2009

Visual Representations of Meta-Information

Ann M. Bisantz, *SUNY University at Buffalo*
Richard T. Stone, Ph.D., *SUNY University at Buffalo*
Jonathan Pfautz, *Charles River Analytics*
Adam Fouse, *Charles River Analytics*
Michael Farry, *Charles River Analytics*, et al.

Visual Representations of Meta-Information

Ann M. Bisantz
Richard T. Stone
University at Buffalo, State University of New York
Jonathan Pfautz
Adam Fouse
Michael Farry
Charles River Analytics
Emilie Roth
Roth Cognitive Engineering
Allen L. Nagy
Wright State University
Gina Thomas
U.S. Air Force Research Laboratory

Abstract: We conducted two studies that investigated display characteristics related to color (hue, saturation, brightness, and transparency) and contrast with a background for displaying information qualifiers (termed meta-information) such as uncertainty, age, and source quality. Level of detail (or granularity) of the meta-information and task demands were also manipulated. Participants were asked to rank and rate colored regions overlaid on different map backgrounds based on the level of meta-information the regions displayed. Results from Study 1 indicated that participants could appropriately rank and rate levels of meta-information across saturation, brightness, and transparency conditions, and results from Study 1 and Study 2 showed that the natural direction of ordering is complex and dependent on the relevance of different information to the task and the contrast of the overlay region with the background.

Introduction

In many domains, users are confronted with large volumes of information from a variety of sources. In addition to understanding the content of the information, they need to understand and reason about potential qualifiers of the information. These qualifiers, or *meta-information,* include characteristics such as the uncertainty associated with the data, the age of the data, and the source of the data (Pfautz, Fouse, Shuster, Bisantz, & Roth, 2005). For example, in military command and control tasks, commanders must reason about the location of threats. Information about those threats may come from sensors with associated error and failure rates; it may be several hours (or days) old; and/or it may be derived from intelligence sources with varying degrees of trustworthiness.

ADDRESS CORRESPONDENCE TO: Ann Bisantz, 438 Bell Hall, Department of Industrial and Systems Engineering, University at Buffalo, State University of New York, Amherst, NY 14260, bisantz@buffalo.edu.
Visit the *JCEDM* Online Companion at *http://cedm.webexone.com.*

Journal of Cognitive Engineering and Decision Making, Volume 3, Number 1, Spring 2009, pp. 67–91.
DOI 10.1518/155534309X433726.

These critical data qualifiers are not usually presented as part of the primary information displays, and users have to take additional action, such as selecting or rolling over primary information, to obtain qualifier values. Although this may be acceptable in some situations, in many situations it would be preferable to integrate the meta-information directly into the primary information display. This includes time-pressured situations, situations in which the ability to interact is limited (e.g., in vehicles, or with handheld devices), and high-risk situations in which failure to ask for or receive information qualifiers may lead to serious errors.

This paper reports research that examined alternative visualization mechanisms for directly embedding meta-information in information displays to facilitate assimilation of that meta-information by the user. There has been limited empirical research on the factors that influence how representations of uncertainty, information age, or other forms of meta-information are understood and used; instead, research has primarily focused on visualization techniques. However, before the role of meta-information in general decision-making tasks and visualization can be studied, one must first assess the usefulness of meta-information representations at a more basic level. Therefore, we conducted two studies that examined the efficacy of different graphical techniques for communicating meta-information such as uncertainty, information age, and source quality.

Previous Research

Much of the existing research base has focused on methods to visualize uncertainty. There is general consensus that it is important to communicate errors or uncertainty associated with information such as scientific data or geospatial regions. In these applications, the term *uncertainty* is used to describe errors or imprecision in measurement as well as conflicting outputs from different scientific models of the same phenomenon. This research has tended to focus on developing novel techniques for visually presenting uncertainty information, with little attempt to evaluate their effectiveness for users. For example, a recent review of papers from the geographical information systems (GIS) and scientific visualization fields, MacEachren et al. (2005, p. 139) concluded that "we do not have a comprehensive understanding of the parameters that influence successful uncertainty visualization." They also noted that "the research seems to take for granted that visual depictions of uncertainty are useful for decision making" (p. 141).

Within GIS, Buttenfield and Weibel (1988) suggested a set of mappings from types of data qualifiers to appropriate forms of visual representation, such as using object shape to code uncertainty in position and using color saturation to code uncertainty in attribute accuracy. Gershon (1998) introduced a useful distinction between intrinsic and extrinsic methods to represent uncertainty regarding an object. An intrinsic representation incorporates the representation into the object itself, whereas an extrinsic representation uses forms that annotate the object (e.g., pie graphs, error bars, numeric tags). Linguistic representations such as the words *probable* and *unlikely* are one type of extrinsic representation that has been extensively studied (Budescu, Weinberg, & Wallsten, 1988; Wallsten & Budescu, 1995) and compared with graphical representations (Bisantz, Marsiglio, & Munch, 2005).

Common intrinsic manipulations include the use of various dimensions of color (e.g., hue, saturation, brightness), texture (Botchen, Weiskopf, & Ertl, 2005; Interrante, 2000; Urness, Interrante, Marusic, Longmire, & Ganapathisubramani, 2003), and object shape. Objects whose shape (size and configuration) encodes values of variables (e.g., bubbles whose size corresponds to uncertainty level) are commonly referred to as *glyphs* (Aigner, Miksch, Thurnher, & Biffl, 2005; Pang, Wittenbrink, & Lodha, 1997; Wittenbrink, Pang, & Lodha, 1996).

Other intrinsic representations include using blurriness (Bisantz et al., 2005; Finger & Bisantz, 2002; MacEachren, 1992), animation or "vibration" (Brown, 2004; Ehlschlaeger, Shortridge, & Goodchild, 1997; Howard & MacEachren, 1996; MacEachren, 1992), extra dimensionality (Brown, 2004; Gershon, 1998), and transparency. MacEachren et al. (2005) described two contrasting techniques for utilizing transparency to convey uncertainty. In one technique, a "fog" would float over the representation, with less certain information obscured by a more opaque fog (MacEachren, 1992). In another technique (Drecki, 2002), less certain objects are more transparent, letting more background show through.

Color, an obvious candidate for visual coding of various types of meta-information, has several aspects that can be manipulated (Wyszecki & Stiles, 1982)—for example, *hue,* which distinguishes among colors such as red, green, purple, and yellow; *saturation,* which refers to how far a color is from a gray of equal intensity; *lightness,* which embodies the achromatic notion of perceived intensity of a reflecting object; and *brightness,* which is used instead of lightness to refer to perceived intensity of a self-luminous (i.e., light-emitting) object (Foley, van Dam, Feiner, & Hughes, 1995).

Color variables have been used by a number of researchers to code meta-information. For instance, Nadav-Greenberg, Joslyn, and Taing (2008) used hue to code a range or the worst-case value of predicted wind speeds, along with the median speeds, on a map; larger ranges (more uncertainty) or greater values were shown using redder hues. They compared these representations with box plots and showed that users were better able to assess the size of the range when it was presented directly through color coding, rather than when the worst-case data were presented.

Pfeiffer (2002) used brightness and saturation to represent belief and confidence in the presence of an object known to a robot. Howard and MacEachren (1996) used color saturation to indicate uncertainty, either on a map different from the one indicating data values or combined in one display, with color lightness indicating data values. Slocum, Cliburn, Feddema, and Miller (2003) used the brightness of three hues (red, green, and blue) to represent uncertainty about three variables related to water availability. Darker regions were more certain, and users had to interpret mixtures of these colors. Aerts, Clarke, and Keuper (2003) displayed uncertainty in urban growth predictions using color lightness (lighter red corresponded to areas of more uncertain growth, and darker red indicated more certainty).

Interestingly, there is a general tendency to assume a natural mapping between level of salience and level of certainty. Researchers have generally assumed that darker or more intense (e.g., more saturated) colors should be associated with

greater levels of certainty, whereas less certain information should be paler or faded and, therefore, may be less salient (Aerts et al., 2003; Endsley, 2003; Evans, 1997; Hengl, 2003; MacEachren, 1992; Slocum et al., 2003). Making glyphs or regions more transparent, or obscuring them with "fog," has a similar effect.

Other researchers have explicitly noted that some uncertainty displays call attention to areas of less certainty. For instance, in some cases, displays were created to call attention to differences across simulated or model outputs, through techniques such as animation (Ehlschlaeger et al., 1997), 3-D glyphs (Schmidt et al., 2004), or contrast with the background (Howard & MacEachren, 1996). In other cases, researchers have suggested mitigating this phenomenon. Interrante (2000), in discussing the use of texture regularity to code uncertainty, noted that using irregularity in hue distribution as a coding scheme for uncertainty may inappropriately draw attention to the more uncertain areas of a display (because regular changes in hue would be less salient). Schaefer, Gizdavu, and Nicholls (2004) used ovals to reflect the degree of uncertainty regarding potential air traffic conflicts and the time of the potential conflict. The authors adjusted the color of the ovals so that larger (and therefore more uncertain) bubbles were lighter, whereas smaller and more certain ovals were more intensely colored, to reduce the salience of the larger, more uncertain glyphs.

There appears to be an implicit assumption among researchers that uncertain information should be visually coded to be less salient (e.g., by making it more faded, gray, lighter, blurry, or transparent). However, although this assumption is near universal, it has not been empirically tested. A primary objective of the current studies was to provide an empirical test of this widely assumed belief.

In addition, the studies were designed to examine visual coding of other types of meta-information beyond degree of certainty. For example, if information that is less certain should be represented in a less salient manner, does this also hold for other types of meta-information (e.g., information that is older or from a less reliable source)? Finally, the studies were designed to examine the role of task context on how meta-information should be visually coded. Specifically, the studies examined whether task context should influence decisions about whether to highlight the higher (e.g., more certain, more recent, more trustworthy) or lower (e.g., less certain, less recent, less trustworthy) values of meta-information.

Method: Study 1

Purpose

The first study investigated the utility of alternative graphical techniques for representing different types of meta-information on maps. Specific questions addressed in the first study were as follows:

- Can participants reliably rank and/or rate levels of meta-information along different perceptual attributes of visual stimuli (hue, transparency, brightness, or saturation)?

- Do rankings/ratings change depending on the type of meta-information being coded? Specifically, is there a "natural mapping" between level of certainty and the graphical code used (e.g., more certain information should have higher saturation, brightness, or opacity), as past researchers have generally assumed? Are there also "natural mappings" for other types of meta-information, such as information age, and if so, is the direction of the mapping the same across different types of meta-information?
- Can participants reliably rank/rate across differing numbers of levels of detail or granularity? We tested 4, 8, and 12 levels of each type of meta-information.
- Can participants reliably rank/rate against a multicolored map representative of U.S. military maps as well as they can against a neutral background? This last question was included to increase the ecological validity and generalizability of the results.

Participants

Thirty volunteers (21 men and 9 women, aged 20 to 29 years) from a university population were paid $15. They were not color deficient (verified via the Ishihara Test), and all had computer experience and self-reported normal or corrected-to-normal visual acuity.

Experimental Stimuli

Participants were shown a 30.5- × 30.5-cm area on a CRT computer monitor that displayed multiple 2.5- × 2.5-cm square colored elements. Elements were placed on randomly selected centers of a 7 × 7 grid overlaid on the area to ensure consistent minimum distances between the elements. (This grid was not displayed to the participants.)

Independent Variables

Level of granularity, background, display attributes, and task framing were manipulated.

Level of Granularity. Four, 8, or 12 levels of meta-information were displayed on the map, using 4, 8, or 12 square colored elements (2.5 × 2.5 cm).

Background. The background of the display consisted of either a neutral gray (hexadecimal code #CCCCCC) overlaid with a black grid (with a few grid lines emphasized and labeled to simulate roads on a map) or a map showing geographical features and political boundaries, sampled from a standard U.S. military map (selected to ensure that the colors and texture on the map were representative of military land maps). These backgrounds were selected to differentiate between a neutral and a representative map background and were not intended to represent the same location.

Display Attributes. Four sets of colored elements were systematically developed to include changes in hue (multiple hues), level of brightness (for a lavender hue),

level of saturation (for a magenta hue), and level of transparency (for a red hue). Each set was combined with the two backgrounds, at 4, 8, or 12 levels of meta-information, to create a total of 24 maps. The hues and levels of brightness, saturation, and transparency used were chosen based on principles of color models and past research in this area (Ware, 2004). Issues relevant to stimulus selection included natural mappings (Bennett, Nagy, & Flach, 1987), recognizable ranges, perceptual distinctness, equidistance across heterogeneous backgrounds (Berlin & Kay, 1969; Hering, 1964; MacAdam, 1942; Munsell Color Company, 1976; Post & Greene, 1986; Rosenholtz, Nagy, & Bell, 2004), and accurate reproduction capability on computer hardware (Glassner, 1995; Khang & Zaidi, 2002).

The constant hues for brightness, saturation, and transparency were selected so that they were significantly different from the average background color because the study focus was on assigned rankings and ratings, rather than visual search for the display elements. An opaque border was also used around the colored squares in the transparency condition so that the location of the squares was apparent. The colors were also chosen to maximize the number of perceptually distinct colors as brightness and saturation varied (certain hues have a greater number of perceptually distinct variations along brightness and saturation dimensions; MacAdam, 1942; Munsell Color Company, 1976).

For the hue condition, the set of hues was chosen by using the 11 nameable colors (Ware, 2004) and adding cyan to reach the desired 12 levels of granularity. To choose the set of eight hues, we eliminated pink, brown, gray, and cyan; for the set of four hues, we chose red, green, yellow, and blue. Figure 1 shows task stimuli for each display attribute, for both background conditions, for the 8-level condition of granularity.

Task Framing. The experiment instructions were framed in two ways: display elements were described as representing either information uncertainty or information age (latency).

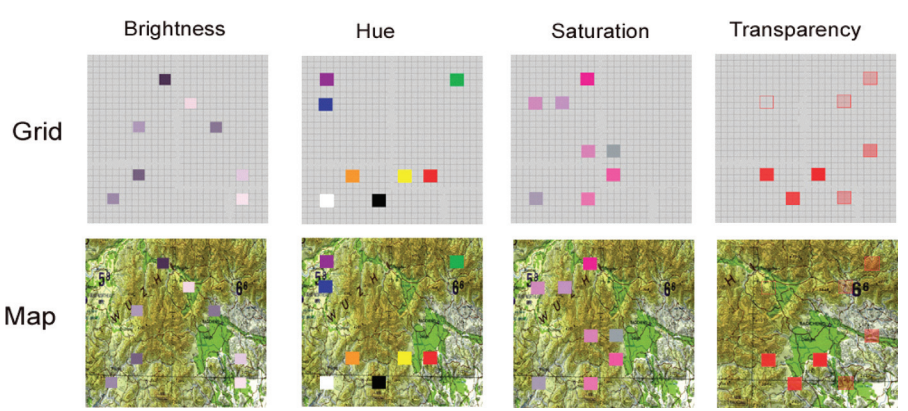

Figure 1. Display areas created for the 8-level condition.

Experimental Design

The experiment was conducted using a 4 (display attribute) × 2 (background) × 3 (level of granularity) × 2 (framing condition) mixed design. Display attribute and background were within-subject conditions, whereas level of granularity and framing were between-subject conditions. Five participants were assigned to each between-subject combination. Eight maps (4 display attributes × 2 backgrounds) were presented to each participant.

Experimental Tasks and Procedure

All participants performed two tasks—a ranking task and a rating task—in a private laboratory. The lighting in the room was controlled to reduce the ambient light in the room and eliminate any glare on the screens.

The experiment took about 1 hr, on average. After participants gave informed consent and were screened for color deficiency, they completed a short training task that allowed them to practice ranking and rating the display elements, using colors that were not used in the actual experiment. Participants then performed the ranking task, were offered a short break, and then performed the rating task. For both tasks, the eight maps were presented to participants in random order.

In the ranking task, participants were asked to rank the elements according to either latency or probability, depending on the framing condition, by dragging numbers (i.e., the integers 1 to 4, 8, or 12) from the side of the display onto the elements. Participants ranked all elements on each display before moving to the next map, and they could change the rankings as they wished before they moved on to the next map. For the rating task, a circle randomly appeared around one of the elements, and participants were asked to move a slider to a position between 1 and 100 to assign a value of latency or probability to that element. Then a new element was circled, until all elements were rated. Participants could not view the ratings they had already provided. Participants rated all the elements on one map before another map was shown.

For the ranking task, participants were shown the following instructions on the screen: "Please rank the regions according to the chance that a thunderstorm will occur in the region [or according to how old the information about the potential thunderstorm is]. Use 1 to equal Most Likely [or most recent] and 4 [or 8 or 12] to equal Least Likely [or oldest]." For the rating task, participants were shown the following: "Please rate the circled region according to the chance that a thunderstorm will occur in the region [or how old the information about the potential thunderstorm is]. 0 = No Chance [or 0 hours old]: 100 = Certain [or 100 hours old]." Participants were not provided with normative endpoints or legends in either task (i.e., they were not told if the most saturated or brightest element corresponded to the most recent, or most certain, information) so that we could determine if there was a "natural" or stereotypical direction to the ranking.

Ranks and ratings assigned by participants for each element were automatically logged. Time (in seconds) to complete the rating or ranking task was also recorded.

Results: Study 1

Assigned Rankings and Ratings

The primary measures of interest were the ranking and rating data. To analyze the degree to which participants could order regions consistent with expectations, rank orders for each participant were correlated (using the Spearman's rank order correlation coefficient) with a normative, or expected, order for each display type, based on the level of saturation, brightness, or transparency (similar to the methodology used by Bisantz et al., 2005, for analyzing uncertainty icons). For ratings, expected ratings (based on the number of levels) were computed by dividing the range 0 to 100 by the number of intervals required and taking the midpoint of those intervals (e.g., for the 4-level condition, there were four intervals of 25, with midpoints of 12.5, 37.5, 62.5, and 87.5).

Ratings from each participant were correlated (using Pearson's correlation coefficient) with these expected ratings for each display type. Note that for the hue display condition, the initial normative ordering was somewhat arbitrary, resulting in almost no significant correlations. Therefore, post hoc analysis (through inspection and through trial and error) was performed to determine an ordering that was representative of the orders produced by the participants. To best reflect participants' orderings of hue, different orders were constructed for the ranking and rating tasks.

As shown in Table 1, correlations between these orders and participants' orders were generally high. Table 1 shows the range of the absolute value of correlation coefficients for all significant correlations. Correlation results were summarized according to the total number of significant correlations (at $\alpha = .05$). Significant correlations provide an indication of the degree to which participants could order the areas consistently (with respect to a standard order) and how this degree differed based on display type, background, or number of levels. Results were similar in terms of the number of significant correlations across display type (other than hue), framing, background, and number of levels for both the ranking and rating tasks (see Table 2).

Additionally, we analyzed the direction of the correlations to understand if there were "natural" endpoints or ordering directions (i.e., whether participants tended to

TABLE 1. Maximum and Minimum Significant Correlation Coefficients, by Display Condition, for the Ranking and Rating Tasks

	Ranking Task		Rating Task	
	Min.	Max.	Min.	Max.
Hue	.615	.786	.612	.975
Saturation	.58	1	.649	1
Brightness	.734	1	.596	.999
Transparency	.748	1	.621	1

equate low saturation with older, or highly uncertain, data). There was an interesting result regarding the direction of ordering. Participants in the uncertainty framing condition tended to assign more certain rankings or ratings to more "intensely" colored squares, whereas the direction was mixed when participants' ranks were based on information age (see Table 3). For instance, for the brightness condition there were 22 significant correlations in which participants ordered pale pink to dark purple squares in the direction from least likely to most likely, but there was only 1 in the reverse direction (mapping pale pink to most likely and dark purple to least likely). When participants were assigning ranks according to information age, half of the significant correlations were in one order and half in the other.

Another way to assess the degree to which regions could be effectively rank ordered or rated is to examine the mean, median, and standard deviation of the rank or rating for each region. Because we did not force participants to rank in a particular direction (e.g., instruct them that very transparent corresponded to the endpoint of oldest), participants did vary in the direction with which they ordered the regions. Therefore, a simple measure of central tendency for rankings would not be appropriate. To account for these differences in ordering direction, we first transformed all rankings or ratings for participants and trials in which the measured correlation was negative. Transformed ranks were computed as $(N + 1) - R$, in which N was the number of levels (4, 8, or 12) and R was the assigned rank. Thus, a rank ordering of regions of 7 – 8 – 5 – 4 – 2 – 1 – 3 transforms to 2 – 1 – 4 – 5 – 7 – 8 – 6. Transformed ratings were computed as $100 - R_t$, in which R_t was the assigned rating.

Figure 2 shows, for each experimental condition, the maximum and minimum of the average rankings assigned to any region in the condition, along with the mean and median of those average ratings. Note that the minimum of the average rankings is close to 1 as expected; maximum rankings for the 8- and 12-level conditions are somewhat lower than the expected maximums in some conditions; and the mean and median are consistent with the expected values of 2.5, 4.5, and 6.5. Figure 3 shows, for each experimental condition, the maximum and minimum of the average ratings assigned to any region in the condition, along with the mean and median of those average ratings. Note that the mean and median values are for the most part near 50, as expected, and that the 8-level condition tends to have a greater range. Overall, however, standard deviations for mean ratings were high (ranging from 3.27 to 44 over all ratings); further research using appropriate psychophysical methods would be necessary to develop a validated rating scale associated with these stimuli.

Task Time

Task time was also measured for each display/background combination (eight ranking times and eight rating times per participant). Mean times (with standard deviations) were 21.65 s (11.11), 40.71 s (22.44), and 69.16 s (48.90) for the ranking task and were 33.07 s (17.14), 57.55 s (29.56), and 97.97 s (54.91) for the rating task, for the 4-, 8-, and 12-level granularity conditions, respectively.

TABLE 2. Number of Significant Correlations, for Each Background × Display × Level Condition, for the Ranking and Rating Tasks

	Ranking Task							Rating Task						
	Map Background			Grid Background				Map Background			Grid Background			
Levels:	4	8	12	4	8	12	Ranking Total	4	8	12	4	8	12	Rating Total
Hue	2	3	7	2	2	7	23	0	5	8	1	6	8	28
Saturation	5	9	7	5	8	8	42	6	10	7	5	8	7	43
Brightness	6	9	8	7	9	8	47	6	9	7	5	10	10	47
Transparency	8	8	8	7	9	9	49	6	10	9	4	10	10	49
Total by level	21	29	30	21	28	32		18	34	31	15	34	35	
Total		80			81		161		83			84		167

TABLE 3. Number of Significant Correlations in Each Direction, for Each Framing × Display Condition, for the Ranking and Rating Tasks

	Ranking Task				Rating Task			
	Least Likely–Most Likely	Most Likely–Least Likely	Oldest–Newest	Newest–Oldest	Least Likely–Most Likely	Most Likely–Least Likely	Oldest–Newest	Newest–Oldest
Saturation Gray–bright pink	20	2	15	5	20	0	16	7
Brightness Pale pink–dark purple	22	1	12	12	23	0	12	12
Transparency Transparent red–opaque red	20	4	18	7	24	0	12	13
Hue Pale–dark	15	0	8	0	11	2	13	2

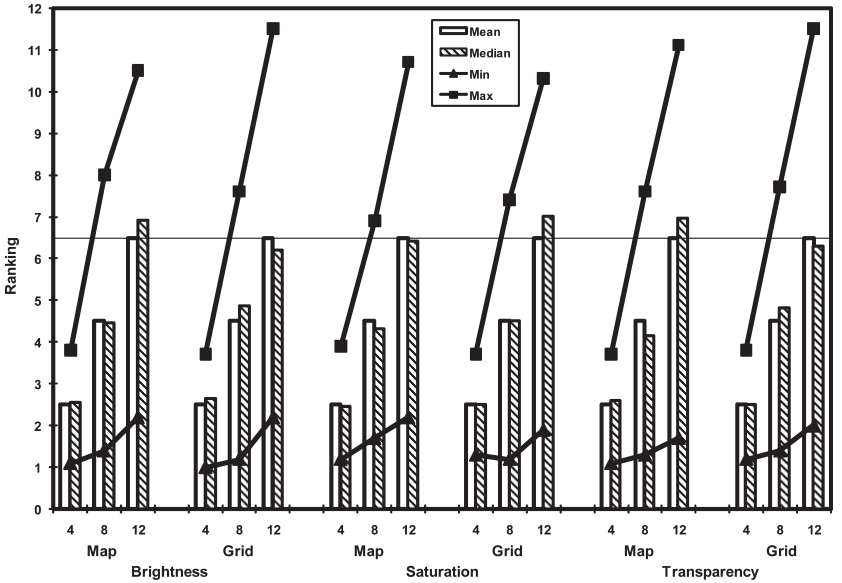

Figure 2. Mean, median, maximum, and minimum average rankings, for all conditions.

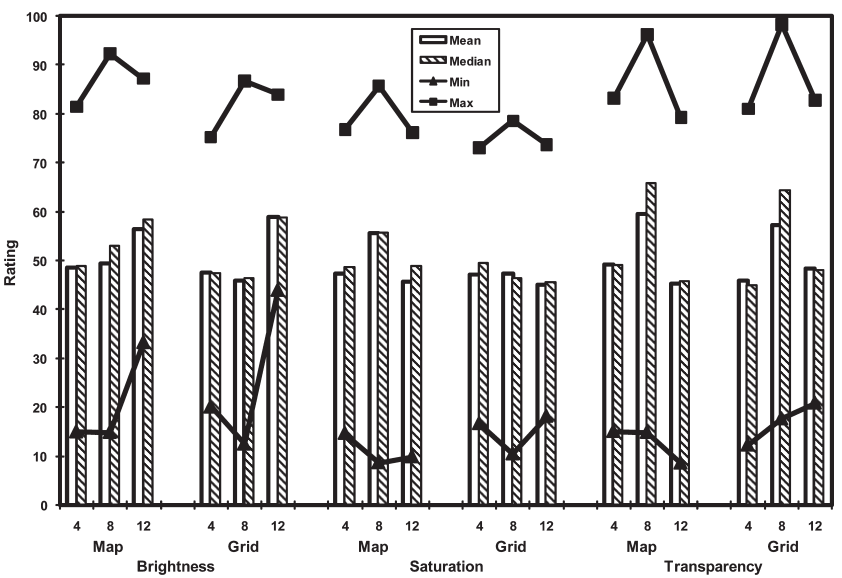

Figure 3. Mean, median, maximum, and minimum average ratings, for all conditions.

Note that the task times on a per-item basis were consistent across levels—that is, it did not take longer to rate or rank each item when there were more items in the set.

Times were analyzed using a mixed-effects ANOVA, with framing condition used as a between-subject variable and display type and map background as

within-subject variables. Separate ANOVAs were run for rating and ranking tasks and for the 4-, 8-, and 12-level conditions (because of obvious differences in task content across the levels and task type). Analysis showed almost no significant differences attributable to framing condition, display type, or map background and no significant interactions. The sole exception was that of the ranking task, for Level 8, for which there was a main effect of background, $F(1, 8) = 8.502$, $p = .019$. The lack of any strong pattern across analyses indicates that there were no reliable differences in task time attributable to the independent variables tested.

Summary

These results indicate that participants can rank and rate the representations in the expected order for the saturation, brightness, and transparency conditions. This provides empirical confirmation that saturation, brightness, and transparency can be effective means to encode different types of meta-information.

The analysis of average rankings and average ratings provided some indication that fewer levels are preferred: in terms of ratings, the 8-level condition had the largest spread between average minimum and maximum ratings, and the 4- and 8-level conditions had minimum and maximum rankings close to the expected levels. The 12-level condition had a smaller overall spread in terms of ratings (whereas it should have had the largest, from 4.1 to 95.8, if all levels had been evenly distributed).

The difference between the 12-level condition and the conditions with fewer levels may be attributable to working memory limitations (e.g., participants may have had more difficulty remembering previously assigned ratings when the number of levels was greater than the typical working memory span). From a design standpoint, selection of an appropriate number of levels should be guided first by task demands, such as the level of detail necessary for people to differentiate among potential actions. Information should not be displayed at a greater level of detail than is required by the task.

There was no apparent effect of background, which may indicate some flexibility in terms of design. However, this conclusion warrants further study, for several reasons. First, the map background was chosen to avoid areas with water (normally colored blue). Additionally, the hues selected for the saturation, brightness, and transparency condition were carefully selected to maximize their difference from the hues contained in the map: however, we did not control the degree to which individual levels of hue, saturation, or brightness contrasted with the background. Thus, although the results of this experiment established a tendency to assign higher levels of certainty to darker, more saturated colors, it is not clear if this is attributable to a mapping between uncertainty level and color (e.g., high certainty is always indicated by high saturation) or to the impact of the salience of the colors against the background map (e.g., high certainty is always indicated by highly salient colors).

To disambiguate these alternative hypotheses we conducted a second study, in which we manipulated the background, so that the same level of color saturation would be more or less salient depending on the background it was presented against. If what matters is level of saturation (or brightness, or transparency) and

not saliency, then the pattern of results should be the same across the different backgrounds. If what matters is visual saliency, then the pattern of results should reverse, depending on background.

Additionally, Study 1 indicated some differences in the direction of mappings across meta-information types. Whereas participants consistently assigned less intensely colored areas to less certain information, the directionalities assigned for latency were mixed. It is not clear whether these results are attributable to fundamental differences in how participants treat different types of meta-information or if other factors, such as differences in how participants interpreted the relevance of the meta-information to the task, influenced their performance.

It is possible that perceived task relevance affects how participants assign meta-levels to visual codes. Participants may have uniformly assumed that more "certain" information was more important to highlight (through the use of either more intense or more salient colors). There may have been more variability in how they interpreted the relevance or usefulness of information age to the task. Some participants may have thought that "older" information was more useful to the task (perhaps because the information about the storm had remained stable over time and thus was more reliable) and should therefore be highlighted, whereas others may have thought that "more recent" information was more relevant (as it represented a newer storm forecast) and should therefore be highlighted.

To explicitly test whether the mapping of meta-information levels to visual codes is a function of task context, we manipulated task context in the second study, so that in some cases the low end of the meta-information dimension (older, less certain, less reliable source) was more task relevant and in other cases the higher end (newer, more certain, more reliable) was more task relevant.

Method: Study 2

Purpose

The purpose of this experiment was to clarify the findings from Study 1 by testing specific hypotheses regarding the mapping of levels of meta-information to the visual characteristics of a region, as described previously. Thus, Study 2 examined whether higher values of meta-information are always assigned to higher values of a graphic attribute such as saturation (e.g., higher certainty is always indicated by high saturation) or whether the "natural" mapping of meta-information to graphic codes varies with the contrast of the meta-information representation with the background (i.e., the salience of the representation), with the objectives of the task, or with a combination of salience and task relevance. To do this, we carefully controlled the background and foreground colors so that we could differentiate whether people assigned colors to meta-information based on saturation level or salience, and we manipulated the task instructions so that high levels of meta-information were either more or less task relevant.

This experiment also examined a larger set of meta-information types by including source quality as well as uncertainty and information age. Because the pattern of

results across the attributes of brightness, saturation, and level of transparency was similar to that in Study 1, saturation was selected for use in Study 2. We designed the experimental stimuli and task instructions to allow the following hypotheses to be differentiated:

A. Participants will assign regions that are more saturated to higher levels of meta-information (in this case, more certain, newer, or a more trusted source).
B. Participants will assign regions that contrast most strongly with the background to higher levels of meta-information (in this case, more certain, newer, or a more trusted source).
C. Participants will assign regions that are more saturated to information that is more relevant to the task.
D. Participants will assign regions that contrast most strongly with the background to information that is more relevant to the task.

Participants

Thirty-six volunteers (23 men and 13 women, 19–67 years old, mean of 26 years) from the general university population were paid $15 each or were given class credit to participate. They were not color deficient (verified via the Ishihara Test) and had self-reported normal or corrected-to-normal visual acuity.

Experimental Stimuli

Participants were shown a 16.5- × 17.5-cm map (on a computer monitor) displaying four 2- × 2-cm white squares showing the numbers 1 to 4, onto which participants were asked to drag colored elements.

To generate the maps, eight nameable colors (Ware, 2004) were selected (red, green, blue, yellow, orange, purple, pink, and brown). These colors were used to create eight sets of four colored squares, which ranged from a relatively saturated level of color to a relatively unsaturated level. The color difference (measured with a Euclidean distance algorithm in the CIELAB color space; Ware, 2004; Wyszecki & Stiles, 1982) across these levels of saturation was consistent across the eight nameable colors.

These sets were paired with both a saturated and a neutral (gray, black, or white) background, overlaid with a simulated street grid, to create 16 map stimuli. For maps with the neutral background, the most saturated square exhibited the most contrast with the background. For maps with the saturated background, the least saturated square exhibited the most contrast with the background. Different neutral backgrounds were used to ensure that the color difference between each set of stimuli and its background was kept constant across nameable colors. For instance, a set of four blue squares was developed in which the squares ranged in saturation from very saturated blue to an unsaturated (grayish) blue. These squares were paired with either a saturated blue or a white background.

Independent and Dependent Variables

There were two within-subjects variables tested: meta-information type (uncertainty, source quality, and information age) and task framing (whether more or less

meta-information was task relevant), for a total of six within-subject combinations. Table 4 shows the instructions. Color-to-rank assignment and trial time were recorded.

Experimental Design

Meta-information type and task framing were crossed to form six within-subjects conditions. One experimental trial consisted of one map showing four

TABLE 4. Instructions for Study 2, for Three Types of Meta-Information and Two Types of Task Framing

More Meta-Information Is Task Relevant	Less Meta-Information Is Task Relevant
All Conditions	
You are designing map displays to help the city property tax assessment office. You will be responsible for selecting colors to represent qualities of information on a map. For each trial, drag appropriate colors onto the map, using each color once. [This appeared on every task screen for every trial.]	
Uncertainty	
The certainty of assessments can vary based on the information available from recent home sales. The map needs to show the neighborhoods where the assessed property values are MOST CERTAIN, so tax bills can be prepared and mailed to the houses in those neighborhoods. Assign colors to the regions according to the certainty of the property assessments that were assigned. "1" is the most certain assessment, and "4" is the least certain assessment.	The certainty of assessments can vary based on the information available from recent home sales. The map needs to show the neighborhoods where the assessed property values are LEAST CERTAIN, so more research on comparable properties can be performed. Assign colors to the regions according to the certainty of the property assessments that were assigned. "1" is the most certain assessment, and "4" is the least certain assessment.
Recency	
Each year the city reassesses different neighborhoods. The map needs to show which neighborhoods have been reassessed MOST RECENTLY so new tax bills can be generated. Assign colors to the regions according to their reassessment date. "1" is the most recent, and "4" is the least recent reassessment.	Each year the city reassesses different neighborhoods. The map needs to show which neighborhoods have been reassessed LEAST RECENTLY so new assessments can be planned. Assign colors to the regions according to their reassessment date. "1" is the most recent, and "4" is the least recent reassessment.
Information Source	
Neighborhoods have been assessed by assessors that have different levels of experience. The map needs to show which neighborhoods have been surveyed by assessors with MORE EXPERIENCE so that they can be used as training examples for new employees. Assign colors to the regions according to the experience of the assessor who did the work. "1" is the neighborhood assessed by the most experienced assessor, and "4" was assessed by the least experienced assessor.	Neighborhoods have been assessed by assessors that have different levels of experience. The map needs to show which neighborhoods have been surveyed by assessors with LESS EXPERIENCE so that the assessments can be checked. Assign colors to the regions according to the experience of the assessor who did the work. "1" is the neighborhood assessed by the most experienced assessor, and "4" was assessed by the least experienced assessor.

regions for the participant to rank. To reduce the number of experimental trials, only half of the nameable colors were used for each participant (e.g., maps generated from four of eight nameable colors). Participants experienced either the green, orange, purple, and red stimuli (Color Set 1) or the blue, brown, pink, and yellow stimuli (Color Set 2), each with both the saturated and neutral backgrounds, for a total of eight map types. Within each Meta-Information × Framing condition, participants completed 16 trials (2 trials with each background map). To avoid confusing participants with different instructions on each trial, trials were blocked by the six Meta-Information × Framing conditions. Trials were randomly ordered within each block, and block orders were balanced using a Latin square design. Three participants were assigned to each of six orders, in each color set (18 participants per color set).

Experimental Task and Procedure

Participants performed a ranking task in a setting similar to that in Study 1. After participants gave informed consent and were screened for color deficiency, they completed a short training exercise and then performed the experimental task. The experiment took about 45 min, on average. Participants were asked to assign colors to the numbered regions according to the instructions provided by dragging colored squares from the right of the screen onto the uncolored numbered squares on the maps. Breaks were offered between blocks.

Results

Assignment matrices were created for each Color Set (2) × Meta-Information Type (3) × Framing (2) × Background condition (2), for a total of 24 matrices. Each matrix showed the mapping between participant rankings and colored squares (see Table 5 for an example). Entries represent the number of times that participants assigned a particular colored square to a particular ranking, across all trials and participants.

Entries on the diagonals reflect situations in which participants put the colored squares in order (e.g., from least to most saturated, or vice versa). Entries on the diagonal running from the top right to bottom left reflect an ordering in which more saturated squares were assigned higher levels of meta-information, whereas entries on the opposite diagonal reflect the reverse. In the example shown in

TABLE 5. Example Assignment Matrix

	1 = Grayish	Color 2	Color 3	4 = Bright
1 = Most certain	44	2	3	**95**
2	4	45	**94**	1
3	3	**94**	42	5
4 = Least certain	**93**	3	5	43

Note. This assignment matrix is for the Color Block 1/uncertainty/saturated background/less is relevant framing condition. There were 576 total assignments, of which 65% were on the top right to bottom left diagonal (shown in boldface type).

Table 5, 95% of the assignments were on one of the two diagonals (indicating that participants could effectively order the colored squares), and the majority were on the top right to bottom left diagonal, indicating a mapping in which more-saturated squares were assigned to higher levels of certainty.

Table 6 shows the number of assignments on each diagonal for each of the assignment matrices. The percentage of assignments on the diagonal was high in all cases, ranging from 92% to 97%. We conducted chi-square goodness-of-fit tests to determine whether there was a nonchance assignment of rankings to one diagonal as compared with the other. Results indicated that assignment was significantly different from chance assignment in 21 out of 24 cases. Counts shown in bold in Table 6 indicate the preferred diagonal for each matrix.

For each experimental condition, it is possible to determine whether the rankings were consistent or inconsistent with Hypotheses A through D. For instance, in Table 5, the results were consistent with Hypothesis A, that more saturated colors will be assigned to higher levels of meta-information, as well as Hypothesis D, that colors that contrast most with the background (in this case, least saturated colors) will be assigned to more relevant levels (in this case, lower levels of meta-information). Table 6 shows whether the preferred diagonal was consistent or inconsistent with each hypothesis, for each matrix.

Hypothesis D was supported in 18 out of 24 comparisons, with only 3 significant comparisons showing inconsistent evidence. In contrast, Hypotheses A, B, and C were supported in 12, 11, and 13 cases, respectively, but also had 9, 10, and 8 significant, inconsistent comparisons each. Note that the pattern of results shown in Table 6 was consistent across types of meta-information and generally across color set, although the instances in which assignment to one or the other diagonal was not different from chance occurred in the second color set.

The four hypotheses can also be directly compared. For each pair of hypotheses, 12 of 24 cases should be the same (e.g., the same diagonal preferred) and the other 12 should differ. Table 7 shows the number of cases in which hypothesis listed in the row was preferred over that in the column (the maximum number in each cell is 12). Similar to the results in Table 6, this comparison shows stronger evidence for Hypothesis D (that more relevant levels of meta-information will be assigned colors that contrast most strongly with the background), in terms of both greater support for Hypothesis D and less support against Hypothesis D, than for the other hypotheses.

Task time was also analyzed across the meta-information type, framing, and background conditions. A three-way, within-subjects ANOVA showed no main effects or interactions for analysis of logarithmically transformed trial times (to meet normality assumptions of the ANOVA). When color set was added as a between-subjects variable, we were unable to transform the data to meet normality assumptions, so instead we used a Kruskal-Wallace test to compare conditions within each color set. Conditions were not significantly different within Color Block 1; however, there was a significant difference ($p < .05$) across conditions within Color Block 2. Median times for Color Block 2 are plotted in Figure 4 and show a decrease across

TABLE 6. Overall Results for Study 2

Color Block	Framing	Background	Type	Sat. = Most, Gray = Least	Gray = Most, Sat. = Least	% on Diagonal	χ^2	*p*	Hypothesis			
									A	B	C	D
1	H	Neu.	U	**322**	230	96%	15.3	.0001	C	C	C	C
1	H	Neu.	R	**324**	214	93%	22.1	<.0001	C	C	C	C
1	H	Neu.	S	**429**	113	94%	183.1	<.0001	C	C	C	C
1	H	Sat.	U	**324**	216	94%	21.2	<.0001	C	I	C	I
1	H	Sat.	R	**329**	213	94%	24.4	<.0001	C	I	C	I
1	H	Sat.	S	241	**315**	97%	9.6	.002	I	C	I	C
1	L	Neu.	U	171	**369**	94%	71.9	<.0001	I	I	C	C
1	L	Neu.	R	215	**329**	94%	23.5	<.0001	I	I	C	C
1	L	Neu.	S	189	**349**	93%	47.0	<.0001	I	I	C	C
1	L	Sat.	U	**376**	174	95%	73.5	<.0001	C	I	I	C
1	L	Sat.	R	**432**	118	95%	178.1	<.0001	C	I	I	C
1	L	Sat.	S	**353**	191	94%	47.6	<.0001	C	I	I	C
2	H	Neu.	U	**372**	172	94%	72.8	<.0001	C	C	C	C
2	H	Neu.	R	**390**	156	95%	99.4	<.0001	C	C	C	C
2	H	Neu.	S	**329**	201	92%	30.4	<.0001	C	C	C	C
2	H	Sat.	U	174	**372**	95%	71.1	<.0001	I	C	I	C
2	H	Sat.	R	164	**386**	95%	88.8	<.0001	I	C	I	C
2	H	Sat.	S	172	**368**	94%	70.4	<.0001	I	C	I	C
2	L	Neu.	U	235	**315**	95%	11.4	0.0008	I	I	C	C
2	L	Neu.	R	**276**	270	95%	0.0	*ns*	—	—	—	—
2	L	Neu.	S	**278**	266	94%	0.2	*ns*	—	—	—	—
2	L	Sat.	U	**302**	254	97%	4.0	.046	C	I	I	C
2	L	Sat.	R	244	**294**	93%	4.5	.0347	I	C	C	I
2	L	Sat.	S	258	**284**	94%	1.2	*ns*	—	—	—	—

Note. For framing, H = high meta-information is task relevant and L = low meta-information is task relevant. For background, Neu. = neutral and Sat. = saturated. For meta-information type, U = uncertainty, R = recency, and S = source. Counts shown in boldface type indicate the preferred diagonal for each matrix. Percentages are the percentage of assignments on either diagonal. Statistical results indicate whether assignment to one versus the other diagonal was different from change. For the hypotheses, C = finding consistent with the hypotheses and I = finding that was inconsistent.

TABLE 7. Number of Cases in Which the Hypothesis Listed in the Row Was Preferred Over That in the Column

	A	B	C	D	For
A	—	6	4	2	12
B	5	—	4	1	10
C	5	6	—	3	14
D	8	8	8	—	24
Against	18	20	16	6	

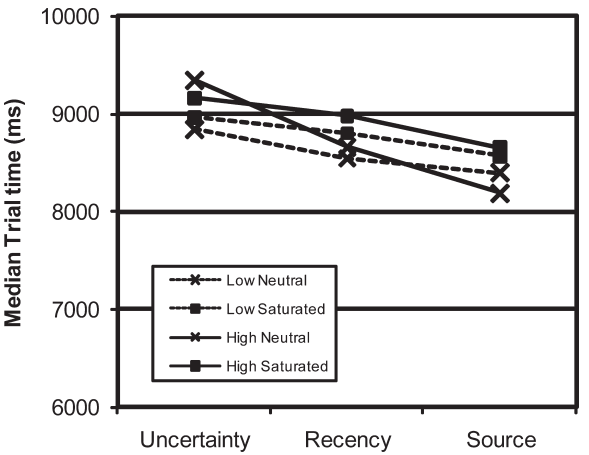

Figure 4. Median trial time across conditions for Color Block 2.

meta-information conditions, providing some indication that neutral background may prove advantageous over saturated background in terms of time.

Summary

Results from Study 2 provide support for the hypothesis that regions that contrast most with the background will be assigned to levels of meta-information that are most relevant to the task. When it was necessary to create maps for a task that focused on older, less trustworthy, or uncertain assessments, these were assigned to regions with the most contrast; the reverse was true if the task depended on information about newer, more trustworthy, or more certain assessments. Thus, what was important was not the color of the region or the level of meta-information but, rather, the color relative to the background and the relevance of the meta-information.

Discussion

This set of studies represents an initial investigation into the factors that influence the mapping between visual codes and types of meta-information. It provides some of the first empirical evidence of the usability of different visual codes

to communicate meta-information. The results provide guidance for how designers can represent meta-information in displays.

Study 1 established that saturation, brightness, and transparency are all potentially effective visual attributes for communicating meta-information, including uncertainty and information age. These results are encouraging in that they suggest that designers have flexibility when selecting the type of graphical attribute for representing different meta-information. The findings should be further verified based on research in more operational tasks, in which the impact of display attributes on decision making or task performance can also be assessed. Additionally, effects might differ with background/display element combinations other than those we tested. Using a set of nameable hues, however, is not recommended, given the inconsistencies with which participants ordered the hue sets. There may be hue sets that have more ordinal qualities (e.g., those that fall along a line in the color space model), and these should be assessed in future studies.

The results of the studies also clarified the factors that influence how people tend to map levels of meta-information to visual codes. Whereas past research implicitly assumed a natural mapping between higher levels of meta-information (e.g., high certainty) and more intense visual codes (e.g., darker, more saturated), the results of the present study suggest the relationship is more complex. The assignment of visual code to level of meta-information depends on task context as well as the salience of the representation. People tend to assign the most visually salient code to the most task-relevant information. Study 2, which carefully controlled for task context and visual salience independent of specific visual code, indicated that participants were most likely to choose a mapping that accounted for both information relevance and the salience of the region against the background. That is, participants tended to assign rankings based on display characteristics as well as the demands of the task. It should be noted that the participants did not have to use the output of their ranking to support further problem solving. In a situation where the assignment of colors supported subsequent tasks, one might expect an even stronger effect.

This finding has important implications for design: It suggests that designers must understand the use of meta-information in context, as well as characteristics of the information display (e.g., characteristics of the background map), when mapping visual cues to meta-information. In situations where backgrounds with heterogeneous characteristics are common (e.g., some maps), it may be possible to automatically determine representations that control for contrast with the background (e.g., provide a range of contrasts, and thus salience, against the background). However, there is potential for ambiguity in interpretation if the ordering of regions with respect to salience differs from the ordering with respect to intensity.

If the task context has clearly defined and unchanging operational goals, then more task-relevant information should be mapped to more salient visual codes. In cases where an operator's goals or task contexts vary, there is a trade-off between providing information in a standardized manner (e.g., always making more certain information more salient) versus representing the meta-information in a way

that is reflective of the changing task demands. Errors in interpretation could arise because of lack of consistency with the natural mapping (relevant information should be salient) or because of inconsistent mapping over time. If an operator uses a display frequently and must respond rapidly, standardization may be more appropriate. Alternatively, in cases where the task is infrequent or more deliberative, it may be desirable to allow the operator control over the direction of mapping, so that the most task-critical information can be displayed in a salient way.

Information display designers are regularly faced with decisions about the representation of meta-information, particularly uncertainty, without empirical guidance. These studies are among the first to empirically investigate the use of different visual coding techniques to represent meta-information or information qualifiers. In the tasks here, the qualifiers modified regions in space. In other contexts modifiers might be required for information values (e.g., a sensor value has an associated age) or categories (e.g., there is uncertainty regarding whether a contact is a submarine). Empirical investigations of effective techniques for communicating meta-information in these other types of context need to be performed.

One question is whether the results here are particular to meta-information or relate more generally to information display. Although our focus has been on communication of meta-information, we would expect the results to apply more generally. Certainly, it is appropriate to recommend that more task-relevant information be displayed in a more salient way, and whether one calls uncertainty or latency "information" or "meta-information" is context specific (Pfautz et al., 2005).

Study Limitations and Implications for Future Work

The research reported here was intended as a foundational step upon which future studies addressing the impact of meta-information display on decision making could build. Therefore, it focused on controlled tasks in a laboratory setting. The results indicated that multiple levels of meta-information granularity and different color attributes have potential value and should be further studied in more operational task environments that require decision makers to act based on the information provided.

The results also suggest additional directions for investigation. For example, research is needed to better understand the impact of "natural mapping" under conditions in which participants are explicitly told how to map the meta-information to the visual codes. In our study the participants were not given any guidance with respect to direction of mapping, but in most real-world applications a legend or training would be provided. The results of Study 2 suggest that information displays that map task-relevant levels of meta-information to the most strongly contrasting (or salient) regions should produce the best performance, even when legends and/or training are provided. This hypothesis needs to be tested using more realistic tasks.

There is also a need to test the generality of the conclusions of Study 2 across a wider range of visual coding techniques. The impact of other methods that can be

used to manipulate information salience (e.g., texture, animation) should be investigated with respect to meta-information representation and task demands. Finally, the results suggest that representations such as glyphs and uncertainty bubbles, which often use size to code the range of data (e.g., variance), may be more difficult to use in circumstances where it is the more certain information that needs to be highlighted, because larger graphical forms are inherently more salient. Research is needed to explicitly test this hypothesis and to explore visual coding techniques that can compensate for the inherent salience of larger graphical forms.

In addition, our studies considered only one graphical variable and meta-information type at a time. That is, participants were judging only the age, uncertainty, or trustworthiness of the information. They were not making combined judgments (e.g., age and uncertainty), nor were they judging the information value as well as a meta-information level (e.g., the value of the assessed property as well as the uncertainty regarding the assessment). Such multifaceted judgments are necessary in many tasks and should be the basis for additional research.

Other researchers have implemented displays in which more than one graphical code was used to show multiple types of information. For instance, Howard and MacEachren (1996) used color saturation to indicate uncertainty, either on a different map from the one indicating data values or combined in one display, with color lightness indicating data values. Slocum et al. (2003) used the brightness of three hues (red, green, and blue), as well as glyph and transparency manipulations, to represent uncertainty. They explored the usability of their system with domain experts, and they suggested that the method of combining colors might be difficult for some decision makers and that the intrinsic (colored) method was more suited to policy decision makers (who wanted a "big picture"), whereas glyphs, which may lead to more visual complexity on the display, may be more suitable for experts who want to access specific information. These claims constitute interesting hypotheses that should be empirically investigated.

Finally, the representational method we studied did not allow users to interactively adjust or compare aspects of the display. Other researchers have allowed users to toggle between displays or to select thresholds for displaying levels of uncertainty (e.g., Howard & MacEachren, 1996). Adding such flexibility provides an additional promising direction in which to explore in future research.

Conclusions

This research presents a systematic study of the effectiveness of different types of graphical representations using color to communicate levels of meta-information. Results indicate similar performance on rating and ranking levels of information across color variables (e.g., saturation, transparency, and brightness) and types of meta-information (uncertainty, information age, and source quality), indicating that designers have some flexibility when selecting display characteristics and their assignment to meta-information types. Results from both studies, however, indicate that the natural mapping of levels of meta-information to color intensity

is not straightforward. When developing information displays, designers need to consider both the salience of the representation against the background and the task relevance of meta-information levels.

References

Aerts, J. C. J. H., Clarke, K. C., & Keuper, A. D. (2003). Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science, 30*, 249–261.

Aigner, W., Miksch, S., Thurnher, B., & Biffl, S. (2005). PlanningLines: Novel glyphs for representing temporal uncertainties and their evaluation. In *Proceedings of the Ninth International Conference on Information Visualisation* (pp. 57–63). Piscataway, NJ: Institute of Electrical & Electronics Engineers.

Bennett, K. B., Nagy, A. C., & Flach, J. M. (1987). Visual displays. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 659–696). New York: Wiley.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Bisantz, A. M., Marsiglio, S., & Munch, J. L. (2005). The effects of format and level of specificity for displaying uncertainty. *Human Factors, 47*, 777–798.

Botchen, R. P., Weiskopf, D., & Ertl, T. (2005). *Texture-based visualization of uncertainty in flow fields.* In *Proceedings of IEEE Visualization 2005.* Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Brown, R. (2004). Animated visual vibrations as an uncertainty visualisation technique. In *Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (pp. 84–89). New York: Association for Computing Machinery.

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.

Buttenfield, B. P., & Weibel, R. (1988). *Visualizing the quality of cartographic data*. Paper presented at the Third International Geographic Information Systems Symposium (GIS/LIS 88), San Antonio, Texas.

Drecki, I. (2002). Visualisation of uncertainty in geographical data. In W. Shi, P. Fisher, & M. Goodchild (Eds.), *Spatial data quality* (pp. 140–159). London: Taylor & Francis.

Ehlschlaeger, C. R., Shortridge, A. M., & Goodchild, M. F. (1997). Visualizing spatial data uncertainty using animation. *Computers and Geosciences, 23*, 387–395.

Endsley, M. (2003, August). *Supporting assessments of uncertainty in complex worlds.* Paper presented at the XVth International Ergonomics Association Triennial Congress, Seoul, South Korea.

Evans, B. J. (1997). Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers and Geosciences, 23*, 409–422.

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.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1995). *Computer graphics: Principles and practice.* Reading, MA: Addison-Wesley.

Gershon, N. (1998). Visualization of an imperfect world. *IEEE Computer Graphics and Applications, 19*(4), 43–45.

Glassner, A. (1995). *Principles of digital image synthesis.* San Francisco: Morgan Kaufmann.

Hengl, T. (2003). Visualisation of uncertainty using the HIS colour model: Computations with colours. In D. Martin (Ed.), *Proceedings of the 7th International Conference on GeoComputation* (pp. 8–17). Southampton, UK: University of Southampton.

Hering, E. (1964). *Outlines of a theory of the light sense*. Cambridge, MA: Harvard University Press.
Howard, D., & MacEachren, A. M. (1996). Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Science, 23*, 59–77.
Interrante, V. (2000). Harnessing natural textures for multivariate visualization. *IEEE Computer Graphics and Applications, 20*(6), 6–11.
Khang, B., & Zaidi, Q. (2002). Cues and strategies for color constancy: Perceptual scission, image junctions, and transformational color matching. *Vision Research, 42*, 211–226.
MacAdam, D. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America, 32*, 247–274.
MacEachren, A. M. (1992). Visualizing uncertain information. *Cartographic Perspective, 13,* 10–19.
MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., et al. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science, 32*, 139–160.
Munsell Color Company. (1976). *Book of color.* Baltimore: Author.
Nadav-Greenberg, L., Joslyn, S. L., Taing, M. U. (2008). The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making, 2*, 24–27.
Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *Visual Computing, 13*, 370–390.
Pfautz, J., Fouse, A., Shuster, K., Bisantz, A., & Roth, E. (2005). Meta-information visualization in geographic information display systems. In *Digest of Technical Papers—SID International Symposium* (Vol. 36, pp. 1406–1409). San Jose, CA: Society for Information Display.
Pfeiffer, J. J. (2002). Using brightness and saturation to visualize belief and uncertainty. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrams 2002* (pp. 272–289). Heidelberg, Germany: Springer-Verlag.
Post, D., & Greene, F. (1986). Color name boundaries for equally bright stimuli on a CRT. In *Proceedings of the Society for Information Display, Digest of Technical Papers* (pp. 70–73). San Jose, CA: Society for Information Display.
Rosenholtz, R., Nagy, A., & Bell, N. (2004). The effect of background color on asymmetries in color search. *Journal of Vision, 4*, 224–240.
Schaefer, D., Gizdavu, A., & Nicholls, D. (2004). The display of uncertainty information on the controller working position. In *Proceedings of the 23rd Digital Avionics Systems Conference* (pp. 51–61). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
Schmidt, G. S., Chen, S. L., Bryden, A. N., Livingston, M. A., Rosenblum, L. J., & Osborn, B. R. (2004). Multidimensional visual representations for underwater environmental uncertainty. *IEEE Computer Graphics and Applications, 24*(5), 56–65.
Slocum, T. A., Cliburn, D. C., Feddema, J. J., & Miller, J. R. (2003). Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science, 30*, 299–317.
Urness, T., Interrante, V., Marusic, I., Longmire, E., & Ganapathisubramani, B. (2003). Effectively visualizing multi-valued flow data using color and texture. In *Proceedings of the 14th IEEE Visualization Conference (VIS'03;* pp. 115–121). Piscataway, NJ: IEEE Computer Society.
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.
Ware, C. (2004). *Information visualization: Perception for design*. Amsterdam: Elsevier.
Wittenbrink, C. M., Pang, A. T., & Lodha, S. K. (1996). Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics, 2*, 266–279.
Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

Ann M. Bisantz is an associate professor of industrial and systems engineering at the University at Buffalo, State University of New York. She conducts research in areas of uncertainty representation, decision making, interface design, and cognitive engineering methods in a variety of domains, including military command-and-control, health care, and emergency management. She received her PhD from the Georgia Institute of Technology.

Richard T. Stone, PhD, AEP, is an assistant professor at Iowa State University in the Department of Industrial and Manufacturing Systems Engineering and the Department of Mechanical Engineering. Stone's current research focuses on enhancing human performance with consideration to cognitive, physiological, environmental, and technological factors. Current domain areas include augmented reality and telerobotic systems.

Jonathan Pfautz is a principal scientist and vice president of cognitive systems at Charles River Analytics, Inc., in Cambridge, Massachusetts. His major research interests include cognition and perception-based visualization and user interface design, cognitive work analysis, social network analysis, human behavior modeling, and the applications of cognitive engineering to software system development across a variety of military and nonmilitary domains. He received his PhD from Cambridge University.

Adam Fouse is a PhD student in the Department of Cognitive Science at the University of California, San Diego. Previously, he was a scientist at Charles River Analytics, where he directed research on multimodal interfaces. He received a BA in cognitive science and computer science from Brown University in 2004.

Michael Farry is a senior scientist at Charles River Analytics, Inc. There he manages and leads project teams that blend well-grounded cognitive engineering techniques, innovative design ideas, state-of-the-art software engineering practices, and thorough evaluation methods to deliver technology that is effectively and soundly tailored to user requirements. Farry received MEng and SB degrees from the Massachusetts Institute of Technology, where his research focused on the construction of social networks from sensor network data.

Emilie Roth is the principal scientist of Roth Cognitive Engineering. She has been involved in cognitive systems analysis and design in a variety of domains, including nuclear power plant operations, railroad operations, military command and control, medical operating rooms, and intelligence analysis. She serves on the editorial board of the journal *Human Factors* and is editor of the Design of Complex and Joint Cognitive Systems track of the *Journal of Cognitive Engineering and Decision Making*.

Allen L. Nagy is a professor of psychology at Wright State University, Dayton, Ohio, where he has been a faculty member of the Psychology Department and the Human Factors/Industrial Organizational Graduate Program since 1986. He has served as the human factors area leader, graduate program director, and chair of the department. His areas of research include human vision, color perception and color discrimination, abnormal color vision, color naming and color codes, and the use of stimulus features, such as color, to guide selective visual attention. His research has been concerned with the visual mechanisms underlying performance of visual tasks as well as applications of models of visual performance to the design of visual displays.

Gina Thomas is a research psychologist for the Air Force Research Laboratory's Human Effectiveness Division, where she focuses on perceptual and cognitive issues for command and control displays. She obtained her PhD in human factors psychology from Wright State University.