Representing Uncertainty in Graph Edges: An Evaluation of Paired Visual Variables

Hua Guo, Jeff Huang, and David H. Laidlaw, *Fellow, IEEE*

Abstract—When visualizing data with uncertainty, a common approach is to treat uncertainty as an additional dimension and encode it using a visual variable. The effectiveness of this approach depends on how the visual variables chosen for representing uncertainty and other attributes interact to influence the user's perception of each variable. We report a user study on the perception of graph edge attributes when uncertainty associated with each edge and the main edge attribute are visualized simultaneously using two separate visual variables. The study covers four visual variables that are commonly used for visualizing uncertainty on line graphical primitives: lightness, grain, fuzziness, and transparency. We select width, hue, and saturation for visualizing the main edge attribute and hypothesize that we can observe interference between the visual variable chosen to encode the main edge attribute and that to encode uncertainty, as suggested by the concept of dimensional integrality. Grouping the seven visual variables as color-based, focus-based, or geometry-based, we further hypothesize that the degree of interference is affected by the groups to which the two visual variables belong. We consider two further factors in the study: discriminability level for each visual variable as a factor intrinsic to the visual variables and graph-task type (visual search versus comparison) as a factor extrinsic to the visual variables. Our results show that the effectiveness of a visual variable in depicting uncertainty is strongly mediated by all the factors examined here. Focus-based visual variables (fuzziness, grain, and transparency) are robust to the choice of visual variables for encoding the main edge attribute, though fuzziness has stronger negative impact on the perception of width and transparency has stronger negative impact on the perception of hue than the other uncertainty visual variables. We found that interference between hue and lightness is much greater than that between saturation and lightness, though all three are color-based visual variables. We also found a compound relationship between discriminability level and the degree of dimensional integrality. We discuss the generalizability and limitation of the results and conclude with design considerations for visualizing graph uncertainty derived from these results, including recommended choices of visual variables when the relative importance of data attributes and graph tasks is known.

Index Terms—Visual variable, perception, uncertainty visualization, graph visualization

1 Introduction

Information often carries all kinds of uncertainty, and it is usually desirable or even essential that a visualization presents this uncertainty explicitly to the users to help them make more informed decisions. Previous work has shown that visualizing uncertainty can lead to higher-quality decision making [1], [2]. While research has focused on developing techniques for visualizing uncertainty [2], [3], [4] and evaluating techniques [5] or basic visual variables [6], [7] for their effectiveness in conveying uncertainty, only a few have focused on handling uncertainty in graph visualizations (e.g. [7]), which, unlike many other visualization types, rely heavily on line-based marks instead of point- or area-based ones. More fundamentally, to the best of our knowledge, no previous work has conducted a systematic empirical investigation on the influence of the inherent interference between pairs of visual variables on the effectiveness of the common visual variables used for visualizing uncertainty. We see this as an important aspect of uncertainty visualization since uncertainty is often depicted alongside other data attributes that are of interest to the users.

• *The authors are with the Department of Computer Science, Brown University, Providence, RI 02912. E-mail: {huag, jeff, dhl}@cs.brown.edu.*

Manuscript received 5 Oct. 2014; revised 12 Mar. 2015; accepted 8 Apr. 2015. Date of publication 19 Apr. 2015; date of current version 4 Sept. 2015. Recommended for acceptance by J. Heer.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2015.2424872

We attempt here to answer the following questions: 1) is the effectiveness of a visual variable in encoding uncertainty in a graph strongly influenced by the presence of other visual variables, 2) is the influence of the additional visual variables strong enough to alter the effectiveness ranking for a set of visual variables, and 3) how do other factors in the visualization affect the degree of interference between a pair of visual variables? To answer these questions, we conducted a controlled experiment in which participants were asked to complete a series of graph tasks. The edges of the graphs in the experiment each has one generic ordinal attribute and one attribute representing uncertainty of the other attribute. The tasks involve making decisions about the values of the two edge attributes. For the first two questions, we chose brightness, fuzziness, grain, and transparency to encode the uncertainty attribute, and width, hue, saturation to encode the other edge attribute. The choices of the visual variables are based on previous theoretical and empirical work on visual variables and uncertainty visualization, as detailed in Sections 2 and 3. For the third question, we chose to investigate the effect of two visualization-related factors: 1) discriminability of each visual variable, i.e. the perceptual distance between two consecutive levels of a visual variable, and 2) the graph task performed by a user.

Our contribution is threefold. First, our work provides empirical evidence that with line-based marks, interference between pairs of visual variables can alter the effectiveness ranking for visualizing uncertainty among a set of candidate visual variables. Second, we show empirically that both discriminability and graph-task type can influence the degree of interference between a pair of visual variables. Finally, we derive design recommendations for choosing visual variables to visualize graph with uncertainty.

2 Related Work

This work builds upon research on visual variables, uncertainty visualization, and perceptual studies on graph visualizations. We detail how our work relates to and extends previous work in these three areas in the following subsections.

2.1 Visual Variables

Our work is motivated by the concepts of "disassociativity" and "dimensional integrality" of visual variables in uncertainty visualization. We thus start with an overview of research on visual variables and these two related concepts.

The system of visual variables was first developed by Bertin [8]. He identified seven visual variables: position, size, shape, color brightness, color hue, orientation, and grain. Morrison [9] suggested the addition of color saturation and arrangement, and MacEachren [10], [11] proposed three more: fuzziness, resolution, and transparency. The list can be further expanded by including some of the visual primitives examined by Cleveland and McGill in their seminal work on graphical perception [12], including angle, volume, and curvature. Visual variables going beyond traditional static, 2D displays such as motion, depth, and occlusion are also touched upon in more recent work [13].

While Bertin provided no empirical assessment of the visual variables, the effectiveness (usually measured using accuracy) and efficiency (usually measured using response time) of different visual variables have been examined in much other work. Cleveland and McGill [12] ranked selected visual variables on the degree of accuracy users can achieve when working with quantitative information encoded by these visual variables, and Mackinlay [14] extended the ranking to ordinal and nominal data types using previous psychophysical results. Garlandini and Fabrikant [15] empirically assessed four visual variables (size, color brightness, color hue, and orientation) for their effectiveness and efficiency in the design of 2D maps. Bezerianos and Isenberg [16] evaluated how a user's perception of angle, area, and length is affected by viewing distances and angles when working with tiled wall-sized displays. In the same vein but from a different perspective, John [17] assessed the effectiveness and efficiency of color hue at different perceptual levels proposed by Bertin (i.e. associativity, selectivity, order, and quantity) on the basis of experimental evidence from previous research in psychology and design research in thematic cartography. Similarly, Filippakopoulou et al. [18] provided empirical data verifying the perceptual levels of certain visual variables proposed by Bertin.

Despite these efforts to assess the effectiveness and efficiency of basic visual variables, little attention has been paid to Bertin's notion of "disassociativity" [19]. According to Bertin, a visual variable is "associative" if it allows the viewer to differentiate a set of symbols for other visual variables while ignoring variations in this variable. Bertin argued that brightness and size are disassociative: since these variables affect symbol visibility, it would be impossible or very difficult to ignore variations in them.

Bertin's notion of disassociativity is related to the concept of "dimensional integrality" as discussed in Garner's book on perception and information processing [20]. This concept is supported by much empirical work. For example, Callaghan's work on texture-segregation [21] showed that shape-based texture-segregation is impaired by task-irrelevant variance in hue, but not the other way around. In [22], two-way interference between hue and brightness was observed in the texture-segregation task. Dimensional integrality was later discussed by Colin Ware [23] in a way tailored more towards visual stimuli. While Bertin implied that a visual variable is either "associative" or "disassociative," the concept of dimensional integrality describes a continuum from "integral" to "separable" and applies to pairs of visual variables instead of individual ones. The concept of dimensional integrality is the basis for the hypotheses tested here, though our hypotheses also aim to be more specific about factors that may affect the degree of dimensional integrality between a pair of visual variables in the context of visualizing uncertainty in graphs.

2.2 Evaluation of Uncertainty Visualization

There has been active research on uncertainty visualization techniques and design studies (e.g. [3], [24], [25], [26]) in the areas of geographic information systems, scientific visualization, and information visualization, but they are seldom accompanied by in-depth evaluations of the techniques.

Among the works that evaluate uncertainty visualization techniques, some address the challenge of presenting data and its uncertainty simultaneously with minimal mutual interference, but through interactions instead of choices of visual variables. Evans [27], studying combining data and reliability information using static composition and animation, found that these two methods show no statistically significant difference in terms of either viewing time or accuracy. In static composition, color hue was used to represent land-use class and color saturation was used to present reliability. Interestingly, Aerts et al. [6], also comparing static and toggling methods, found that subjects are more accurate estimating uncertainty values using the static method than the toggling method, and that more subjects prefer the static methods than the toggling method.

Other work has focused on the use of visual variables to convey uncertainty. In [10], MacEachren discussed nine candidate graphical variables for representing uncertainty, including Bertin's seven variables (location, size, brightness, texture, color, orientation, and shape) and two further variables, color saturation and focus. The discussion, however, is theoretical rather than empirical. Later in [28], MacEachren et al. evaluated common visual variables for their intuitiveness of representing uncertainty, and assessed the accuracy of fuzziness and brightness. In [1], Leitner and Buttenfield tested three visual variables for encoding uncertainty, brightness, texture, and saturation, and concluded that brightness leads to the highest number of correct answers, followed by texture, with saturation coming last. Recently,

TABLE 1
The Factors Included in the Experiment and their Levels

Factor	# of levels	Levels
vCertainty	4	lightness, fuzziness, grain, transparency
vStrength	3	width, hue, saturation
Discriminability	2	low, high
Task type	2	visual search, comparison

Boukhelifa et al. [7] proposed sketchiness as a new visual variable for depicting uncertainty information using line-based marks and evaluated its effectiveness compared to blur, dash, and grayscale. Sanyal et al. [5] evaluated techniques for visualizing 1D and 2D data that represent uncertainty using either color-mapping or glyph sizes together with the traditional error bars.

The work by MacEachren et al. [29] is probably the most relevant to the present study. Comparing two methods for depicting data and data uncertainty, they discovered that using hue for data and texture overlay for data uncertainty led to more accurate task performance than using hue for data and saturation for data uncertainty, possibly because the former is more visually separable than the latter.

2.3 Perceptual Studies on Graph Visualization

Several user studies have evaluated how the basic visual properties of edges in node-link visualizations can affect the user's performance on graph tasks. As far as we know, however, user studies to date have focused mostly on edge geometry rather than on other visual attributes as considered here.

In [30], Holten and van Wijk evaluated six alternative representations of directed edges with varying shape and color on performance in path-finding tasks, and found that a tapered representation in which edge width varies gradually along its length led to the best performance. Most relevant to the current study, Holten and van Wijk compared multi-cue representations where shape and color were used simultaneously to encode the edge direction with single-cue representations, but found no statistically significant difference between the two in terms of performance. Xu et al. [31] studied the effect of edge curvature on graph readability with varying graph size; also on the effect of curvature, Telea et al. [32] conducted a qualitative study comparing hierarchical edge bundling with node-link diagrams.

3 Experimental Design

To simplify the experiment, we consider only the case in which each edge has one ordinal attribute and one attribute representing uncertainty. In the experiment instructions, we call the ordinal attribute the "strength" of the edge, and its uncertainty "certainty"; we use these terms here for the two attributes henceforth.

This experiment has four factors: the visual variable representing certainty (hereafter *vCertainty*), the visual variable representing strength (hereafter *vStrength*), discriminability, and task type. Table 1 lists the four factors and the levels for each. We chose the four visual variables for *vCertainty* since the experiment in [28] suggests they are the most intuitive

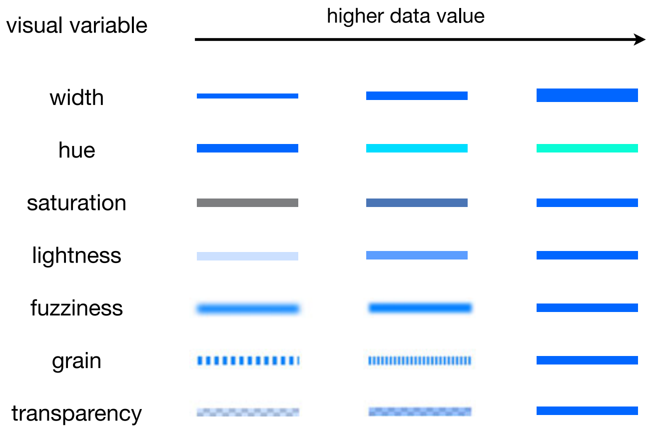


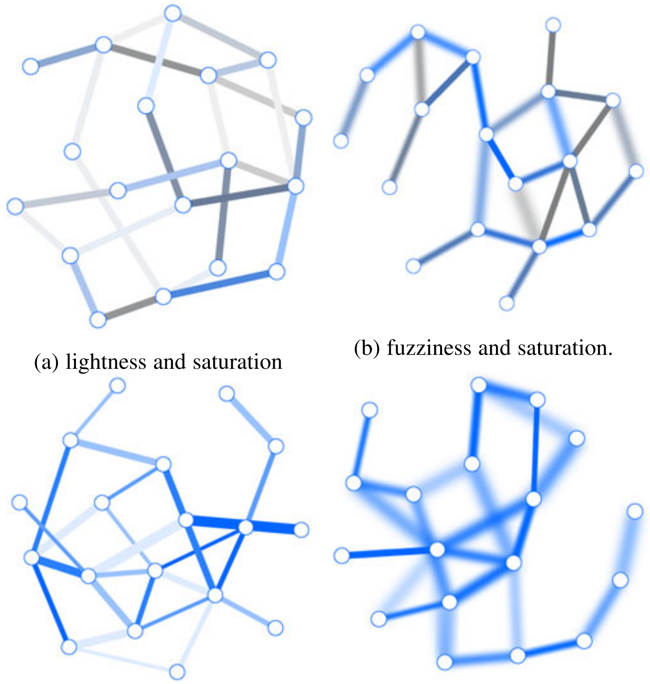
Fig. 1. An illustration of how each of the seven visual variables progress with increasing data value.

representations of uncertainty that can be applied to line-based marks. The three visual variables for *vStrength* were chosen according to the theoretical ranking of visual variables for depicting ordinal data in [14]. Fig. 1 illustrates how the appearance of an edge varies with each of the seven visual variables. In particular, we use the HSL color space in the experiment and define hue, saturation, and lightness accordingly. We assign five encoding levels to each visual variable, and use two discriminability levels and two task types (visual search and comparison). We allocate a fixed amount of time for each type of task. Section 3.3 describes how the discriminability levels and the time limitations for the tasks were determined. We use participants' accuracy as the measure of effectiveness.

Each trial in the experiment concerns either the certainty attribute or the strength attribute. Henceforth we call the attribute the participant needs to focus on in a trial the *primary attribute* and the other attribute the *secondary attribute*. In each trial, the participant was asked to make judgments about the primary attribute while trying not to be distracted by variations in the secondary attribute.

Below we list the hypotheses tested in the experiment. *H1-H4* are concerned with interactions among visual variables while *H5-H8* relate discriminability and graph task type to user perception.

- *H1*. When participants are working on tasks concerning the certainty attribute, the effectiveness of a visual variable in encoding certainty is mediated by the visual variable used for encoding strength, i.e. there will be an interaction effect between *vCertainty* and *vStrength* when certainty is the primary attribute.
- *H2*. Taking *H1* further, the effectiveness of fuzziness, grain, and transparency will not change significantly with different *vStrength*s. Lightness will be more accurate when paired with width than with hue or saturation.
- *H3*. In tasks concerning the strength attribute, the visual variable encoding certainty will mediate the effectiveness of the visual variable encoding strength, i.e. there will be an interaction effect between *vCertainty* and *vStrength* when strength is the primary attribute.



(a) lightness and saturation

(b) fuzziness and saturation.

(c) lightness and width.

(d) fuzziness and width.

Fig. 2. Node-link visualizations using four of the 12 pairs of visual variables examined in this experiment. The above graphs have two attributes that are each encoded as different visual variables (specified in the subcaption).

- *H4*. Taking *H3* further, the accuracy of width will not vary significantly with different *vCertainty*s. Hue and saturation will have much lower accuracy when certainty is encoded using lightness compared to other alternatives.

  H2 and *H4* are based on grouping of the seven visual variables. We group lightness, hue, and saturation as color-based, fuzziness, grain, and transparency as focus-based (that is, affecting the overall clarity and visibility of a mark), and width as geometry-based. This grouping is similar to that in [7]. The rationale behind *H2* and *H4*, then, is that focus-based visual variables will have similar degrees of interference with either color-based or geometry-based visual variables, and color-based visual variables will have much greater interference with color-based visual variables than with geometry-based ones.
- *H5*. Accuracy will be lower under the low-discriminability condition than the high-discriminability condition.
- *H6*. Accuracy will be the same on the visual search tasks as on the comparison tasks.
- *H7*. The level of distraction a secondary visual variable has on the same primary visual variable will not be affected by discriminability, i.e. there will be no significant interaction effects between difficulty and *vStrength* in edge certainty tasks or between difficulty and *vCertainty* in edge strength tasks.
- *H8*. The relative accuracy for a visual variable can generalize across task types; i.e. there are no significant interaction effects between task type and *vStrength* or between task type and *vCertainty*.

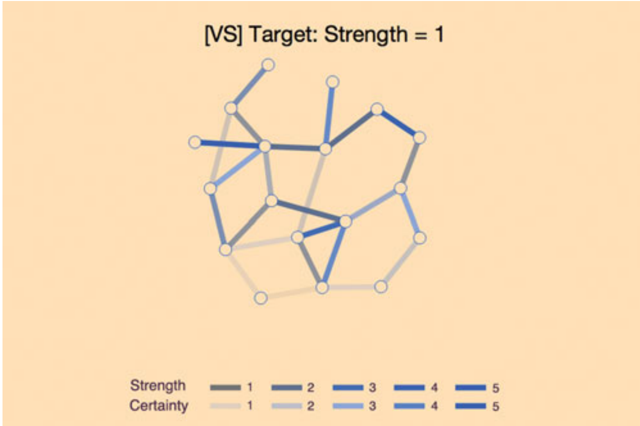


Fig. 3. An example screen for a visual search trial.

The following subsections describe the stimulus design, tasks, participants and procedure in this experiment.

3.1 Design of Stimuli

All graph stimuli used in this study are visualizations of randomly generated graph datasets with 18 nodes and 25 edges. We use random graphs instead of real-world datasets to remove the confounds of dataset size and the distribution of data attribute values. We generated all graphs using NetworkX [33], a Python library for network creation and manipulation. The graph generator implements the $G(n, M)$ variant of the Erdős-Rényi model and produces a graph by picking randomly out of the set of all graphs with 18 nodes and 25 edges.

Each graph edge has a "strength" attribute and a "certainty" attribute; both ranging in value between 1 and 5. In half of the graphs, "strength" is the primary attribute and "certainty" is the secondary attribute; the reverse is true for the other half. We generated graphs separately from the visual search task and the comparison task with different distributions of edge attribute values.

For the visual search tasks, values for the secondary attribute were always drawn from a discrete uniform distribution on {1,2,3,4,5}. The values for the primary attribute differ between "positive" graph stimuli and "negative" graph stimuli. A positive stimulus contains exactly one edge with the target value, i.e., the value that the participant needs to

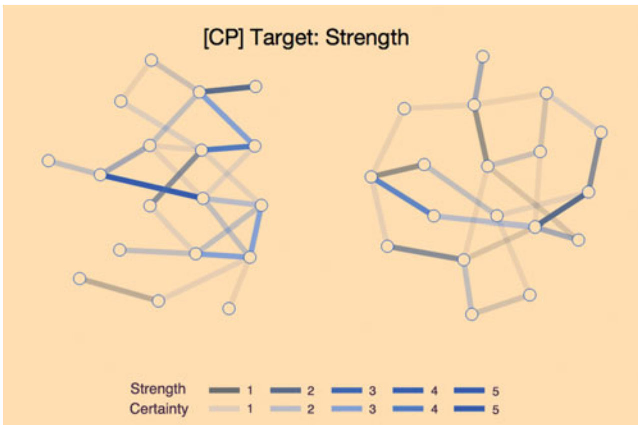


Fig. 4. An example screen for a comparison trial.

TABLE 2
Default Visual Attributes for the Edges

width	hue	saturation	lightness	fuzziness	grain	transparency
0.24 cm	216	1	0.5	none	solid	1 (opaque)

search for, for the primary attribute. The remaining edges were randomly assigned an integer value between 1 and 5 excluding the target value for the primary attribute. A negative stimulus does not contain any edge with the target value for the primary attribute, and each edge was randomly assigned an integer value between 1 and 5 excluding the target value.

In generating graphs for the comparison tasks, values for the secondary attribute were also drawn from a discrete uniform distribution on {1,2,3,4,5}. The sums of the values for the primary attribute must differ unambiguously between each graph pair, so we defined three distributions representing low, medium, and high overall strength or certainty, as detailed in Table 3. Using these three distributions, we defined three configurations for each graph pair: low versus medium, medium versus high, and low versus high. We then assigned one configuration to each graph pair and generated values for each graph accordingly.

The graphical stimuli were generated using JavaScript and D3.js [34] with an adapted version of the built-in force-directed algorithm in D3. Fig 2 shows example stimuli using four of the 12 pairs of visual variables. To reduce the possible interference of line length with other visual variables, we constrain the length of each line to between 5 and 15 pixels. While the possible interference of line length can be completely removed by using lines of uniform lengths, that would impose unrealistic constraints on the graph layout algorithm. Table 2 lists the default visual attributes (i.e., the default value for each visual variable when it is neither the primary nor the secondary visual variable in a condition) for each edge in the visualizations. Colors are defined in the HSL color space.

3.2 Tasks

Participants were given two types of tasks, visual search tasks (**VS**) and comparison tasks (**CP**). Both tasks require binary responses. In visual search tasks, participants look for an edge of a specific certainty or strength value. They need to determine whether or not the target edge is present in a stimulus, and they press one of two keys to indicate their response. In comparison tasks, participants see two graphs and must decide which one has higher overall strength or certainty, again indicating their responses by pressing one of two keys. The overall strength or certainty is defined as the average of all the strength or certainty attributes in a graph, and participants were explicitly told that they should try to estimate the average by looking at the overall appearance of the graph rather than attempting to combine estimates for individual edges. (The amount of time allocated to the comparison task also makes it impossible to complete by examining individual edges.). Fig. 3 and Fig. 4 show example screens for VS and CP tasks respectively.

We choose these two tasks for two reasons. First, they are common visualization tasks in the literature. For example, the visual search task combines "Find Extremum" and "Retrieve Value" in the visualization task taxonomy proposed by Amar et al. [35], and the comparison task corresponds to the "Characterize Distribution" proposed in the same taxonomy. At the same time, the distinction between the two tasks is also analogous to the "identification-comparison" dimension proposed in [36] for characterizing exploratory visualization tasks. Second, these two tasks potentially require two distinct types of cognitive and perceptual operations: visual search involves pattern matching for each individual edge, while comparison requires visually aggregating a set of edges.

TABLE 3
Three Distributions of the Primary Attribute as Used in the Graphs for the Comparison Task

	Distribution of the primary attribute				
Attribute value	1	2	3	4	5
Low	7	8	5	3	2
Medium	3	6	7	6	3
High	2	3	5	8	7

3.3 Determining Free Parameters

Two sets of parameters were chosen through pilot studies: 1) encoding levels for each visual variable under the two discriminability conditions, and 2) time allocated for each task type. In addition, we also chose the target values for the visual search task through the pilot.

We used two criteria in choosing the encoding levels: 1) the perceptual distance between consecutive levels should be as nearly uniform as possible, and 2) encoding choices for different visual variables should be equally difficult for the same discriminability level.

To pick two encoding levels for each visual variable satisfying these two criteria, we devised an experiment consisting of simple identification tasks. In an identification task, the participant is shown a page with a "target" line and five "candidate" lines, each corresponding to one encoding level for the visual variable. The stimulus page disappears after 2 seconds, and the participant is then instructed to select the candidate line matching the target line by pressing the corresponding key.

Five participants took part in this pilot study. Each participant was first shown a set of encodings with maximal distance between every two consecutive levels, and went on to encoding sets with gradually decreasing distance between consecutive levels. Participants would perform 10 repetitions for each encoding set; they worked through all encoding sets for one visual variable first, and was instructed to move on to the next visual variable when the accuracy dropped below 40 percent or when no more encoding sets were available. Finally, we calculated the average accuracy for each set of encoding levels across all five participants. We chose the set of encoding levels with just above 98 percent accuracy as the "high discriminability" level and the set of encoding levels with just above 80 percent accuracy as the "low discriminability" level for each visual variable. For the encoding sets used here, values for individual levels were determined using either

Fechner's law [37] (transparency), Stevens' power law [38] (lightness, saturation, width, fuzziness and grain), or human judgment (hue). We restricted the range of hue to within 170 to 216 (roughly between green and blue) in the HSL color space so that the hue values would always have a natural ordering. We varied saturation between 0.01 and 1, value between 0.5 and 1, and transparency between 0.1 and 1 to ensure the marks would always be clearly visible.

We used stimuli and tasks similar to the main study to determine how much time to allocate for each task in the main study. Participants in the pilot study were asked to perform visual search tasks and comparison tasks on graphs with 25 edges. However, they performed the tasks with the secondary visual variable being absent. They were also asked to complete the task as accurately and as quickly as possible but were not given a limited amount of time per task. We collected data from three participants with 10 repetitions for each condition, discarded trials with incorrect responses, and took the rounded average of response time in the remaining trials for each type of task as the time limit in the main study.

During the pilot, we discovered that participants found it significantly more difficult to locate middle values for the visual search task than boundary values. We started with three variations of the visual search task: users needed to look for an edge with value 1, 3, or 5 for the primary attribute. It turned out, however, that participants took on average two to three times longer to report presence/absence correctly when the target value is 3 than in the other two cases. We therefore decided to exclude value 3 in the main study since we are primarily interested in comparing performance across different visual variables, not the relative ease of extracting different values using the same visual variable.

3.4 Participants

We recruited 20 participants (11 female) for this study, all undergraduate or graduate students at Brown University. Participants were recruited using both flyers and internal university mailing lists. Participant ages ranged from 19 to 27 years old ($M = 22.63$, $SD = 3.69$). Every participant had normal or corrected-to-normal vision and normal color vision. None of the participants had extensive experience with graph visualizations.

3.5 Experimental Procedure

We conducted the experiments in a laboratory setting. In each session, the participant first filled out an informed-consent form and read the learning task instructions. The participant was then given a set of practice trials, during which he or she received feedback on the correctness of each trial and could ask any question about the tasks. Having finished the practice trials, the participant continued to finish the main trials independently and filled out a demographic questionnaire at the end of the session. Each session lasted 50-70 minutes. All sessions were completed using the same computer with a 15.4-inch display. The display was configured to have resolution $1{,}440 \times 900$ and brightness $120\ \mathrm{cd/m^2}$.

The goal of the practice trials was to familiarize the participants with the rules and procedures for the tasks, not with each individual pair of visual variables. Therefore, all practice trials in a given session used the same pair of visual variables, which was the same as the pair of visual variables that appeared last during the main trials in that session. Each session had 24 practice trials.

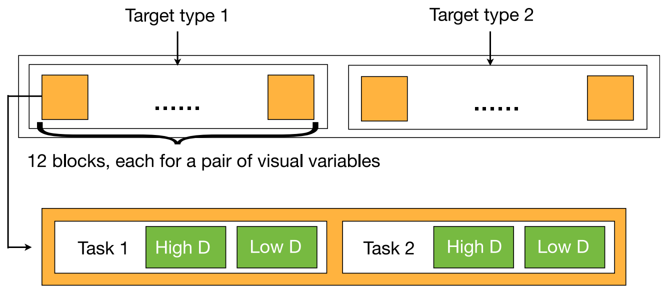


Fig. 5. Schematic representation of trial ordering. At the highest level, the trials are grouped by "target type", which can be either strength or certainty. Within each target type, trials are further grouped into 12 blocks (orange rectangles) where stimuli within the same block use the same pair of visual variables. Within each of the 12 blocks, trials are grouped by task type and ordered by discriminability level.

Fig. 5 illustrates the arrangements of the main trials. These trials were first grouped based on the target type. Half of the participants were asked to make decisions based only on the strength of an edge in the first half of all trials, and based only on the certainty of the edge in the remaining trials. The other half of the participants completed trials asking about the certainty of the edges first, followed by trials asking about edge strength. Within each group of trials with the same target type, the trials were further grouped into 12 major blocks. Stimuli in each block used the same set of visual variables. The order of the 12 blocks was counterbalanced using a Latin square design across all the participants. Trials within each of the 12 blocks were further grouped by task types into two minor blocks. The order of task types within each of the 12 major blocks was also counterbalanced using a Latin square design both within and across the participants. Trials within each minor block were further divided into two difficulty levels: easy trials always appeared before difficult trials. There were four repeated trials for each combination of condition, task type, and difficulty level. For visual search tasks, half of the four repeated trials in each combination were positive stimuli. For comparison tasks, the graph on the left had the higher overall strength or certainty in half of the four repeated trials. The ordering of each group of four repeated trials was completely randomized. In total, we collected $288 \times 20 = 5{,}760$ trials across all sessions.

At the beginning of each major trial block, the participant was shown a page with legends for each of the two visual variables used in this trial block. The legends were accompanied by short explanations to help the participant understand the encoding levels for each visual variable. For example, the explanation for the visual variable "lightness" was "edges with higher certainty levels are darker".

In each trial, the participant was first shown the target for the current trial as well as the legends for the visual variables for the trial. For **VS**, the target consisted of a target type ("certainty" or "strength") and a target value (either 1 or 5). For **CP**, the target specification included only the target type. The participant could spend as much time as needed

TABLE 4
Primary Analysis Results

Target type	Effects type	Variable	Condition	F	df	p-value
certainty	simple	task type	all	118.965	1,19	0.001
certainty	simple	vStrength	all	3.722	2,38	0.033
certainty	interaction	vStrength x vCertainty	all	3.327	6,114	0.005
strength	simple	task type	all	116.942	1,19	0.001
strength	simple	vCertainty	all	2.908	3,57	0.042
strength	simple	discriminability	all	22.404	1,19	0.001
strength	simple	vCertainty	strength = hue	3.911	3,57	0.013
strength	simple	discriminability	strength = saturation	8.307	1,19	0.01
strength	interaction	task type x vStrength	all	8.216	2,38	0.001
strength	interaction	discriminability x vStrength x vCertainty	all	2.570	6,114	0.023
strength	interaction	vStrength x vCertainty	discriminability = high	3.779	6,114	0.002
strength	interaction	discriminability x vCertainty	strength = width	9.2	3,57	0.001

on the target page and could proceed to the stimuli page when ready. The stimuli page was presented for a fixed amount of time: 5 seconds for **VS** and 3 seconds for **CP**. This amount of time was determined through the pilot study described earlier. When the stimuli page disappeared, the participant had to make a response using the keyboard before proceeding to the next trial.

4 Results

We analyzed the data from the experiment using RM-ANOVA in SPSS. We computed a single number for each participant's accuracy by aggregating the responses from the four repetitions under the same condition; accuracy ranges from 0 to 1, with 1 being correct in all four repetitions and 0 incorrect in all four repetitions. The data was split along target type (strength attribute versus certainty attribute) before the analysis and RM-ANOVA was performed for each target type. Mauchly's test showed that the sphericity assumption was not violated for any of the sources discussed below. To assess if the accuracy is influenced by response bias, we repeated the above analysis with d-prime. While the exact values of p-values differ between accuracy-based and d-prime-based analyses, the significance of the p-values is consistent. We therefore report all results in terms of accuracy as it is a more conventional measure of the effectiveness of visualizations.

After the initial ANOVA analysis, we found a significant three-way interaction among discriminability level, *vStrength*, and *vCertainty* when the target type is the strength attribute. To further investigate this complex interaction, we split the data along the discriminability and *vStrength* factors and performed additional two-way ANOVAs on the subsets obtained.

We also performed pairwise comparisons to explore all the significant interaction effects observed. We applied the Bonferroni correction to adjust the alpha for all pairwise comparisons, and we report the corrected p-value calculated by SPSS for these comparisons. SPSS calculates the corrected p-value by applying the Bonferroni correction backwards: the corrected p-value equals the actual p-value multiplied by the total number of possible pairwise comparisons. Therefore, all p-values reported in this section can be directly compared to the experiment-wise alpha (0.05) to determine significance.

The following subsections report important findings supported by statistically significant results. The complete lists of statistically significant results for the primary analyses and the pairwise comparisons are provided in Tables 4 and 5, respectively.

4.1 *vStrength* Mediates the Relative Effectiveness of Certainty Visual Variables

In tasks concerning the certainty attribute, we found a significant interaction between *vStrength* and *vCertainty* ($p = 0.005$, $F_{6,114} = 3.327$), suggesting that the relative effectiveness of the four certainty visual variables is conditional upon the choice of *vStrength*. To further investigate how the accuracy in the *vCertainty* is mediated by the choice of *vStrength*, we performed pairwise comparisons among the four certainty variables while holding *vStrength* constant and vice versa. The results showed that when *vStrength* was hue, participants were 12.1 percent more accurate in making judgments about certainty using grain than lightness ($p < 0.001$; see Fig. 6, cross marks). When holding *vCertainty* constant, we found that accuracy in interpreting lightness was the lowest when strength was encoded using hue; participants were 14.6 percent more accurate ($p = 0.001$; Fig. 6, star marks) when *vStrength* was width and 11.7 percent more accurate when *vStrength* was saturation ($p = 0.012$; Fig. 6, plus marks).

These results suggest that *H1* is valid: the effectiveness of *vCertainty* is mediated by the choice of *vStrength* in tasks regarding edge certainty. However, *H2*, a stronger version of *H1*, is only partly valid. Indeed, for fuzziness, grain, and transparency, we found no significant variation in accuracy with different strength encoding choices, as hypothesized. However, while we hypothesized that lightness would be more accurate when paired with width than with either hue or saturation, our results show that the participants were similarly accurate with lightness when either width or saturation was present, and in both cases they were much more accurate than when *vStrength* was hue.

4.2 *vCertainty* Mediates the Relative Effectiveness of Strength Visual Variables

When the target type is strength, we found a significant interaction between *vStrength* and *vCertainty* when the discriminability level was high ($p = 0.002$, $F_{6,114} = 3.779$).

TABLE 5
All Significant Pairwise Differences

Target type	Fixed factors	Across	level 1	level 2	p-value	difference	figure
certainty	vStrength = hue	vCertainty	grain	lightness	0.001	12.1 percent	Fig. 6
certainty	vCertainty = lightness	vStrength	width	hue	0.001	14.6 percent	Fig. 6
certainty	vCertainty = lightness	vStrength	saturation	hue	0.012	11.7 percent	Fig. 6
strength	discriminability = high, vCertainty = lightness	vStrength	width	hue	0.004	18.3 percent	Fig. 7
strength	discriminability = high, vCertainty = fuzziness	vStrength	hue	width	0.018	16.7 percent	Fig. 7
strength	discriminability = high, vStrength = width	vCertainty	lightness	fuzziness	0.001	21.7 percent	Fig. 7
strength	discriminability = high, vStrength = hue	vCertainty	fuzziness	lightness	0.035	13.3 percent	Fig. 7
strength	discriminability = high, vStrength = hue	vCertainty	grain	lightness	0.051	15.0 percent	Fig. 7
strength	task = visual search	vStrength	width	saturation	0.05	8.8 percent	Fig. 9
strength	task = comparison	vStrength	hue	width	0.022	6 percent	Fig. 9
strength	vStrength = width, vCertainty = lightness	discriminability	high	low	0.001	19.2 percent	Fig. 10a
strength	vStrength = width, vCertainty = transparency	discriminability	high	low	0.042	10.0 percent	Fig. 10a
strength	vStrength = width, vCertainty = fuzziness	discriminability	low	high	0.004	15.8 percent	Fig. 10a
strength	vStrength = hue, vCertainty = grain	discriminability	high	low	0.025	13.3 percent	Fig. 10b
strength	vStrength = saturation, vCertainty = transparency	discriminability	high	low	0.019	13.3 percent	Fig. 10c

"Fixed factors" indicates which factors are held constant in the pairwise comparisons. The "Across" Column indicates which factor varies in a pairwise comparison. The "figure" column indicates the figure showing the pairwise comparisons.

The accuracy for each combination of strength variable and certainty variable is shown in Fig. 7. We did not, however, find a significant interaction between *vStrength* and *vCertainty* when the discriminability level is low.

To explore the interaction effect between vStrength and vCertainty under high discriminability, we performed pairwise comparisons among the four certainty variables when holding *vStrength* constant and vice versa. The results showed that when certainty was encoded using lightness, participants were 18.3 percent more accurate in judging strength levels when *vStrength* was width rather than hue ($p = 0.004$; Fig. 7, star marks). When certainty was encoded using fuzziness, however, participants were 16.7 percent more accurate estimating

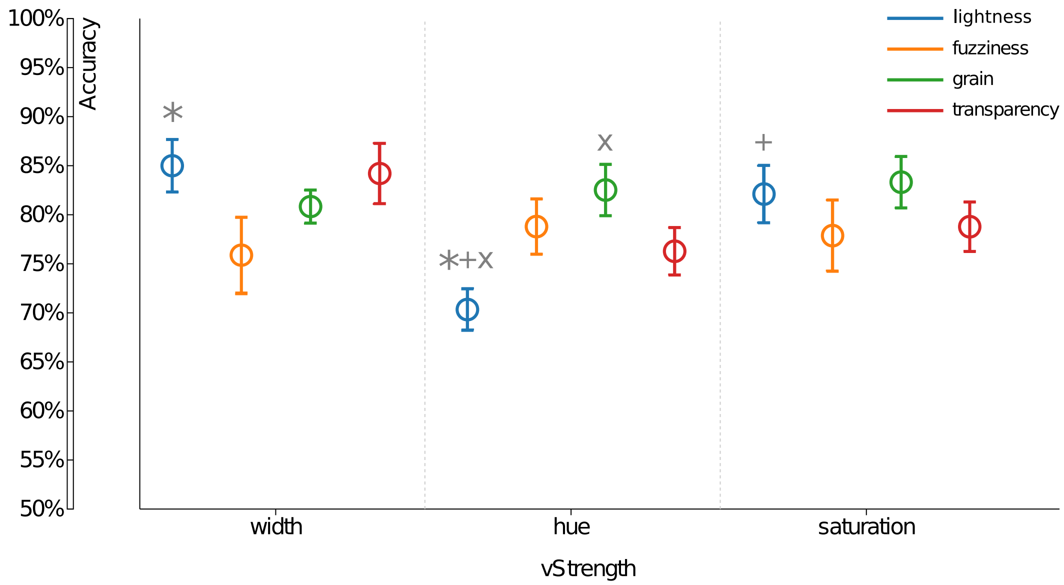


Fig. 6. Accuracy for each pair of strength visual variable and uncertainty visual variable when tasks concern the certainty attribute. Error bars show ± 1 standard error. Paired symbols indicate significant pairwise differences: accuracy on lightness given width versus hue(*); accuracy on lightness given saturation versus hue(+); accuracy on grain versus lightness given hue($\times$).

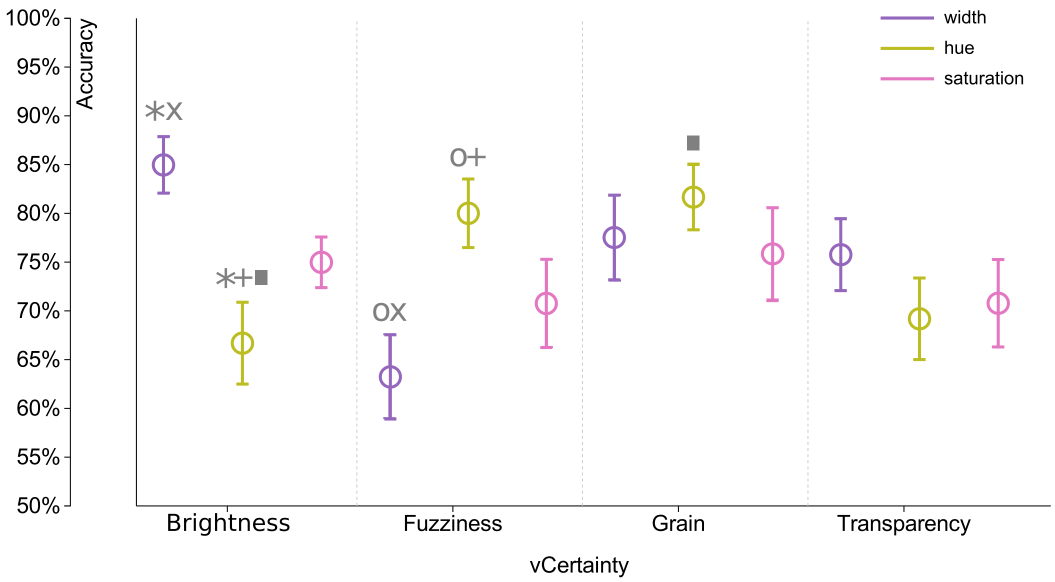


Fig. 7. Accuracy for each pair of strength visual variable and certainty visual variable in tasks concerning the strength attribute and high discriminability. Error bars show ± 1 standard error. Paired symbols indicate significant (including borderline) pairwise differences: accuracy on width given lightness versus fuzziness($\times$); accuracy on hue given fuzziness versus lightness(+); accuracy on hue given grain versus lightness(square); accuracy on width versus hue given lightness(*); accuracy on hue versus width given fuzziness(o).

strength levels when *vStrength* was hue rather than width ($p = 0.018$; Fig. 7, circle marks). When holding *vStrength* constant, participants were 21.7 percent more accurate in interpreting width when *vCertainty* was lightness instead of fuzziness ($p = 0.001$; Fig. 7, cross marks). Participants were 13.3 percent more accurate in interpreting hue when hue was paired with fuzziness rather than lightness ($p = 0.035$; Fig. 7, plus marks). There was also an insignificant trend ($p = 0.051$) that participants were 15.0 percent more accurate in interpreting hue when *vCertainty* was grain instead of lightness (Fig. 7, square marks).

These results suggest that *H3* is valid: the effectiveness of *vStrength* is mediated by the choice of *vCertainty* in tasks concerning edge strength. However, *H4* is also only partially valid. As predicted, the effectiveness of hue for encoding strength was reduced by the variation in lightness. While we predicted that all four *vCertainty* would have similar impacts on the perception of width, however, fuzziness turned out to have a stronger negative impact on the perception of width than the other three certainty visual variables.

4.3 Task Type Matters

We found a main effect of task type when participants were working on tasks about the strength attribute ($p < 0.001$, $F_{1,19} = 116.942$), and also when they were focusing on the certainty attribute ($p < 0.001$, $F_{1,19} = 118.965$). This means that participants' accuracies differ significantly between the visual search and the comparison tasks.

Looking at Figs. 8 and 9, we see that participants were generally more accurate on the comparison tasks than on the visual search tasks. Averaging over all other conditions, the difference in accuracy between the two types of tasks is 20.45 percent. These results reject *H6*, that "accuracy will be the same on the visual search tasks as on the comparison tasks".

More interestingly, we also observed an interaction effect between task type and *vStrength* ($p = 0.001$, $F_{2,38} = 8.216$) in trials concerning edge strength. To investigate this interaction effect further, we performed pairwise comparisons among strength visual variables for each task type. The results showed that in the visual search task, participants were most accurate with width and were significantly more accurate at interpreting width than saturation ($p = 0.05$; see Fig. 9, star marks). However, participants were least accurate with width in the comparison task and were significantly less accurate at interpreting width than hue ($p = 0.022$; see Fig. 9, circle marks). These results reject *H8*, in which we hypothesized that the way visual variables interfere with each other would not differ across task types.

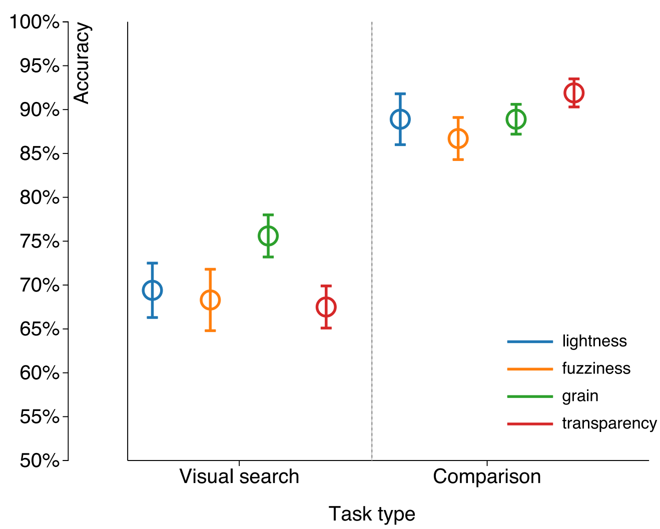


Fig. 8. Accuracy for each certainty visual variable for the two types of tasks when the target type is certainty. Error bars show ± 1 standard error.

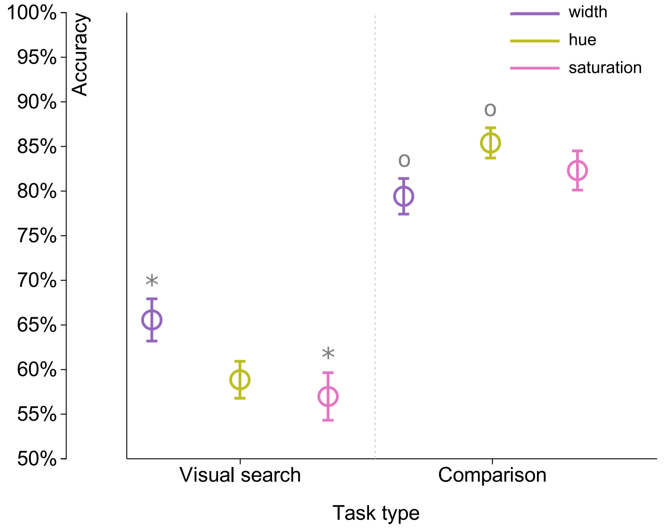


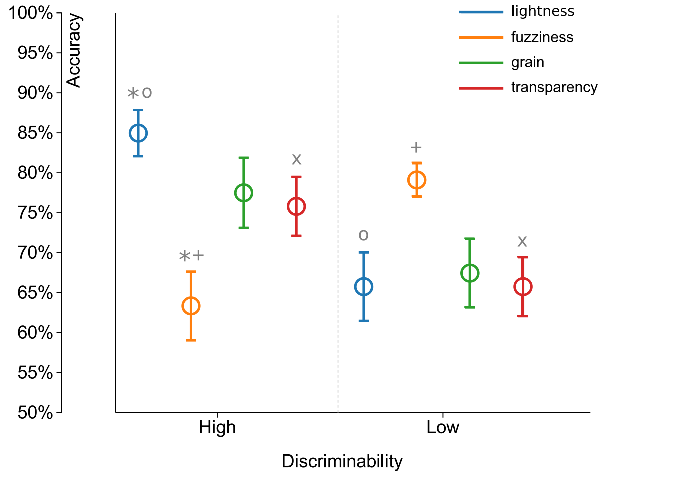
Fig. 9. Accuracy for each strength visual variable for the two types of tasks when target type is strength. Error bars show ± 1 standard error. Paired symbols indicate significant (including borderline) pairwise differences: accuracy between width and saturation in visual search(*); accuracy between hue and width in comparison(o).

4.4 Lower Discriminability does not Always Lead to Lower Accuracy

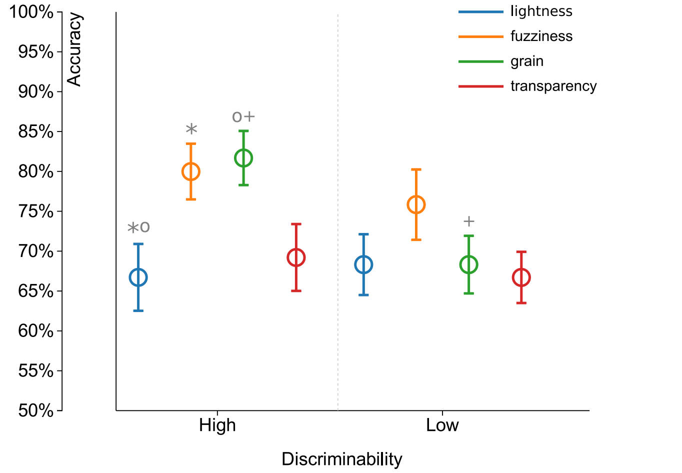
When the target type is certainty, discriminability level did not have a significant effect on accuracy. However, when the target type is strength, we observed a three-way interaction effect among discriminability, *vStrength*, and *vCertainty* ($p = 0.023$, $F_{6,114} = 2.570$). To further investigate the impact of discriminability level on tasks concerning edge strength, we performed a set of two-way ANOVAs within each of the three *vStrength* levels. The results showed a significant interaction effect between discriminability and *vCertainty* ($p < 0.001$, $F_{3,57} = 9.2$) when strength is encoded using width, a main effect of *vCertainty* ($p = 0.013$, $F_{3,57} = 3.911$) when strength is encoded using hue, and a main effect of discriminability when strength is encoded using saturation ($p = 0.01$, $F_{1,19} = 8.307$).

We performed pairwise comparisons to explore the interaction effect between discriminability and vCertainty when strength is encoded using width. Pairwise comparisons showed that when strength was encoded using width, accuracy was 19.2 percent higher with higher discriminability when *vCertainty* was lightness ($p < 0.001$; Fig. 10a, circle marks) and 10.0 percent higher when *vCertainty* was transparency ($p = 0.042$; Fig. 10a, cross marks), but 15.8 percent lower ($p = 0.004$) with higher discriminability when *vCertainty* was fuzziness (see Fig. 10a, plus marks). When strength was encoded using hue, accuracy was 13.3 percent higher with higher discriminability when *vCertainty* was grain ($p = 0.025$; Fig. 10b, plus marks) but had no significant difference across discriminability levels with other *vCertainty*. When strength was encoded using saturation, accuracy was 13.3 percent higher with higher discriminability when *vCertainty* is transparency ($p = 0.019$; Fig. 10c, star marks) but had no significant difference across discriminability levels with other *vCertainty*.

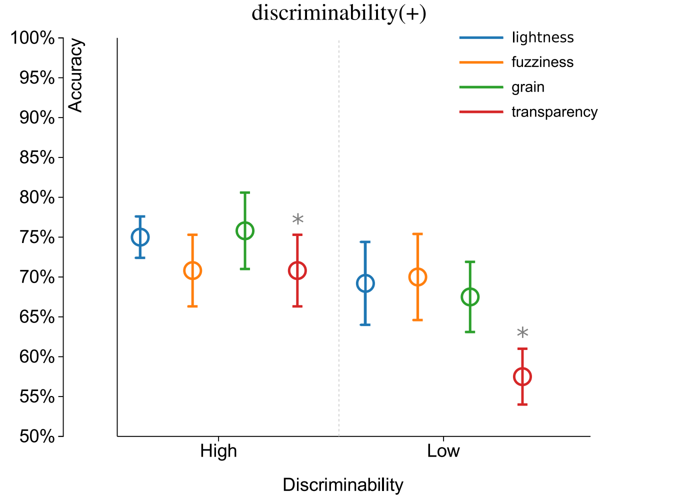
These findings reject *H5*, which states that accuracy will always be lower under the low-discriminability condition,



(a) Strength encoded using width. Significance: accuracy between lightness and fuzziness under high discriminability(*); accuracy on lightness(o) across discriminability; accuracy on transparency across discriminability($\times$); accuracy on fuzziness(+) across discriminability



(b) *vStrength* = hue. Significance: accuracy between fuzziness and lightness under high discriminability(*); accuracy between grain and lightness under high discriminability(o); accuracy on grain across discriminability(+)



(c) *vStrength* = saturation. Significance: accuracy on transparency across discriminability(*)

Fig. 10. Accuracy with the strength attribute in the presence of each uncertainty visual variable, plotted against the two discriminability levels. Error bars show ± 1 standard error. Paired symbols show significant pairwise differences.

and *H7*, that the level of distraction on the same primary visual variable introduced by a secondary visual variable will not be affected by discriminability.

5 Discussion

Here we summarize and provide possible explanations for the more important findings described in Section 4, discuss limitations in the experimental design, and recommend design considerations for visualizing network uncertainty.

5.1 Findings and Interpretations

We discuss each of the important findings below.

5.1.1 Effectiveness of the Certainty Visual Variable

Which visual variable is more effective in visualizing uncertainty? In the context of this experiment, the question of effectiveness is two-fold: 1) does a certainty visual variable have high accuracy given the presence of a strength variable, and 2) does variation in the certainty visual variable negatively impact the user's perception of the strength variable? Fig. 6 shows that the answer to the first question depends on the choice of *vStrength*. Lightness is as good as fuzziness, grain, and transparency when *vStrength* is width or saturation, but is the least accurate when *vStrength* is hue. In particular, we note that, while all are color-based, lightness and saturation seem to be much more visually separable than lightness and hue. Fig. 7 answers the second question. Fuzziness has the strongest negative impact on the effectiveness of width: in Fig. 7, the accuracy of interpreting width is the lowest with the distraction of fuzziness than with the other three certainty visual variables. Similarly, lightness has the strongest negative impact on the effectiveness of hue.

Putting the interference among visual variables aside for a moment, we note that our result is consistent with the findings in [7] on the relative effectiveness of fuzziness and grain. In [7], participants were more accurate at estimating uncertainty encoded using grain than using fuzziness. While not statistically significant, we also found grain to have higher accuracy than fuzziness while averaged over all *vStrength*, task, and discriminability conditions, despite the differences in tasks and in the presence of secondary visual variables between the two experiments.

5.1.2 Differing Effectiveness across Task Types

We measured and averaged the time the pilot participants took to accomplish each type of task correctly when the appearance of the graph edges only vary in one visual variable. The average durations were then used as the time limit for the tasks in the main study. This gives the two task types a comparable baseline difficulty. However, we observed in the study that participants were significantly more accurate on comparison tasks than on the visual search tasks. This suggests that the difficulty of visual search tasks may increase more than that of comparison tasks with the addition of variation in another visual variable. One possible explanation is that in the visual search tasks, the participant needed to scan multiple edges and spend effort on each to separate the primary visual attribute from the secondary visual attribute, while for the comparison tasks, separation of the two visual attributes was done in parallel for all the edges.

The effect of task type on accuracy also seems to be mediated by the visual variable used: in Section 4.3, we report that width led to more accurate responses than hue and saturation in the visual search task, but not in the comparison task. It is possible that participants may have used different strategies when visually aggregating values represented by width versus by hue or saturation. This implies that empirical results obtained on the effectiveness of visual variables may be task-specific for some visual variables, and thus that in evaluating visual variables it is worthwhile to include multiple tasks that cover distinct cognitive or perceptual operations.

5.1.3 Discriminability Mediates the Level of Integrality between Visual Variables

Section 4.2 reports that when the target attribute is strength, we found an interaction effect between *vStrength* and *vCertainty* under high discriminability, but not under low discriminability. This unexpected result might be explained by a compound effect of discriminability level. When the discriminability level is low, the discriminability between the encoding levels for each of the strength visual variables is reduced, and so is the discriminability for those of the certainty visual variables. Decreased discriminability of the primary visual variable makes it more difficult to complete the tasks accurately, while decreased discriminability of the secondary attribute makes the tasks easier by reducing the variation that must be filtered out when working on the task. This compound effect of discriminability could also explain the observed interaction between discriminability and *vCertainty* when strength was encoded using width.

5.1.4 Discriminability May Reflect the Inherent Degree of Integrality between Visual Variables

Looking at the change in accuracy between high overall discriminability to low overall discriminability, we observed three patterns: a) significant decrease in accuracy (observed when *vStrength* = width and *vCertainty* = lightness or transparency; when *vStrength* = hue and *vCertainty* = grain; and when *vStrength* = saturation and *vCertainty* = transparency), b) no significant difference in accuracy, c) significant increase in accuracy (observed when *vStrength* = width and *vCertainty* = fuzziness). We hypothesize that the shift from pattern (a) to pattern (c) corresponds to a shift from weaker to stronger interference between pairs of visual variables. A significant decrease in accuracy suggests that the positive contribution to accuracy of reduced variation in the secondary visual variable is not enough to cancel out the negative contribution of increased ambiguity in the primary visual variable. A significant increase in accuracy suggests that the effect of reduced variation in the secondary visual variable outweighs the effect of increased ambiguity in the primary visual variable. This hypothesis is also consistent with the degree of interference between visual variable pairs discussed in the previous section.

5.1.5 Participants Could More Easily Judge the Presence or Absence of the Extreme Values than of the Middle Values

Another interesting observation related to discriminability during the pilot study is that searching for a middle target value seemed much more difficult than searching for one that is either the minimum or maximum in the data. In the pilot, we asked the participants to look for an edge with value 1, 3, or 5 for the primary attribute. It turned out that participants took on average two to three times longer to report presence/absence correctly when the target value is 3 than in the other two cases. This observation, though not tested in the main study, seems to indicate that marks with extreme values on a visual dimension more easily attract attention among a group of stimuli than values in the middle.

5.2 Open Questions

A further question is how the results may generalize with varying graph size and density. Previous work [39] has shown that the readability (measured using both accuracy and response time) of a graph tends to decrease dramatically with increasing graph density and size. We predict that performance will drop for all pairs of visual variables with increasing graph density and size, but the relative level of interference between pairs of visual variables observed in this study will remain.

We would like to emphasize that task and visual mark types need to be considered when generalizing the results of this study. We expect our results to apply to a visualization as long as it uses line-based marks to convey information and its users may need to search for specific visual marks or judge the aggregated value of a group of visual marks. However, it is worth noting that our results may not generalize to different visualization tasks: as we discuss in Section 5.1.2, the degree to which one visual variable influences the user's performance with another variable seems to vary between tasks. Therefore, it is important to consider whether the key perceptual-level tasks a visualization needs to support match with the tasks tested here when applying the results. A number of tasks that are more specific to graphs [40] have not been tested in this study, and more experiments may be needed to better understand how the results generalize to those graph tasks. Such experiments will also provide opportunities to test how users process multiple visual variables differently across tasks. We also speculate that the results may not completely generalize to other types of visual marks. It is possible that certain visual variables, e.g. width, may become more salient and more robust to interference when being applied to visual marks that occupy more space.

One factor not considered here is the learning effect and how it may differ across different visual variables or tasks. In our experiment, none of the participants had previous experience with graph visualizations and the practice trials were not designed to give each participant extensive training on each individual visual variable. Therefore, results of the experiment apply only to novice users of node-link visualizations. It is possible that, given the same amount of training, users might improve more substantially in accuracy on some visual variables than on others. Future work could also investigate the maximum accuracy users can achieve for each pair of visual variables with plenty of practice.

It also remains an open question whether, and to what extent, participants can make judgments by taking the two attributes encoded using separate visual variables as a whole. Example tasks in the context of the experiment presented here would be to find a graph edge that has high strength and high certainty, or to determine which graph is "better" by weighing both strength and certainty. This probably boils down to whether the participant can treat a pair of visual variables as a new compound variable and what the properties, e.g. discriminability between adjacent levels, are of this new compound variable. While reminiscent of research on redundancy gain (e.g., [22]), this is a different problem and needs to be tested in future work.

Finally, the analysis on the effect of discriminability on the interference between paired visual variables here has been qualitative. It would be interesting to quantitatively relate the effect of discriminability to the level of dimensional integrality between a pair of visual variables.

5.3 Design Recommendations

We now distill our results and discussion into the following design recommendations:

- Lightness is an effective visual variable for depicting uncertainty; however, we echo previous work [19] in advising strongly against using lightness and hue to encode distinct data dimensions simultaneously.
- Fuzziness, grain, and transparency are all robust to the choice of visual variables to encode the secondary dimension. However, fuzziness has a strong negative impact on the perception of width and other alternatives should be considered when lines with different widths are to be distinguished.
- In addition to taking user tasks into consideration when designing visualization layouts and interactions, it may also be worthwhile to consider user tasks at the earlier stage of choosing visual variables.
- When two visual variables are employed simultaneously to visualize two data dimensions, perception of one of the variables can be made easier either by increasing its discriminability or by reducing the discriminability of the other visual variable.

6 Conclusion

This paper presents an experiment investigating the effectiveness of four visual variables—lightness, fuzziness, grain, and transparency—in the context of secondary visual variables and common graph-related tasks. Part of our goal was to quantify the effectiveness of a visual variable in representing uncertainty by not only how accurate people can be in interpreting the uncertainty it represents but also by the degree to which it influences the perception of other visual variables present. Our results show that fuzziness, grain, and transparency, the three visual variables that change the overall visibility of the mark, are robust to variation in the secondary visual variable. However, fuzziness has strong negative impact on the perception of width. Lightness, on the other hand, has strong bidirectional interference with hue but not with saturation, though all three are color-based dimensions. Our results also show that the negative effect on accuracy of interference between a pair of visual variables may be reduced by increasing the discriminability of the primary visual variable and decreasing the discriminability of the secondary visual variable. Finally, our results provide some evidence that the effectiveness of visual variables may depend on task type.

ACKNOWLEDGMENTS

This work was funded partially by IIS-10-18769. The authors would like to thank the reviewers for their valuable feedback.

REFERENCES

[1] M. Leitner and B. P. Buttenfield, "Guidelines for the display of attribute certainty," *Cartography Geograph. Inf. Sci.*, vol. 27, no. 1, pp. 3–14, 2000.

[2] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, "Visualizing geospatial information uncertainty: What we know and what we need to know," *Cartography Geograph. Inf. Sci.*, vol. 32, no. 3, pp. 139–160, 2005.

[3] A. Pang, "Visualizing uncertainty in geo-spatial data," in *Proc. Workshop Intersections Between Geospatial Inform. Info. Technol.*, 2001, pp. 1–14.

[4] K. Potter, P. Rosen, and C. R. Johnson, "From quantification to visualization: A taxonomy of uncertainty visualization approaches," in *Uncertainty Quantification in Scientific Computing*. New York, NY, USA: Springer, 2012, pp. 226–249.

[5] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. Moorhead, "A user study to compare four uncertainty visualization methods for 1D and 2D datasets," *IEEE Trans .Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1209–1218, Nov./Dec. 2009.

[6] J. C. Aerts, K. C. Clarke, and A. D. Keuper, "Testing popular visualization techniques for representing model uncertainty," *Cartography Geograph. Inf. Sci.*, vol. 30, no. 3, pp. 249–261, 2003.

[7] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J.-D. Fekete, "Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2769–2778, Dec. 2012.

[8] J. Bertin, *Semiology of Graphics: Diagrams, Networks, Maps*. Madison, WI, USA: Univ. Wisconsin Press, 1983.

[9] J. L. Morrison, "A theoretical framework for cartographic generalization with the emphasis on the process of symbolization," *Int. Yearbook Cartography*, vol. 14, no. 1974, pp. 115–127, 1974.

[10] A. M. MacEachren, "Visualizing uncertain information," *Cartographic Perspectives*, vol. 13, no. 13, pp. 10–19, 1992.

[11] A. M. MacEachren, *How Maps Work: Representation, Visualization, and Design*. New York, NY, USA: Guilford Press, 2004.

[12] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *J. Amer. Statist. Assoc.*, vol. 79, no. 387, pp. 531–554, 1984.

[13] M. Carpendale, "Considering visual variables as a basis for information visualisation," Dept. Comput. Sci., Univ. Calgary, Calgary, AB, Canada, Tech. Rep. TR#2001-693, 2003.

[14] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Trans. Graph.*, vol. 5, no. 2, pp. 110–141, 1986.

[15] S. Garlandini and S. I. Fabrikant, "Evaluating the effectiveness and efficiency of visual variables for geographic information visualization," in *Spatial Information Theory*. New York, NY, USA: Springer, 2009, pp. 195–211.

[16] A. Bezerianos and P. Isenberg, "Perception of visual variables on tiled wall-sized displays for information visualization applications," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2516–2525, Dec. 2012.

[17] V. John, "Functional efficiency, effectiveness, and expressivity of Bertin's visual variable colour hue in thematic map design," *J. Humanities Soc. Sci.*, vol. 8, pp. 46–55, 2013.

[18] V. Filippakopoulou, B. Nakos, E. Michaelidou, and L. Stamou, "Evaluation of the selectivity of visual variables," THALES Project, no. 65/1216, 2004.

[19] A. Reimer, "Squaring the circle? bivariate colour maps and Jacques Bertins concept of disassociation," in *Proc. Int. Cartograph. Conf.*, 2011, pp. 3–8.

[20] W. R. Garner, *The Processing of Information and Structure*. New York, NY, USA: Psychol. Press, 2014.

[21] T. C. Callaghan, "Interference and dominance in texture segregation: Hue, geometric form, and line orientation," *Perception Psychophys.*, vol. 46, no. 4, pp. 299–311, 1989.

[22] T. C. Callaghan, "Dimensional interaction of hue and brightness in preattentive field segregation," *Perception Psychophys.*, vol. 36, no. 1, pp. 25–34, 1984.

[23] C. Ware, *Information Visualization: Perception for Design*. New York, NY, USA: Elsevier, 2012.

[24] R. A. Boller, S. A. Braun, J. Miles, and D. H. Laidlaw, "Application of uncertainty visualization methods to meteorological trajectories," *Earth Sci. Inf.*, vol. 3, no. 1-2, pp. 119–126, 2010.

[25] A. Slingsby, J. Dykes, and J. Wood, "Exploring uncertainty in geodemographics with interactive graphics," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2545–2554, Dec. 2011.

[26] D. Spiegelhalter, M. Pearson, and I. Short, "Visualizing uncertainty about the future," *Science*, vol. 333, no. 6048, pp. 1393–1400, 2011.

[27] B. J. Evans, "Dynamic display of spatial data-reliability: Does it benefit the map user?" *Comput. Geosci.*, vol. 23, no. 4, pp. 409–422, 1997.

[28] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan, "Visual semiotics and uncertainty visualization: An empirical study," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2496–2505, Dec. 2012.

[29] A. M. MacEachren, C. A. Brewer, and L. W. Pickle, "Visualizing georeferenced data: Representing reliability of health statistics," *Environ. Planning A*, vol. 30, no. 9, pp. 1547–1561, 1998.

[30] D. Holten and J. J. van Wijk, "A user study on visualizing directed edges in graphs," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 2299–2308.

[31] K. Xu, C. Rooney, P. Passmore, D.-H. Ham, and P. H. Nguyen, "A user study on curved edges in graph visualization," *IEEE Trans. Vis. Comput. Graph.* vol. 18, no. 12, pp. 2449–2456, Dec. 2012.

[32] A. Telea, O. Ersoy, H. Hoogendorp, and D. Reniers, "Comparison of node-link and hierarchical edge bundling layouts: A user study," in *Visualization and Monitoring of Network Traffic*. Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2009.

[33] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proc. 7th Python Sci. Conf.*, Pasadena, CA USA, Aug. 2008, pp. 11–15.

[34] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.

[35] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 111–117.

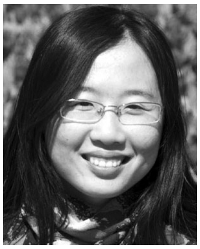
[36] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: An analytical review," *J. Vis. Languages Comput.*, vol. 14, no. 6, pp. 503–541, 2003.

[37] L. L. Thurstone, "Three psychophysical laws," *Psychol. Rev.*, vol. 34, no. 6, pp. 424–432, 1927.

[38] S. S. Stevens, *Psychophysics*. New Brunswick, NJ, USA: Transaction Publishers, 1975.

[39] M. Ghoniem, J.-D. Fekete, and P. Castagliola, "On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis," *Inf. Vis.*, vol. 4, no. 2, pp. 114–135, 2005.

[40] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proc. AVI Workshop BEyond Time Errors: Novel Evaluation Methods Inform. Vis.*, 2006, pp. 1–5.



Hua Guo is a PhD candidate in the Department of Computer Science, Brown University. She is a member of the Visualization Research Lab, Brown University. Her research interest is to inform and automate visualization design and evaluation using human-centered approaches.

Jeff Huang received the PhD degree in information science from the University of Washington. He is an assistant professor in the Computer Science Department, Brown University. His research interests include the intersection between human-computer interaction and information retrieval, involving the decoding of user behavioral data to drive novel applications.

David H. Laidlaw received the PhD degree in computer science from the California Institute of Technology, where he also did post-doctoral work in the Division of Biology. He is a professor in the Computer Science Department, Brown University. His research centers on applications of visualization, modeling, computer graphics, and computer science to other scientific disciplines. He is a fellow of the IEEE and the IEEE Computer Society and recipient of the 2008 IEEE VGTC Visualization Technical Achievement Award.