Published in final edited form as:

Risk Anal. 2014 August ; 34(8): 1495–1505. doi:10.1111/risa.12161.

The effect of cognitive load on decision making with graphically displayed uncertainty information

Pamela M. Allen¹, John A. Edwards², Frank J. Snyder³, Kevin A. Makinson^{4,*}, and David M. Hamby⁴

¹School of Social and Behavioral Health Sciences, Oregon State University, Corvallis, OR, USA

²School of Psychological Science, Oregon State University, Corvallis, OR, USA

³Department of Health and Kinesiology, College of Health and Human Sciences, Purdue University, West Lafayette, IN, USA

⁴Nuclear Engineering and Radiation Health Physics, Oregon State University, Corvallis, OR, USA

Abstract

An experiment examined the ability of 5 graphical displays to communicate uncertainty information when end users were under cognitive load (i.e., remembering an 8-digit number). The extent to which people could accurately derive information from the graphs and the adequacy of decisions about optimal behaviors based on the graphs were assessed across eight scenarios in which probabilistic outcomes were described. Results indicated that the load manipulation did not have an overall effect on derivation of information from the graphs (i.e., mean and probability estimation) but did suppress the ability to optimize behavioral choices based on the graph. Cognitive load affected people's use of some graphical displays (basic probability distribution function) more than others. Overall, the research suggests that interpreting basic characteristics of uncertainty data is unharmed under conditions of limited cognitive resources, whereas more deliberative processing is negatively affected.

Keywords

graphical communication; cognitive busyness; uncertainty; probability; decision making

1. INTRODUCTION

A great deal of work in recent years has sought to create methods to quantify uncertainty, typically in order to help decision-makers plan for probabilistic events such as the accidental release of hazardous substances.^(1–4) Such probabilistic estimation is generally carried out to provide the end-user with a sense of the range of potential consequence, for identifying the diagnosticity of a particular point estimate, or to help direct limited resources. Although uncertainty estimation models have grown increasingly sophisticated, human decision-

Address correspondence to Pamela M. Allen: School of Social and Behavioral Health Sciences, Oregon State University, Waldo Hall 402, Corvallis, OR, 97331, USA; allenpa@onid.oregonstate.edu..

*Kevin A. Makinson's current affiliation is with the Radiation Safety Office, University of Washington, Seattle, WA, USA

makers are the ultimate end users of their output. Studies have repeatedly shown that people err in their interpretation and use of probability information for decision making.^(5–7)

Uncertainty information is typically presented to users visually, most commonly in graphical format. There is a great variety of methods of graphically presenting uncertainty information. These range from the relatively common (e.g., boxplots, use of error bars around means) to the more specialized (e.g., Probability Density Functions, Cumulative Distribution Functions) to the esoteric (e.g., 3-D Probability Density Function). It seems intuitively likely that some of these graph types are easier for people to understand and thus more likely to lead to effective decisions. However, there has been little work examining the effectiveness of different types of graphical displays of probability. There are exceptions,^(8–11) but few studies have compared commonly used graphical display methods or make recommendations for their use.

Ibrekk and Morgan (1987)⁽¹²⁾ and Edwards, Snyder, Allen, Makinson, and Hamby (2012)⁽¹³⁾ are rare examples of studies examining the relative effectiveness of the types of graphical displays that are frequently used in scientific disciplines for communicating probability information. Ibrekk and Morgan explored the effectiveness of nine graphical display formats for communicating weather forecasts. Among other things, they found that cumulative distribution functions can severely mislead users who must estimate the mean probability of an event from the information portrayed in the graph. In their study, even participants with a good knowledge of basic statistics did not perform well at this task. Edwards et al. expanded on Ibrekk and Morgan by examining a different, more comprehensive set of displays based on a systematic review of contemporary methods for the presentation of uncertainty across a variety of scientific disciplines.⁽¹⁴⁾ Displays included 10 graph types: box plots, standard error bars, scatter plot, probability density function (PDF), cumulative distribution function (CDF), complementary cumulative distribution function (CCDF), multiple PDFs, multiple CDFs, multiple CCDFs, and 3-dimensional PDFs. In addition, Edwards et al. included measures designed to assess skill in graph interpretation such as mean estimation, or, identification of the mean value of a variable at a certain probability level, and probability estimation, or identification of the probability that a criterion variable will be at a certain level. Edwards et al. also assessed the accuracy of behavioral choices based on the presented probabilistic output. To do this, participants were told that certain actions should be taken if the value of a target variable was a certain criterion probability level, and then asked what actions they would take. Further, Edwards et al. explored the impact of time pressure on decisions based on the graphs. Their results indicated that accuracy differed by graphical display type such that error bars and boxplots led to the most accurate mean estimation and the best behavioral choices. CCDFs were associated with the highest probability estimation accuracy. When participants were put under time pressure, the quality of their behavioral choices worsened, but boxplots and error bars remained the best-performing graph types. However, simple interpretation of the graphs (i.e., mean estimation and probability estimation) was unaffected by time pressure.

Edwards et al. argued that the time pressure manipulation affected behavioral decision-making but not mean or probability estimation because time pressure reduces the amount of

effortful thought a person can put into a judgment. The underlying assumption was that simple identification of a point on a graph is a largely automatic cognitive process but that effortful thinking is required when people must then translate that into action.⁽¹⁵⁾ More specifically, behavioral choices of the type used in the Edwards et al. studies require several judgments in combination. First, the decision requires the identification of a point on the graph that represents the current circumstances. Second, it requires identifying a criterion value. Third, it requires comparing those two points. Fourth, it involves assessing the likely outcomes of the behavior for their effects. When making behavioral decisions with potentially negative outcomes, people typically take such outcomes into account in weighing whether the behavior is worth the risk or not. That is, people are as or more concerned with the valence and magnitude of possible effects as they are with the probabilities of such effects, at least in some circumstances.⁽⁵⁾ Overall, the behavioral choices made in the Edwards et al. studies corresponds to *synthesis* in Carswell's (1992)⁽¹⁶⁾ taxonomy of graphical tasks. Mean and probability estimation, on the other hand, correspond to *point reading*. The latter is considered to be a low-level, focused task involving the simple extraction of information from the graph, whereas the former involves integrating information so as to go beyond the data given, which is more processing-intensive. Although there are studies examining the relative processing complexity of the perceptual operations involved in simple comprehension of various graph types,^(17,18) there is little that focuses on synthesis-level operations, and, to our knowledge, none that examine the effects of cognitive load on graphical interpretation. However, there is evidence that, in general, increased complexity of the mental tasks involved in reading and interpreting a graph necessitate greater mental effort, as indicated by subjective report and increased reaction times to respond to questions about the graph.⁽¹⁹⁾

The current study investigates Edwards et al.'s (2012) explanation for the effects of time pressure on the quality of behavioral choices using a cognitive load manipulation in which the amount of attention that participants can give to a target task is limited by a concurrent task.⁽²⁰⁾ A vast number of studies have shown that under cognitive load individuals rely on relatively automatic processes (as compared to conscious, effortful processes) to make judgments and decisions in a variety of domains.^(21–24) Here we adapted the procedure of Edwards et al. to determine the affect of increased cognitive load (and thus increased reliance on automatic processing) on performance of the behavioral choice task and mean and probability estimation.

The main predictions were that, first, cognitive load would decrease the accuracy of behavioral choices based on the graphs. This should occur regardless of graph type. Second, cognitive load should *not* affect the accuracy of the simpler judgments of mean and probability estimation. The possible exception to this would be for graphs that have poor task-display compatibility,⁽²⁵⁾ such that the point estimate in question is not easily discerned. This would require the participant to attempt to effortfully determine the quantity in question from whatever information is contained in the graph, a more complex judgment than simple point reading. In the Edwards et al. paradigm, an example of this case is when people are trying to estimate mean values from PDF displays, which do not overtly portray

such values, but for which such values can be deduced with some thought (i.e., by identifying the point at which half of the area of the PDF fall to either side).

2. METHODS

The methods of Edwards et al. (2012) were replicated for the current experiment, with two exceptions. First, in order to test the primary hypothesis, a cognitive load manipulation was added. Second, the number of graph types tested was reduced from ten to five, in order to focus on those found by Edwards et al. to be the best performers.

2.1. Participants

Participants were 191 individuals (91 male, 3 missing gender information) who were either paid \$20 for participation or were given extra credit in a psychology course. The median age was 21 and the range was 18–76 years ($M=24.1$, $SD=9.6$). Response times (explained below) were examined to identify participants who may not have attended to the instructions and experimental materials. Seven outlier participants who spent very little time reading the scenarios or answering the questions (e.g., less than several seconds) were identified and excluded from further analysis. Background characteristics of the retained 184 participants are shown in Table I.

2.2. Materials and Measures

Participants read eight hypothetical scenarios depicting situations involving probabilistic output (see Appendix 1 for example). There were no predictions associated with scenario type; rather, they functioned as within-subjects conceptual replications which helped to increase power. A paragraph described each scenario, which entailed one of eight possible negative events: venting of chlorine emissions from a research center, an aging biohazard filtration system, an airport screening device, treatment for exposure to radioactive substances, an *E. coli* outbreak, exposure to a dangerous mold strain, storage of nuclear waste, and radon contamination. The likelihood of the event depended on situational variables, and this probability was presented using five of the ten probabilistic graphical display methods from Edwards et al. (2012) Figure 1 illustrates the five graphs types (using the radon scenario). The four multivariate display types used by Edwards et al. were excluded on the basis of poor performance in past studies (Multiple Probability Density Function [MPDF]; Multiple Cumulative Density Function [MCDF]; Multiple Complementary Cumulative Density Function [MCCDF]; 3-D Probability Density Function [3-D PDF]). We also excluded boxplots from the current study because of their close similarity to error bars. Following the procedures of Edwards et al., ten distributions of data, defined by degree of positive or negative skewness, were generated for each graph type. These distributions were created to be unique and easily distinguishable by the participant by manipulating a known distribution and performing various operations to make it visibly different from the others. For example, a Gaussian distribution was stretched by multiplying by a constant, moved to the right by adding a second constant, and manipulated again by adding a whole second Gaussian distribution to the left of the first. This would create a two hump density function that after normalization, could be used to ask about the probability of the various deleterious events from the scenarios. Edwards et al. implemented different

distributions in order to assure that the results were not specific to a particular pattern of data distribution. Because there was no evidence that the distribution manipulation was related to any of the criterion measures, either in Edwards et al. or in the current study, we do not discuss this further. In all, 400 graphs—one graph for each unique combination of eight scenarios, five graphical displays, and ten distributions of data—were used. A different distribution was randomly assigned to each of the eight scenarios so that participants used eight of the ten distributions during the study.

Participants were asked several questions designed to assess their past experience with graphs of the type examined in this study. Specifically, participants were asked whether they had seen probability information presented with this type of graph in the past (yes or no), and whether they used uncertainty information at work.

Three questions (Appendix 2), repeated in each scenario, assessed participants' graph interpretation as well as their behavioral choice in response to the situation based on that understanding. For graph interpretation, participants were asked to supply the mean for each variable (e.g., “What is the mean (average) radon concentration?”) and estimate the probability that an observed measurement of each variable was more or less than a given amount (e.g., “How likely is it that the radon concentration is more than 0.3 pCi/L?”). For the CDF and CCDF, the probability estimation question was phrased in a way consistent with what the graph showed (probability estimation; i.e., “How likely is it that a value is *more than* x?” for the CCDF and “How likely is it that a value is *less than* x?” for the CDF). Additionally, participants were asked to make a binary behavioral choice (i.e., act; do not act) in response to each hazard. This question contained parameters to guide optimal decisions. To illustrate, the example in Appendix 2 indicates that action should take place if radon concentration exceeds 0.3 pCi/L. Demographic information was collected, as were additional questions regarding perceptions of the graphical displays (e.g., easy to read, seemed accurate) and prior experience with quantitative uncertainty information.

To implement the cognitive load manipulation, half of participants were randomly assigned to receive an 8-digit number that they were asked to remember while answering the questions about each scenario. Eight 8-digit numbers (one per scenario) were created using a random number generator. One of the eight numbers was randomly assigned to each scenario so that each scenario involved a different 8-digit number. After each scenario, participants recalled the number in a free response format prompt. A question was included as a manipulation check of the cognitive load task: all participants were asked, “When answering questions, how much of your attention did you have available for interpreting the graph?” (1 = *not much at all*, 9 = *all of my attention*). Participants under load were expected to perceive that they had less attention available for that task.

2.3. Procedure

The experiment was conducted using DirectRT 2008, a computer-based software developed for administering psychology studies.⁽²⁶⁾ The software presented all information to participants and recorded all responses. Each participant was randomly assigned to one of the five graphical display conditions (Figure 1) and one of the two cognitive load conditions (number memorization task vs. no number memorization task). Each participant worked

through all eight scenarios (a repeated measure) in random order. Working through each scenario entailed reading a paragraph description of the relevant variables in the hypothetical scenario, viewing a graph that depicted one of the ten (randomly selected) data distributions of those variables, and answering the questions described above. Graphical displays included a title and a note explaining what was portrayed in the graph (e.g., for the CDF in Figure 1, the title was “The cumulative probability density function (CDF) of radon concentration at 50–60 blocks East from city center” and the note was “Depicts the probability that the radon concentration will be less than or equal to a given level at 50–60 blocks East from city center”) in keeping with the style of published figures.

Overall, the experiment had a 5 (Graphical Display: error bars, scatterplot, PDF, CDF, CCDF) $\times 2$ (Cognitive Load: number memorization task vs. no number memorization task) design with 8 repeated measures (each participant responded to the same questions in all 8 scenarios). Participants worked at their own pace through all portions of the experiment and finished in less than 60 minutes.

Participants in the cognitive load condition were given the eight-digit number to remember after reading the paragraph description of each scenario but before proceeding to use the graph to answer the graph interpretation and behavioral choice questions of each scenario.

2.4. Answer Generation

Correct answers were calculated according to established procedures.⁽¹²⁾ Answers to the interpretation questions (mean estimation, probability estimation), and behavioral choice questions were defined for each combination of scenario, graphical display, and data distribution. Mean estimation answers were simply generated from the raw data displayed in the graph. Probability estimation answers were determined from the CDF or CCDF plots, which provide the probability of a data point being more or less than a specified value.

Correct answers for the behavioral choice questions (act; do not act) depended on the respondent identifying a particular criterion point on the display. For some combinations of question and data, the optimal behavioral choice answer was unclear because the data supported more than one answer to that question. For example, if a participant was shown the error bars in Figure 1 and read, “If the radon concentration is more than 0.3 pCi/L at your new house you will need to take precautionary measures to avoid exposure. Your real estate agent just confirmed your purchase of the 50th block house, will you take precautionary measures?” The correct answer is “yes, take precautionary measures” because more than 75 percent of the data at a distance of 50 blocks supports this conclusion (i.e., is above 0.3 pCi/L ppm). However, if the question was phrased to ask about a house 80 blocks East from city center, the correct answer is ambiguous because around 50 percent of the data support either conclusion. Therefore, any decision which 25 – 75 percent of the data supported was treated as ambiguous and coded as missing by design.

2.5. Dependent Variables

As in Edwards et al. (2012), an accuracy index was computed for the probability estimation question by subtracting the correct answer from the participant's estimate and taking the absolute value so that lower scores indicated greater accuracy, and 0 indicated perfect

accuracy. A similar index was computed for the mean estimation question; however, because the units of the portrayed data differed from scenario to scenario (e.g., pCi/L in the radon scenario, parts per million in the chlorine emission scenario), it was important to rescale the variable so that the index was on a common scale across scenarios. Therefore, an accuracy index analogous to a z-score was computed by taking the absolute value of the difference between the participants' estimates and the correct answer and dividing by the standard deviation of this difference score (computed within scenario). The resulting variable had units defined in terms of the standard deviation. For this variable lower scores indicated greater accuracy. Behavioral choice accuracy was simply coded as correct (1) or incorrect (0) depending on whether the participant's response matched the correct answer.

There were some cases of missing data in these accuracy data (<2%), mostly due to the ambiguous behavioral choice answers noted above. Because these cases of ambiguous accuracy resulted from the chance intersection of a randomly assigned variable (distribution) with a specific criterion value, missingness was due to the experimental design and Missing Completely at Random.⁽²⁷⁾ Multiple Imputation procedures in SPSS were used for missing data.

3. RESULTS

3.1. Manipulation Check

As a manipulation check, participants were asked how much of their attention they felt they were able to use for interpreting the graphs (1= Not Much, 9 = All). There was a significant difference in ratings such that participants felt they could devote less of their attention in the cognitive load condition ($M=6.09$) than in the no load condition ($M=6.93$), $F(1, 174) = 8.78$, $p = .003$, $\eta^2 = .05$, suggesting that the load manipulation was effective. The main analyses reported below were also conducted without participants who made errors in recalling the number from the cognitive load task (52% of Cognitive Load participants made no errors), but this did not affect the results.

3.2 Graphical Interpretation

For both graph interpretation questions, General Linear Models were estimated using graphical method of presentation and cognitive load condition as between-subjects categorical predictors of correct response and scenario type as a within-subjects variable. The Graphical Display main effect and effects involving cognitive load were of chief interest. For the mean estimation question there was a significant main effect for Graphical Display ($F[4, 174] = 12.81$, $p < .001$, $\eta^2 = .23$). Consistent with the results of Edwards et al. (2012), post hoc tests (Bonferroni adjusted) showed that, collapsed across cognitive load condition, error bars ($M=.28$) were significantly more accurate than any of the other displays, which did not differ from each other (PDF $M=.68$; CCDF $M=.95$; CDF $M=.82$; scatterplot $M=.75$). Results of post hoc tests within load conditions are reported in Table II. There was also a Graphical Display $\times$ Cognitive Load interaction ($F[4, 174] = 2.62$, $p = .04$, $\eta^2 = .06$). Load led to significantly reduced accuracy for participants using the PDF (Table II), but did not affect the other graphical displays.

Analyses of the probability estimation question revealed a significant effect for Graphical Display ($F [4, 174] = 3.44, p = .01, \eta^2 = .07$; Table III). PDFs ($M=14.41$) led to significantly better accuracy than CDFs ($M=24.99$) or error bars ($M=25.37$). PDFs, CCDFs, and scatterplots led to the best performance. In support of our first hypothesis, there were no effects involving cognitive load.

3.3. Behavioral Choice/Correct Action

For behavioral choice, a dichotomous variable (1 = correct; 0 = incorrect) was created to indicate if the participant's answer matched the correct answer (act; do not act) for each scenario. Because of the dichotomous nature of this dependent variable, repeated measures logistic regression was used (as implemented in the Generalized Estimating Equations procedures in SPSS, with a binomial distribution and logit link specified). For behavioral choice accuracy, there was a significant Graphical Display effect, $\chi^2 (4, N=184) = 16.16, p = .003$. Post hoc tests indicated that CCDFs ($M=.82$) performed significantly better than CDFs ($M=.66$) and results for within load conditions are reported in Table IV. As predicted by hypothesis 1, there was a Cognitive Load main effect on behavioral choice accuracy, $\chi^2 (1, N=184) = 9.09, p = .003$, such that accuracy was worse under cognitive load ($M=.70$) than without it ($M=.79$). There was also a Graphical Display $\times$ Cognitive Load interaction, $\chi^2 (4, N=184) = 15.13, p = .004$. As can be seen in Table IV, the nature of this interaction was that all of the graphs showed reduced accuracy under load (significantly so for the PDF, marginally so for the CCDF [$p=.06$], and nonsignificantly so for the scatterplot and error bars), except for the CDF, which showed a slight and nonsignificant increase in accuracy under load. Another way of investigating this interaction is to examine the relative efficacy of the graphical displays within load condition, which suggests that under load the graphical displays do not significantly differ in accuracy, but with no load the CDF performs significantly worse than the other graphical displays, which do not differ among themselves.

Compared to the graph interpretation questions, behavioral choice results in this study replicated those of Edwards et al (2012). Under time pressure and cognitive load the error bars were among the best and CDF was among the worst for accuracy. Results of the control condition of the current study and the control condition in Edwards et al. (2012) corresponded well, as error bars were the best performing in the original study and in the current study error bars, along with PDF and CCDF, were significantly better than the other display types.

Main effects of Scenario or interactions involving Scenario were not hypothesized for any of our three accuracy measures. Nonetheless, these analyses were conducted to ensure that unexpected systematic effects of Scenario were not present. There were several significant tests, however, the results were not systematic or theoretically interpretable.

3.4. Other Variables

Separate analyses were conducted adding participants' use of uncertainty information at work and familiarity with the graphs to the analyses predicting the three types of accuracy. Out of these six analyses, one showed significant main effects for these demographic variables or interactions between the variables and Graphical Display or Cognitive Load.

There was an interaction between familiarity with the graph and graphical display predicting mean estimation, $F(4, 164) = 2.81, p < .03, \eta^2 = .06$. This showed that familiarity made a difference for the PDF, such that increased experience was significantly related to accuracy (with experience $M = .42$, without $M = .87$), but this was not the case for the other graphical displays.

Participants were also asked to rate on a Likert scale how easy it was to interpret the graph (1 = Not Easy at All, 9 = Extremely Easy). There was a significant effect of Graphical Display for this question, $F(4, 172) = 10.26, p < .001$. Post hoc tests showed that the scatterplot and error bars displays were both seen as significantly easier to interpret than the other three graphical displays. There were no significant effects of Graphical Display or Cognitive Load for the questions regarding how accurately the graphs seemed to present the information or how useful the graphs were.

4. DISCUSSION

Overall, results were consistent with expectations that interpreting basic characteristics of uncertainty data would be unharmed under conditions of limited cognitive resources, whereas more deliberative processing is negatively affected. Because identification of a point on a graph should not take much cognitive effort, the load manipulation was not expected to, and did not, have an overall effect on either mean or probability estimation. Load did reduce accuracy when participants were using the PDF to make their mean estimates, which was not the case for the other graphical displays. This may indicate that the PDF is not as useful as the others for making mean estimates because without explicitly labeling the mean the PDF has low task-display compatibility for mean estimation. Without minimizing the number of perceptual and cognitive mental operations performed for mean estimation⁽²⁸⁾ the PDF was susceptible to the impact of the load manipulation. However, other displays not affected by the load, especially the other density functions, the CDF and CCDF, and the scatterplot, also exhibit relatively low task-display compatibility for mean estimation, but were not affected by the cognitive load. In the CDF and CCDF the median is relatively obvious, and for these displays the median could have served as a rough proxy for the mean. Unlike PDFs, scatterplots did not do particularly well in the no load condition, so unlike PDFs, it appears that effortful interpretation processes are not aiding scatterplot use.

Cognitive load did influence performance in the behavioral choice question, presumably because of the need for more effortful attention and deliberative processes necessary to make that judgment. There was an interaction such that load appeared to affect some graphical displays more than others, but the effect indicated that the worst performing graphs in the no load condition (CDF, scatterplot) were the least influenced by the load manipulation. This appears to be due to an accuracy floor effect as the level of accuracy displayed for these graphs, especially the CDF, was poor. In other words, performance on these graphs was bad enough without load that the addition of load could not worsen them further.

The expected results are consistent with our reasoning that the complexity of the decision making task determines the extent to which decision making quality is suppressed by

Cognitive Load. Simply identifying a point on the graph (as is done for either of the graph interpretation questions) should take little deliberation or conscious effort, so long as the graphical format displays the relevant quantity and the reader understands how to interpret the graph. Consistent with this, performance on mean and probability estimation questions was relatively unaffected by cognitive load. This was expected to and did hold true only for graphical formats that overtly display the quantity in question. For mean estimation, that was error bars (a format that explicitly labels the mean). For probability estimation as implemented here (e.g., “how likely is it that a value is *more/less than* x?”), CCDFs and CDFs should have been and were relatively immune to cognitive load. Graphs that do not overtly display the necessary quantity in an easy-to-read fashion were affected by the load manipulation, as people had to divert cognitive resources to determine how the displayed quantities might provide clues as to the target value.

As can be seen in Tables 2, 3, and 4, results in the no load condition for the three questions indicate that error bars were the best graph for mean estimations, replicating Edwards et al. (2012), and second best for behavioral choice. CCDFs showed comparable levels of performance in probability estimation to Edwards et al.'s similar control condition. PDFs also performed well for probability estimation. CDFs typically did not function well for answering any of the questions, consistent with past studies.^(12,13)

The CCDF outperformed the CDF for probability estimation. These two graph types seem very similar. CDFs portray the probability of a data point being less than or equal to a value, whereas CCDFs indicate the probability of a data point being greater than or equal to a value. As mentioned previously, the probability estimation question was phrased in a way consistent with what these graph showed (e.g., “How likely is it that a value is *more than* x?” for the CCDF and “How likely is it that a value is *less than* x?” for the CDF). This means that CCDFs led to more accurate answers to the questions they are designed to answer than did CDFs, a result also found by Edwards et al. (2012). The reason for this finding is unclear, but it suggests that people are somehow more comfortable thinking about a situation of “less than or equal to” as opposed to “greater than or equal to.” Unfortunately, we know of no research that examines this.

Several limitations of the study deserve attention. The study did not attempt to examine the need to optimize multiple outcomes when making behavioral choices under uncertainty, something that is often the case in real-life decision making situations. Second, we cannot ensure that participants did not draw on cues other than the graphical displays. Because it is not clear what other cues could have been, we feel confident in the assumption that participants followed instructions asking them to rely on the graphical displays as primary source of cues. Third, we assume that observed performance differences between graphical displays hold within individuals. However, the current study did not directly test this possibility as graphical display was randomly assigned between subjects and each participant used only one display type.

Nonetheless, the current study offers several important contributions. At a practical level, the results reported here are of importance for better understanding the constraints to quality decision making when probabilistic output is interpreted by end users who prefer or

are required to engage in multitasking. The cognitive load manipulation mimics the effect of trying to juggle multiple concerns at once, which is a likely situation for many decision makers. This should include personnel called on to respond quickly to the types of chemical, biological, radioactive substance, or security related emergency scenarios simulated in this study. With respect to the related scientific literature, research on using graphical displays and other visuals to communicate uncertainty has focused on only one or two graph types and it is sometimes not clear why particular graphs or visuals are chosen as the focus of study.⁽²⁸⁾ We have addressed the latter by basing selection of displays on a systematic review of the literature.⁽¹⁴⁾ Of main importance in the current study, the load manipulation did not have an overall effect on graph interpretation but did affect performance on behavioral choice due to the greater need for deliberative processing to answer this question. This suggests that interpreting basic characteristics of the data is unharmed under conditions of limited cognitive resources. However, many decisions require additional cognitive steps that go beyond simple identification of a point on a graph. For this reason, it is of value to continue to investigate the interplay between constraints of the decision making context, deliberative processing and more automatic modes of thought in the use of uncertainty information. In addition, future work could specifically investigate the utility of providing brief heuristics or prior statistical training to enhance effective use of these graphical displays for more complex decisions and under conditions of limited cognitive resources.

Acknowledgments

This work was funded by the Department of the Defense, Defense Threat Reduction Agency through Grant # HDTRA1-08-1-0044. The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

APPENDIX 1: EXAMPLE SCENARIO

Radon is a naturally occurring radioactive substance found in dirt and rocks that becomes airborne. If inhaled, it can cause serious health problems. In fact, radon is the second leading cause of lung cancer in the U.S. behind smoking. Many basements in the U.S. are contaminated with radon without their residents' knowledge. You are about to move into the house of your dreams on Edwards Street. Before you do, you research the concentration of Radon for the entire length of the street.

APPENDIX 2: EXAMPLE QUESTIONS (FROM RADON SCENARIO)

1) Mean estimation

What is the mean (average) radon concentration? Please enter your answer in pCi/L (For example, 0.1, 0.2, 0.25, or 0.4).

2) Probability estimation

How likely is it that the radon concentration is less than 0.3 pCi/L? Please enter your answer as a probability (any number between 0 and 1. For example, 0, .26, .88, or 1).

3) Behavioral choice

If the radon concentration is more than 0.3 pCi/L at your new house you will need to take precautionary measures to avoid exposure. Your real estate agent just confirmed your purchase of the 50th block house, will you take precautionary measures?

1. Yes
2. No

REFERENCES

1. Hamby D. A probabilistic estimation of atmospheric tritium dose. *Health Physics*. 1993; 65:33–40. [PubMed: 8505228]
2. Hamby D, Benke R. Uncertainty of the iodine-131 ingestion dose conversion factor. *Radiation Protection Dosimetry*. 1999; 82:245–56.
3. Harvey R, Hamby D, Palmer T. Uncertainty of the thyroid dose conversion factor for inhalation intakes of ¹³¹I and its parametric uncertainty. *Radiation Protection Dosimetry*. 2006; 118:296–306. [PubMed: 16410291]
4. Simpkins A, Hamby D. Uncertainty in Transport Factors Used To Calculate Historical Dose From ¹³¹I Releases At the Savannah River Site. *Health physics*. 2003; 85:194–203. [PubMed: 12938966]
5. Dawes, RM. Behavioral decision making and judgment. In: Gilbert, DT.; Fiske, ST.; Lindzey, G., editors. *The handbook of social psychology*. McGraw Hill; Boston, MA: 1998. p. 497-548.
6. Kahneman, D.; Slovic, P.; Tversky, A. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press; Cambridge, MA: 1982.
7. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*. 1974; 185:1124–31. [PubMed: 17835457]
8. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological review*. 1995; 102(4):684.
9. MacGregor D, Slovic P. Graphic representation of judgmental information. *Human-Computer Interaction*. 1986; 2:179–200.
10. Sun Y, Li S, Bonini N. Attribute salience in graphical representations affects evaluation. *Judgment and Decision Making*. 2010; 5:151–8.
11. Wickens CD, Gempfer K, Morphet ME. Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*. 2000; 2:99–126.
12. Ibrenk H, Morgan MG. Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*. 1987; 7:519–29.
13. Edwards JA, Snyder FJ, Allen PM, Makinson KA, Hamby DM. Decision making for risk management: A comparison of graphical methods for presenting quantitative uncertainty. *Risk Analysis*. 2012; 32:2055–70. [PubMed: 22616656]
14. Makinson KA, Hamby DM, Edwards JA. A review of contemporary methods for the presentation of scientific uncertainty. *Health Physics*. 2012; 103:714–31. [PubMed: 23281507]
15. Edland, A.; Svenson, O. Judgment and decision making under time pressure: Studies and findings. In: Svenson, O.; Maule, JA., editors. *Time pressure and stress in human judgment and decision making*. Plenum Press; New York, NY: 1993. p. 27-40.
16. Carswell C. Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1992; 34:535–54.
17. Casner SM. A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*. 1991; 10:111–51.
18. Holland JG, Spence I. Judging proportion with graphs: The summation model. *Applied Cognitive Psychology*. 1998; 12:173–90.
19. Huang W, Eades P, Hong S-H. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*. 2009; 8(3):139–52.

20. Gilbert DT, Osborne RE. Thinking backward: Some curable and incurable consequences of cognitive busyness. *Journal of Personality and Social Psychology*. 1989; 57:940–9.
21. Gilbert DT, Pelham BW, Krull DS. On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*. 1988; 54:733–40.
22. Krull DS, Erickson DJ. Judging situations: On the effortful process of taking dispositional information into account. *Social Cognition*. 1995; 13:417–38.
23. Roch SG, Lane JAS, Samuelson CD, Allison ST, Dent JL. Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups. *Organizational Behavior and Human Decision Processes*. 2000; 83:185–212. [PubMed: 11056068]
24. Sivaramakrishnan S, Manchanda RV. The effect of cognitive busyness on consumers' perception of product value. *Journal of Product & Brand Management*. 2003; 12:335–45.
25. Sparrow JA. Graphical displays in information systems: Some data properties influencing the effectiveness of alternative forms. *Behaviour & Information Technology*. 1989; 8:43–56.
26. Jarvis, BG. *DirectRT*. Empirisoft Corporation; New York, NY: 2008.
27. Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. Wiley New York; New York, NY: 1987.
28. Lipkus IM, Hollands J. The visual communication of risk. *Journal of the National Cancer Institute Monographs*. 1999; 25:149–63. [PubMed: 10854471]

Note: Graphs in this figure depict different data distributions. Due to space limitations they are simplified versions of the graphs used in the studies (e.g., no graph titles are included). CDF = cumulative distribution function; CCDF = complementary cumulative distribution function; PDF = probability density function). Data in CDF, CCDF, and PDF displays were presented to participants to be within the range of 50-60 blocks east from city center.

Figure 1.
Five graphical displays communicating uncertainty information.

Table I

Participant demographics (N=184)

Demographics		Total	No Load	Load
Gender ¹	Female	52%	51%	53%
	Male	47%	49%	47%
Education ²	Current Undergraduate	76%	77%	75%
	Undergraduate degree or some graduate school	14%	15%	14%
	Graduate degree	5%	5%	5%
Statistics training ³	High School or less	41%	46%	38%
	1–4 undergraduate courses	46%	44%	48%
	> 4 undergraduate courses or at least 1 graduate course	10%	10%	11%
Use of uncertainty information ⁴	Works in a decision making capacity	10%	7%	12%
	Uses uncertainty information at work	33%	30%	35%
	Uses uncertainty information in daily life	29%	31%	28%
	Has seen probability information in a graph like this before	50%	43%	55%

Note: Numbers for each row represent percent of entire sample, the no load condition (N=88), and the load condition (N=96). None of the chi square tests of gender, education, statistics training or the four measures of use of uncertainty information revealed significant differences between experimental conditions. Percentages may not equal 100% due to rounding.

¹There were 2 missing values for gender in the load condition.

²There were 4 missing values for education (3 in the load condition).

³There were 3 missing values for statistics training in the load condition.

⁴These categories are not mutually exclusive.

Table II

Mean estimation accuracy by graphical display and cognitive load condition (N=184)

Graphical Display	Mean estimation accuracy ^I			
	No Load		Load	
	<i>M</i>	(SE)	<i>M</i>	(SE)
1. Error Bars	0.35 ^a	0.10	0.21 ^a	.10
2. Scatterplot	0.75 ^{bc}	0.10	0.75 ^b	.10
3. CDF	0.86 ^{bc}	0.10	0.77 ^b	.10
4. CCDF	0.87 ^c	0.10	1.03 ^b	.10
5. PDF	0.46 ^{ab}	0.10	0.90 ^b	.10 ⁱ

Note: CDF = cumulative probability distribution function; CCDF = complementary cumulative probability distribution function; PDF = probability density function. Estimates in a column that do not share superscripts have a significantly different pairwise comparison (Bonferroni adjusted) at $p < .05$. Rows containing an ⁱ are significantly different between load conditions.

^I Accuracy in estimating the mean, defined as absolute value of (estimate – correct answer)/SD. Lower value indicates higher accuracy.

Table III

Probability estimation accuracy by graphical display and cognitive load condition (N=184)

Graphical Display	Probability estimation accuracy ^I			
	No Load		Load	
	<i>M</i>	(SE)	<i>M</i>	(SE)
1. Error Bars	28.60 ^{ab}	3.57	22.13 ^a	3.29
2. Scatterplot	19.12 ^{ab}	3.57	22.43 ^a	3.29
3. CDF	23.65 ^{ab}	3.57	26.34 ^a	3.37
4. CCDF	13.98 ^b	3.37	24.25 ^a	3.37
5. PDF	12.47 ^b	3.47	16.34 ^a	3.47

Note: CDF = cumulative probability distribution function; CCDF = complementary cumulative probability distribution function; PDF = probability density function. Estimates in a column that do not share superscripts have a significantly different pairwise comparison (Bonferroni adjusted) at $p < .05$.

^I Accuracy in estimating probability, defined as absolute value of (estimate – correct answer). Estimates were percentages. Lower value indicates higher accuracy.

Table IV

Behavioral choice by graphical display and cognitive load condition (N=184)

Graphical Display	Correct action ^I			
	No Load		Load	
	<i>M</i>	(SE)	<i>M</i>	(SE)
1. Error Bars	0.79 ^a	0.05	0.72 ^a	0.05
2. Scatterplot	0.75 ^{ab}	0.04	0.70 ^a	0.05
3. CDF	0.62 ^b	0.04	0.69 ^a	0.04
4. CCDF	0.86 ^a	0.03	0.76 ^a	0.05
5. PDF	0.87 ^a	0.03	0.64 ^a	0.06 ^I

Note: CDF = cumulative probability distribution function; CCDF = complementary cumulative probability distribution function; PDF = probability density function. Estimates in a column that do not share superscripts have a significantly different pairwise comparison (Bonferroni adjusted) at $p < .05$. Rows containing an ^I are significantly different between load conditions.

^I 0 = incorrect action chosen; 1 = correct action chosen. Means are given for illustrative purposes; actual statistical analyses were on a binomial basis.