Journal of Experimental Psychology: Applied
2016, Vol. 22, No. 4, 381–392

1076-898X/16/$12.00 http://dx.doi.org/10.1037/xap0000103

Where Are You? The Effect of Uncertainty and Its Visual Representation on Location Judgments in GPS-Like Displays

Mary Hegarty
University of California, Santa Barbara

Alinda Friedman
University of Alberta

Alexander P. Boone and Trevor J. Barrett
University of California, Santa Barbara

Two experiments revealed how nonexperts interpret visualizations of positional uncertainty on GPS-like displays and how the visual representation of uncertainty affects their judgments. Participants were shown maps with representations of their current location; locational uncertainty was visualized as either a circle (confidence interval) or a faded glyph (indicating the probability density function directly). When shown a single circle or faded glyph, participants assumed they were located at the center of the uncertain region. In a task that required combining 2 uncertain estimates of their location, the most common strategy—integration—was to take both estimates into account, with more weight given to the more certain estimate. Participants' strategies were not affected by how uncertainty was visualized, but visualization affected the consistency of responses, both within individuals and in relation to models of individual's preferred strategies. The results indicate that nonexperts have an intuitive understanding of uncertainty. Rather than arguing for a particular method of visualizing uncertainty, the data suggest that the best visualization method is task dependent.

Keywords: visualization, uncertainty, reasoning, strategies, integration

Imagine you are visiting a foreign city with a friend. You take the wrong bus by mistake and realize you are lost. Both you and your friend each have a smartphone with GPS software and consult your mobile map displays to determine where you are. However, the two GPS displays show you at two somewhat different locations. They also differ in the size of the circular region or "blue dot" indicating the uncertainty of the estimates of your current location. How do you and your friend figure out where you really are?

This type of judgment, which involves combining information from two uncertain estimates, is central to many real-world decisions. When meteorologists predict the path of a hurricane or the temperature for tomorrow, they combine predictions from different atmospheric models, each of which has some uncertainty. Similarly, when political pundits predict the outcome of an election or investment managers advise their clients on which stock to buy or sell, they are typically basing their judgments on uncertain predictions from different sources. In all of these examples, the models may make different predictions, which is one aspect of uncertainty; in addition, there may be different amounts of uncertainty (e.g., expressed by confidence intervals) about each prediction. Although these examples involve experts who deal with uncertain predictions on an everyday basis, in this research we examine how nonexperts reason about uncertainty.

One question addressed by the present research is whether nonexperts have an intuitive understanding of uncertainty. It is well-known that people are not good at reasoning under uncertainty (Tversky & Kahneman, 1974) and in the literature on communication of medical and environmental risk, there has been much debate about whether and how to present uncertainty to the public (e.g., Johnson & Slovic, 1995, 1998; Politi, Han, & Col, 2007). Many scientists endorse a deficit model which claims that the public cannot understand uncertainty, so that they advocate presenting only deterministic predictions to nonexperts (Frewer et al., 2003). However recent research has documented that people often reason better with than without uncertainty (Joslyn & LeClerc, 2013). That is, if the comparison of reasoning with uncertainty is against reasoning without uncertainty, rather than against rational choice models, research argues for including measures of uncertainty when presenting data to the public (Joslyn & LeClerc, 2012). More generally, researchers have argued that we need more basic information on the process of reasoning with visualizations of uncertainty (Kinkeldey, MacEachren, Riveiro, & Schiewe, 2015).

This article was published Online First October 10, 2016.

Mary Hegarty, Department of Psychological & Brain Sciences, University of California, Santa Barbara; Alinda Friedman, Department of Psychology, University of Alberta; Alexander P. Boone and Trevor J. Barrett, Department of Psychological & Brain Sciences, University of California, Santa Barbara.

This research is based upon work supported by Grant 1212577 to Mary Hegarty by the National Science Foundation and a grant to Alinda Friedman by the Natural Sciences and Engineering Council of Canada. The authors thank Sarah Creem-Regehr for comments on an earlier draft of the manuscript.

Correspondence concerning this article should be addressed to Mary Hegarty, Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. E-mail: hegarty@psych.ucsb.edu

Another question addressed by this research is how to best present information about uncertainty. Here we focus on graphical representations of uncertainty, although uncertainty is often conveyed verbally (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986) or numerically (Savelli & Joslyn, 2013). A common method of presenting information about uncertainty is to use error bars to show confidence intervals, but error bars are often misunderstood, even by researchers who use them in interpreting their own data (Belia, Fidler, Williams, & Cumming, 2005). One misconception is the assumption that the estimate is equally likely to be anywhere within the error bars and not at all likely to be outside the error bars (e.g., Correll & Gleicher, 2014; Zwick, Zapata-Rivera, & Hegarty, 2014). Because of misconceptions in interpreting error bars, researchers have advocated alternative forms of uncertainty visualizations, including violin plots and faded representations that show graded probability of estimates with more distance from the mean (Correll & Gleicher, 2014; Cumming, 2007).

Common visualizations of positional uncertainty in two-dimensional space generalize the idea of error bars to two dimensions. The "blue dot" on smartphone displays is a case in point. Typical GPS-based map displays, such as Google maps (Google, 2007), represent positional uncertainty using a transparent blue circle that surrounds the "you-are-here" marker. The blue circle can be interpreted as the 95% confidence interval for the estimate of one's current location, assuming a bivariate Gaussian distribution.[1] Another example of a two-dimensional visualization is the *cone of uncertainty*, used to convey predictions of the path of a hurricane over the coming days. Although the cone of uncertainty shows the 66% confidence interval of the predicted path of the hurricane, people often misinterpret this visualization; for example, they assume there is little or no likelihood of the hurricane moving outside the area of the cone (Broad, Leiserowitz, Weinkle, & Steketee, 2007; Cox, House, & Lindell, 2013; Ruginski et al., 2016). Because of these misconceptions, researchers have sought methods of identifying better ways of visualizing uncertainty. One approach, adopted by MacEachren et al. (2012) is to examine the intuitiveness of different visual encodings of uncertainty, with the assumption that visual encodings that convey a sense of uncertainty are also more effective for reasoning about uncertainty. In studies in which participants were asked to judge the appropriateness of different symbols for representing uncertainty and compare the amount of uncertainty indicated in different map regions, the conclusion was that fuzziness and graded point size were the most intuitive visual encodings of uncertainty.

Here, we examine how people reason about uncertainty in the task of judging one's location based on a GPS-like display. We chose this task because it is an everyday situation in which people view visualizations of uncertainty. Because research has indicated that confidence intervals are not well understood (Belia et al., Broad et al., 2007; Correll & Gleicher, 2014) we contrasted the interpretation of the familiar blue circle, where the border indicates a confidence interval, to a faded glyph (compound symbol) that displayed the complete Gaussian distribution of possible locations (see Figure 1 for examples of our stimuli). The blue-circle display was designed to be generic (rather than the specific display used on Google Maps, e.g.) so as not to bias responses of participants who used different displays on their smart phones. The faded glyph provided more information about the likelihood of one's position at different locations in space, did not show an arbitrary boundary (which has been misinterpreted in other visualizations of uncertainty), and incorporated the visual variable of fuzziness, which has been found to give people an intuitive sense of uncertainty (MacEachren et al., 2012).

In the main experimental task, people were presented with images showing either two circles or two faded glyphs on a map. Participants were instructed that they and a friend each had a cell phone with different GPS software, giving different indications of their current location on one device. The task was to indicate where they thought they actually were by clicking that location with the computer mouse. Participants were further told that the larger the glyph, the more uncertain was the indicated location estimate. In Experiment 1, glyph type was manipulated between-subjects; in Experiment 2, it was manipulated within-subjects. In each experiment, we also included a secondary task (see Figure 2) in which participants were shown just one glyph (circle or fade) on a map indicating their current location and had to click on the map to indicate their best guess of their current location. In the two-glyph task, there are two estimates of position, and each estimate carries its own amount of uncertainty, indicated by its size. This task is analogous to estimating a population mean from independent sample means, each with its own standard error.

Research Questions

Our first goal was to examine, at a basic level, how people interpret displays of positional uncertainty. For example, in the case of a single blue circle, do people assume that they are most likely to be at the center of the region (interpreting it as a confidence interval of a Gaussian-type distribution) or do they believe that they are equally likely to be anywhere inside the circle? Does the way in which uncertainty is visualized affect their judgment, and, if so, how? For example, the faded glyph makes the center more salient and visually indicates that positions near the center of the glyph are more likely so that people might be more likely to click the center when they see this glyph.

We were most interested in how people would judge their location when two circles or fades were shown. There are several strategies that might be used to estimate where one is, given two different distributions. The first is to integrate the information represented by the different distributions. From a Bayesian perspective, this strategy involves computing a weight for each glyph size using Equations 1 and 2 and then multiplying the center x value for each glyph by these weights, as in Equation 3. This method takes account of uncertainty by giving more relative weight to the glyph with more certainty (in this case, the smaller glyph; see Cheng Shettleworth, Huttenlocher, & Rieser, 2007; Friedman, Ludvig, Legge, & Vuong, 2013).

$$W_{Small} = \frac{\sigma^2_{Large}}{(\sigma^2_{Small} + \sigma^2_{Large})} \tag{1}$$

$$W_{Large} = \frac{\sigma^2_{Small}}{(\sigma^2_{Small} + \sigma^2_{Large})} \tag{2}$$

$$\textit{Estimated X Location} = W_{Small} * \mu_{Small} + W_{Large} * \mu_{Large} \tag{3}$$

[1] Ed Parsons, personal communication, December 9, 2014.

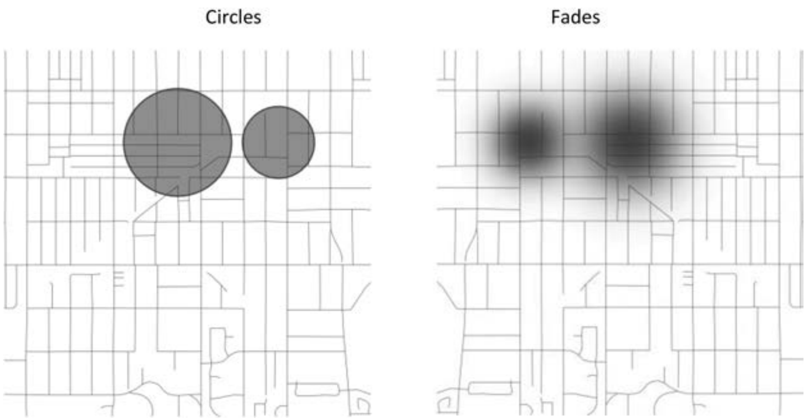

Figure 1. Example of stimuli in the two-glyph task. In each case the larger glyph is from a distribution with a standard deviation of 80 and the smaller glyph is from a distribution with a standard deviation of 60.

It should be noted that because the glyphs never overlapped in our stimuli, from a purely Bayesian perspective the information from the glyphs should not be integrated because in this circumstance it is more likely that the circles represent different mean locations that are far from each other instead of two representations of the same location (Cheng et al., 2007). However, participants were also instructed to assume that they and their friend were in the same location so they should take this information into account; pilot data indicated that some participants did integrate information from the two distributions. Thus, we examined the data with respect to a Bayesian integration strategy.

A second strategy that might be used to estimate where one is located, given two uncertainty distributions, is to select the center of the smaller glyph. This strategy takes account of uncertainty information by essentially ignoring the less certain estimate altogether and giving all the weight to the smaller glyph.

A third strategy that might be used to estimate where one is located is to ignore uncertainty information altogether. One way this can be done is by selecting a location in the middle between the edges of the glyphs.

In Experiment 1, we investigated what strategies are used spontaneously by undergraduate students with little or no formal training in statistics, but with some familiarity with smartphone apps. Half of the participants were presented with circular glyphs showing the confidence interval of the location estimate, whereas the others were presented with faded glyphs showing the probability density function of the bivariate Gaussian distribution of locations. In addition to documenting strategies, we investigated whether the way uncertainty is visualized (as a circle or fade) affects participants' strategies and whether it affects the consistency of their responses, given a strategy.

Experiment 1: Between-Subjects

Method

Participants. Participants were 88 students in an introductory psychology course at the University of California, Santa Barbara (UCSB; M age = 18.64, range = 18–27) who participated in return for college credit. Participants were randomly assigned to view either the circle glyphs (n = 44) or the fade glyphs (n = 44). Two participants were not included in the data analysis because they were color blind, and one did not follow instructions, leaving 85 participants (37 male, 47 female, 1 declined to state) in the final sample. All but 2 reported that they owned a smartphone, and all but 3 reported that they used a navigation app on their cellphone.

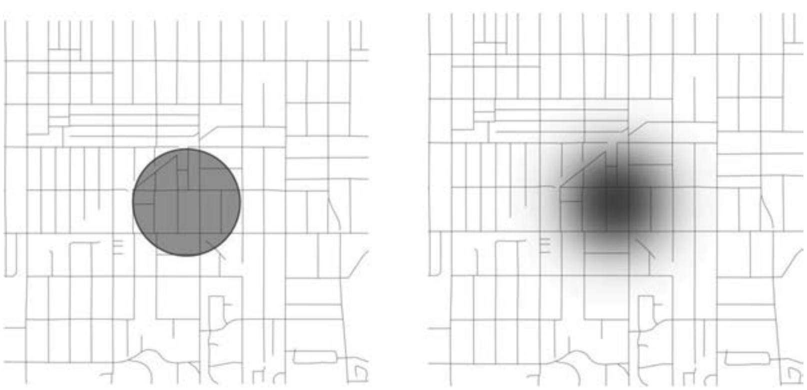

Figure 2. Example of stimuli in the one-glyph task.

A large minority of the students ($n = 36$) had taken one or more classes in statistics or probability.

Materials

On each trial of the main task, participants were shown a map display with one of two types of representations of their locations (glyphs; see Figure 1). Each glyph was a circular image representing a location with uncertainty. Circles had a clearly drawn boundary, indicating the 95% confidence interval of a Gaussian distribution of locations. The diameter was thus ± 2 standard deviations around the mean of the distribution. Fades directly represented the Gaussian probability density function with decreasing density toward the periphery.

The glyphs were shown in one of five different sizes (based on a standard deviation of 40, 50, 60, 70 or 80 pixels) on an 800 $\times$ 800 pixel display. The glyph with a standard deviation of 60 was paired with each of the other sizes to create five different relative size pairs; 40–60, 50–60, 60–60, 70–60, and 80–60. The two circles were always placed side by side so that the *Y* coordinate of the center was the same for both glyphs and the space between the edges of the two circles was 20 pixels. The fades had no edges but their centers were the same distance apart as the five circle pairs and they were matched in apparent size to the circles on the basis of a preliminary psychophysics study.[2] In the canonical set of stimuli, the glyph with the variable size (40, 50, 60, 70, 80) was placed on the left, and the standard (60) was placed on the right. This canonical set of stimuli was flipped (placing the standard on the left in all flipped stimuli), and both the canonical and flipped sets were placed in four different locations of the map, so that there were 40 trials (5 Relative Sizes $\times$ 2 Flips $\times$ 4 Map Locations) for each glyph condition (circles vs. fades).

We also included a one-glyph (see Figure 2) task to examine how participants interpreted the circle and fade glyphs without the need to combine information from two estimates. Participants were shown a map with a single glyph (circle or fade), with the glyph size based on a standard deviation of 40, 50, 60, 70, or 80 pixels. Each glyph was shown in each of the four quadrants of the map for a total of 20 trials, and we also created mirror images (flips) of each map for a total of 40 trials (5 Sizes $\times$ 4 Locations $\times$ 2 Flips) for each glyph.

Participants were administered an online posttask questionnaire that first presented images from the Ishihara Compatible Pseudoisochromatic Plate (PIPIC) Color Vision test (Waggoner, 2005) to test for color blindness. The questionnaire was used to collect demographic information, information about participants' educational background, and the strategies that they used for the two-glyph and one-glyph tasks.

Procedure

Up to 3 participants at a time took part in a windowless laboratory room. After providing informed consent, they were given instructions for the main task (the two-glyph task), which were read aloud by the experimenter while the participant followed the same instructions on the computer display. The participants were told that the blue regions showed location readings from two smartphones, with the size of the blue region representing the amount of uncertainty (larger blue regions indicating more uncertainty). Participants were instructed to imagine that they and a friend were in the same location and each had a cell phone with different GPS software, giving potentially different indications of where they were on one device. The task was to use the computer mouse to click on the map location indicating where they thought they were actually located.

On each trial, they were shown a blank screen for 1 s, followed by the two-glyph image. Trials were self-paced, ending when the participant made a mouse click on the display. Participants were given four practice trials without feedback and were given an opportunity to ask questions about their task before completing the 40 experimental trials, which were presented in a different random order to each participant.

Then participants were given instructions for the one-glyph task, which were read aloud while the participant read along on the computer display. They were told that the images they would see in the next part of the experiment showed location readings from a single smartphone and that again their task would be to indicate their best estimate of their current location by clicking with a mouse on the relevant location of the display. They were reminded that larger glyphs indicated more uncertainty. They were given four practice trials without feedback before completing the 40 experimental trials in the same glyph condition (circle or fade) that they had been assigned in the two glyph task. Finally, participants completed the online questionnaire, before being debriefed, thanked, and dismissed.

Results

Two-Glyph Task

We took the following approach to analyzing the data for the main (two-glyph) task:

First, we examined the verbal reports together with scatterplots of each participant's individual responses across all trials to identify the strategies that were used to respond and to categorize each participant accordingly. The first two authors did this independently for 50 of the 85 participants, and their interrater reliability was 100%. The other trials were coded by one of these authors.

Second, we examined the participants' consistency of responding (variable error) by computing the standard deviations across

[2] When we initially created the circle and fade glyphs based on circular Gaussian distributions with the same standard deviation, it was evident that the fades appeared smaller than the circles. A psychophysics experiment was conducted in order to match the apparent size of the fade and circle glyphs. Participants viewed trials that displayed circle and fade glyphs side by side. Circle glyphs (*SD* of 25, 45, 65, 85, and 105) were paired with fade glyphs that varied in size with intervals of 5. For example, the circle glyph with a standard deviation of 65 was paired with fade glyphs that ranged in standard deviation from 40 to 110 with intervals at every 5 pixels. Four repetitions (half flipped horizontally) of each trial produced 304 total trials that were presented randomly. Thirty-one participants completed the task by indicating whether the left or right glyph was larger. Results indicated that, in general, participants crossed over to selecting the fade glyphs as larger when the fade glyphs standard deviation was increased by 20, 25, 22, 23, and 22 when matched with circle glyphs of 25, 45, 65, 85, 105 standard deviations, respectively. On the basis of the similar values across the five sized circles, the decision was made to take the average and increase the standard deviation of the fade glyphs uniformly by 22.5 pixels to match the apparent visual size of the circle glyphs.

each participant's responses, separately for each glyph and size pair. This measure indicates how consistently participants were able to choose their desired locations as a function of conditions, ignoring strategy.

Third, we used the strategies that were identified from the verbal reports and scatterplots to create three models with predicted values for each of the size pairs. We then compared the estimated with the predicted values for each participant and model. This process is described further in the following text.

Strategy categorization. As noted, we first examined participants' verbal reports to identify the strategies they used for the location estimates. We found five strategies that were clearly stated (see the example statements of each strategy in Table 1). Three of these indicated a more or less correct (valid) understanding of the instructions (pick the smaller, pick the middle, integrate toward the smaller) and two indicated an incorrect understanding (pick the larger, integrate toward the larger).

We next determined the predicted *x*-axis locations for each size pair for each of the three valid strategies. Two of these (pick smaller and pick middle) are shown by the gray lines in Figure 3. We did not consider the *y*-axis values in the models because the stimuli in each pair did not vary in the *y* dimension and comparison of the mean *y* values across conditions did not show a difference greater than 4 pixels in the 800×800 displays.

For the pick the smaller strategy, the predicted values were the center *x*-axis value of the smaller glyph in each size pair when sizes were different and the center between glyphs when the size was the same; for the pick the middle strategy the center *x* value between the putative edges of the two glyphs in each size pair was used as the predicted value; and for the integration strategy Bayesian weights were computed using Equations 1 and 2 and the predicted values were computed using Equation 3. We then computed the root mean square error (RMSE) for each participant using the mean estimated and predicted values from each of the three valid strategies. Finally, we categorized each participant based on the minimum value of his or her RMSEs.

We also examined the scatterplots of each participant's estimates against the strategy that was selected algorithmically. Some of the scatterplots showed that the strategy selected as the minimum RMSE was either (1) very close in absolute size to another strategy (e.g., by 1 or 2 pixels)—in other words, the algorithm was not distinguishing between the two, as when the participant was using multiple strategies, or (2) the participant was responding either invalidly (e.g., pick larger) or roughly randomly (the random responders were categorized as "other"). For both the invalid and random responders, the algorithm selected "pick center"; mathematically, this was the closest possibility. In these cases, the first two authors overrode the algorithmic selection. Thus, we did not further analyze data from a total of 15 (17.6%) participants for one of three reasons that were, coincidentally, evenly distributed: The scatterplots showed that their responses were either consistent with more than one strategy (5.9%) or their strategy statements and/or plots indicated an incorrect understanding of the instructions (i.e., "pick the larger" and "integrate toward the larger"; 5.9%), or we could not categorize their strategies from their verbal reports or scatterplots ("other"; 5.9%). Table 2 shows the frequency and percentage of participants who fell into each valid and invalid category, and the average *x* values for participants coded as using each valid strategy are presented in Figure 3.

It is notable that the most frequent strategy—one used by half the participants—was to find a location between the two glyphs but closer to the smaller one; in other words, to integrate. Neither glyph (circle vs. fade), $\chi^2(2, N = 70) = 2.25$, $p = .33$, nor statistics training (i.e., whether participants had taken a statistics class), $\chi^2(2, N = 70) = 2.19$, $p = .33$, significantly affected students' adoption of the three primary strategies.

Variable error. We examined the consistency of responses, within individuals, by computing the standard deviations of the responses for the 70 participants who used one of the three valid strategies as a function of glyph and pair size. We analyzed these in a 2 (glyph) $\times$ 5 (size) mixed-design analysis of variance (ANOVA). There was a main effect of size, $F(4, 272) = 5.38$, $p < .001$, $\eta_p^2 = .07$, and an interaction between glyph and size, $F(4, 272) = 6.10$, $p < .001$, $\eta_p^2 = .08$. The means for each size pair for the circle stimuli, in order of the glyphs' standard deviations were 14.8, 17.9, 15.7, 21.6, and 18.7 pixels, respectively; and for the fade stimuli the means were 12.1, 18.0, 36.0, 20.3, and 17.6 pixels.

It became obvious when we looked at the individual scatterplots that participants sometimes had a great deal of variability among the responses they made to the same-size pairs (60–60), especially for the fades. For the pick smaller and integrate strategies, participants had to select the smaller glyph or choose the center between glyphs when the size was the same. For 60–60 pairs, it is relatively easy to see that they are the same size with the circle glyphs. In contrast, for same-size fade glyphs, the scatter plots indicated that participants sometimes misperceived one of the glyphs as being smaller and responded incorrectly on that basis. Thus, we reanalyzed the data without the 60–60 pairs. Now, only the main effect of size was reliable, $F(3, 204) = 6.75$, $p < .001$, $\eta_p^2 = .09$. In

Table 1
Examples of Strategy Reports

Strategy	Example(s) of verbal report(s) stating that strategy
Pick the smaller	"I decided by the smaller uncertainty circle, but if they were equal, I just chose the center." "I went with the smaller circle most of the time, if they were the same size I chose an area between both circles."
Pick the middle	"I picked a spot evenly between the two displays." ". . . in between the middle of the two blue dots."
Integrate	"I assumed I would be closer to the smaller circle where the amount closer was proportional to the difference in circle sizes." "I chose the point in between the two, and if one of the circles was larger than the other I shifted the center point between the two closer to the smaller circle."
Pick the larger	"I clicked on the sphere that was bigger."
Integrate larger	"I clicked towards the side of the larger display because I thought it would even out the distance."

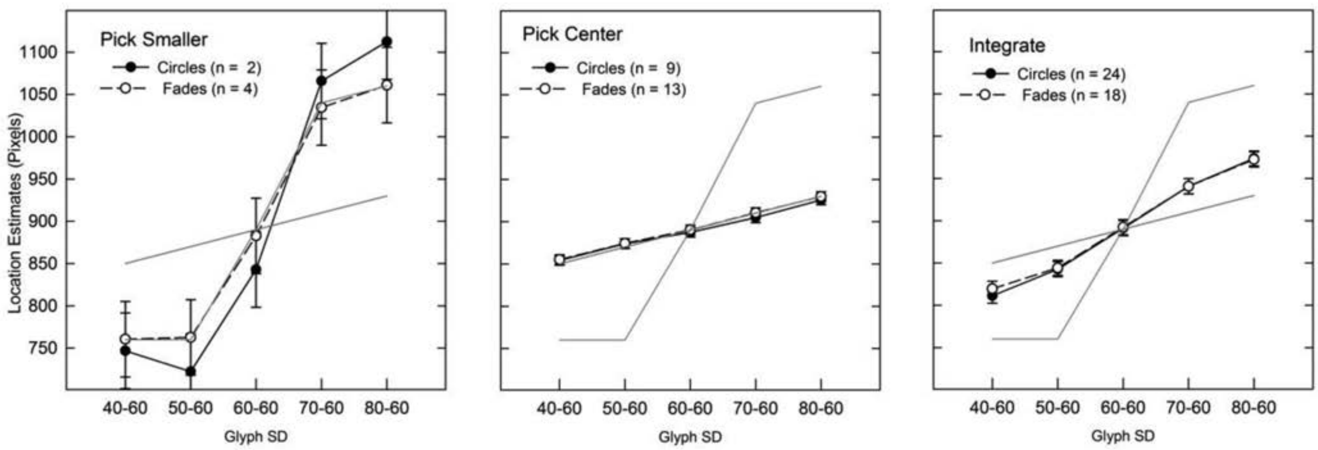

Figure 3. Mean *x* values of the estimates for students coded as using the three primary strategies in Experiment 1. The gray lines indicate the predicted locations for pick center and pick smaller. Error bars indicate 95% confidence interval computed with the mean squared error from the interaction term of each analysis of variance.

general, participants were more variable in their responses to all but the smallest glyph pair.

RMSE as a function of glyph and size. We analyzed the variability around the predicted values for participants who used each of the three main strategies (pick smaller, pick center, integrate) in separate Glyph × Size Pair ANOVAs, both with and without the 60–60 pairs. As is shown in Figure 4, the results were mixed.

The effect of glyph was not significant for any of the strategies, either with or without the 60–60 pairs. The effect of size pair was significant for all three strategies, $F(4, 16) = 19.24$, $p < .0001$, $\eta_p^2 = .83$, for pick smaller, $F(4, 80) = 5.24$, $p < .001$, $\eta_p^2 = .21$, for pick center, and $F(4, 160) = 11.18$, $p < .0001$, $\eta_p^2 = .22$, for integrate, and remained significant when the 60–60 pairs were removed. Finally, the interaction was significant for pick smaller, $F(4, 16) = 5.70$, $p < .005$, $\eta_p^2 = .59$, and integrate, $F(4, 160) = 11.33$, $p < .0001$, $\eta_p^2 = .22$, but not for pick center, and neither remained reliable without the 60–60 pairs.

The interactions between glyph and size pair for the pick smaller and integrate strategies are probably because it is easier to perceive that the two glyphs are the same size in the 60–60 case for circles than for fades. Participants were more likely to erroneously judge that one of the glyphs was smaller, when they were actually the same size, in the case of fades than they were in the case of circles. Hence, the interactions disappear with the removal of the 60–60 pairs. In the case of the pick smaller strategy, the interaction also reflects that the centers of the glyphs are more salient because of increased shading for fades but not for circles. Thus, it is relatively easy to pick the center of the smaller glyph, as required by this strategy.

Table 2
Frequency (Percentage) of Participants Coded as Using Each Strategy in Experiments 1 and 2

Strategy	Experiment 1	Experiment 2	Overall
Valid			
Pick smaller	6 (7.1%)	4 (7.7%)	10 (7.3%)
Pick middle	22 (25.9%)	10 (19.2%)	32 (23.4%)
Integrate	42 (49.4%)	22 (42.3%)	64 (46.7%)
Invalid			
Pick larger	1 (1.2%)	1 (1.9%)	2 (1.5%)
Integrate larger	4 (4.7%)	1 (1.9%)	5 (3.7%)
Multiple	5 (5.9%)	8 (15.4%)	13 (9.5%)
Other	5 (5.9%)	6 (11.5%)	11 (8.0%)
Total	85 (100%)	52 (100%)	

One-Glyph Task

The first two authors independently coded the verbal reports of strategies for the one-glyph task and had 100% agreement. The majority of participants (67.2%) indicated that they responded in the center of the circle or the darkest location of the fade. Others indicated that they responded in the center but placed themselves on a street (4.2%), whereas some participants (18.6%) indicated that they placed themselves closer to the center for small circles than for large circles. Finally, 10% of participants' statements were not interpretable.

We examined both variable error and RMSE. All responses were compared with the center location to compute the RMSEs. For variable error, there was a nonsignificant trend, $F(1, 68) = 3.29$, $p < .10$, $\eta_p^2 = .05$, of more error for circles ($M = 18.5$) than for fades ($M = 12.2$). There was a main effect of size, $F(4, 272) = 26.02$, $p < .001$, $\eta_p^2 = .28$, and an interaction between glyph and size, $F(4, 272) = 3.76$, $p < .0001$, $\eta_p^2 = .05$. Variability increased with size (means across glyphs were 8.2, 12.0, 14.9, 18.7, and 22.9, respectively, from smallest to largest) and more so for the circles than for the fades.

A similar pattern was found for the RMSE data. Although the glyph main effect was not significant ($p < .15$), circles produced more variability against the model than did fades (19.0 pixels vs. 13.6 pixels, respectively). Further, there was a main effect of size, $F(4, 272) = 27.76$, $p < .0001$, $\eta_p^2 = .29$, and the glyph by size interaction approached significance, $F(4, 272) = 3.26$, $p = .05$, $\eta_p^2 = .05$. The means for the different sizes across glyphs were 8.9,

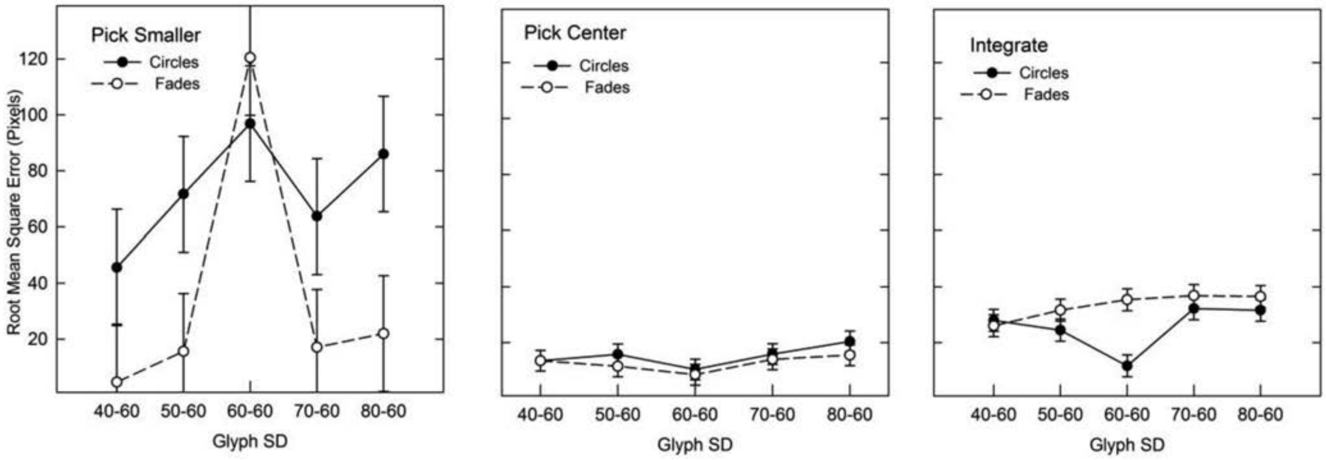

Figure 4. Root mean square error (RMSE) from the predicted locations for each of the three main strategies in Experiment 1. Error bars indicate 95% confidence interval computed with the mean squared error from the interaction term of each analysis of variance.

12.8, 15.6, 20.3, and 23.8 pixels, respectively, from smallest to largest. Figure 5 shows the interaction.

Notably, the means for both standard deviations and RMSE measures are similar, which reflects that because most participants were selecting the center on each trial, both measures are capturing the same variability; variability of the responses around the participant's mean response (self-consistency) versus variability of the responses around the actual center *x*-axis value. The increase in standard deviation and RMSE error with size likely reflects the increased difficulty in judging the center of the blue region as it gets larger, and the fact that some subjects reported that they deliberately placed their estimates farther from the center when the glyph was larger. The greater effect of size for the circle glyphs compared to the fades probably reflects greater difficulty in estimating the center of the circles, because circles are uniformly shaded, whereas the centers of the fades have increased shading. In contrast, the greater effect of size for the circles does not reflect differences in perceived sizes of the circle and fades, as these were matched.

Discussion

Experiment 1 revealed important differences in strategies between individuals in how they combine different uncertain estimates of position. Most participants adopted a consistent strategy throughout, and the majority took uncertainty into account in their estimates. The most common strategy was some version of integration in which both location estimates were taken into account and more weight was given

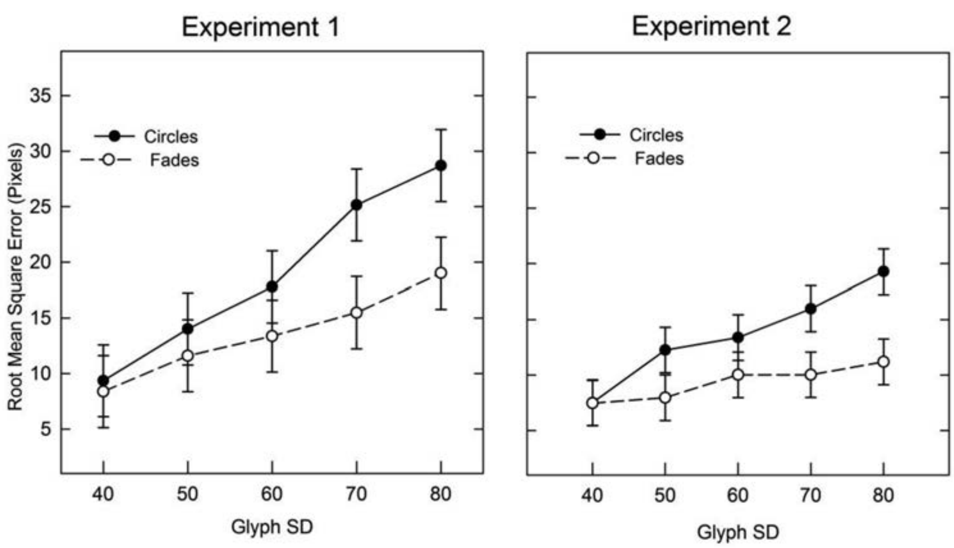

Figure 5. Root mean square error (RMSE) from the center of the circle for the one-glyph task in Experiments 1 and 2. Error bars indicate 95% confidence interval computed with the mean squared error from the interaction term of each analysis of variance.

to the more certain estimate. Although most people using the integration strategy were "qualitative integrators" in that they chose a position weighted toward the more certain estimate, remarkably, 4 participants were almost perfect Bayesians in that their estimates were almost perfectly fit by the Bayesian integration models (RMSE errors of less than 10 pixels from this model).

In both the one- and two-glyph tasks, participants' estimates were more variable for larger glyphs. In the one-glyph task, size interacted with glyph such that with increasing size, participants' estimates of the center were less variable (and closer to the actual center) for the fade glyphs than for the circle glyphs, likely because the center is darker than the surrounding regions of these glyphs, and was therefore more salient in the display. In contrast, there were no significant effects of glyph (circle or fade) in the two-glyph task.

Experiment 1 had low power to detect significant glyph effects, because glyph was manipulated between-subjects and because there were relatively few trials. We address these limitations in Experiment 2.

Experiment 2: Within-Subjects

In Experiment 2, we manipulated glyph within-subjects and included more experimental trials, to increase the power to detect any differences in interpretation of the two different methods of visualizing uncertainty. This design also enabled us to examine consistency of responding within individuals for the circle and fade glyphs.

Another goal of Experiment 2 was to further investigate possible effects of knowledge on performance of this task, especially the strategy adopted. Specifically, we recruited students from geography and psychology classes and examined the effects of geography knowledge and statistics knowledge on students' strategies.

Method

Participants. Participants were 26 students (11 male, 15 female; M age = 20.38 years, range = 18–25) in geography courses at UCSB and 26 students (7 male, 19 female; M age = 20.54, range = 18–29) in an introductory psychology course at the University of Alberta, who participated in return for college/course credit. All had normal or corrected-to-normal vision. In the UCSB sample, all but one student owned a smartphone and all but two students used a navigation app on their smartphone; 17 (65.4%) had taken a statistics course. In the Alberta sample, all students owned a smartphone, and all but two used the navigation app; 16 (61.5%) had taken a statistics course. The UCSB students had taken an average of 4.54 geography courses (SD = 3.83, range = 1–13); two of the Alberta students had taken one geography course, whereas the others had taken none.

Materials

The materials (tasks and trials) were the same as in Experiment 1 with two minor exceptions. First, for the UCSB sample we used the PIPIC (Waggoner, 2005) to measure color blindness, whereas for the Alberta sample, items from this scale were included in the questionnaire (as in Experiment 1). Second, in the questionnaire the UCSB participants were asked a single question about their strategies on the experimental tasks, whereas the University of Alberta participants were asked two questions: one about their strategy for the two-glyph task and the other about their strategy for the one-glyph task.

Procedure

Up to 3 participants at a time took part in the experiment. After giving informed consent, UCSB students were first given the PIPIC test of color blindness. Then they were given instructions for the main task (the two-glyph task) and completed four practice trials (two trials from each glyph type), and a total of 160 trials (two replications of the 40 circle and 40 fade trials) in a random order. Then they were given instructions for the one-glyph task, completed four practice trials for that task, and then completed two replications of the 40 circle and 40 fade trials for this task, followed by the questionnaire. Alberta students first performed the two-glyph task, then the one-glyph task, and then they were administered the online color blindness test (as in Experiment 1), followed by the questionnaire.

Results

Because there were no differences in the experimental procedure, except for the small difference in timing of the color blindness test and the questionnaire given to the two groups, and because preliminary analyses indicated no differences between participants at UCSB and at University of Alberta, these data were combined for analyses.

Two-Glyph Task

Strategy categorization. We used the same procedure as in Experiment 1 to categorize the participants' strategies, separately for each type of glyph. There were three cases in which the algorithm did not distinguish between two of the strategies, and in these cases two raters assigned the strategy on the basis of examination of the scatterplots. Table 2 shows the results of the strategy categorization and Figure 6 shows the mean x-axis values of the locations indicated by participants who were classified as using each strategy. The distribution of strategies for Experiment 2 was very similar to that for Experiment 1, and, again, our analyses focused on the participants who used one of the three valid strategies (pick the smaller, pick the center, or integrate). These students were all classified as using the same strategy for the circle and fade glyphs. The strategy that a participant adopted was not dependent on statistics training (whether students had taken a statistics course), $\chi^2(2, N = 36) = 2.66$, $p = .26$, geography training (whether students had taken one or more geography courses), $\chi^2(2, N = 36) = 1.61$, $p = .45$, or university attended (UCSB vs. University of Alberta), $\chi^2(2, N = 36) = 2.33$, $p = .31$.

Variable error. We examined the consistency of responses as a function of glyph and pair size by computing the standard deviations of the responses for each participant in each of these conditions and examining the results in an ANOVA. There was a main effect of glyph, $F(1, 35) = 59.05$, $p < .0001$, $\eta_p^2 = .63$, a main effect of size, $F(4, 140) = 6.07$, $p < .001$, $\eta_p^2 = .15$, and an interaction between the factors, $F(4, 140) = 3.21$, $p < .02$, $\eta_p^2 = .08$. All of these effects remained reliable when the 60–60 pairs

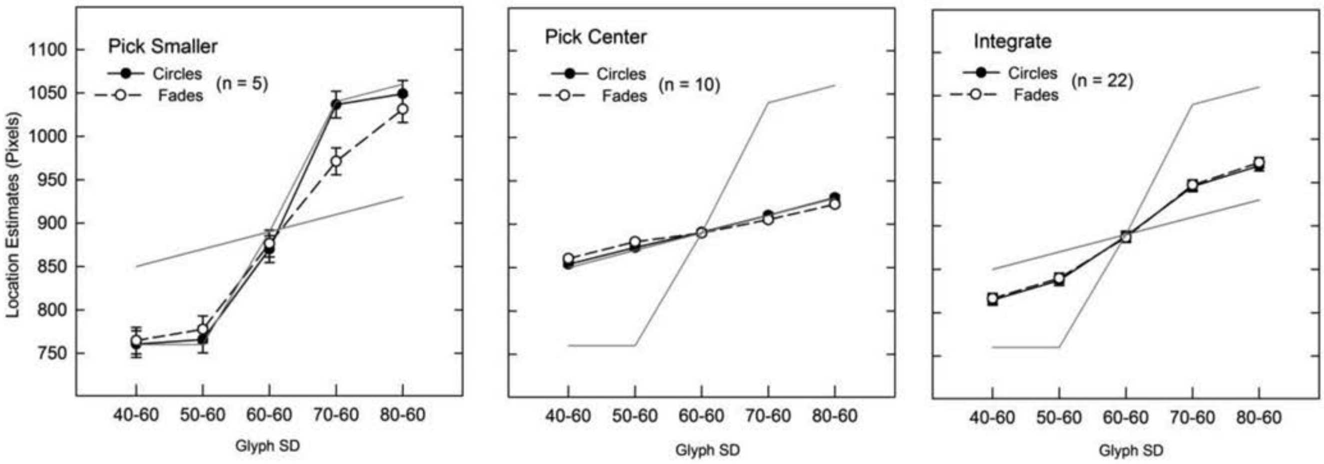

Figure 6. Mean *x* values of the estimates for students coded as using the three primary strategies in Experiment 2. The gray lines indicate the predicted locations for pick center and pick smaller. Error bars indicate 95% confidence interval computed with the mean squared error from the interaction term of each analysis of variance.

were removed from the analysis. Participants were less self-consistent in their responses to fades ($M = 22.7$) than they were to circles ($M = 13.1$). The means for the individual size pairs for circles were 9.9, 10.3, 20.5, 11.8, and 13.3; and for fades the means were 12.3, 18.9, 31.0, 27.8, and 23.6. The means for the fades were more variable than were those for the circles at the larger sizes.

RMSE as a function of glyph and size. We analyzed the RMSEs in a Glyph × Pair Size ANOVA for each of the three valid strategies. Descriptive statistics are presented in Figure 7. Importantly, the effect of glyph was significant in all three cases: for pick smaller, $F(1, 3) = 19.26$, $p < .03$, $\eta_p^2 = .87$, for pick center, $F(1, 9) = 12.99$, $p < .01$, $\eta_p^2 = .59$, and for integrate, $F(1, 21) = 7.37$, $p < .02$, $\eta_p^2 = .26$. Further, all of these effects remained reliable when analyzed without the 60–60 pairs. The means across all size pairs for circles and fades were 38.0 and 61.6 for pick smaller; 9.6 and 16.3 for pick center, and 21.3 and 29.2 for integrate.

There were also significant effects of size pair for each strategy; for pick smaller, $F(4, 12) = 38.18$, $p < .0001$, $\eta_p^2 = .93$; for pick center, $F(4, 36) = 6.46$, $p < .001$, $\eta_p^2 = .42$, and for integrate, $F(4, 84) = 16.57$, $p < .001$, $\eta_p^2 = .44$. The effects remained reliable when analyzed without the 60–60 pairs.

Finally, the interaction was significant only for the pick smaller, $F(4, 12) = 5.18$, $p < .02$, $\eta_p^2 = .63$, and integrate strategies, $F(4, 84) = 6.35$, $p < .001$, $\eta_p^2 = .23$. However, the effect was not significant for the integrate strategy without the 60–60 pairs, suggesting that the interaction in this case reflected misperception of one of the glyphs as being smaller in the case of the fades that

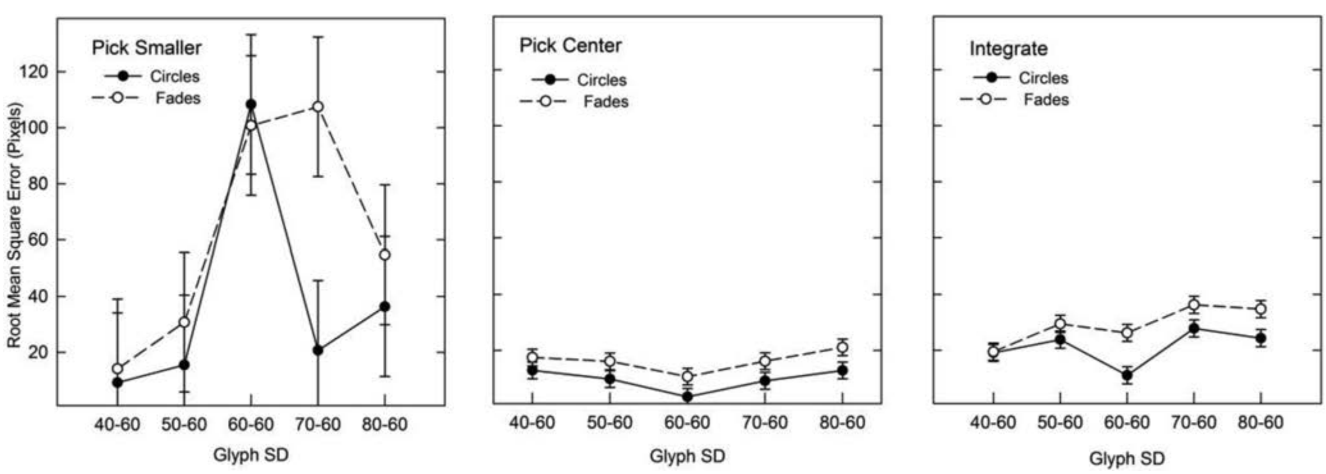

Figure 7. Root mean square error (RMSE) from the predicted locations for each of the three main strategies in Experiment 2. Error bars indicate 95% confidence interval computed with the mean squared error from the interaction term of each analysis of variance. Note that though the effects of glyph for "pick smaller" differ from Experiment 1, the number of participants choosing this strategy was very small overall (6 in Experiment 1; 5 in Experiment 2).

were in fact equal in size (60–60). In the case of the pick smaller strategy, it appears that participants might also have had difficulty in perceiving the relative size of the fades for the 70–60 case, sometimes misperceiving these glyphs as the same size. However, for the majority of participants the glyph effect was not mitigated by size.

One-Glyph Task

For the one-glyph task, we again examined variable error and RMSE across all participants using the glyphs' center x-axis values as the predicted values; the questionnaire data indicated that the majority of participants (19 of the 26 participants who were asked about their strategy for the single glyph trials) were again trying to select the center of the glyph on each trial for this task.

For variable error, there was a main effect of glyph, $F(1, 35) = 8.81$, $p < .01$, $\eta_p^2 = .20$. Within each participant, there was more variability among responses to the circles than to the fades; the means were 13.7 and 8.9. There was also a main effect of size, $F(4, 140) = 8.36$, $p < .0001$, $\eta_p^2 = .19$, and a Glyph × Size interaction, $F(1, 140) = 4.03$, $p < .003$, $\eta_p^2 = .10$.

As in Experiment 1, the results from the analysis of the RMSE data were similar to those for variable error. There was a main effect of glyph, $F(1, 35) = 8.10$, $p < .008$, $\eta_p^2 = .19$. The means were 13.7 and 9.3, respectively, for circles and fades. There was also a main effect of size, $F(4, 140) = 8.47$, $p < .0001$, $\eta_p^2 = .19$, and a Glyph × Size interaction, $F(1, 140) = 4.41$, $p < .003$, $\eta_p^2 = .11$. Figure 5 shows the interaction.

As in Experiment 1, the means for the one-glyph task were very similar across both measures, most likely because most participants were using the same strategy (i.e., pick the middle of the glyph) and the effects were likely due to greater difficulty in judging the center for larger glyphs, especially when the center was not marked (in the circle glyphs).

Discussion

The distribution of strategies in Experiment 2 was very similar to that of Experiment 1, and the within-subjects design of Experiment 2 revealed that strategy is a function of person rather than being influenced by glyph. Neither statistics nor geography knowledge significantly affected strategy, indicating that although we have characterized the strategies adopted for this task, the present research does not reveal any knowledge differences that are correlated with strategy choice.

With the increased statistical power in this experiment, there were reliable effects of how uncertainty was visualized (as a circle or fade) in both the two-glyph and one-glyph tasks. In the case of the two-glyph task, glyphs influenced within-subject response variability in general and for each strategy in particular, with fade glyphs being more variable in each case and despite the fact that they were created to be perceptually the same size as the circles. In contrast, in the one-glyph task, within-subject response variability, and variability with respect to the pick the middle strategy were greater for circles than fades. Assuming that less variable performance is superior, these results indicate that the more effective visualization for locational judgments depends on the task.

General Discussion

In two experiments, we examined how people combine two uncertain estimates of their position in two-dimensional space. Different people adopted different strategies, but once they adopted a strategy, most were consistent in using that strategy throughout the trials. The strategy adopted was not affected by how uncertainty was visualized (as a circle or a fade) or by knowledge of either statistics or geography. Although the visualization of uncertainty did not affect participants' strategies, it did affect the self-consistency of participants' responses, given a strategy. It also affected the RMSE computed against the models' predicted values. Moreover, the effect of visualization (glyph) depended on the task (see Figures 5 and 7). For the two-glyph task, both variable error and RMSE were larger for the fade glyph, but for the one-glyph task, both measures were larger for the circle glyph, despite the fact that the circles and fades were matched in apparent size. We first discuss how nonexperts perform this task in general, and then discuss the effects of how uncertainty was visualized.

How Nonexperts Combine Uncertain Estimates

A striking result of our experiments is that the majority of individuals took uncertainty into account in their location estimates, despite the fact that they were given minimal instructions about the display conventions (the only information they were given was that larger circles or fades indicted more uncertainty). Across the two experiments, only a minority of participants using valid strategies (23.4% see Table 2) picked the middle location (between the two distributions), which is the pattern of responding that does not take uncertainty into account. The result that most participants took uncertainty into account indicates good intuitive understanding and argues that we should not underestimate the ability of the general public to reason with uncertain information (cf. Joslyn & LeClerc, 2013).

The most common strategy, adopted by almost half of the participants (see Table 2), was to integrate the information represented by the different distributions. This result indicates that people generally assumed that both glyphs contained valid information. In contrast only a small minority (7.3% of participants) consistently chose the center of the smaller glyph, ignoring the less certain estimate. Although the most common strategy was to integrate, it should be noted that most individuals who used this strategy were qualitative integrators. That is, they chose a location closer to the more certain estimate but did not typically respond as if they used the correct weights indicated by a Bayesian analysis. However, a few participants (4 in Experiment 1 and 1 in Experiment 2) made responses that were very well fit by the Bayesian model, indicating that they were intuitive Bayesians.

Although participants in general had good intuitive understanding of uncertainty, it should also be noted that the responses of a minority of participants revealed misconceptions. Specifically, in the two-glyph task, a small number of participants chose the center of the larger (less certain) glyph as their most likely location, or integrated toward the larger glyph, although they were informed that larger glyphs indicated more uncertainty. These responses might reflect inattention to the instructions or failure to inhibit a more general "bigger is better" heuristic.

Another notable misconception was identified in the one-glyph task. Although the majority of participants correctly assumed that they were most likely to be in the center of the blue region in this task, a sizable minority responded that they deliberately placed themselves farther from the center for larger glyphs. That is, the increasing variability in responses for larger glyphs was not merely due to more difficulty in estimating the center of larger glyphs but was due to a deliberate strategy to bias estimates more toward the periphery of the distribution when the glyph was larger.

Effects of How Uncertainty is Visualized

The second goal of this research was to examine how the visualization of uncertainty (as a circle showing a confidence interval or a faded glyph showing the actual probability density function) affects people's location judgments. The first important result was that the visualization of uncertainty did not affect the strategy adopted. There were no significant effects of glyph on strategy when glyph was manipulated either between participants (in Experiment 1) or within participants (in Experiment 2). Moreover, in Experiment 2, almost all participants adopted the same strategy for fade and circle trials.

The second important result is that type of visualization did affect the consistency of responding, but its effects were task dependent. Specifically responding was less variable in the one-glyph task for the fades and generally less variable in the two-glyph task for the circles.

In the one-glyph task, the dominant response was to assume that one is most likely to be in the center of the blue region. Less variable responding in this task for the fade glyphs is likely due to the fact that in the fade visualizations, the center of the blue region was easier to judge because it was darker (the glyph was more opaque in this region). In contrast, the circles were of uniform opacity. Indeed, because the fade glyphs visualize the actual probability density function, they directly represent the fact that in a bivariate Gaussian distribution, the most likely location is at the center of the glyph. In contrast, the centers of the circle glyphs were not visibly marked and needed to be estimated, perhaps by mentally superimposing crosshairs on the circle (Huttenlocher, Hedges, & Duncan, 1991). This estimation process would likely result in more variability in responses with the circles compared to the fades.

Although fades have clearer centers than do circles, they have vaguer boundaries. This results in more variable responding in the two-glyph task. For example, executing the integration strategy involves the following cognitive processes: (a) judging which glyph is smaller; (b) judging the centers of the two glyphs; (c) judging the boundaries of the two glyphs; (d) choosing the parameters by which to weight the two estimates, presumably on the basis of the judged relative sizes of the two glyphs; and (e) placing the estimated location between the two boundaries but closer (by the weighting factors) to the smaller glyph. Although the centers are darker in the fade glyphs, facilitating the second step above, the other steps are facilitated by the circle glyphs, for which there are clearly marked boundaries. These boundaries make it easier to see which circle is larger, judge the relative size of the two distributions, and judge where to place the location estimate between the two distributions, whether this is weighted toward the smaller glyph (for the integrate strategy) or midway between the two glyphs (for the pick the center strategy).

Implications and Future Directions

This research informs the debate about whether information about uncertainty in data should be communicated to the public (Frewer et al., 2003; Johnson & Slovic, 1995, 1998; Politi et al., 2007). Although there are many aspects of this debate, one critical issue is whether nonexperts can understand and reason effectively with displays of data that include information about uncertainty. Our research indicates relatively good intuitive understanding of uncertainty visualizations, despite the fact that participants received minimal information on how to interpret the displays. Although much psychological research has been concerned with revealing errors in reasoning under uncertainty (Tversky & Kahneman, 1974) or interpretation of visualizations of uncertainty (Belia et al., 2005; Correll & Gleicher, 2014; Zwick et al., 2014), the present research is more hopeful and suggests that people can be good intuitive statisticians in certain circumstances. With more explanation of what the displays show, for example, indicating the exact confidence interval shown (e.g., 66% vs. 95%), we might expect people to be even more facile in making judgments under uncertainty. Moreover, our research has identified some misconceptions about these displays that could be directly addressed in instructions that accompany displays of uncertain predictions. Examining the effects of additional instructions on the comprehension of uncertainty visualizations is an important goal for future research.

Our research does not argue for the superiority of one method of visualizing uncertainty in general. The fade glyphs provided more information about the likelihood of one's position in space (Correll & Gleicher, 2014) and had fuzzy boundaries, which give an intuitive sense of uncertainty (MacEachren et al., 2012), whereas the circle glyphs are more familiar in this context. However, neither glyph was superior for all judgments. Consistent with much research on the design of graphics (Hegarty, 2011; Nadav-Greenberg, Joslyn, & Taing, 2008), our results suggest that the best way to show uncertainty depends on the task or "use case." Specifically, for tasks that depend primarily on estimating the central tendency of the uncertain distribution, faded glyphs should be preferred, but for tasks that emphasize the extent of uncertainty, or the differences between distributions, such as their relative size or the distance between them, glyphs that show the confidence interval as a hard boundary should be preferred. More generally, the results suggest that the ease of perceiving different properties of the data with different visualizations is an important factor that should be considered when designing information displays (cf. Padilla et al., 2015).

Although we were successful in characterizing the strategies used by the majority of participants in this task, our research did not identify any characteristics of individuals that were predictive of the strategy they used. Neither statistics nor geography training was predictive of strategy. However, the participants did not vary much in statistics training (no participant had taken more than two statistics courses) and many of the geography courses that students had taken were not quantitative in nature. Understanding the characteristics of people who adopt different strategies is an important goal for future research.

In terms of the effects of prior knowledge, a limitation of the present research is that almost all of our participants were very familiar with smartphone displays of positional uncertainty. In future research it will be important to test participants who are less familiar with the specific displays and task(s) used in the research. Although this research focused on studying reasoning about smartphone displays as an everyday example of reasoning about uncertainty, it will also be important to examine whether our conclusions apply to other situations requiring information from two or more predictive models to be combined (e.g., models of tomorrow's temperature, the path of a hurricane, or economic forecasts). Future work should examine the applicability of our conclusions to other domains.

References

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods, 10,* 389–396. http://dx.doi.org/10.1037/1082-989X.10.4.389

Broad, K., Leiserowitz, A., Weinkle, J., & Steketee, M. (2007). Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society, 88,* 651–667. http://dx.doi.org/10.1175/BAMS-88-5-651

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin, 133,* 625–637. http://dx.doi.org/10.1037/0033-2909.133.4.625

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics, 20,* 2142–2151. http://dx.doi.org/10.1109/TVCG.2014.2346298

Cox, J., House, D., & Lindell, M. (2013). Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification, 3,* 143–156. http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2012003966

Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics, 29,* 89–93. http://dx.doi.org/10.1111/j.1467-9639.2007.00267.x

Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M., & Ritson, C. (2003). The views of scientific experts on how the public conceptualize uncertainty. *Journal of Risk Research, 6,* 75–85. http://dx.doi.org/10.1080/1366987032000047815

Friedman, A., Ludvig, E. A., Legge, E. L., & Vuong, Q. C. (2013). Bayesian combination of two-dimensional location estimates. *Behavior Research Methods, 45,* 98–107. http://dx.doi.org/10.3758/s13428-012-0241-x

Google. 2007. *Google announces launch of google maps for mobile with my location technology*. Retrieved from http://googlepress.blogspot.in/2007/11/google-announces-launch-of-google-maps_28.html

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3,* 446–474. http://dx.doi.org/10.1111/j.1756-8765.2011.01150.x

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376. http://dx.doi.org/10.1037/0033-295X.98.3.352

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: Initial studies of its effects on risk perception and trust. *Risk Analysis, 15,* 485–494. http://dx.doi.org/10.1111/j.1539-6924.1995.tb00341.x

Johnson, B. B., & Slovic, P. (1998). Lay views on uncertainty in environmental health risk assessment. *Journal of Risk Research, 1,* 261–279. http://dx.doi.org/10.1080/136698798377042

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied, 18,* 126–140. http://dx.doi.org/10.1037/a0025185

Joslyn, S., & LeClerc, J. (2013). Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science, 22,* 308–315. http://dx.doi.org/10.1177/0963721413481473

Joslyn, S., Savelli, S., & Nadav-Greenberg, L. (2011). Reducing probabilistic weather forecasts to the worst-case scenario: Anchoring effects. *Journal of Experimental Psychology: Applied, 17,* 342–353. http://dx.doi.org/10.1037/a0025901

Kinkeldey, C., MacEachren, A. M., Riveiro, M., & Schiewe, J. (2015). Evaluating the effect of visually represented geodata uncertainty on decision-making: Systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science, 42,* 1–21. http://dx.doi.org/10.1080/15230406.2015.1089792

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics, 18,* 2496–2505. http://dx.doi.org/10.1109/TVCG.2012.279

Nadav-Greenberg, L., Joslyn, S., & Taing, M. U. (2008). The effect of weather forecast uncertainty visualization on decision-making. *Journal of Cognitive Engineering and Decision Making, 2,* 24–47. http://dx.doi.org/10.1518/155534308X284354

Padilla, L. M., Hansen, G., Ruginski, I. T., Kramer, H. S., Thompson, W. B., & Creem-Regehr, S. H. (2015). The influence of different graphical displays on nonexpert decision making under uncertainty. *Journal of Experimental Psychology: Applied, 21,* 37–46. http://dx.doi.org/10.1037/xap0000037

Politi, M. C., Han, P. K., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making, 27,* 681–695. http://dx.doi.org/10.1177/0272989X07307270

Ruginski, I. T., Boone, A. P., Padilla, L. M., Liu, L., Heydari, N., Kramer, H. S., . . . Creem-Regehr, S. H. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition and Computation, 16,* 154–172. http://dx.doi.org/10.1080/13875868.2015.1137577

Savelli, S., & Joslyn, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology, 27,* 527–541. http://dx.doi.org/10.1002/acp.2932

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Waggoner, T. (2005). *Ishihara Compatible Pseudoisochromatic Plate (PIPIC) Color Vision Test*. Elgin, IL: Good-Lite Company.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365. http://dx.doi.org/10.1037/0096-3445.115.4.348

Zwick, R., Zapata-Rivera, D., & Hegarty, M. (2014). Comparing graphical and verbal representations of measurement error in test score reports. *Educational Assessment, 19,* 116–138. http://dx.doi.org/10.1080/10627197.2014.903653

Received May 10, 2016
Revision received August 21, 2016
Accepted August 22, 2016 ■