DEVELOPING A THEORETICAL FRAMEWORK OF TASK COMPLEXITY FOR RESEARCH ON VISUALIZATION IN SUPPORT OF DECISION MAKING UNDER UNCERTAINTY

Stephen M. Fiore[1], Samantha F. Warta[1], Andrew Best[2], Olivia Newton[1], and Joseph J. LaViola Jr.[1]
[1]University of Central Florida, [2]The University of North Carolina at Chapel Hill

This paper describes initial validation of a theoretical framework to support research on the visualization of uncertainty. Two experiments replicated and extended this framework, illustrating how the manipulation of task complexity produces differences in performance. Additionally, using a combinatory metric of workload and performance, this framework provides a new metric for assessing uncertainty visualization. We describe how this work acts as a theoretical scaffold for examining differing forms of visualizations of uncertainty by providing a means for systematic variations in task context.

INTRODUCTION

Uncertainty in decision-making continues to be a pressing challenge for operators in complex environments because of the ambiguous and abstract nature of the parameters making up the decision context. In developing decision aids, cognitive engineers must operationalize these parameters and determine how to best present that information. Visualization of uncertainty has consistently been used as a means of helping operators make more informed decisions. Much research has gone into the study of visualizing uncertainty (Andrienko et al., 2010; Bisantz et al., 2011; Sedig & Parsons, 2013) and examining uncertainty visualization methods (Kehrer & Hauser, 2013; Kirschenbaum et al., 2013; MacEachren et al., 2005; Pham et al., 2009; Sanyal et al., 2009). Despite gains in understanding the visualization of uncertainty, an important gap remains in how environmental complexity, termed *task complexity*, alters the decision context, in addition to the complexity produced by uncertain information itself.

Understanding this interaction would provide insight, not only to decision-making, but also on visualization efficacy, a concept objectively quantifiable through the assessment of workload and decision accuracy. We present a theoretical framework to operationalize concepts capturing aspects of task complexity relevant to uncertainty. Our theory-grounded approach enables cognitive engineers to consistently evaluate how system manipulations will function in varied settings.

We operationalize task context using variations in complexity. These parameters readily lend themselves to manipulations in visualization research. Complexity represents an information rich environment containing multiple forms of data, requiring assimilation and integration. We use Wood's (1986) model of complexity as it provides a set of orthogonal dimensions enabling operationalization of task complexity. We varied decision scenarios along component and coordinative complexity dimensions. Component complexity addresses the number of distinct acts and/or cues an individual must process. Coordinative complexity concerns the degree to which integration of task variables must occur for task completion. In sum, we distinguish between tasks varying in complexity (i.e., differing numbers of variables and levels of integration; see Wood, 1986), and how these interact with each other and the amount of uncertainty.

In addition to traditional performance metrics, we used unique assessments of workload and performance. We leveraged developments in Cognitive Load Theory (CLT) and gains in understanding human information processing as well as its interaction with complex systems (e.g., Paas et al., 2003). CLT attends to how the inherent complexity of a given domain can alter workload. For example, when dealing with multiple forms of complex data, the quantity may overwhelm the human information processing system. Further, the nature or quality of those elements can vary such that they require either a high or low degree of integration themselves; thus, additionally influencing cognitive load.

We also leveraged the concept of *cognitive efficiency*, a metric evolving out of measures of instructional efficiency. This describes the relationship between a learner or operator's subjective assessment of workload and their overall task performance (Fiore et al., 2006; Paas & Van Merrienboer, 1993). A cognitive efficiency measure standardizes measures of performance and workload and computes the difference between scores. Positive scores indicate that relative performance was greater than relative workload and suggests that some intervention led to more efficient cognitive processing. Negative scores indicate relative performance was less than relative workload, meaning that some intervention was not efficient. Johnston et al. (2013) found that decision support augmented with graphical displays, produced higher cognitive efficiency scores when compared to those not using such displays. We utilized the Fiore et al. (2006) and Johnston et al. (2013) "cognitive efficiency" metric to determine how task context affects workload and performance.

In sum, our framework incorporates both operational variations as well as measurement diagnostics. Ultimately, our goal is to provide cognitive engineering research with a solid foundation on which to test variations in visualization to determine how task context may influence the efficacy of visualizations. To that end, we now detail our general hypotheses and two experiments developed to test this framework.

Hypotheses

We focused on how task complexity variations influence performance and cognitive efficiency. Thus, we hypothesized that trials with the highest complexity (high component/high coordinative), would produce the lowest performance and lower cognitive efficiency, relative to other items. Similarly, test trials with the lowest complexity (low component/low coordinative), we hypothesized, would produce the highest performance and greater cognitive efficiency, relative to other items.

GENERAL METHOD

Materials

Each trial consisted of a decision between two side-by-side grids that represented abstractions of different map regions. The decision was to judge which grid displayed more uncertainty. Trials were categorized as either easy or difficult judgments. Judgment difficulty was established by assigning ascending point values to each symbol, dependent upon the amount of uncertainty being visualized. Each grid was scored and difficulty was determined by comparing the difference in scores between the two side-by-side grids. Problem grids with a greater difference in point values were classified as easy (e.g., 8 versus 16 points) and those with similar values were classified as hard (e.g., 8 versus 10 points). In brief, when there was little difference between grids as to the 'amount' of uncertainty represented, the judgment as to which was 'more' uncertain, was classified as a hard decision (and vice versa). Amazon's Mechanical Turk (AMT) and Qualtrics, an online experiment platform, were used to collect data from participants.

Procedure

First, participants were redirected from AMT to the experiments on Qualtrics. Following informed consent, participants completed a series of background items (not reported here due to space constraints). Participants were provided with the details of the experimental scenario and an explanation of the task they were expected to complete. Following this, a short training session, with feedback, functioned as a tutorial to introduce the judgment task and the visualizations used. On each trial, participants were given up to 60 seconds to determine which of the side-by-side grids showed greater uncertainty. If no response was given, the trial auto-advanced. After making each judgment, respondents were asked to rate judgment difficulty on a 7-point Likert scale to assess cognitive workload.

EXPERIMENT ONE

We adopted a set of discrete symbols developed by MacEachren et al. (2012) in their study of visual semiotics. We focused on two elements of uncertainty examined with their typology – space and time. Our purpose here was not to test the efficacy of a particular form of visualization and its superiority in conveying uncertainty. Rather, our purpose was to test the utility of our theoretical framework as a methodology for varying task complexity. From this, any form of uncertainty visualization can be systematically tested to ascertain its efficacy when complexity varies. This experimental task asked participants to compare two abstract map regions and determine which had the greatest amount of spatial location uncertainty and/or temporal information uncertainty.

Method

Participants

Eighty participants were recruited (28 *female*, 51 *male*, 1 *other*, and M_{age} = 31.88). The majority of participants were Caucasian (n = 56), located in the USA (n = 76), and reported English as their first language. All participants were compensated with a base payment of \$3.00. To incentivize performance, a bonus was awarded based on participants' points earned. Each correct judgment was worth one point, and \$0.10 was awarded for every 20 points earned. The experiment took approximately 30 minutes.

Design

Independent Variables. This experiment used a within-subjects design with two independent variables, the *type of visualization* and *task complexity*. The types of visualizations were divided into *spatial* and *temporal*. *Spatial location uncertainty* indicated the degree of uncertainty that an object is in a given location and is represented by a solid circle (see Figure 1). Each solid circle has varying amounts of fuzziness around the edges to reflect levels of uncertainty. *Temporal information uncertainty*, represented by open circles, is an indication of the degree of uncertainty that an object is somewhere at a given time. Each open circle has a horizontal, orange dashed "timeline" in the middle, and a vertical, black line to indicate a point in time. A longer rectangle across the timeline indicates greater uncertainty around the time point represented by the black line. These uncertainty types were manipulated within our task complexity framework.

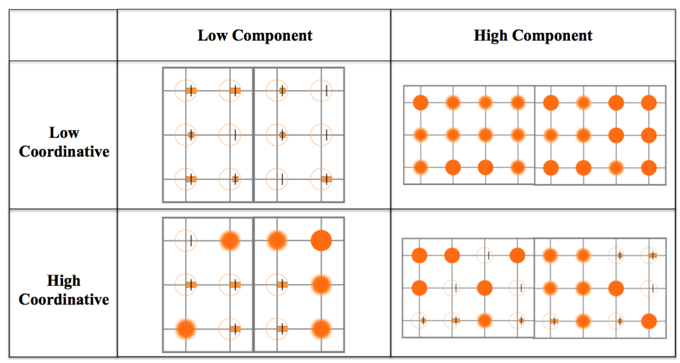

Figure 1. Experiment 1 Stimuli by Condition.

Complexity was divided into *coordinative complexity* and *component complexity*. Both types of complexity have two levels (low, high). Component complexity involves the number of items a decision maker must evaluate. In our study, *low component complexity* stimuli utilized map grids with only six symbols each and *high component complexity* stimuli utilized map grids with twelve symbols each. Coordinative complexity involves the degree to which the items in a decision context must be 'integrated' by the decision maker. In our study, *low coordinative complexity* stimuli involved only spatial or only temporal visualizations whereas *high coordinative complexity* stimuli used combined visualizations requiring uncertainty judgments that integrate both spatial and temporal symbols on the same map grid.

As shown in Figure 1, each quadrant illustrates what would be one of the uncertainty judgments presented in a trial where uncertainty is compared in the side-by-side grids. Note that the low coordinative quadrants show grid comparisons that were either all temporal (top left) or all spatial (top right), and the high coordinative show grid comparisons requiring an integration of spatial and temporal uncertainty visualizations. Blocks were organized by levels of complexity to form four experimental conditions: low component/low coordinative (LL), high component/low coordinative (HL), low component/high coordinative (LH), and high component/high coordinative (HH). Low coordinative blocks (four blocks) contained 12 judgments whereas high coordinative blocks (two blocks) had 24 judgments, for a total of 96 judgments. The presentation of blocks and problem grids were randomized and counter-balanced between participants.

Dependent Variables. The dependent variables reported are the accuracy of judgments based upon the percent correct (performance) and cognitive efficiency (CE) score. CE is derived by taking standardized workload scores and combining them with standardized performance scores. As described in Fiore et al. (2006), such scores can be represented as the perpendicular distance from a line representing a level of zero efficiency (CE = $[z_p - z_w]/\sqrt{2}$). Because these are standardized scores, this results in positive and negative values that hover around a mean of 0. Positive scores indicate cognitive efficiency in that there is relatively better performance in proportion to reported workload, whereas negative scores indicate cognitive inefficiency (i.e., relative performance is less than relative workload). Due to space limitations, we only report a portion of the data from this experiment.

Results

Performance Accuracy. A repeated-measures ANOVA compared the effects of component complexity, coordinative complexity, and judgment difficulty on participants' performance accuracy across the LL, LH, HL, and HH conditions. Mauchly's test of Sphericity was met. Participant performance by experimental condition is illustrated in Figure 2.

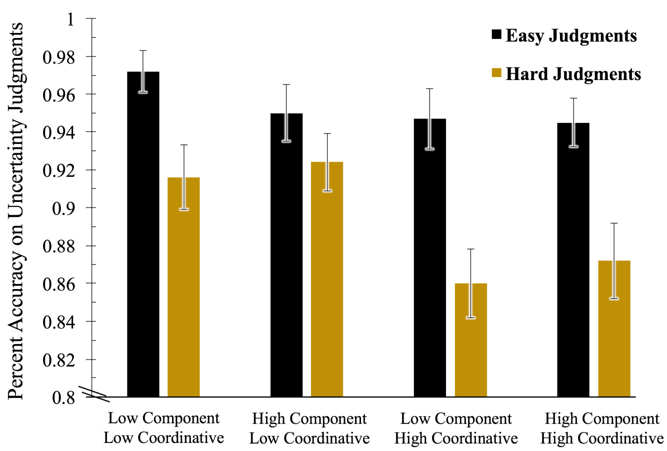

Figure 2. Experiment 1 Participant Performance by Condition.

First, there was a significant main effect of coordinative complexity on performance accuracy, $F(1, 78) = 22.26$, $p < .0001$, $\eta_p^2 = .222$, observed power = .997. Participants performed better in low coordinative ($M = .94$, $SE = .01$) as opposed to high coordinative complexity conditions ($M = .91$, $SE = .01$). Second, there was a significant main effect of judgment difficulty on performance accuracy, $F(1, 78) = 69.56$, $p < .0001$, $\eta_p^2 = .471$, observed power = 1.00. Participants performed better on easy judgments ($M = .95$, $SE = .01$), where the side-by-side grids had a greater degree of difference in uncertainty, as opposed to hard judgments ($M = .89$, $SE = .02$), where the side-by-side grids were very similar in uncertainty.

Third, there was a significant interaction between component complexity and judgment difficulty, $F(1, 78) = 3.83$, $p = .05$, $\eta_p^2 = .047$, observed power = .489. The difference in performance between easy and hard judgments was larger in the low component complexity ($M = .96$, $SE = .01$ versus $M = .89$, $SE = .02$), than the high component complexity conditions ($M = .95$, $SE = .01$ versus $M = .90$, $SE = .02$). Last, there was a significant interaction between coordinative complexity and judgment difficulty, $F(1, 78) = 15.75$, $p < .0001$, $\eta_p^2 = .168$, observed power = .975. Participants performed equally well on easy judgments in both low ($M = .96$, $SE = .01$) and high coordinative complexity conditions ($M = .95$, $SE = .01$), but the difference was larger between hard judgments in the low ($M = .92$, $SE = .01$) and high ($M = .87$, $SE = .02$) coordinative complexity conditions. There were no other significant results.

Cognitive Efficiency. A repeated-measures ANOVA compared the effects of component complexity, coordinative complexity, and judgment difficulty on participants' cognitive efficiency (CE) scores across the LL, LH, HL, and HH conditions (see Figure 3). Cognitive efficiency was calculated as described earlier. Mauchly's test of Sphericity was met.

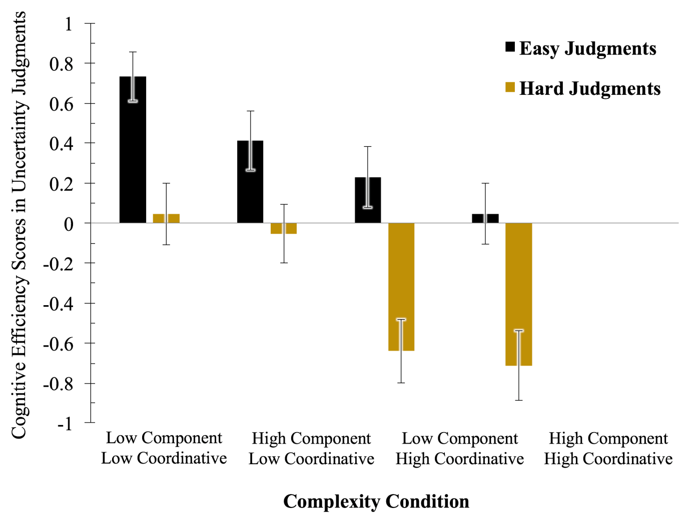

Figure 3. Experiment 1 CE Scores across Conditions.

First, there was a significant main effect of component complexity on CE scores, $F(1, 79) = 5.91$, $p < .05$, $\eta_p^2 = .070$, observed power = .671. Overall, CE scores were highest in low component ($M = .09$, $SE = .137$) as opposed to high component complexity conditions ($M = -.08$, $SE = .14$). Second, there was a significant main effect of coordinative complexity on CE scores, $F(1, 79) = 59.80$, $p < .0001$, $\eta_p^2 = .431$, observed power = 1.00. Overall, CE scores were highest in low coordinative ($M = .29$, $SE = .14$) as opposed to high coordinative complexity conditions ($M = -.27$, $SE = .14$). This can be generally viewed when comparing the left and right sides of Figure 3. Overall, for the low coordinative trials, CE scores were either positive, meaning performance was high relative to workload, or near zero, meaning relatively equal performance and workload. For the high coordinative trials, CE scores were, on average, largely negative, meaning performance was lower relative to workload. Third, there was also a significant main effect of judgment difficulty on CE scores, $F(1, 79) = 220$, $p < .0001$, $\eta_p^2 = .736$, observed power = 1.00. CE scores were higher on easy judgments ($M = .36$, $SE = .13$) than hard judgments ($M = -.34$, $SE = .14$).

Fourth, there was a significant interaction effect between component complexity and judgment difficulty, $F(1, 79) = 4.27$, $p < .05$, $\eta_p^2 = .051$, observed power = .533. CE scores on easy judgments were positive for both low component ($M = .48$, $SE = .13$) and high component complexity ($M = .23$, $SE = .14$), with a large difference between them. But, for hard judgments, CE scores were negative on both low coordinative ($M = -.30$, $SE = .15$) and high coordinative ($M = -.38$, $SE = .15$) and more closely equal to each other.

Lastly, there was also a significant interaction between coordinative complexity and judgment difficulty, $F(1, 79) = 13.00$, $p < .01$, $\eta_p^2 = .141$, observed power = .945. CE scores were positive and highest on easy judgments in both low ($M = .57$, $SE = .13$) and high coordinative complexity conditions ($M = .14$, $SE = .14$). For the hard judgments, CE scores were negative, with low coordinative near zero ($M = -.003$, $SE = .141$). But, the greatest cognitive inefficiency was found with high coordinative complexity ($M = -.68$, $SE = .15$). As such, the greatest difference in CE was between easy judgments that were low in coordinative complexity (indicating very cognitively *efficient* responses) and hard judgments that were high in coordinative complexity (indicating very cognitively *inefficient* responses). There were no other significant results.

EXPERIMENT TWO

Experiment two was designed as a follow-up to experiment one, replicating and extending the test of our complexity framework. Specifically, the visualization symbols in this study progressed from more abstract to more *ecologically valid*. Participants were asked to identify the map region possessing the greatest amount of uncertainty in relation to the probability that an area on the map grid, containing a boat, would be potentially affected by hurricane force winds.

Method

Participants

One hundred participants were recruited through Amazon's Mechanical Turk (51 *female*, 49 *male*). Ages ranged from 18 to 61 years ($M_{age} = 34.60$ years). The majority were Caucasian ($n = 70$) and located in the United States ($n = 90$). All participants, except three, reported English as their first language. Participants were compensated as in the first experiment, and participation also took approximately 30 minutes.

Design

Independent Variables. This experiment also used a within-subjects design with two independent variables (*type of visualization* and *task complexity*). The types of visualizations were divided into two categories, *intensity* and *integrated*. Both intensity and integrated symbol visualizations were judged on their degree of uncertainty using the concept of color shades. Uncertainty was illustrated by the varying shades of red within the symbols (lower uncertainty shown by darker shades; higher uncertainty by lighter shades). Located within each image, there was a small gray icon meant to represent a boat. In the *intensity* visualizations, the boat remains in the same location between symbols varying in levels of uncertainty across map grids. However, the *integrated* visualizations introduce changes in location of the boat between symbols varying in levels of uncertainty dependent upon color shade.

As in Experiment 1, complexity was divided into coordinative and component complexity, each having two levels. Low component complexity map-grids contained six symbols each and high component complexity map grids contained twelve. Low coordinative complexity conditions involved only map grids with intensity symbols whereas high coordinative complexity map grids used the integrated symbols. This required integrating both boat location within the symbol and shade surrounding the boat. Blocks were organized by levels of complexity, forming four experimental conditions as in Experiment 1 (LL, LH, HL, HH). Each block contained 12 judgments (total 48). The presentation of blocks and problem grids were randomized and counter-balanced between participants. See Figure 4 for an example of the stimuli by condition; each quadrant represents what would be one of the two map grids presented in a trial for the uncertainty judgment.

Dependent Variables. The dependent variables examined in this study were the same as those in experiment one.

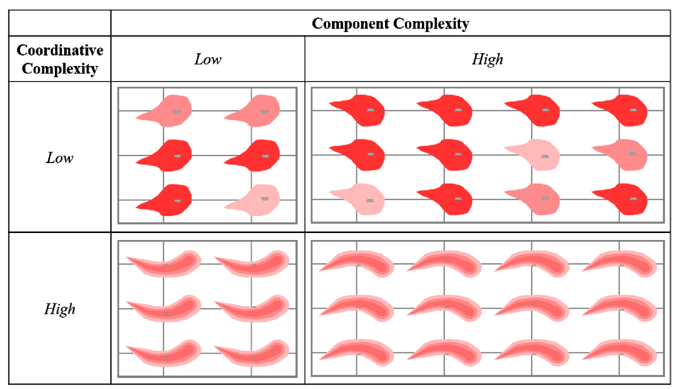

Figure 4. Experiment 2 Stimuli by Condition.

Results

Performance Accuracy. A repeated-measures ANOVA compared the effects of component complexity, coordinative complexity, and judgment difficulty on performance accuracy across the LL, LH, HL, and HH conditions. Mauchly's test of Sphericity was met. First, there was a significant main effect of coordinative complexity on performance accuracy, $F(1, 98) = 36.49$, $p < .0001$, $\eta_p^2 = .271$, observed power = 1.00. Participants performed better in low coordinative ($M = .94$, $SE = .01$) as opposed to the high coordinative complexity condition ($M = .91$, $SE = .01$). Second, there was also a significant main effect of judgment difficulty on participants' performance accuracy, $F(1, 98) = 30.84$, $p < .0001$, $\eta_p^2 = .239$, observed power = 1.00. Participants performed better on easy judgments ($M = .95$, $SE = .01$) than hard judgments ($M = .89$, $SE = .025$). There were no other significant results.

Cognitive Efficiency. A repeated-measures ANOVA compared the effects of component complexity, coordinative complexity, and judgment difficulty on participants' cognitive efficiency (CE) scores across conditions (see Figure 5). Mauchly's test of Sphericity was met. First, there was a significant main effect of coordinative complexity on CE scores, $F(1, 98) = 66.70$, $p < .0001$, $\eta_p^2 = .405$, observed power = 1.00. Participants' CE scores were higher in low coordinative ($M = 1.20$, $SE = .311$) as opposed to high coordinative complexity conditions ($M = -1.26$, $SE = .371$). Second, there was also a significant main effect of judgment difficulty on CE scores, $F(1, 98) = 101.13$, $p < .0001$, $\eta_p^2 = .508$, observed power = 1.00. Participants' CE scores were higher on easy judgments ($M = .410$, $SE = .315$) than hard judgments ($M = -.470$, $SE = .306$).

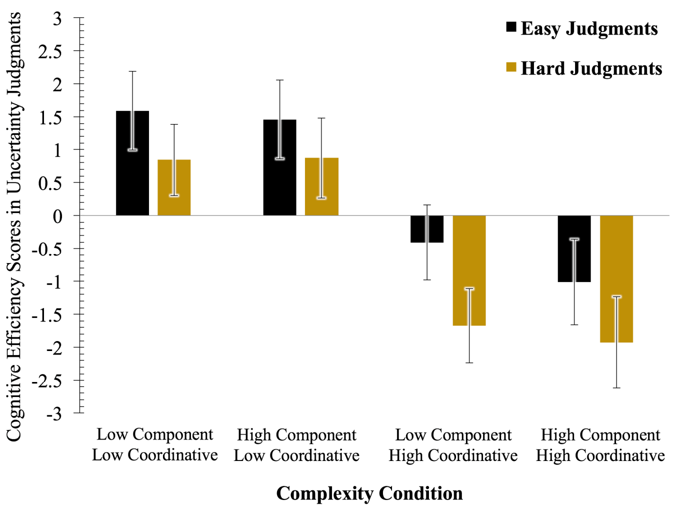

Figure 5. Experiment 2 CE Scores Across Conditions.

Lastly, there was a significant interaction between coordinative complexity and judgment difficulty, $F(1, 98) = 6.18$, $p < .05$, $\eta_p^2 = .059$, observed power = .692. CE scores were positive and highest for low coordinative in both easy ($M = 1.53$, $SE = .31$) and hard judgments ($M = .86$, $SE = .32$). For high coordinative, CE scores were negative, for both easy ($M = -.71$, $SE = .39$) and hard judgments ($M = -1.80$, $SE = .37$). Again, the greatest difference in CE was between easy judgments that were low in coordinative complexity (indicating cognitively *efficient* responses) and hard judgments that were high in coordinative complexity (indicating cognitively *inefficient* responses). There were no other significant results.

DISCUSSION

We presented initial testing of a theoretical framework meant to support research on uncertainty visualization. Two experiments replicated and extended how task complexity can be manipulated. Using a combinatory metric of workload and performance, we demonstrate how to utilize a synergistic combination of measurement approaches. Cognitive efficiency provides a single metric combining subjective assessments of cognitive load with actual performance. As such, across two different forms of stimuli, similar patterns of performance and cognitive efficiency were found.

This framework provides an innovative level of diagnosticity for evaluating visualization manipulations. Our broader context for this involves the operational environment where decision makers are faced with variations of uncertainty in determination of "courses of action" (COA). As such, we propose a theoretical scaffold for examining factors influencing COA selection with differing visualizations. We add methodological value to such research by illustrating how to systematically vary task context. In sum, this framework details a set of orthogonal dimensions and metrics that present an elegant means of operationalizing and assessing task complexity. From this, research in cognitive engineering has a solid foundation on which to test variations in visualizations to determine how task context may influence their efficacy. With these, cognitive engineers can enhance understanding of how visualizations impact COA selection through context variations that alter the complexity of the decision environment.

ACKNOWLEDGEMENTS

This research was partially supported by Grant N000141512708 from the Office of Naval Research. The views, opinions, and findings contained in this article are the authors and should not be construed as official or as reflecting the views of the University of Central Florida or the Office of Naval Research.

REFERENCES

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... & Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science, 24*(10), 1577-1600.

Bisantz, A. M., Cao, D., Jenkins, M., Pennathur, P. R., Farry, M., Roth, E., ... & Pfautz, J. (2011). Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making*, *5*(3), 277-293.

Fiore, S. M., Scielzo, S., Jentsch, F., & Howard, M. (2006). Effects of discrimination task training on X-ray screening decision. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 2610-2614). Santa Monica, CA: Human Factors and Ergonomics Society.

Johnston, J., Fiore, S. M., Paris, C., & Smith, C. A. P. (2013). Application of Cognitive Load Theory to developing a measure of team cognitive efficiency. *Military Psychology, 25*(3), 252-265.

Kehrer, J. & Hauser, H. (2013). Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics, 19*(3), 495-513.

Kirschenbaum, S. S., Trafton, J. G., Schunn, C. D., & Trickett, S. B. (2013). Visualizing uncertainty: The impact on performance. *Human Factors, 56*(3), 509-520.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, *32*(3), 139-160.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics and uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), 2496-2505.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743.

Pham, B., Streit, A., & Brown, R. (2009). Visualisation of information uncertainty: Progress and challenges. In *Trends in Interactive Visualization* (pp. 19-48). Springer London.

Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P., & Moorhead, R. J. (2009). A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE Transactions on Visualization and Computer Graphics, 15*(6), 1209-1218.

Sedig, K., & Parsons, P. (2013). Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, *5*(2), 84-133.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37*, 60-82.