Visual Semiotics & Uncertainty Visualization: An Empirical Study

Alan M. MacEachren, Member, IEEE, Robert E. Roth, James O'Brien, Bonan Li, Derek Swingley, Mark Gahegan

Abstract—This paper presents two linked empirical studies focused on uncertainty visualization. The experiments are framed from two conceptual perspectives. First, a typology of uncertainty is used to delineate kinds of uncertainty matched with space, time, and attribute components of data. Second, concepts from visual semiotics are applied to characterize the kind of visual signification that is appropriate for representing those different categories of uncertainty. This framework guided the two experiments reported here. The first addresses representation *intuitiveness*, considering both visual variables and iconicity of representation. The second addresses relative performance of the most intuitive abstract and iconic representations of uncertainty on a map reading task. Combined results suggest initial guidelines for representing uncertainty and discussion focuses on practical applicability of results.

Index Terms — uncertainty visualization, uncertainty categories, visual variables, semiotics.

1 INTRODUCTION

Uncertainty is a fact of information; all information contains uncertainty, usually of multiple kinds. While there have been many calls for research about uncertainty visualization as a method to help information users understand and cope with uncertainty [e.g., 1, 2] and a large number of potential strategies and tools for representing uncertainty visually have been developed (see the Background section below), empirical research to assess uncertainty visualization methods has been relatively limited [exceptions include 3, 4, 5, 6, 7, 8, 9]. As a result, our understanding of when and why one uncertainty visualization strategy should be used over others remains incomplete.

Here, we address this gap by reporting on two experiments that provide insights on how to signify different categories of uncertainty. We focus on discrete symbols that could be used to signify uncertainty of individual items within information graphics, maps, or even tables or reports. The experimental design integrates theory from Visual Semiotics, Cartography, Information Visualization, and Visual Perception. Specifically, the experiments examine relative effectiveness of a set of uncertainty representation solutions—differing in the visual variable leveraged and level of symbol iconicity—when used to represent three types of uncertainty (due to accuracy, precision, and trustworthiness) matched to three components of information (space, time, and attribute). The paper is organized in four sections: Background, Experiment #1, Experiment #2, and Conclusion/Discussion.

2 BACKGROUND

Uncertainty representation and visualization has been addressed by a wide range of authors from many disciplinary perspectives. Research on uncertainty visualization has a long history [e.g., 10, 11, 12, 13, 14] and remains an active research topic within both Information Visualization and Cartography [9, 15, 16, 17, 18, 19, 20, 21]. There are multiple contemporary reviews of extant techniques for visualizing uncertainty, including MacEachren et al. [1], Zuk [7], and Bostrom [22]. Rather than summarize or repeat these reviews, we confine background to three topics that underpin the experiments reported. First, we discuss conceptualizations / taxonomies of uncertainty that link components of information (space, time, and attribute) with the types of uncertainties that may be present in these components. Then, we summarize two visual semiotic frameworks used to inform the uncertainty visualizations examined in the experiments. We first review the *visual variables*, or basic building blocks of a graphic representation, and summarize extant visual variable typologies. Next we describe the difference between iconic and abstract symbols, or the degree to which the sign-vehicle mimics its referent. The background reviews on each of these three topics were used to structure the design of the pair of experiments.

2.1 Conceptualizing Uncertainty

Uncertainty has long been recognized as a multifaceted concept [23]. A typology of uncertainty initially proposed by Thomson et al. [24], and subsequently extended by MacEachren et al. [1], underpins the research presented here. To provide context, we review the core components of the extended typology. It is organized around two primary axes: components of information and types of uncertainty (Table 1). A fundamental distinction typically is made among three components of geographic information: (1) space, (2) time, and (3) attribute; this distinction underlies most efforts to develop efficient and effective information structures for spatiotemporal information and is basic to the human understanding of the world [25].

MacEachren, et al [1] match nine types of uncertainty to these three components of information: (1) accuracy/error, (2) precision, (3) completeness, (4) consistency, (5) lineage, (6) currency, (7) credibility, (8) subjectivity, and (9) interrelatedness. This results in 27 unique conditions of information uncertainty (Table 1). In a case study focusing on spatial uncertainty visualization to support decision making within the domain of floodplain mapping, Roth [26] found accuracy/error to be the most influential of the nine types of uncertainty on decision making, with precision and currency having a secondary influence. Additional empirical investigation across uncertainty conditions has been limited.

2.2 Visual Semiotics

Visual semiotics offers a theoretical framework to conceptualize the mechanisms through which graphic representations can signify both information and its associated uncertainty. In its simplest definition, semiotics is the study of sign systems; the core goal is to understand how a symbol (the sign-vehicle) becomes imbued with meaning (the interpretant) to represent a thing or concept (the referent) [27]. Semiotics provides a framework for understanding both why graphic representations work and how to revise graphic representations for optimal signification. Important to a semiotic theory of information visualization is the identification and articulation of the basic visual variables that can be manipulated to encode information (uncertainty or otherwise).

- *Alan M. MacEachren, Penn State University, maceachren@psu.edu.*
- *Robert E. Roth, University of Wisconsin-Madison, reroth@wisc.edu*
- *James O-Brien, Risk Frontiers, Macquarie University, james.obrien@mq.edu.au*
- *Bonan Li, ZillionInfo, bonan.li@zillioninfo.com*
- *Derek Swingley, Penn State University, swingley@gmail.com*
- *Mark Gahegan: University of Auckland, m.gahegan@auckland.ac.nz*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

Table 1: Conditions of Information Uncertainty. 3 components of information (space, time, and attribute) paired with 9 uncertainty types (accuracy/error, precision, completeness, consistency, lineage, currency/timing, credibility, subjectivity, and interrelatedness). Table updated from MacEachren et al. [1]

Category	Space	Time	Attributes
Accuracy/ error	coordinates., buildings	+/- 1 day	counts, magnitudes
Precision	1 degree	once per day	nearest 1000
Completeness	20% cloud cover	5 samples for 100	75% reporting
Consistency	from / for a place	5 say M; 2 say T	multiple classifiers
Lineage	# of input sources	# of steps	transforma-tions
Currency/ timing	age of maps	C = Tpresent - Tinfo	census data
Credibility	knowledge of place	reliability of model	U.S. analyst vs. informant
Subjectivity	local ←→ outsider	expert ←→ trainee	fact ←→ guess
Interrelatedness	source proximity	time proximity	same author

The concept of visual variables was originally outlined by Bertin (under the label of "retinal variables") in 1967 and made available in an English translation in 1983 [28]. Bertin's contention—one that is still generally accepted in Information Visualization and Cartography—was that there are a set of fundamental visual variables, or manipulable primitives of graphic sign vehicles, from which any information graphic can be built. Bertin identified seven visual variables: (1) location, (2) size, (3) color hue, (4) color value, (5) grain, (6) orientation, and (7) shape. Morrison [29] suggested the addition of two more visual variables: (8) color saturation and (9) arrangement. Subsequently, MacEachren [11, 27] proposed adding three variables made practical by advances in graphics technology: (10) clarity (fuzziness) of sign vehicle components, (11) resolution (of boundaries and images), and (12) transparency (each is potentially relevant for signification of uncertainty).

Bertin and others used the concept of visual variables to develop a syntactics of graphic sign vehicles. Syntactics often are described as the 'grammatical rules' of a sign system, detailing how and when the primitive elements of a sign-system should be used for signification. Bertin based his graphical syntactics upon the level of measurement of the signified dataset, giving a rating of acceptable or unacceptable to each visual variable for numerical, ordinal, and categorical data. MacEachren [27] describes the syntactics for the above twelve visual variables, giving a three-step rating of good, marginal, and poor for use with numerical, ordinal, and categorical data. The usefulness of such syntactics of visual variables was demonstrated in Mackinlay's [30] early implementation of an expert system for automating the design of graphical presentations.

The syntactic relations of eleven of the twelve visual variables for representing uncertainty were examined in the first series of each experiment. Figure 1 provides examples of variation in the eleven tested visual variables. Resolution, as presented by MacEachren [27], is omitted because it is applicable to line symbols and images only, while the experiments reported here focus on point symbols only.

2.3 Symbolic Iconicity

Based on accepted information visualization and cartographic principles, we can predict that symbols with a dominant perceptual order will be more effective in tasks that take advantage of pre-attentive visual processes (e.g., visual search tasks, symbol comparison tasks, visual aggregation and region comparison tasks) [27]. Thus, highly abstract symbols that vary only a single visual

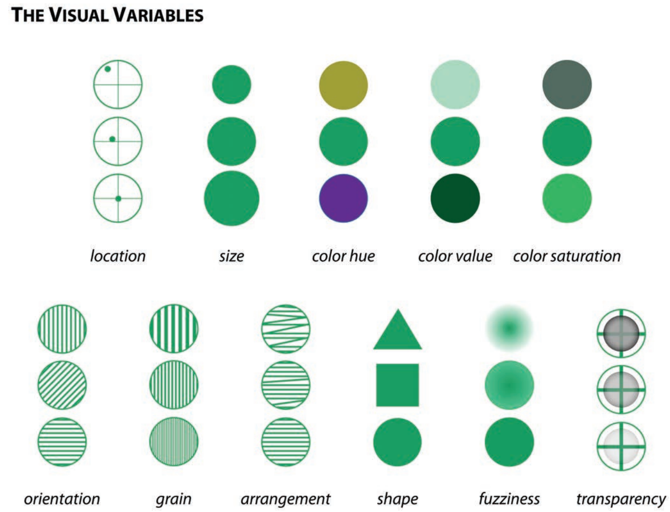

Fig 1. Visual variables applied to point symbol sets.

variable should be effective at these tasks. In contrast, we also can predict that sign vehicles prompting appropriate metaphors will be easier to match correctly with qualitatively different aspects of information, such as different categories of uncertainty. To prompt metaphors, the variation in symbols needs to incorporate a high degree of iconicity (thus be associative or pictorial rather than geometric; see Figure 2). The characteristics of sign-vehicles that make them iconic, however, often interfere with pre-attentive processing because they are more visually complex.

Ideal symbols, then, are likely to be ones that are easily understood (i.e., that are logically associated with the concept they represent) while also being effective for map reading tasks that require visual aggregation or visual search (i.e., that support pre-attentive processing). These symbol goals represent a fundamental trade-off between abstract sign vehicles, which rely on a single visual variable to communicate differences in the information, and iconic sign vehicles, which are designed to prompt particular interpretants through commonly understood metaphors.

The experiments described below addressed aspects of these two criteria separately; Experiment #1 addressed symbol intuitiveness (i.e., extent to which symbols are directly apprehended or readily understood) while Experiment #2 addressed task performance in situations in which multiple symbols appear on a display.

3 Experiment #1: Assessing Intuitiveness

Experiment #1 required participants to judge suitability of symbol sets for representing variation in a given category of uncertainty. Experimental design was informed by the framework of uncertainty conditions introduced in Table 1 and the principles of visual semiotics relating to the visual variables and symbol iconicity. To make the experiment practical, we narrowed the nine-part

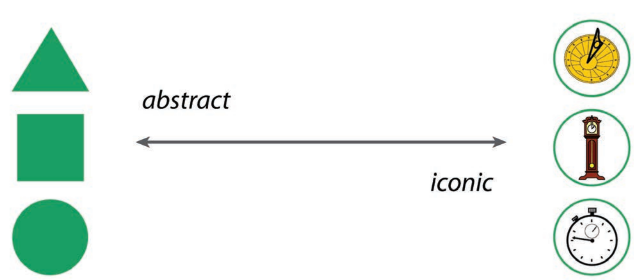

Fig 2. Symbol Iconicity. Abstract symbols (those that are geometric, varying only a single visual variable) are good for tasks that take advantage of pre-attentive processing. However, iconic symbols (those that are associative or pictorial, prompting metaphors) are potentially easier to match correctly with qualitatively different aspects of data, such as uncertainty conditions.

uncertainty typology detailed by MacEachren et al. [1] to three high-level types: (1) accuracy, defined as correctness or freedom from mistakes, conformity to truth or to a standard or model, (2) precision, defined as exactness or degree of refinement with which a measurement is stated or an operation is performed, and (3) trustworthiness, defined as source dependability or the confidence the user has in the information, thus a broad category that includes aspects of the final seven categories in Table 1. This leaves nine conditions of uncertainty for examination in the experiment (space + accuracy, space + precision, space + trustworthiness, time + accuracy, time + precision, time + trustworthiness, attribute + accuracy, attribute + precision, and attribute + trustworthiness).

Below we describe: (1) design of the symbol sets used in both experiments and (2) design, analysis, and results of Experiment #1.

3.1 Symbol Set Design

Each symbol set contained three symbols matched to a range from high to low certainty; the 3-step scale matched the typology cited above. Symbol sets designed were either iconic (resembling or having similarity with the referent) or abstract (having an arbitrary link with referent, here varying only a single visual variable). The individual symbol sets were grouped into 10 series: one for the general representation of uncertainty and one for each of the nine categories of uncertainty described above. The general series included only abstract symbol sets based upon variation in visual variables. The remaining nine series included both abstract and iconic symbol sets, allowing for comparison between two levels of iconicity. The iconic symbol sets were designed to prompt metaphors specific to the condition of uncertainty represented by the series. For the remainder of the manuscript, we use the term *symbol set* to mean a group of three symbols that could be used to depict three ordinal levels of uncertainty and the term *series* to refer to a group of symbol sets that are compared for a specific condition of uncertainty.

The Series #1 symbol sets conveyed variation in uncertainty by manipulating only a single visual variable; see Figure 1. In Experiment #1, this series of eleven symbol sets were presented in two different directions, with opposite ends 'up' in each variant, resulting in 22 symbol sets. We adopted two design constraints when designing the Series #1 symbol sets. First, color attributes (hue, value, saturation, transparency) were controlled, except when they were the visual variable under consideration. For example, all symbols used the same green hue, except the symbol set relying on color hue to convey information. The use of transparency differed from the others because it is not possible to recognize transparency unless there is an additional feature under the symbol that can be seen through it [31]. Second, all symbol sets, excepting the one using shape, had a circular outline that, excepting the symbol set using size, was the same size. Results from Series #1 provided input to decisions not only about symbolization of uncertainty on its own. Results are relevant to application of each visual variable to redundant signification (e.g., to enhance contrast of iconic symbols that might be logical but not easily located on a map) and to multivariate signification (signification of information plus its uncertainty and/or multiple aspects of uncertainty).

Design of the symbol sets for Series #2-10 focused on determining an appropriate metaphor for each of the nine uncertainty categories. We constructed 10 symbol sets for each of the nine conditions of uncertainty (90 total, subsequently narrowed to 60, see below). We adopted three design constraints. First, within the 10 total symbol sets for a category, five were abstract and five were iconic. Second, the abstract symbol sets, due to their generic design, were included in multiple series to provide a basis for comparability; this approach was not possible for the iconic symbols due to the pictorial customization for each condition of uncertainty. Of the five abstract symbol sets for each series, one abstract symbol set (the color saturation set from Series #1; see Figure 1) was included for all nine conditions of uncertainty. This decision was based on multiple suggestions in the literature that color saturation provides an intuitive method to signify uncertainty [11, 32]. Of the remaining four

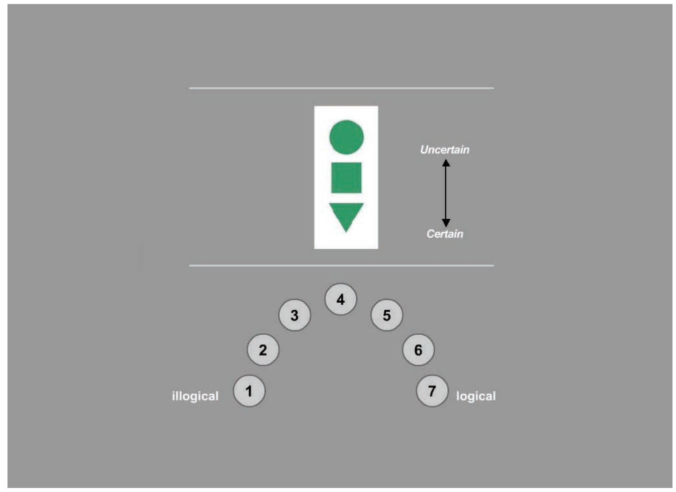

Fig 3. The Experiment #1 trial interface.

abstract symbol sets, two were common to each component of information (space, time, and attribute) and two were common to each type of uncertainty (accuracy, precision, and trustworthiness). Finally, while the logic behind the design of iconic symbol sets for each series was much more difficult to formalize, it can be noted that each series included iconic symbols emphasizing both confidence ranges and ambiguity. The final 76 symbol set designs are illustrated in Figure 4 along with descriptive statistics from Experiment #1.

3.2 Experimental Design

Experiment #1 focused on assessing symbol set *intuitiveness* (logic) for each uncertainty category and for uncertainty generally. Because inclusion of 102 symbol sets (22 in Series #1 and 90 in Series #2-10) would make the experiment prohibitively lengthy, a pilot study was run with 31 undergraduate students from Penn State University. Participants were asked to rate on a scale of 1-7 the intuitiveness of a symbol set to represent an explicitly defined category of uncertainty from Series #2-10. The top three rated abstract and iconic symbol sets in each series were selected for inclusion in Experiment #1, narrowing each series from 10 symbol sets to 6. The number of symbol sets in Series #1 was left unaltered so that syntactic relations for uncertainty visualization could be formalized for the full set. Thus, the number of tested symbols sets for Experiment #1 was reduced to 76 (22 in Series #1 and 54 in Series #2-10).

Due to inclusion of map-like displays in Experiment #2 (which drew on Experiment #1 results to determine the included symbol sets), participants were purposefully sampled to ensure they had some knowledge of maps and mapping. Therefore, undergraduate students with a GIScience major, graduate students researching a GIScience topic, and professionals working in GIScience and related fields were recruited for participation in Experiment #1. Seventy-two (n=72) participants completed timed suitability ranking tasks with the 76 symbol sets.

An experimental apparatus was created that presented instructions and tasks consistently and to record answers and *response time (RT)*. Participants in Experiment #1 worked in a computer lab with an experiment proctor present, but all instructions were embedded in the experiment application. Each session began with a descriptive overview of the experiment purpose. This was followed by a practice question to introduce the experimental interface. The experiment then progressed through the 10 series of symbol sets described above (thus 76 trials), in each case starting with the Series #1 symbol sets representing uncertainty generally. Between Series #1 and the rest, the components of information (space, time, and attribute) and types of uncertainty (accuracy, precision, and trustworthiness) were introduced in separate screens. Then, prior to beginning a new series, a preview screen containing all symbol sets to be tested in that series appeared for 10 seconds to familiarize participants with the range of symbols in the series. Order of Series #2-10 as well as order of tasks within all series was randomized to prevent order effects.

After each preview screen, the trial interface was loaded (Figure 3). The interface had two primary components: (1) a symbol set and (2) a set of intuitiveness ranking responses. For each symbol set, the top symbol was labeled as uncertain and the bottom as certain.

Participants specified the intuitiveness of the symbol set by selecting one of the seven interactive ranking buttons. Intuitiveness ranking responses were presented as a discrete visual analog scale (DVAS) from 1 (illogical) to 7 (logical). A DVAS is similar to the more commonly known Likert scale in that they both rely upon evenly-spaced integers to provide quantifiable metrics of participant assessment or preference [33]. However, a Likert scale is presented as a diverging scheme with a central middle point representing the neutral state, with each step in either direction explicitly labeled. The more generic DVAS is presented as a sequential scheme with no neutral middle-point, requiring the labeling of only the poles of the continuum. The DVAS ranking buttons were presented in a half circle, rather than the more traditional horizontal alignment, so that all buttons are an equal distance from this repositioned cursor location. Intuitiveness rankings and RTs were collected for each trial.

Following selection of a intuitiveness ranking, an update screen appeared. The update screen served four purposes: (1) notify about number of trials left in the series and number of series left in the experiment, (2) remind the user about the uncertainty condition for which they are rating each symbol set in the current series, (3) afford a mental break between trials, and (4) ensure that the mouse cursor was at a neutral location prior to every trial.

3.3 Data Analysis

Inferential statistical analysis was applied to the Experiment #1 results in two stages. In the first stage of analysis, differences in intuitiveness and RT were examined within and across series. This stage of analysis was designed to identify the most intuitive symbol set for each condition of uncertainty; this was done for abstract symbols, iconic symbols, and symbols overall. In addition, results of the first stage of analysis provided input for the delineation of syntactic relations among the visual variables for representation of ordinal levels of uncertainty.

In the second stage of analysis, differences between the abstract and iconic symbol sets were examined within and across series. This round of inferential hypothesis testing was designed as a first step to determine the relative merits of abstract versus iconic symbolization for visualizing uncertainty. Series #1 was excluded from the second stage of analysis because of its focus on abstract symbolization only.

For both stages of analysis, nonparametric statistics were applied to intuitiveness rankings, as the recorded random variable is non-continuous when using a DVAS, and parametric testing was applied to the RTs, which were continuous [34]. For the first stage of analysis, the Kruskal-Wallis test (nonparametric) was applied to the intuitiveness rankings and the ANOVA test (parametric) was applied to the RTs; both tests examine statistical difference across three or more groupings. For the second stage of analysis, the Mann-Whitney test (nonparametric) was applied to the intuitiveness rankings and the independent two-group t-test with Welsh df modification (parametric) was applied to the RTs; the Mann-Whitey and t-test are nonparametric and parametric equivalents for examining statistical difference between two unmatched groups. All statistical analysis, descriptive and inferential, was performed using R.

3.4 Results

Results for the first stage of analysis are summarized in Supplement-Table A. Differences in intuitiveness rankings for the Series #1 symbol sets were found to be significant at alpha='0.01'. This confirmed expectation that not all visual variables are intuitive for visualizing ordinal uncertainty information. There was no significant difference in RT, suggesting that participants found the task of judging intuitiveness to be similarly easy/difficult.

Further patterns were identified within Series #1 by looking at descriptive statistics (see Figure 4). Three symbol sets (fuzziness, location, and value) received a mean intuitiveness ranking over '5.0', with fuzziness and location both having a mode of '7' (the highest value on the DVAS). Based on this evidence, we find fuzziness, location, and value to be good for visualizing discrete entity uncertainty reported at the ordinal level. Three symbol sets (arrangement, size, and transparency) received a mean intuitiveness ranking between '4.0' and '5.0' and a modal intuitiveness ranking of '5.0' or higher (with means and medians at the scale midpoint or better), suggesting that they were deemed by participants as somewhat logical for the visualization of uncertainty. Therefore, we find arrangement, size, and transparency to be acceptable for visualizing discrete entity uncertainty reported at the ordinal level. The remaining symbol sets (saturation, hue, orientation, and shape) had mean, median, and modal intuitiveness rankings below '4.0' and were therefore deemed as unacceptable for visualizing discrete entity uncertainty reported at the ordinal level. This is particularly interesting for saturation, which is a commonly cited variable thought to be intuitively related to uncertainty [1].

It is important to note that the presented directionality of both good and marginal symbol sets mattered in their intuitiveness for visualizing uncertainty, as only one direction was deemed intuitive by participants (fuzziness: more fuzzy=less certain; location: further from center=less certain; value: lighter=less certain; arrangement: poorer arrangement=less certain; size: smaller=less certain; transparency: more obscured=less certain).

Returning to Supplement-Table A, a significant difference at alpha='0.01' was found in the intuitiveness ratings across Series #2-10. There are two possible explanations for this finding. The first is that it was more difficult for participants to conceptualize one or several of the uncertainty conditions compared to the rest (e.g., they understood how uncertainty is present in the space and attribute components, but not the time component, or, they understood the accuracy and precision categories of uncertainty, but not the trustworthiness category). The participants may miss the metaphor prompted by a given symbol set if they have a poor conceptualization of the associated condition of uncertainty. The second possible explanation is a difference in logic of symbol sets by series, thus participants may have understood the concepts to be represented, but they did not find the symbol sets in some categories to be logically matched with those concepts. Because a significant difference in RT was not found across Series #2-10—showing that participants did not need to spend more time interpreting some series compared to others—the second explanation is more likely.

Six of the nine conditions of uncertainty (space + accuracy, space + precision, space + trustworthiness, time + trustworthiness, attribute + precision, and attribute + trustworthiness) reported a significant difference in intuitiveness ratings at alpha='0.05' for the symbol sets within the given series (four of these are significant at alpha='0.05'). Thus, all space conditions and all trustworthiness conditions exhibit differences in symbol set intuitiveness ratings. In only one case (see below) is the difference attributable to differences between iconic and abstract symbol sets generally. When examining the descriptive statistics for individual symbol sets within each series (Figure 4), the difference in intuitiveness rankings for the three trustworthiness series is caused by one symbol set receiving distinctly higher ratings while the difference in intuitiveness rankings for space series (aside from space + trustworthiness) is caused by one symbol set receiving distinctly lower ratings.

Only three of nine conditions of uncertainty (space + precision, space + trustworthiness, and attribute + precision) reported a significant difference in RTs at alpha='0.05' (time + trustworthiness is significant at alpha='0.10'). However, all series exhibiting a significant difference in RT also exhibited a significant difference in intuitiveness ranking. This relationship is to be expected, as symbol sets that are not logical or do not invoke proper metaphors will likely take longer to interpret, and therefore longer to rate for intuitiveness. Because this match between differences in ratings and RT was not

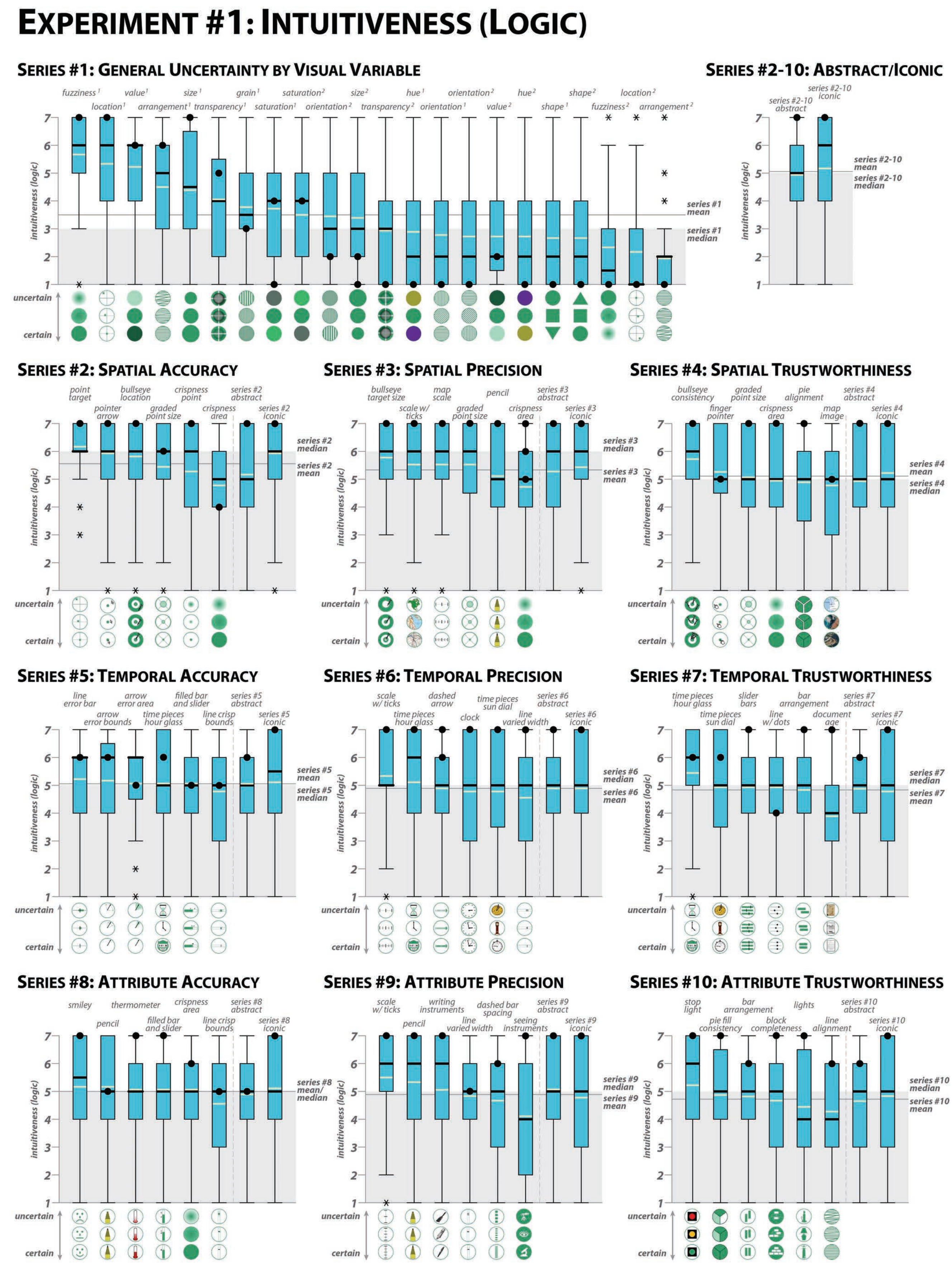

Figure 4. Descriptive statistics by series and symbol set with results for abstract symbols based on visual variables (Series 1) at the top followed by Series 2-10.On box-plots mean is shown as a black line, median as a gray line, and mode as a black dot.

exhibited in Series #1, in which each symbol set isolated a single visual variable, the relationship between intuitiveness and RT is perhaps only apparent with an increase in symbol iconicity.

As shown in Figure 4 descriptive statistics, the average intuitiveness scores for the Series #2-10 symbol sets were generally higher than those from Series #1, with a large majority of symbol sets receiving a score over ‘5.0’ (the threshold used for Series #1 for marking a particular symbol set as good for visualizing uncertainty). This finding was expected, as the Series #1 set of symbols were designed without any particular category of uncertainty in mind and participants were asked to judge their intuitiveness for general uncertainty signification.

Using the descriptive statistics in Figure 4 with the inferential statistics in Supplement-Table A, it is possible to recommend an abstract and iconic symbol set ‘intuitiveness winner’ for each condition of uncertainty and to determine if this ‘win’ is significant (i.e., if the lowest ranking symbol sets can be discredited or if it remains viable). Identifying a intuitiveness winner is useful for the actual application of the symbol sets, but was also essential for administration of Experiment #2. Table 2 summarizes the winning abstract and iconic symbol set for each condition of uncertainty, identifying the symbol sets by the names given in Figure 4, and an asterisk if the win was significant.

The second stage of analysis examined the difference between abstract and iconic symbolization in Series #2-10. The results of this round of analysis are provided in Table 3. Looking at Series #2-10 pooled together, there was a significant difference in intuitiveness rankings at alpha=‘0.01’ between abstract and iconic symbol sets. The descriptive statistics for abstract and iconic symbol sets provided in Figure 4 reveal that iconic symbol sets received a slightly higher mean intuitiveness ranking overall (‘5.13’) than their abstract counterparts (‘4.98’). However, the difference was significant at alpha=‘0.01’ for only one of the individual series, space + accuracy is, with attribute + accuracy significant at alpha=‘0.10’. This mismatch may be caused by the added statistical power provided when pooling Series #2-10 symbol sets, allowing for the detection of smaller differences between groups with the same level of statistical significance. For three of the nine series (time + precision, time + trustworthiness, and attribute + precision), the abstract symbol sets scored slightly higher than the iconic symbol sets. Thus, it is not possible to state that the iconic symbolization is consistently more intuitive regardless of uncertainty condition.

There was a significant difference in RT between abstract and iconic symbolization at alpha=‘0.01’ when Series #2-10 were grouped. Unlike intuitiveness rankings, however, this relationship also was present when looking at the difference between abstract and iconic symbol sets within a majority of individual series. Five of the nine series (space + precision, space + trustworthiness, time + accuracy, time + precision, and time + trustworthiness) showed a significant difference between RTs at alpha=‘0.05’; an additional two series (attribute + precision and attribute + trustworthiness) were significant at alpha=‘0.10’. For all but one of the series (attribute + accuracy), participants required more time to determine the intuitiveness of iconic symbol sets than their abstract counterparts.

The overall result that iconic symbol sets are rated slightly higher on intuitiveness for uncertainty representation but require slightly longer to rate matches theoretically-grounded expectations. Abstract symbol sets should be fast to judge since the process of interpreting order and directionality (i.e., which end means more and which means less) is largely a perceptual task. Iconic symbol sets will require more cognitive processing to identify the intended metaphorical relationship with the uncertainty condition signified. But, since the iconic symbol sets have been designed explicitly to prompt a metaphorical relationship with the uncertainty condition signified, when the design is successful, the rating of intuitiveness should be higher. The fact that iconic symbol sets were not overwhelmingly rated as more intuitive suggests that: (a) the uncertainty conditions are hard for users to conceptualize, thus the match with any metaphor will be weak, (b) differences among uncertainty conditions, while understood by users, do not have obvious visual analogs, or (c) we were simply not successful in designing symbol sets that prompt a metaphor that fits the conceptualization of different uncertainty conditions. The latter was a factor in the *seeing instruments* and *lights* symbol sets for attribute precision and trustworthiness, respectively and in the *document age* set for temporal trustworthiness (Figure 4).

Table 2: Abstract and iconic intuitiveness choices for each series. Groupings with significant differences in intuitiveness ranking (from Table 2) at alpha=‘0.05’ are marked (*)

Series #	Abstract Winner	Iconic Winner
Series #2. Space + Accuracy	graded point size*	point target
Series #3. Space + Precision	scale w/ ticks*	bullseye target size
Series #4. Space + Trustworthiness	crispness area	consistency bullseye*
Series #5. Time + Accuracy	line error bar	arrow error bounds
Series #6. Time + Precision	scale w/ ticks*	time pieces hour glass
Series #7. Time + Trustworthiness	line w/ dots	time pieces sun dial*
Series #8. Attribute + Accuracy	filled bar and slider	smiley
Series #9. Attribute + Precision	scale w/ ticks*	pencil*
Series #10. Attribute + Trustworthiness	pie fill consistency	stop light

Table 3: Results for the second stage of analysis, assessing statistical differences between abstract and iconic symbolization. The Mann-Whitney test was applied to the intuitiveness rankings and the independent two-group t-test with Welsh df modification was applied to the RTs. Significant results at alpha=‘0.10’, alpha=‘0.05’, and alpha=‘0.01’ are marked in increasing shades of red.

Series #	Intuitiveness Ratings		Response Times		
	W	p-value	t	df	p-value
Series #2. Space + Accuracy	16370.0	*0.0000*	1.4303	341.947	0.1535
Series #3. Space + Precision	21939.0	0.2706	-2.7179	394.349	*0.0069*
Series #4. Space + Trustworthiness	21530.5	0.1574	-3.3146	421.223	*0.0010*
Series #5. Time + Accuracy	22087.5	0.3293	-2.0233	317.988	*0.0439*
Series #6. Time + Precision	23085.5	0.8493	-2.8751	354.435	*0.0043*
Series #7. Time + Trustworthiness	23702.5	0.7696	-2.4773	373.571	*0.0137*
Series #8. Attribute + Accuracy	21150.0	*0.0873*	1.4040	356.348	0.1612
Series #9. Attribute + Precision	24016.0	0.5896	-1.7144	405.775	*0.0872*
Series #10. Attribute + Trustworthiness	22070.5	0.3254	-1.8319	351.631	*0.0678*
Across Series #2-10	1763637.0	*0.0002*	-4.4664	3731.04	*0.0000*

4 Experiment #2: Symbol Sets in Map Displays

4.1 Experimental Design

Experiment #2 complements the focus on symbol intuitiveness from Experiment #1 with a focus on symbol effectiveness for a typical map use task: assessing and comparing the aggregate uncertainty in two map regions. Thirty participants completed the assessment of aggregate uncertainty tasks in a computer lab with a proctor present. As with Experiment #1, undergraduate students with a GIScience major, graduate students researching a GIScience topic, and professionals working in GIScience and related fields were purposefully recruited for participation in Experiment #2 to ensure they had some knowledge of maps and mapping.

The assessment of aggregate uncertainty tasks was completed using the most intuitive abstract and iconic symbol sets identified in Series #2-10 of Experiment #1 (Table 3). In two cases (Series #4 abstract and Series #7 iconic), we used the 2nd highest scoring symbol set for Experiment #2 because the winner already was selected for a different uncertainty type. We included two additional abstract symbol sets from Series #1 of Experiment #1 (fuzziness and color value) that were not identified as the winner for any condition of uncertainty (i.e., in Series #2-10 of Experiment #1), giving us a total of 20 symbol sets for examination, two per series from Experiment #1. Each of the 20 symbol sets was tested in 12 different map region configurations (details below), producing 240 total trials.

Like Experiment #1, Experiment #2 began with a descriptive overview of the experiment, followed by a practice question using the experimental interface. The experiment then progressed in 10 series of 24 trials each. Each trial included two screens shown individually in sequence: (1) a legend showing the three symbols in the tested symbol set with an indication of their order from uncertain to certain (Figure 5) and (2) the map region trial itself (Figure 6). The symbol set legend screen served the secondary purpose as an update screen (as described above for Experiment #1) that offered a mental break and repositioned the mouse cursor to a neutral location.

As shown in Figure 6, the second screen of the Experiment #2 trial interface presented the participants with a map-like display containing nine locations in each of three regions for which uncertainty was indicated by one of the symbols in the trial set. These were presented to the geographically knowledgeable participants as "maps" with two "regions," but the maps were abstract enough to represent information displays more generally. The participant's task was to select the region of the pair for which information is least certain overall. Thus, participants had to conceptually combine nine symbols in each region into an assessment of aggregate uncertainty. Participants submitted their choices by clicking directly on the chosen map region.

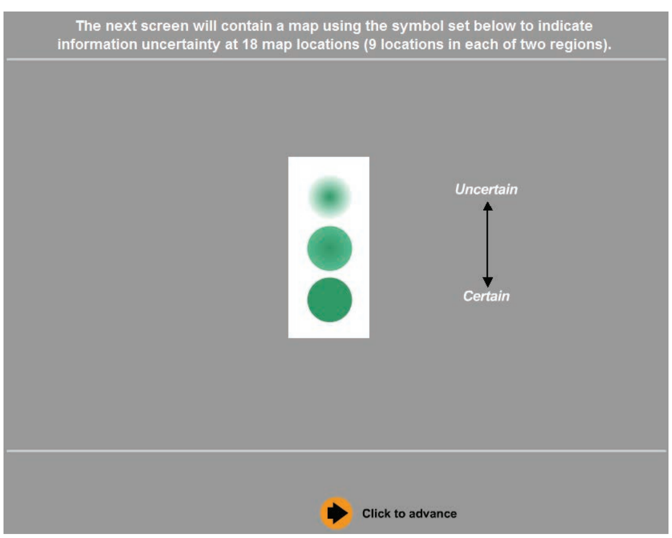

Fig 5. Example screen #1 of an Experiment #2 trial.

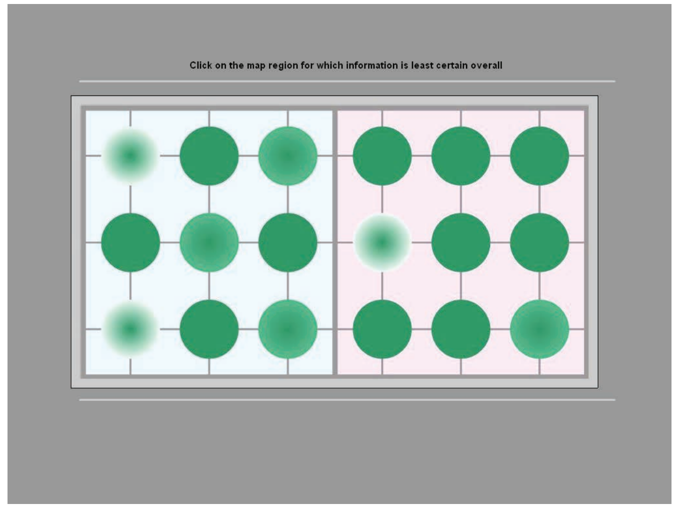

Fig 6. Example screen #2 of an Experiment #2 trial. The trial interface presents two map regions to the participant, each with uncertainty signified for nine locations. The participant must conceptually aggregate the uncertainty of each region and select the region that is least certain by directly clicking on the map.

The spatial configuration of the uncertainty symbols is likely to influence aggregate judgments. This was controlled for by devising a spatial configuration strategy that prevented participants from being able to memorize the configuration of uncertainty (which might influence their accuracy and speed in responding to the map region comparison task), yet kept the task functionally equivalent from one series to the next (so that the overall level of difficulty for each series of trials was the same). We designed the symbol configurations so that each map region fell into one of four degrees of aggregate uncertainty selected to generate tasks covering a range of difficulty: (1) Highly Uncertain: 7-H + 1-M + 1-C (where H = most uncertain symbol, M = middle symbol, and C = most certain symbol in symbol set); (2) Moderately Uncertain: 4-H + 3-M + 2-C; (3) Moderately Certain: 2-H + 3-M + 4-C; (4) Highly Certain: 1-H + 1-M + 7-C. There are 12 non-equivalent configuration pairings when each individual map region is allowed to fall into one of four degrees of aggregate uncertainty (see Figure 7). We removed configurations where both map regions have equal amounts of uncertainty (i.e., the 1-1, 2-2, 3-3, and 4-4 pairings) so that each trial had a 'correct' answer. All 12 configurations were tested in an individual trial for each of the included symbol sets (20 symbols sets for 12 map region configurations produced 240 total trials).

As in Experiment #1, the first series of trials focused on general uncertainty, without a particular uncertainty condition mentioned. Series #1 included all map region configurations for the crispness (12 trials) and color value (12 trials) symbol sets. After participants completed Series #1, background information on the nine conditions of uncertainty (as reviewed above) was provided to the participants (the same background information as used define the conditions in Experiment #1). The order of the remaining nine series of trials was randomized. Each of the subsequent series included the map region configurations for the abstract (12 trials) and iconic (12 trials) symbol set winners from Experiment #1 for the associated condition of uncertainty; the series numbering is the same in both Experiment #1 and Experiment #2 in the following analysis and reporting. The viewing order of individual trials was randomized within each of the remaining 10 series, as with Series #1. Suitability rankings and RTs were collected for each trial.

4.2 Data Analysis

As with Experiment #1, inferential statistical analysis was applied to the Experiment #2 results in two stages. In the first stage of analysis, differences in accuracy and RT were examined across Series #2-10. This analysis provided insight into the nature of

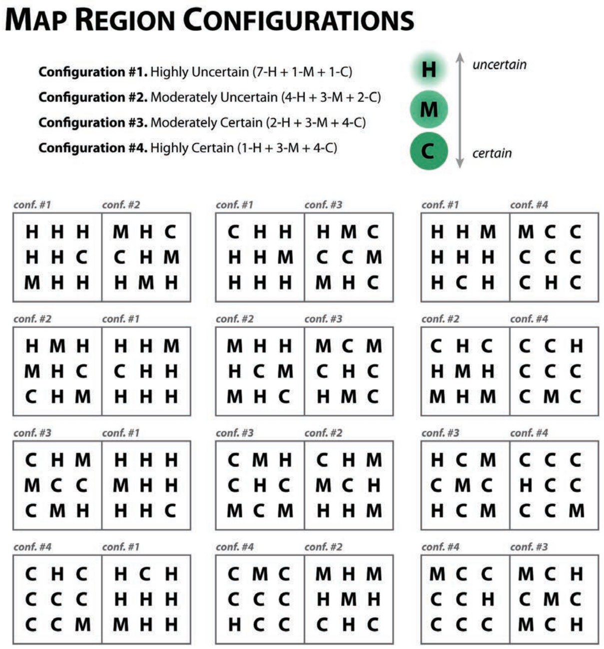

Fig 7. The 12 map region configurations. Each individual map region was allowed to fall into one of four degrees of aggregate uncertainty, producing twelve possible map region configurations.

geospatial uncertainty and the relative difficulties exhibited when performing map reading tasks under different uncertainty conditions. The inferential statistical analysis considered all Series #2-10 symbol sets together, as well as the abstract and iconic symbol sets individually; Series #1 was not included in this analysis, as this pair of symbol sets was designed for general uncertainty. Descriptive summary statistics were used to identify the symbol sets and uncertainty conditions that garnered the best and worst performance.

In the second stage of analysis, differences between the abstract and iconic symbol sets were examined within and across series. This stage included examining differences between the two Series #1 symbol sets to determine if either supported more accurate or faster assessment of aggregate uncertainty generally. It also included analysis of differences between abstract and iconic symbols in Series #2-10 symbol sets, both pooled together and within each series individually. This step was designed to determine the relative merits of abstract versus iconic symbolization for visualizing uncertainty. The inferential statistical analysis in both stages provided performance measures to complement the intuitiveness measure provided in Experiment #1.

For both analysis stages, nonparametric statistics were applied to assessment accuracy, as the recorded random variable was binary and therefore non-continuous, and parametric testing was used for RTs, which were continuous [34]. For the first stage, the Pearson's chi-square test with Yates' continuity correction (nonparametric) was applied to the accuracy recordings and the ANOVA test was applied to the RTs. For the second stage of analysis, the Pearson's chi-square test with Yates' continuity correction (non-parametric) was applied to the accuracy recordings and the independent two-group t-test with Welsh df modification (parametric) was applied to the RTs. As with Experiment #1, all analysis for Experiment #2 was performed using the statistical software package R.

4.3 Results

The results from the first stage of analysis for Experiment #2 provided the most clear and consistent set of results from either experiment. As shown in Table 4, significant differences in both assessment accuracy and RT were reported at alpha='0.01' across the nine series. The same level of significance was found when examining abstract or iconic symbol sets in isolation or when pooling all symbol sets together. This finding suggests that participants were not equally comfortable making assessments of aggregate uncertainty for all uncertainty conditions.

Table 4: Statistical results for stage 1 analysis, Experiment #2, differences across uncertainty condition. Pearson's chi-square test with Yates' continuity correction was applied to accuracy recordings and ANOVA was applied to RTs. Significant results at alpha='0.01' are marked in increasing shades of red

Subset	Assessment Accuracy			Response Time		
	x2	df	p-value	F	df	p-value
Series #2-10 all	31.4829	8	*0.000*	36.271	8,6471	*0.0000*
Series #2-10 all abstract	35.2147	8	*0.000*	24.182	8,3231	*0.0000*
Series #2-10 all iconic	25.7732	8	*0.001*	34.838	8,3231	*0.0000*

Results for the second analysis stage for Experiment #2 are summarized in Table 5 and Figure 8. Pooled data for Series #2-10 exhibited no significant difference in assessment accuracy between abstract and iconic symbol sets. Participants were more accurate using iconic symbols for five of the nine series (space + precision, space + trustworthiness, time + accuracy, time + precision, time + trustworthiness), but only two of these had significant differences (space + trustworthiness, alpha='0.01'; and time + precision, alpha = '0.05'). One series resulted in the abstract symbol set being significantly more accurate (space + accuracy, alpha='0.05'). Overall, the level of iconicity did not have a consistent influence on accuracy of aggregate uncertainty assessment.

Table 5: Results for stage 2, Experiment #2, analyzing differences within and across symbol sets. Pearson's chi-square with Yates' continuity correction is applied to accuracy recordings and the independent two-group t-test with Welsh df modification is applied to RTs. Significant results at alpha='0.10', alpha='0.05', and alpha='0.01' marked in increasing shades of red

Series #	Accuracy			Response Time		
	x2	df	p-value	t	df	p-value
Series #1. General	0.9976	1	0.318	-0.4745	717.68	0.6353
Series #2-10	0.0549	1	0.459	-5.3275	6231.70	*0.0000*
Series #2. Space + Accuracy	4.8774	1	*0.027*	-5.8958	680.60	*0.0000*
Series #3. Space + Precision	0	1	1.000	-1.511	701.27	0.1312
Series #4. Space + Trustworthiness	11.9707	1	*0.001*	-8.5933	426.44	*0.0000*
Series #5. Time + Accuracy	0.2009	1	0.654	-1.5461	717.967	0.1225
Series #6. Time + Precision	6.3712	1	*0.0116*	2.9178	717.99	*0.0036*
Series #7. Time + Trustworthiness	1.6911	1	0.194	7.7868	679.033	*0.0000*
Series #8. Attribute + Accuracy	0.25	1	0.617	-1.2987	710.974	0.1945
Series #9. Attribute + Precision	2.1879	1	0.139	-6.4604	641.259	*0.0000*
Series #10. Attribute + Trustworthiness	2.6585	1	0.103	1.9579	618.503	*0.0507*

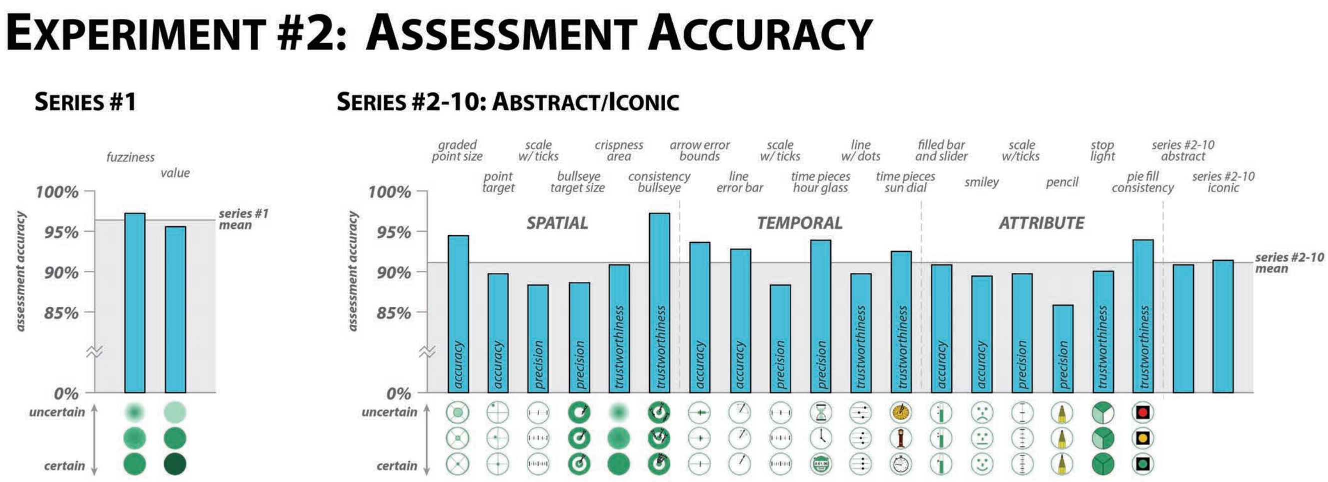

Fig 8. Experiment #2 descriptive statistics by series and symbol set.

As with Experiment #1, RT is related to degree of iconicity in Experiment #2 overall. Pooled results for Series #2-10 exhibited a significant RT difference between abstract and iconic symbol sets at alpha='0.01'. Within series, five of the nine series (space + accuracy, space + trustworthiness, time + precision, time + trustworthiness, and attribute + precision) also reported significant responses time differences at alpha='0.01', with a sixth (attribute + trustworthiness) significant at alpha='0.10'. Like Experiment #1, it generally took longer for participants to compare regions of iconic symbols (mean RT = 3800.41 milliseconds) than regions of abstract symbols (average RT = 3147.81 milliseconds). However, this was not consistent across all series, as significantly more time was taken to respond to abstract symbols for three of the nine series (time + precision, time + trustworthiness, and attribute + trustworthiness). Finally, no significant difference in accuracy or RT was found for the two tested Series #1 symbol sets.

5 Conclusions & Discussion

Like any controlled experiments, this pair necessarily constrained the problem of uncertainty visualization in multiple ways to enable valid analysis. Thus, applicability of results needs to be considered in relation to the constraints. Within these constraints, the research produced several potentially generalizable conclusions. One is that there is a clear difference in intuitiveness for representing uncertainty among abstract sign-vehicles based upon individual visual variables. Fuzziness and location work particularly well; value and arrangement are also rated highly and both size and transparency are potentially usable. As noted above, saturation, often cited as intuitively related to uncertainty, was ranked quite low. These results, since they relate to fundamental visual variables, may prove to be applicable well beyond the kinds of displays tested here.

Another generalization is that, while iconic sign-vehicles can be more intuitive and more accurately judged when aggregated (than are abstract sign-vehicles), the abstract sign-vehicles can lead to quicker judgments. Plus, and not surprisingly, iconic sign vehicles only work well if users understand both the aspect of uncertainty being signified and the metaphor upon which the sign-vehicles are based (this conclusion is our intuition about how to explain the evidence, but needs further research to assess in depth). Finally, while Experiment #2 focused on "maps", these maps were generic enough that results should generalize to other information displays with multiple points-per-region (e.g., displays depicting cluster results for documents). More importantly, the combined experiments allowed for key principles of sign-vehicle design to be assessed and provide input into guidelines for methods to represent various kinds of uncertainty (individually or in combination) in a range of contexts.

As with any empirical research, many things were not tested, thus results can be considered only a step toward comprehensive understanding of the important parameters for effective uncertainty visualization. Multiple questions remain unanswered. Building on the conceptual framework outlined plus empirical results, the following questions are ones that we feel are particularly relevant to address:

- What symbolization methods work best if there is a need to integrate both data and data uncertainty representation into the same sign-vehicles?
- How scalable are the point symbols (sign-vehicles) tested here? Will they work if reduced in size for use on mobile devices?
- How much impact does the background display have on speed and accuracy of sign-vehicle interpretation?
- How does the spatial distribution of symbolized information impact interpretation?
- Do insights about visual signification of uncertainty at discrete locations (as tested here) extend to linear or area (field) data?

More broadly, the experiments reported here were limited to very simplistic display that was non-interactive for tasks that were simple judgments of suitability or information retrieval. These limitations highlight two important additional next steps in research. First, attention needs to be directed to signification of uncertainty in interactive environments in which users have the ability to control factors such as when uncertainty signification is visible and the relative visual balance between data and data uncertainty in displays showing both at once. Second, once design guidelines are developed to specify the best strategy to signify (or interact with) information uncertainty, an equally important question to answer is how the visualization of uncertainty influences reasoning and decision making in problem context for which uncertainty matters. In spite of experimental limitations and open questions, we believe that the approach to considering uncertainty presented here is a general one that can serve as a framework for deeper understanding of visual signification of uncertainty.

Acknowledgments

The authors wish to thank James Macgill and Isaac Brewer for input on early discussions and on a pilot study leading to the experiment reported here. We would also like to thank Todd Peto, who generated many of the symbols tested. This material is based in part upon work supported by the U.S. Department of Homeland Security under Award #2009-ST-061-CI0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

REFERENCES

[1] 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 and Geographic Information Science, vol. 32, no. 3, pp. 139-160, 2005.

[2] J. J. Thomas, and K. A. Cook, "A visual analytics agenda," Computer Graphics and Applications, vol. January/February, pp. 10-13, 2006.

[3] B. Evans, "Dynamic display of spatial data-reliability: Does it benefit the map user?," Computers & Geosciences, vol. 23, no. 4, pp. 409-422, 1997.

[4] A. M. MacEachren, C. A. Brewer, and L. W. Pickle, "Visualizing georeferenced data: Representing reliability of health statistics," Environment and Planning A, vol. 30, pp. 1547-1561, 1998.

[5] M. Leitner, and B. P. Buttenfield, "Guidelines for the display of attribute certainty," Cartography and Geographic Information Science, vol. 27, no. 1, pp. 3-14, 2000.

[6] J. C. Aerts, K. C. Clarke, and A. D. Keuper, "Testing popular visualization techniques for representing model uncertainty," Cartography and Geographic Information Science, vol. 30, no. 3, pp. 249-261, 2003.

[7] T. D. Zuk, "Visualizing Uncertainty," Dissertation, Department of Computer Science, University of Calgary, Calgary, Canada, 2008.

[8] R. E. Roth, "The impact of user expertise on geographic risk assessment under uncertain conditions," Cartographic and Geographic Information Science, vol. 36, no. 1, pp. 29-43, 2009.

[9] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. J. Moorhead, "A user study to compare four uncertainty visualization methods for 1d and 2d datasets," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1209-1218, 2009.

[10] H. Ibrekk, and M. G. Morgan, "Graphical communication of uncertain quantities to non-technical people," Risk Analysis, vol. 7, no. 4, pp. 519-529, 1987.

[11] A. M. MacEachren, "Visualizing uncertain information," Cartographic Perspectives, vol. 13, fall, pp. 10-19, 1992.

[12] N. D. Gershon, "Visualization of fuzzy data using generalized animation," in Visualization, Boston, Massachusetts, 1992, pp. 268-273.

[13] B. Buttenfield, "Representing data quality," Cartographica, vol. 30, no. 3, pp. 1- 7, 1993.

[14] P. F. Fisher, "Visualizing uncertainty in soil maps by animation," Cartographica, vol. 30, no. 3, pp. 20-27, 1993.

[15] T. Zuk, and S. Carpendale, "Theoretical anaylsis of uncertainty visualization," in Proc. of SPIE, Visualization and Data Analysis 2006, San Jose, CA, USA, 2006, pp. 66-79.

[16] M. Riveiro, "Evaluation of uncertainty visualization techniques for information fusion," in Proceedings of The 10th International Conference on Information Fusion (FUSION 2007), Quebec, Canada, 2007.

[17] C. Collins, S. Carpendale, and G. Penn, "Visualization of uncertainty in lattices to support decision-making," in Eurographics/IEEE VGTC Symposium on Visualization, Norrköping, Sweden, 2007, pp. 51-58.

[18] A. Streit, "Encapsulation and abstraction for modeling and visualizing information uncertainty," Dissertation, Information Technology, Queensland University of Technology, Brisbane, Australia, 2008.

[19] M. Skeels, B. Lee, G. Smith, and G. G. Robertson, "Revealing uncertainty for information visualization," Information Visualization, vol. 9, no. 1, pp. 70-81, 2009.

[20] D. Lloyd, and J. Dykes, "Exploring Uncertainty in Geodemographics with Interactive Graphics," IEEE Transaction on Visualization & Computer Graphics, vol. 17, no. 12, pp. 2498-4507, 2011.

[21] D. Spiegelhalter, M. Pearson, and I. Short, "Visualizing Uncertainty About the Future," Science, vol. 333, no. 6048, pp. 1393-1400, 2011.

[22] A. Bostrom, L. Anselin, and J. Farris, "Visualizing seismic risk and uncertainty: A review of related research," Annals of the New York Academy of Sciences, vol. 1128, pp. 29-40, 2008.

[23] D. Sinton, "The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study," Harvard papers in GIS #7, D. Dutton, ed., Cambridge, MA: Harvard University, 1978.

[24] J. Thomson, B. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "Typology for visualizing uncertainty," in IS&T/SPIE Symposium on Electronic Imaging, Conference on Visualization and Data Analysis, San Jose, CA, 2005, pp. 146-157.

[25] D. J. Peuquet, "It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems," Annals of the Association of American Geographers, vol. 84, no. 3, pp. 441-461, 1994.

[26] R. E. Roth, "A qualitative approach to understanding the role of geographic information uncertainty during decision making," Cartography and Geographic Information Science, vol. 36, no. 4, pp. 315-330, 2009.

[27] A. M. MacEachren, How maps work, New York, NY, USA: The Guilford Press, 1995.

[28] J. Bertin, Semiology of graphics: Diagrams, networks, maps, Madison, WI: University of Wisconsin Press, 1967|1983.

[29] J. L. Morrison, "A theoretical framework for cartographic generalization with the emphasis on the process of symbolization," International Yearbook of Cartography, vol. 14, pp. 115-127, 1974.

[30] J. Mackinlay, "Automating the design of graphical presentations of relational information," ACM Transactions on Graphics, vol. 5, no. 2, pp. 110-141, 1986.

[31] R. E. Roth, A. W. Woodruff, and Z. F. Johnson, "Value-by-alpha Maps: An alternative technique to the cartogram," The Cartographic Journal, vol. 47, no. 2, pp. 130-140, 2010.

[32] D. M. Schweizer, and M. F. Goodchild, "Data quality and choropleth maps: An experiment with the use of color," in GIS/LIS, San Jose, CA, 1992, pp. 686-699.

[33] R. F. DeVellis, Scale development: Theory and applications, 2nd edition, Newbury Park, CA: Sage Publications, 2003.

[34] J. E. Burt, G. M. Barber, and D. L. Rigby, Elementary statistics for geographers, 3rd Edition, New York, NY: Guilford Press, 2009.

Supplement - Table A: Statistical results for the first stage of analysis, assessing statistical significance of differences in symbol sets within and across series. The Kruskal-Wallis test was applied to the suitability rankings and the ANOVA test was applied to the response times. Significant results at alpha='0.10', alpha='0.05', and alpha='0.01' are marked in increasing shades of red.

Series #	**Subset**	**Suitability Ratings**			**Response Times**		
		x2	**df**	**p-value**	**F**	**df**	**p-value**
Series #1. General	all	382.7215	21	***0.0000***	1.3540	21,562	0.1305
Across Series #2-10	all	71.3144	8	***0.0000***	0.8892	8,3879	0.5245
	all abstract	17.5867	8	***0.0246***	1.298	8,1935	0.2400
	all iconic	71.3899	8	***0.0000***	2.284	8,1935	***0.0197***
Series #2. Space + Accuracy	all	39.1793	5	***0.0000***	0.5312	5,426	0.7527
	abstract only	7.5315	2	***0.0232***	0.1223	2,213	0.8850
	iconic only	1.6149	2	0.4460	0.2615	2,213	0.7702
Series #3. Space + Precision	all	12.2810	5	***0.0311***	3.3819	5,426	***0.0052***
	abstract only	7.2489	2	***0.0267***	0.4419	2,213	0.6434
	iconic only	3.7198	2	0.1557	3.4159	2,213	***0.0347***
Series #4. Space + Trustworthiness	all	17.6632	5	***0.0034***	7.9894	5,426	***0.0000***
	abstract only	0.7317	2	0.6936	1.9255	2,213	0.1483
	iconic only	15.6625	2	***0.0004***	11.2750	2,213	***0.0000***
Series #5. Time + Accuracy	all	3.4579	5	0.6298	1.2904	5,426	0.2670
	abstract only	2.7157	2	0.2572	0.1407	2,213	0.8688
	iconic only	0.0298	2	0.9852	0.7095	2,213	0.4930
Series #6. Time + Precision	all	7.5734	5	0.1814	2.1273	5,426	0.0612
	abstract only	6.8611	2	***0.0324***	0.2282	2,213	0.7961
	iconic only	0.9645	2	0.6174	0.7379	2,213	0.4793
Series #7. Time + Trustworthiness	all	30.5572	5	***0.0000***	1.8647	5,426	***0.0993***
	abstract only	0.0742	2	0.9636	0.0727	2,213	0.9299
	iconic only	26.9586	2	***0.0000***	1.1227	2,213	0.3273
Series #8. Attribute + Accuracy	all	6.5452	5	0.2567	0.5851	5,426	0.7114
	abstract only	3.1343	2	0.2086	1.1227	2,213	0.3273
	iconic only	0.5508	2	0.7593	0.8520	2,213	0.4280
Series #9. Attribute + Precision	all	22.9447	5	***0.0003***	5.5758	5,426	***0.0001***
	abstract only	10.4678	2	***0.0053***	3.9714	2,213	0.2026
	iconic only	11.9256	2	***0.0026***	7.5969	2,213	***0.0007***
Series #10. Attribute + Trustworthiness	all	11.4268	5	***0.0436***	1.8487	5,426	0.1022
	abstract only	4.4369	2	0.1088	0.0192	2,213	0.9810
	iconic only	5.8229	2	***0.0544***	1.9975	2,213	0.1382