted, iconic representations may be more reliably interpretable because they are systematically developed to map to a range of probabilities and have a natural mapping to the meaning of changing uncertainty.

To accomplish these goals, we conducted four experiments. Experiments 1 through 3 studied differences in participants' decisions to purchase stocks when information about the chance that a stock would be profitable was presented using different formats. In Experiment 1 participants made yes/no decisions, whereas in Experiment 2 participants could allocate an amount of funds to purchase each stock. In contrast to the first two experiments, Experiment 3 used a within-subjects rather than a between-subjects design to compare display formats. Finally, in Experiment 4, participants generated membership functions corresponding to their understanding of the different representations.

EXPERIMENT 1

Method

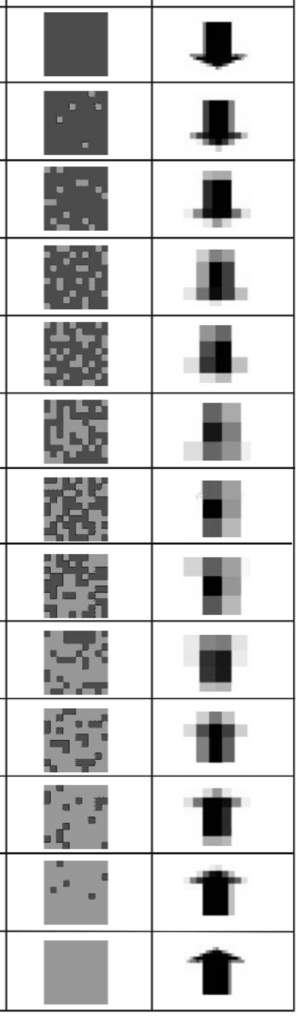
Participants. Sixty participants (47 men, 13 women, mean age 23 years) from the university were paid \$6.50/hr. All reported normal or corrected-to-normal vision and English fluency, and all had completed at least four semesters of college or were at least 20 years of age.

Experimental task. Participants were asked to make yes/no decisions in a simulated stock purchase task based on a display showing information about 24 stocks. The information indicated the probability that the stock would be a profitable investment. (The degree of profitability, or expected profit, was not provided.) The display format varied by condition.

Of the 24 stocks in a trial, 12 were randomly assigned to be profitable and 12 were assigned to be unprofitable. Additionally, each stock was assigned an initial probability to be displayed, which changed every 15 s (20 times in a 300-s trial), as follows. In order to simulate a situation in which information becomes more certain over time, we made the probability associated with each stock tend, with some randomness, to change to reflect the stock's actual profitability. For profitable stocks, the probability increased 70% of the time and decreased 30% of the time; the opposite was true for unprofitable stocks. Thus any one change was not completely predictive of the true nature of the stock. The maximum and minimum probabilities were .95 and .05, so participants never had complete certainty about the nature of a stock.

Scoring. Participants were told to earn as much money as possible over the course of the experimental trial. They were given an initial budget of \$4800. Each purchase cost \$200, and each stock could be purchased only once. Participants gained or lost money on their investment based on the actual status of a stock as profitable or unprofitable: \$10 was gained (or lost) on each purchase at every update after the stock was purchased. Funds earned or lost were therefore directly proportional to the stock purchase time. A profitable stock purchased at the start of a trial would be worth \$400 at the end, whereas an unprofitable stock purchased at the same time would be worth nothing. This scheme was intended to encourage participants to make purchases early in the trial (so they could make more money) but also to wait until they were sure enough, based on the displayed information, that the stock was actually a profitable one.

Independent variables. There were two independent variables: display format and specificity level. Display format was the manner in which the probabilistic information about the stocks was displayed. Six display formats were tested: numeric, linguistic, color icon, arrow icon, arrow icon with linguistic expression, and arrow icon with numeric expression. The linguistic probability expressions were taken from Hamm (1991), who investigated a set of 19 linguistic expressions, selecting phrases that spanned the 0 through 1 range relatively evenly and had relatively small ranges of numeric interpretation across participants (standard deviations in probability ranging from .0 to .14) in prior research. For the current experiment, the phrases that mapped most closely to the ranges used here, based on the mean numeric values reported by Hamm (1991), were selected. The color icon set was identical to that used by Finger and Bisantz (2002), in which red and green pixels were randomly distributed within a square icon in a proportion corresponding to the level of certainty. The arrow icon set was created by blurring up or down arrow icons. Figure 2 shows four of the display formats used. The arrow with



Range (High)	Numeric Expression	Linguistic Expression	Colored Icon	Arrow Icon	
0	0%	Absolutely Impossible			** *
0-.09	5%	Rarely			
.09-.18	14%	Very Unlikely			**
.18-.27	23%	Fairly Unlikely			*
.27-.36	32%	Somewhat Unlikely			**
.36-.45	41%	Uncertain			
.45-.54	50%	Tossup			** *
.54-.63	59%	Better Than Even			
.63-.72	68%	Rather Likely			**
.72-.81	77%	Quite Likely			*
.81-.90	86%	Highly Probable			**
.90-1	95%	Almost Certain			
1.0	100%	Absolutely Certain			** *

Figure 2. Iconic, linguistic, and numeric representations used to convey uncertainty. All 13 representations were used in the high-level specificity condition, those marked with two asterisks were used in the medium-level specificity condition, and those marked with a single asterisk were used in the low-level specificity condition. The numeric expressions shown are those used for the high level of specificity. For the medium and low levels, the numeric expression reflected the midpoint of the appropriate ranges. For print purposes, the color icons are shown in black and gray; they appeared in red and green to participants.

linguistic expression and arrow with numeric expression formats were combinations of the individual formats, in which the numeric or textual labels were placed underneath the icons.

The second independent variable, specificity level, was the size of the ranges of probabilistic information to which icons or expressions were mapped. The size of the ranges was varied across three specificity levels. In the high level, icons or expressions were mapped to 11 possible ranges, along with endpoints of p = .0 and p = 1.0. The span of the ranges was .09. This set of ranges was chosen to allow comparison with and reuse of icons from a previous study (Finger & Bisantz, 2002). For the medium level, every other icon or expression was used, mapping to five possible ranges with a range span of .2, along with the endpoints. For the low level, every third icon or expression from the set of 13 was used, mapping to three possible ranges (spanning .33) and the endpoints. Figure 2 shows the icons and expressions used, and the corresponding ranges, for the highest level of specificity. Starred rows indicate the icon or expressions used for the medium (two stars) and low (single star) levels of specificity. For all representations involving numeric expressions, the number represented the midpoint of the probabilistic range expressed as a percentage. Numbers in the figure correspond to those used in the highest specificity level; for the medium and low levels, the numbers were adjusted accordingly.

Dependent variables. The dependent variables reflected task performance and the information presented to participants at the time they made purchase decisions. Dependent variables were amount of money made (i.e., the score on the task), number of stocks purchased, time period (of the 20 15-s periods in each trial) in which each stock was purchased, the count of icons or expressions (e.g., displayed probability) showing at the time of purchase, and the actual probability associated with a stock at time of purchase (which varied within the range of probabilities represented by the icon or expression).

Experimental design. Participants were randomly assigned to one of six display format conditions. Each participant completed 10 trials under all three levels of specificity, within one display condition. Level of specificity was counterbalanced across six possible orders.

Experimental apparatus. The experiment was coded in Visual Basic 6.0 and used a 300-MHz personal computer with a 17-inch (43.2-cm) color monitor. Participants made selections of stocks to purchase using a standard keyboard and mouse.

Procedure. Each participant performed the experiment for 1 hr per day on 3 consecutive days. Task instructions included the set of icons or expressions to be tested, with the two endpoint icons labeled "0% chance of being a profitable stock" and "100% chance of being a profitable stock"; participants could refer to this during the task. Participants then performed 1 practice trial, after which the results were explained and any questions were answered. Subsequently, they performed 10 5-min trials. The initial probabilities associated with the stocks and true nature of the stocks as profitable or unprofitable were randomly assigned across trials, but they remained consistent across display conditions and participants. Trials were repeated (in different orders) on the 2nd and 3rd days at different specificity levels.

Results

Dependent measures for each trial were analyzed using a mixed effects analysis of variance (ANOVA). Results are summarized in Table 1. Overall, there were no main effects of display format. However, there were effects attributable to the level of specificity as well as limited interaction effects. Level of specificity did have a significant effect on the time interval in which stocks were purchased. Post hoc tests showed that participants made purchasing decisions significantly sooner in the lowest level of specificity than in the two higher levels of specificity (mean periods = 5.36, 5.80, and 5.96 out of 20 periods in each trial, SD = 2.25, 2.31, and 2.32, for the low, medium, and high levels of specificity, respectively). Level of specificity also had a significant impact on the level of uncertainty associated with each stock at the time of purchase. Post hoc tests indicated that as the level of specificity increased, stocks were purchased at increasingly higher levels of probability across the low, medium, and high levels of specificity (mean probabilities = .74, .77, and .79, SD = .078, .097, and .102, respectively). There was a trend toward an interaction showing increases

TABLE 1: *F* Values Across Dependent Measures for Experiment 1

Effect	Money Earned	Time Interval	Level of Uncertainty	No. of Stocks Purchased
Display	*ns*	*ns*	*ns*	*ns*
Specificity	*ns*	21.5***	90.75***	20.895***
Interaction	*ns*	*ns*	1.82*	1.991**

*Marginally significant at $p \leq .10$; **significant at $p \leq .05$; ***significant at $p \leq .01$.

in probability across all display conditions from the lowest to the medium to the highest level of specificity, except for the arrow with numeric condition, for which the medium and high levels were similar.

We also examined the number of stocks purchased. Similar to the results given previously, level of specificity had a significant impact. Post hoc tests indicated that as the level of specificity increased, participants tended to purchase a smaller number of stocks, across all three conditions (mean number = 13.38, 12.75, and 12.43, *SD* = 3.28, 3.62, and 3.51, for the low, medium, and high levels of specificity, respectively). There was also a significant display format by level interaction: Although display formats showed a pattern across specificity levels similar to the main effect of levels, two of the display conditions showed slightly different patterns in that the medium level of specificity either showed a result similar to that of the lowest level (for the arrow condition) or was least overall (for the arrow with numeric condition).

Finally, we investigated the particular icon that was being shown at the time a purchase was made. Figure 3 shows the number of stocks (summed over conditions) that were purchased when their associated icon or representation displayed the probability levels indicated along the *x* axis, for each of the three levels of specificity. Chi-square goodness of fit tests indicated that the number of stocks purchased per representation differed for all three levels of specificity ($p \leq .001$). Figure 3 indicates that participants did wait until the display indicated more certainty and tended to wait until the level of certainty had reached some threshold point. For example, 88%, 93.9%, and 86% of stocks in the low, medium, and high specificity levels, respectively, were purchased when the representation corresponded to a level of uncertainty of .6 or higher.

EXPERIMENT 2

Method

The methodology was identical to that of Experiment 1, apart from the following. There were 30 participants (17 men, 13 women, average age 23 years). The task differed from that of the first experiment, in which all stock purchases were set at $200. In Experiment 2, participants could choose the amount of funds to allocate to a purchase. The intent of this manipulation was to provide a judgment mechanism that could be more sensitive to display manipulations than was the dichotomous judgment tested in Experiment 1. That is, although the display format might not change people's decisions to purchase a stock once the probability reached a particular level, they could express their degree of comfort with the information format by changing the amount they were willing to invest. Correspondingly, scoring was altered so that that the money earned (or lost) on a stock was proportional to the amount of funds invested in the stock. Finally, there were small changes to the independent and dependent variables. Because we saw little difference among display formats, we tested only a subset of them: numeric expressions, arrow icons, and linguistic expressions (as shown in Figure 2). An additional dependent variable measured the amount of money allocated to each stock purchase. The experimental design was identical to that of Experiment 1.

Results

Statistical results are summarized in Table 2. Similar to Experiment 1, there were few effects of display but significant effects of level of specificity. The marginally significant effect of display on time of purchase suggested a trend for stocks in the numeric condition to be purchased soonest, followed by the linguistic and then the arrow

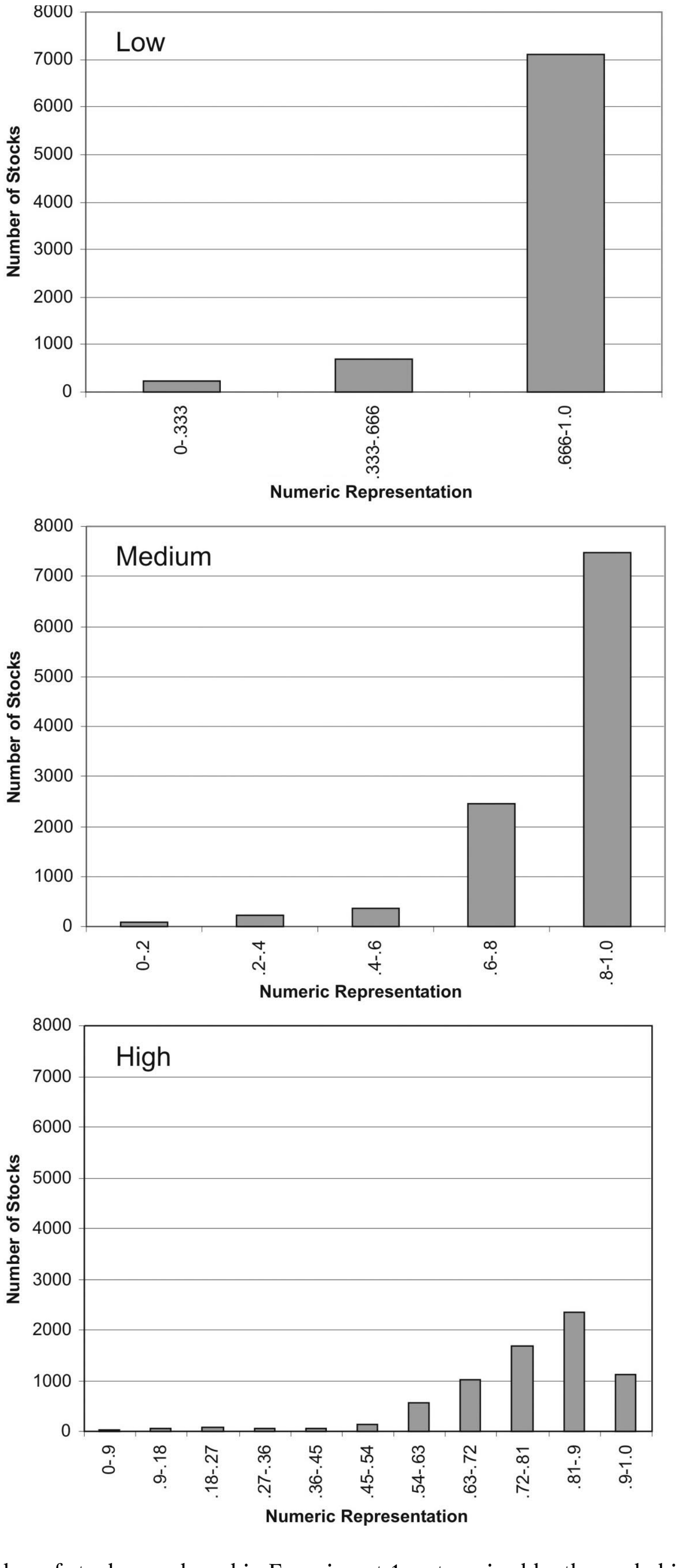


Figure 3. Total number of stocks purchased in Experiment 1, categorized by the probability being displayed at the time of purchase, for all three levels of specificity. Note that depending on the display condition, the probabilities listed on the *x* axis would have been displayed using icons, numeric representations, or linguistic representations.

TABLE 2: *F* Values Across Dependent Measures for Experiment 2

Effect	Money Earned	Time Interval	Level of Uncertainty	No. of Stocks Purchased	Funds Allocated
Display	*ns*	2.793*	*ns*	*ns*	*ns*
Specificity	11.338***	3.993**	55.087***	30.1***	8.59***
Interaction	*ns*	5.379***	2.229*	6.015***	3.88***

*Marginally significant at $p \leq .10$; **significant at $p \leq .05$; ***significant at $p \leq .01$.

condition. Post hoc tests on the effects of level of specificity indicated that money earned increased as specificity increased (mean scores = $831, $1086, and $1379, *SD* = $1379, $1640, and $1408, respectively) and that stocks were purchased at increasingly higher levels of probability as specificity increased (mean probabilities = .72, .77, and .80, *SD* = .126, .146, and .153). Also, participants made purchasing decisions significantly sooner in the lowest level of specificity than in the two higher levels (Figure 4); participants tended to purchase smaller numbers of stocks with increasing specificity (Figure 5); and participants tended to invest less per stock at the lowest level of specificity (Figure 6).

Interactive effects can be summarized as follows. For time of purchase, tests of simple effects and further post hoc tests indicated that for low specificity, the numeric condition differed from the linguistic and arrow conditions; at medium specificity, all three display types differed; and for high specificity, the numeric and arrow conditions differed. As shown in Figure 4, with additional information, participants in the numeric condition tended to wait longer before making stock purchases. The numeric condition tended to be faster than the other two conditions, although this effect diminished with greater specificity. Participants in the arrow condition waited longest overall.

For number of stocks purchased, simple effects and post hoc tests showed that the numeric and linguistic conditions differed from the arrow condition at all three specificity levels. Additionally, for the numeric and arrow condition, low specificity differed from medium and high levels, with no differences for the linguistic condition. Thus the number of stocks purchased declined as specificity increased, particularly for the arrow icon condition, and overall, participants

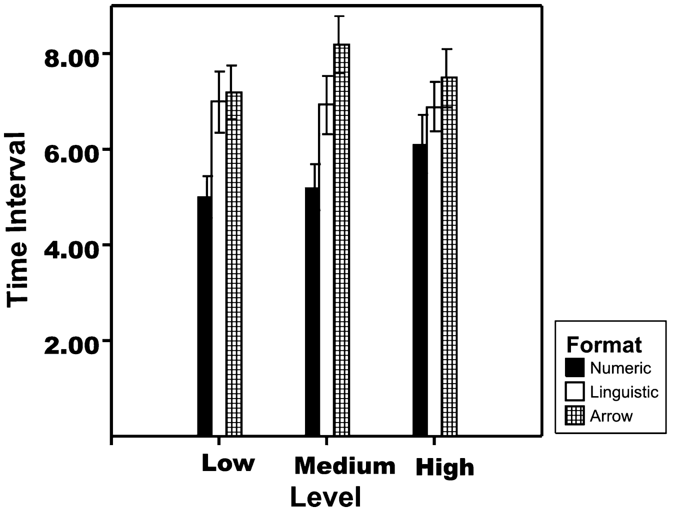


Figure 4. Average time interval and 95% confidence interval at which stock purchases were made in Experiment 2. There was a significant interaction between display format and level of specificity.

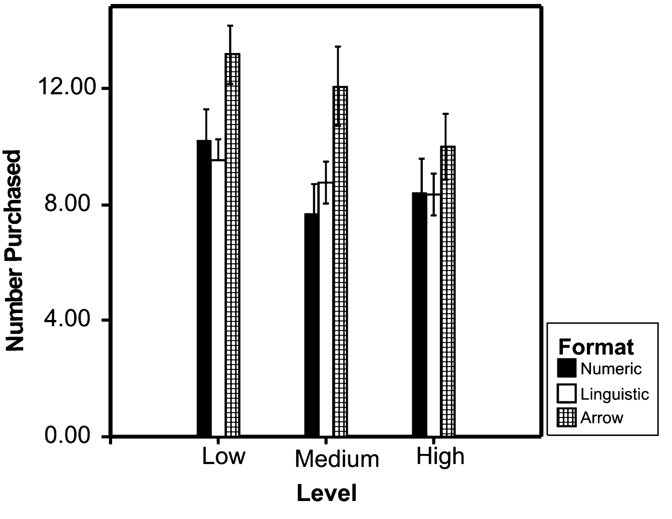


Figure 5. Average number of stocks purchased and 95% confidence interval in Experiment 2. There was a significant interaction between display format and level of specificity.

in the arrow condition purchased more stocks, perhaps indicating a tendency toward more conservative decisions in the graphical condition by spreading the risk over more stocks.

For funds allocated per stock, post hoc tests indicated that for the lowest level of specificity, the numeric and linguistic conditions were greater than the arrow condition; for the medium level of specificity, all display conditions differed; and for the highest level of specificity, the numeric condition was different from the linguistic and arrow conditions. Figure 6 indicates that participants in the numeric and arrow conditions allocated more funds per stock in the conditions in which information was more specific. Also, participants in the arrow condition tended to allocate less than did those in the numeric condition. Additionally, there was a nonsignificant interactive trend: Although average probability increased across all specificity levels, the

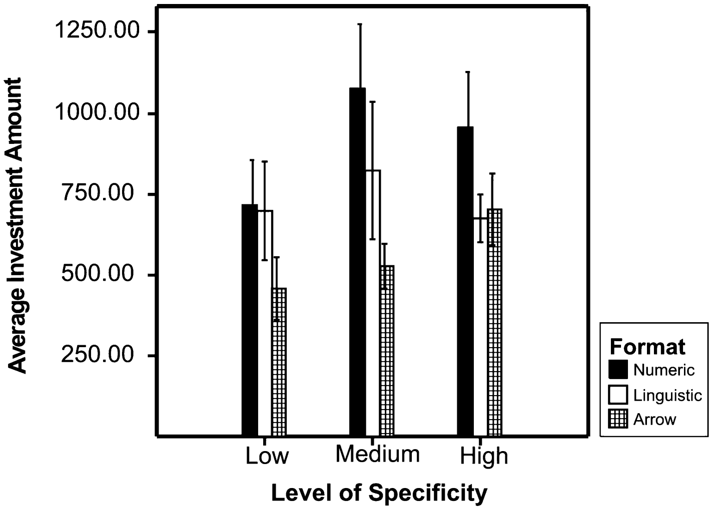


Figure 6. Average investment amount per stock and 95% confidence interval in Experiment 2. There was a significant interaction between display format and level of specificity.

effect was more pronounced for the numeric condition.

Finally, as in Experiment 1, we investigated the particular representation that was being shown at the time a stock was purchased. Figure 7 shows the number of stocks (summed over conditions) that were purchased when their associated icon or representation displayed the probability levels indicated along the *x* axis, for each of the three levels of specificity. Chi-square goodness of fit tests indicated that the number of stocks purchased per representation differed for all three specificity levels ($p < .0005$). Participants did wait until the displayed representation indicated more certainty and, in fashion similar to that seen in Experiment 1, made the majority of purchases when the icon or phrase represented a level of certainty greater than .6.

EXPERIMENT 3

Experiment 3 was identical to Experiment 1, with the following exceptions: Only display format was used as an independent variable (specificity level was fixed at the highest level of specificity), and the six display formats were tested as a within-subjects rather than a between-subjects variable. Twelve participants (5 women, 7 men, average age 24.25 years) were compensated $7.50/hr. Additionally, a small monetary bonus was offered for the highest scoring participant. Participants completed one 2-hr experimental session consisting of 30 experimental trials (5 of each of the six display conditions). Trial order was randomized. Levels of display format and trial were analyzed using a repeated measures ANOVA. Results showed no significant effect of display condition on any of the dependent variables.

EXPERIMENT 4

Method

In Experiment 4, participants generated membership functions associated with a subset of the display formats so that we could better understand how these formats influence performance. Participants provided membership functions for one display format that they were familiar with (through use in a previous experiment) and for two formats that were unfamiliar.

Participants. Ten participants from Experiment 2 (5 who had performed the linguistic condition and 5 who had performed the arrow condition) took part in Experiment 4 and were paid $6.50/hr.

Independent variables. Participants generated membership functions for the three display formats tested previously that did not include a numeric expression: arrow icons, colored icons, and linguistic expressions. All 13 icons or expressions for each format were used.

Experimental materials. Thirteen paper forms were created for each of the icons or expressions tested (169 forms). The icon or expression was printed on each form, along with one of 13 numeric probabilities in the format "probability = .5." The 13 probabilities were equivalent to the numeric expressions shown in Figure 2. Underneath the numeric probability, a line with endpoints labeled "not at all" and "absolutely" was printed. One form associating each icon or expression with each of the 13 numeric probabilities was created.

Experimental procedure. A paper and pencil "direct estimation" method was used to determine the membership functions (Rapaport, Wallsten, & Cox, 1987). Participants marked the point on the line that represented how well the icon or expression described the numeric probability. Participants who had been assigned to the linguistic expression group in Experiment 2 rated the linguistic expressions first, followed by the arrow and then the color icons, whereas participants assigned to the arrow icon group rated the arrow icons first, followed by the linguistic expressions and then the color icons. This ordering scheme was chosen so that the display format that was rated first reflected participants' experience in the experiment only, rather than in the other rating tasks. For each display format, participants assessed each of the 13 icons or expressions, which were presented in random order. For each icon or expression, the forms associated with the 13 probability levels were also randomly ordered. Other than their experience in the previous experiment, participants did not view the icons prior to performing the ratings.

Results

Membership functions were plotted for all

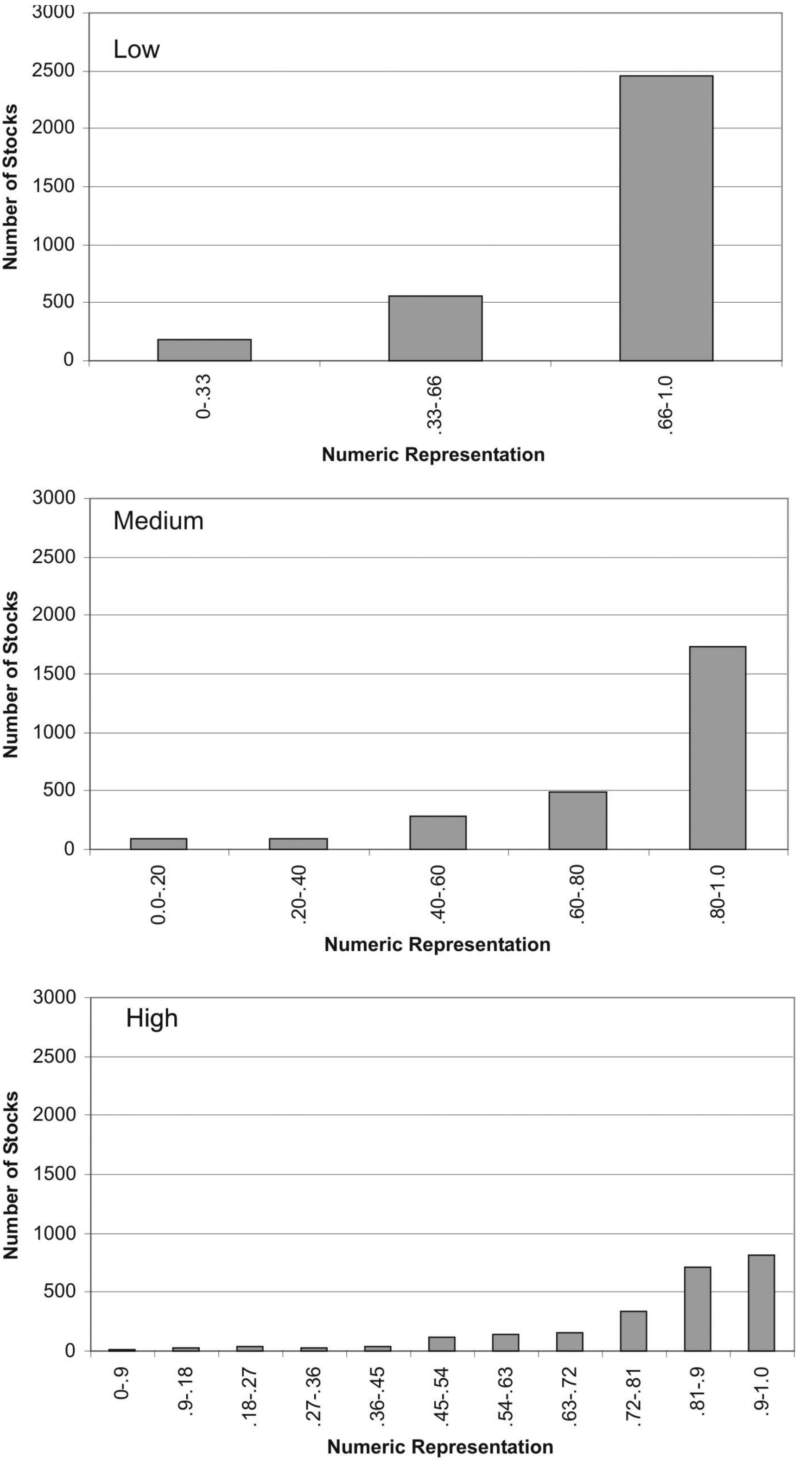


Figure 7. Total number of stocks purchased in Experiment 2, categorized by the probability being displayed at time of purchase, for all three levels of specificity, showing a pattern of responses similar to that in Experiment 1. Note that depending on the display condition, the probabilities listed on the *x* axis would have been displayed using icons, numeric representations, or linguistic representations.

participants for each icon or expression. Representative functions for the three representations tested are shown in Figure 8. Functions for representations intended to display five levels of probabilities (p = .0, .23, .5, .77, and 1.0) are shown (for all 10 participants) to illustrate the trends in changes in shape of the membership functions across the 13 representations tested. Note first that although there is individual variability, there is an overall trend for functions associated with lower probability representations to be monotonically decreasing, for functions associated with higher probability representations to be monotonically increasing, and for functions associated with medium levels of probability to be unimodal, with maximum values that vary based on the representation being evaluated. Figure 9 shows the membership functions

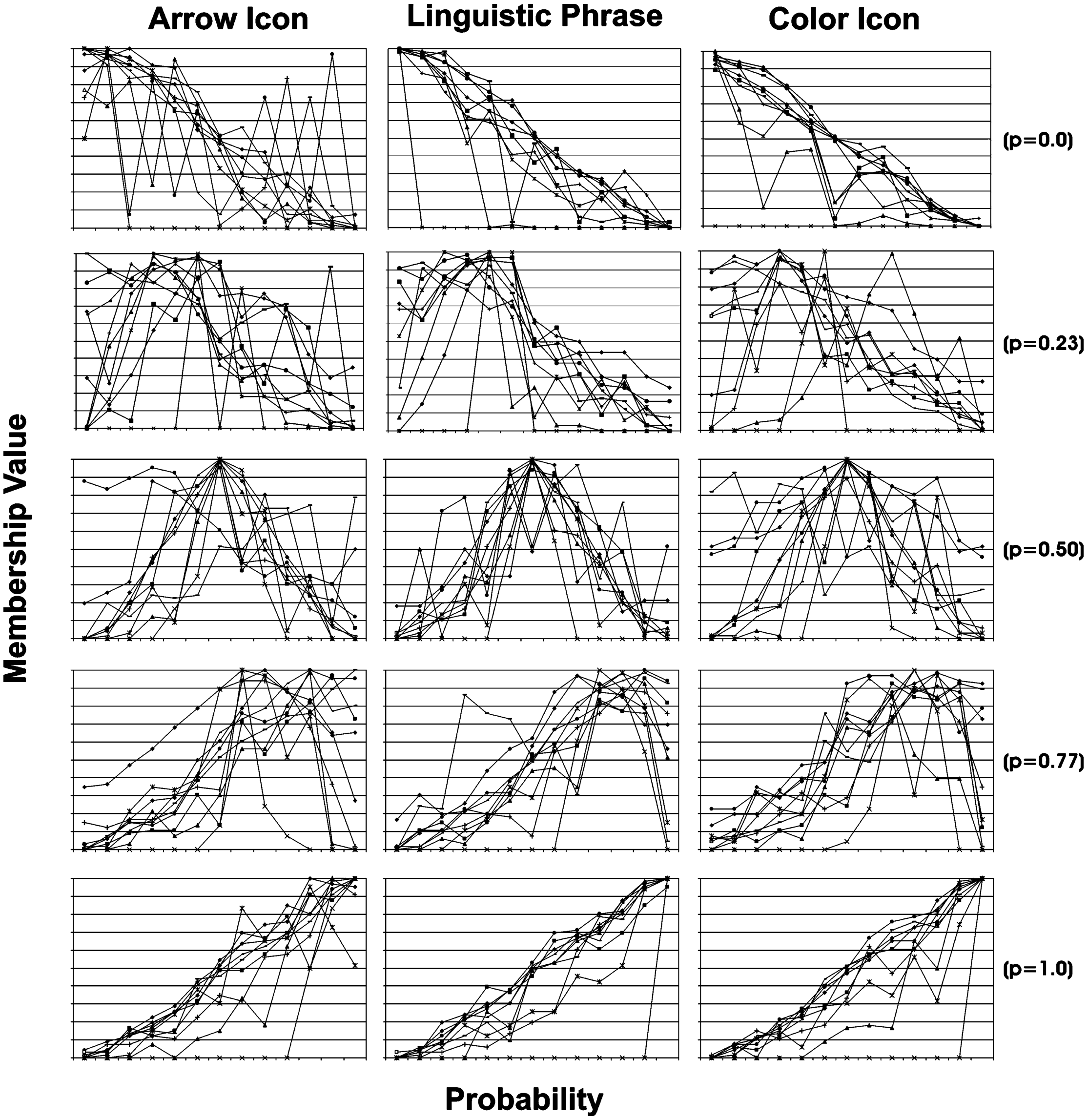


Figure 8. Membership functions for all 10 participants in Experiment 4 for five arrow, linguistic, and color representations. Membership values (on the y axis) ranged from 0 to 1. Probability levels (on the x axis) ranged from 0 to 1. From top to bottom, the graphs represent membership functions for the 1st, 4th, 7th, 10th, and 13th icon in the set. Target probabilities for the icons are listed to the right.

averaged across participants. These patterns are consistent across the three formats, indicating that participants mapped representations to probabilities in a similar fashion, regardless of format.

Additionally, the points of maximum membership values for each representation, for each participant, were analyzed to characterize the adequacy of the icons or phrases in conveying the intended probability level. The probability levels with maximum membership values were correlated with the mean of the probability range associated with each icon or phrase. All correlations were high and significant at the .01 level, indicating a strong relationship between the desired meaning of the representation and the point of maximum membership function value. Average squared correlations across participants,

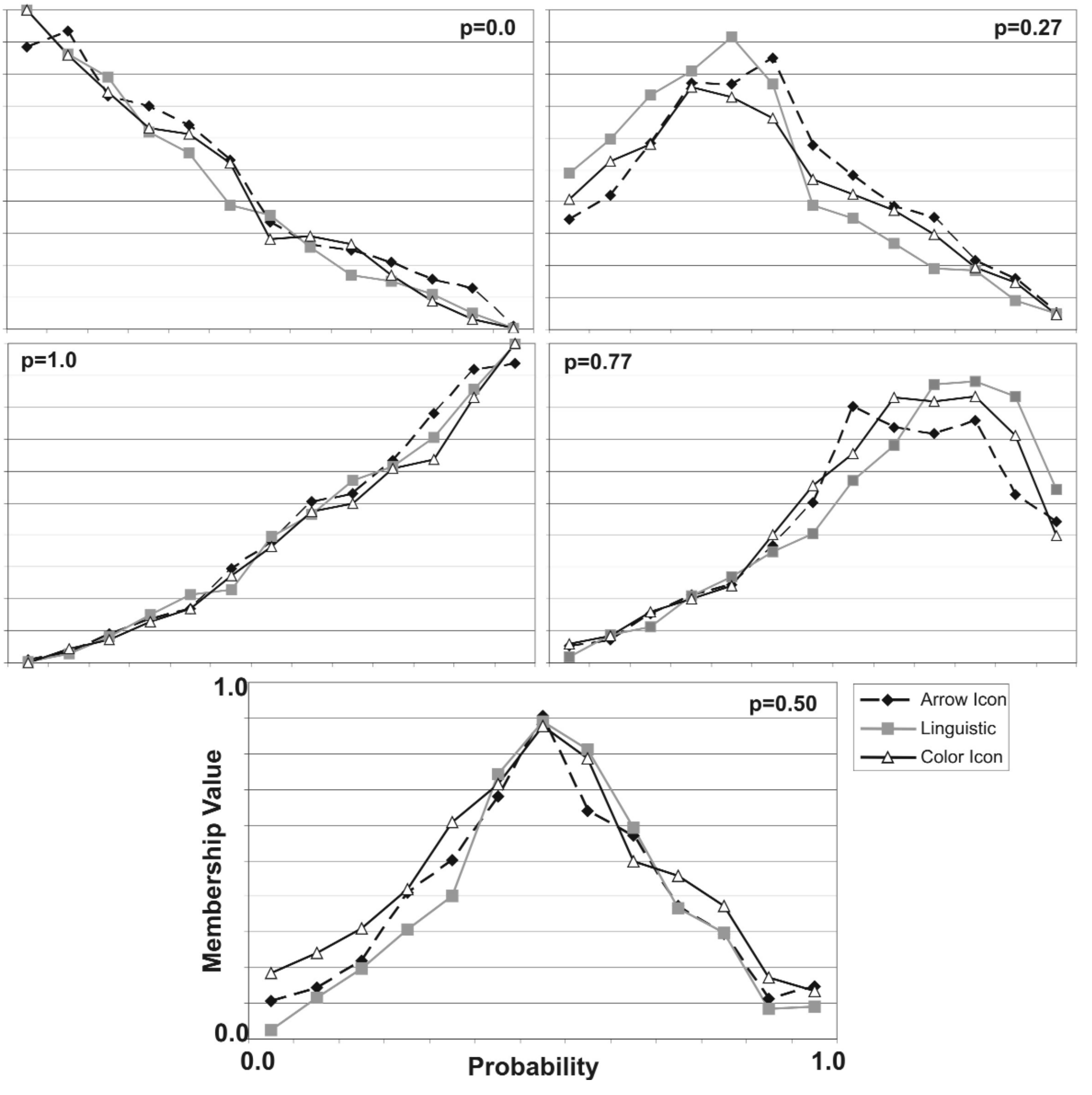


Figure 9. Membership functions averaged across Experiment 4 participants for five arrow, linguistic, and color representations. Membership values (on the *y* axis) ranged from 0 to 1. Probability levels (on the *x* axis) ranged from 0 to 1. The graphs represent membership functions for the 1st (top left panel), 4th (top right panel), 7th (bottom panel), 10th (middle right panel), and 13th (middle left panel) icon in the set. Target probabilities for the icons are listed at the upper right of each panel.

along with maximum and minimum correlations, are shown in Table 3. There was little difference attributable to display format or participants' previous experience (in terms of the display format they had used in Experiment 2).

DISCUSSION

Findings from these studies are consistent with those from related research, and they support recommendations for decision support system design and suggest directions for further research. Similarities in performance across display conditions are consistent with research comparing linguistic with numeric representations of uncertainty and can be explained within a similar theoretical framework. Thus these findings support the important conclusion that degraded graphical icons are a viable method for communicating uncertainty to decision makers. Also, the results indicate that task performance changed with the level of specificity, providing guidance for decision support system design in terms of requirements for information specificity.

Vague Uncertainty Representations and Decision Making

These research findings are consistent with those from a large body of research on linguistic representations of uncertainty, which has shown similarities in decision performance between linguistic and numeric representations (Budescu & Wallsten, 1995; Wallsten & Budescu, 1995; Wallsten, Budescu, & Erev, 1988). Like these studies, our experiments showed similarities in performance across "vague" display formats (graphical and linguistic) and "precise" formats (numeric expressions or graphical expressions annotated with numeric formats). Examination of Table 4, which summarizes effect size statistics (partial eta-squared values) for the display factors across the three experiments, also shows little impact of display format (only three cases were greater than .1). Although some other studies (Kirschenbaum & Arruda, 1994) did find some differences between graphical and other representations of uncertainty, these studies were concerned with spatial uncertainty (e.g., where something is located) rather than state uncertainty (e.g., what something is), indicating task-related differences in the impact of uncertainty representations.

Also, membership functions for graphical and linguistic representations had maximum values that were highly correlated to the intended numeric values. Researchers of linguistic probabilities (Budescu & Wallsten, 1995; Wallsten & Budescu, 1995; Wallsten et al., 1988) have hypothesized that in order to use linguistic representations in making decisions, people convert these representations (theoretically modeled as

TABLE 3: Correlations Between Intended Probability Value and the Maximum Value of Participants' Membership Functions

	Display Format		
Participant Experience	Arrow	Linguistic	Color
Linguistic			
Avg. squared correlation	.88	.90	.87
Max. correlation	.95	.97	.97
Min. correlation	.80	.71	.66
Arrow			
Avg. squared correlation	.87	.92	.96
Max. correlation	.96	.96	.99
Min. correlation	.67	.86	.93
All participants			
Avg. squared correlation	.88	.91	.91
Max. correlation	.98	.98	1.00
Min. correlation	.82	.85	.81

Note. Correlations are provided based on the display condition participants experienced in Experiment 2 and the display format in Experiment 4.

TABLE 4. Summary of Effect Size Statistics (Partial Eta Squared) Across Experiments 1–3

	Experiment		
Measure	1	2	3
Number bought	.045	.111	.087
Money earned	.053	.023	.094
Probability when bought	.077	.054	.142
Time purchased	.080	.171*	.025
Amount invested	n.a.	.092	n.a.

Note: n.a. = not applicable

*F test showed marginal significance at p = .07.

fuzzy membership functions) to numeric estimates with expected values equal to the probability with the maximum membership value.

These theories provide a basis for predicting the current results. If we assume that people's use of graphical representations of uncertainty is similar to their use of linguistic representations, then to use the vague graphical formats people must convert these representations to discrete numeric (i.e., point) estimates. Membership functions tended to be monotonically increasing, decreasing, or single peaked and had maximal values at probability levels very close to the intended level of probability – the numeric representation of uncertainty. Thus the conversion of linguistic and graphical formats to point estimates would have resulted in point estimates similar to each other and very close to the numeric representation. Therefore, task performance across the numeric, graphical, and linguistic representations should have been, and was, similar.

Finally, it is interesting to comment on the similarities among participants in the overall shape, and maximal values, of the membership functions. Researchers have cautioned that interpretations of linguistic probabilities show individual variations (Wallsten & Budescu, 1995), which may be indicated by differently shaped membership functions. Although we saw some individual variation, membership functions tended to be similar across participants, possibly because the arrow and color icons tested were created in a standardized fashion: The amount of blurring and the proportion of red and green pixels were controlled in creating the icons. Also, in this experiment we found similarities in membership functions even for the linguistic expressions. This may be because they were selected from the set investigated by Hamm (1991), which were chosen for their relatively small ranges of numeric interpretation across participants.

Specificity Level and Task Performance

Level of specificity, as manipulated in this study, can be considered a form of ambiguous (Curley & Yates, 1985), imprecise (Honekopp, 2003), or second order probability (Einhorn & Hogarth, 1985) – that is, uncertainty regarding a probability value expressed as a range or distribution. Researchers studying ambiguous probabilities have found that people show preferences for less ambiguity (e.g., smaller ranges) in probability values (Curley & Yates, 1985; Einhorn & Hogarth, 1985; Gardenfors & Sahlin, 1983; Honekopp, 2003; Wallsten, Budescu, & Tsao, 1997). For example, they will express a subjective preference for a gamble in which probabilities are expressed with more precision, or they will pay more for a lottery with lower ambiguity as compared with higher ambiguity, even if the expected values of the lotteries are equivalent. Researchers have suggested that lower levels of ambiguity are better because people may choose actions for which the chance of a desirable outcome is known more precisely, even if that choice has a lower expected value than do options with more ambiguity.

Three related conclusions can be drawn from the results of this study. First, specificity level can impact decision performance, in addition to impacting subjective preference. Participants waited longer to make purchases, purchased fewer stocks, and purchased stocks with higher levels of certainty, on average, as the level of specificity of the information they had increased. Honekopp (2003) examined participant preferences for lotteries in which the level of ambiguity of the probability of winning the lottery was varied. He found that as the level of ambiguity rose, participants tended to base their preferences (or rank orders) for lotteries on the lottery values, rather than on the expected winnings. Thus participants essentially treated all of the lotteries as having equivalent probabilities. This is consistent with the findings of the present study in that participants in the lower specificity (more ambiguous) levels purchased more stocks per trial, thus spreading out their risk across more stocks and thus treating more stocks as identical with respect to their chance of profitability.

Second, participants at all levels of specificity tended to wait until stocks had reached some threshold probability before making purchase decisions. This threshold was higher when more specific information was available. Examination of Figures 3 and 7 indicates that although the majority of purchases in all conditions were made once the probability had reached .6, decisions tended to be made at a higher level of probability when more specific information was available (i.e., when it was possible to differentiate between a probability greater than .6 and probabilities that had reached higher levels; this was not possible, for instance, when there were only three levels of specificity). Also, participants did not simply wait until probabilities had reached their maximal values but made purchases across the range of values above .6. The time interval in which participants made purchases at higher levels of specificity thus would have been longer, on average, as they waited through more time intervals to see the probabilities (on average) increase to levels above .6 for stocks they eventually decided to purchase.

Third, there was evidence that specificity level impacts the efficacy of some display formats. Participants in the arrow condition of Experiment 2 purchased more stocks, investing less per stock, than did participants in the numeric condition, indicating a tendency toward more conservative decisions in the graphical condition by spreading the risk over more stocks. This tendency was reduced (i.e., the number of stocks purchased decreased and the investment amounts increased) as specificity increased. Thus, with more specificity, participants in the more vague, graphical condition acted more similarly to those in the precise, numeric condition.

Implications for Information System Design

Numerous application domains, including military planning, tactical command and control, and emergency planning and response, rely on uncertain situation information produced by sensors, human reports, and algorithms – uncertainty that must be communicated to decision makers. From our results, we can make recommendations for such information system design. First, in our studies, participants were sensitive to and able to take advantage of the increased number of probability levels resulting from conditions with greater specificity. This, combined with results from past research showing subjective preferences for less ambiguity, provides support for information system designs that gather, compute, and display probabilistic information at higher levels of specificity. Future studies would need to confirm whether there is an upper limit of specificity after which additional performance benefits are reduced.

Second, the overall similarities in performance across display modalities indicate that the graphical representations tested here are a viable method for communicating uncertainty to decision makers. Graphical representations may be substituted for numeric annotations (e.g., a standard icon annotated with a numeric probability), particularly when the probabilities are known at higher levels of precision. Alternatively, presentation of uncertain information in graphical formats could be provided to operators as one display configuration option, with numeric information about specific entities available on request.

Icons are used in command and control situation displays to indicate the location of objects as well as to provide information about their identity (through the shape of the icon). Using graphical forms that combine information about uncertainty with information about object state (e.g., identity) has the following advantages for such displays. First, research has shown negative impacts on performance when using displays in which there is close spatial proximity among relevant and irrelevant items (Andre & Wickens, 1988; Eriksen & Eriksen, 1974; Holahan, Culler, & Wilcox, 1978). It follows that graphical forms that eliminate some display items (i.e., numeric annotations) could reduce screen clutter by increasing space among items, thus positively affecting performance. Second, research in visual attention and perception has indicated that properties of visual elements in close proximity or that are part of the same object may be processed in parallel, resulting in performance effects such as redundancy gain, when the elements map to the same response, or response conflict, when the elements map to different responses (Eriksen & Eriksen, 1974; Kramer & Jacobson, 1991; Murphy & Eriksen, 1987). Wickens and Carswell (1995) and Carswell and Wickens (1996) described the proximity compatibility principle, which suggests that when tasks require information sources to be integrated, they are better supported by displays that integrate display components representing the required information. Display integration can be achieved through increased spatial proximity or by combining display components into a single object. Such displays may enhance performance when processing multiple properties of a single object is required (e.g., determining whether an object is a threat based on its location, identity, and information about how certain that identity is).

Extensions of this display technique to related types of uncertainty representations or tasks with different characteristics are also possible. For instance, both this study and that of Finger and Bisantz (2002) utilized a task in which the variable in question (e.g., stock profitability, aircraft identity) had two possible values (profitable or nonprofitable; hostile or friendly). However, in many circumstances an object may take on multiple states. Consider a military identification task in which the problem includes identifying an object as one of several types of weapons platforms. Here, icon sets could be constructed for each possible identity, representing probabilities from .0 to 1.0 that the object had that identity. Operators could elect to display all unknown objects according to their highest associated probability, or they could choose to display the probability that all the objects displayed were of a particular type (i.e., in order to quickly assess the objects that were most likely threatening). Likewise, the current work has focused on representing the probability that an object is in a discrete state (e.g., for the task studied here we did not represent the degree of profitability). To the extent that a continuously valued concept, such as an amount of profit, could be mapped to a discrete set of categories, a similar approach to that suggested for multiple states could be used. Such extensions are grounds for further study.

Implications for Future Research

Numerous directions for further research can be considered. First, the presence of the format by level of specificity interaction in the second experiment indicates that there may be some differences in how people use the different display formats, at lower versus higher levels of specificity, when they can make a continuously valued decision. Further research could investigate these task effects more fully. Also, the graphical representations used here can be considered examples of either natural mappings or population stereotypes (Norman, 1988), given that the notion of blurriness is related to an intuitive understanding of uncertainty. Also, for the stock purchase task, the colored icons with a greater degree of green indicated a good, or probably profitable, stock. Other graphical variables may also allow a natural mapping (e.g., degree of transparency, brightness of hue), whereas some may require more training (e.g., the use of varying graphical textures). Graphical representations of uncertainty may also vary in terms of their salience: For instance, icons that become more transparent or dimmer with greater uncertainty may be less salient, representations that map colors to uncertainty may be more or less salient with increasing uncertainty depending on other colors used in the display, and representations that map increasing size to increasing uncertainty (e.g., the glyphs or error ellipses mentioned in the Introduction) may have higher salience with greater uncertainty. Potential interactions among type of representation, salience, and task requirements require further investigation.

Participants reported little stock purchasing experience. Only 8 participants reported any relevant experience. Implementation and testing of these representations in more realistic decision support systems with experienced operators, along with investigation of operator acceptance of such displays, should be pursued. Also, the task investigated was not time pressured (information was updated only every 15 s). Research regarding response time effects attributable to display formats in time-pressured situations (perhaps because of factors related to screen clutter or object display advantages, as discussed previously) is necessary.

Other research could investigate the degree to which differing representations impact factors other than task performance, such as mental workload, decision confidence, or operators' situation awareness of uncertain variable values. For instance, corresponding to the three levels of situation awareness (SA) that have been described by Endsley (1995), different representations may impact the degree to which operators rapidly and accurately assess uncertainty related to the identity, location, and kinematic parameters of an aircraft; the corresponding threat level of an aircraft; and the future intentions or locations of an aircraft. Critically, in circumstances in which the state variables, interpretations, or projected future states have associated uncertainty, SA must encompass knowledge of that uncertainty. Application of the graphical representations studied here would be most appropriate for attribute or identity information (e.g., object type, threat level, type of future intent) rather than location information.

CONCLUSIONS

Evidence from these studies supports the conclusion that graphical representations are a viable means of communicating uncertainty to decision makers, as performance was similar whether participants were given numeric, linguistic, or graphical representations of the uncertainty regarding the state of an object (in this case, a stock). These results parallel those found in a prior study of degraded icons, and the overall trends that were found, comparing decision performance using linguistic representations with that using numeric representations of uncertainty. Membership functions descriptive of participants' interpretation of iconic, linguistic, and color representations indicated consistency in functions across display formats, supporting these conclusions. Results also indicate performance benefits from presenting uncertain information at higher levels of specificity.

ACKNOWLEDGMENTS

This research was based on work supported by the National Science Foundation under Grant No. IIS9984079 awarded to the first author. Portions were used in partial fulfillment of the M.S. degree of the second author.

REFERENCES

Andre, A. D., & Cutler, H. A. (1998). Displaying uncertainty in advanced navigation systems. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 31–35). Santa Monica, CA: Human Factors and Ergonomics Society.

Andre, A. D., & Wickens, C. D. (1988). The interaction of spatial and color proximity in aircraft stability information displays. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 1371–1375). Santa Monica, CA: Human Factors and Ergonomics Society.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and configural elements: Designing more effective configural displays. *Human-Computer Interaction, 35,* 71–97.

Budescu, D. V., & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior and Human Decision Processes, 46,* 240–263.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. Busemeyer, D. L. Medin, & R. Hastie (Eds.), *Decision making from a cognitive perspective* (pp. 275–318). San Diego, CA: Academic Press.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 281–294.

Carswell, C. M., & Wickens, C. D. (1996). Mixing and matching lower-level codes for object displays: Evidence for two sources of proximity compatibility. *Human Factors, 38,* 1–22.

Curley, S., & Yates, J. (1985). The center and range of the probability interval as factors affecting ambiguity preferences. *Organizational Behavior and Human Decision Processes, 36,* 273–287.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92,* 433–461.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 65–84.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 45,* 1–18.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics, 16,* 143–149.

Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision making task. *Theoretical Issues in Ergonomics Science, 3,* 1–25.

Gardenfors, P., & Sahlin, N. (1983). Decision making with unreliable probabilities. *British Journal of Mathematical and Statistical Psychology, 36,* 240–251.

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes, 48,* 193–223.

Holahan, C. J., Culler, R. E., & Wilcox, B. L. (1978). Effects of visual distraction on reaction time in a simulated traffic environment. *Human Factors, 20,* 409–413.

Honekopp, J. (2003). Precision of probability information and prominence of outcomes: A description and evaluation of decisions under uncertainty. *Organizational Behavior and Human Decision Processes, 90,* 124–138.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, UK: Cambridge University Press.

Kirschenbaum, S. S., & Arruda, J. E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors, 36,* 406–418.

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization of focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics, 50,* 267–284.

Lodha, S. K., Sheehan, R. E., Pang, A. T., & Wittenbrink, C. M. (1996). Visualizing geometric uncertainty of surface interpolants. In W. A. Davis (Ed.), *Proceedings of the Conference on Graphics Interface '96* (pp. 238–245). Toronto: Canadian Information Processing Society.

MacEachren, A. (1992). Visualizing uncertain information. *Cartographic Perspective, 13,* 10–19.

Murphy, T. D., & Eriksen, C. W. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Perception and Psychophysics, 42,* 576–586.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *Visual Computer, 13,* 370–390.

Rapaport, A., Wallsten, T. S., & Cox, J. A. (1987). Direct and indirect scaling of membership functions of probability phrases. *Mathematical Modelling, 9,* 397–417.

Wallsten, T. S. (1990). The costs and benefits of vague information. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 28–43). Chicago: University of Chicago.

Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *Knowledge Engineering Review, 10,* 43–62.

Wallsten, T. S., Budescu, D. V., & Erev, I. (1988). Understanding and using linguistic uncertainties. *Acta Psychologica, 68,* 39–52.

Wallsten, T. S., Budescu, D. V., Rapaport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365.

Wallsten, T. S., Budescu, D. V., & Tsao, C. J. (1997). Combining linguistic probabilities. In R. W. Scholz & A. C. Zimmer (Eds.), *Qualitative aspects of decision making* (pp. 27–55). Miami, FL: Pabst Science.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37,* 473–494.

Ann M. Bisantz is an associate professor of industrial engineering at the University at Buffalo, the State University of New York, in Buffalo, New York. She obtained her Ph.D. in industrial and systems engineering from the Georgia Institute of Technology in 1997.

Stephanie Schinzing Marsiglio is a lead engineer at IBM for the Lean Manufacturing Transformation Team, Microelectronics Division, in Burlington, Vermont. She received an M.S. in industrial engineering from the University at Buffalo, the State University of New York, in 2002.

Jessica Munch is a human factors engineer in the Human Systems Integration Division at the Naval Surface Warfare Center, Dahlgren Division, Dahlgren, Virginia. She obtained her master's degree in human factors engineering from Wright State University in 2003.

Date received: April 17, 2002
Date accepted: May 16, 2005