American Psychological Association

Journal of Experimental Psychology: Applied

© 2019 American Psychological Association
ISSN: 1076-898X

2020, Vol. 26, No. 1, 1–15
http://dx.doi.org/10.1037/xap0000245

The Powerful Influence of Marks: Visual and Knowledge-Driven Processing in Hurricane Track Displays

Lace M. K. Padilla
University of California, Merced

Sarah H. Creem-Regehr and William Thompson
University of Utah

Given the widespread use of visualizations to communicate hazard risks, forecast visualizations must be as effective to interpret as possible. However, despite incorporating best practices, visualizations can influence viewer judgments in ways that the designers did not anticipate. Visualization designers should understand the full implications of visualization techniques and seek to develop visualizations that account for the complexities in decision-making. The current study explores the influence of visualizations of uncertainty by examining a case in which ensemble hurricane forecast visualizations produce unintended interpretations. We show that people estimate more damage to a location that is overlapped by a track in an ensemble hurricane forecast visualization compared to a location that does not coincide with a track. We find that this effect can be partially reduced by manipulating the number of hurricane paths displayed, suggesting the importance of visual features of a display on decision making. Providing instructions about the information conveyed in the ensemble display also reduced the effect, but importantly, did not eliminate it. These findings illustrate the powerful influence of marks and their encodings on decision-making with visualizations.

Public Significance Statement
People use data visualizations with uncertainty to make large-scale policy decisions such as where to allocate resources before a natural disaster and more personal life and death decision such as whether to evacuate before a forecasted hurricane. The current work evaluates how visualization techniques influence reasoning during hazard events and leads to practical recommendations for how to help viewers make their best possible decisions with ensemble hurricane forecast visualizations.

Keywords: ensemble hurricane visualizations, visualization cognition, visual-spatial biases, decision-making, risk

Supplemental materials: http://dx.doi.org/10.1037/xap0000245.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We use data visualizations—visual representations of data—as a vital source of information during hazard events with uncertainty. For example, when making hurricane evacuation decisions, numerous studies find that Americans depend on TV news broadcasts (e.g., Driscoll & Salwen, 1996; Lindell, Lu, & Prater, 2005; Lindell & Perry, 2004), which predominantly use visualizations of hurricane data to communicate the risk associated with a storm. Unfortunately, understanding even simple visualizations of uncertainty is challenging for both trained experts and the general public (Belia, Fidler, Williams, & Cumming, 2005). Given the widespread use of uncertainty visualizations during hazard events, it is vital we understand how they influence our judgments of risk and ultimately our preparatory actions.

Visualization researchers have made significant advancements in developing visualizations that elicit fast and effective judgments. However, despite using best practices, visualizations may still produce unintended judgments (e.g., Belia et al., 2005; Joslyn & LeClerc, 2013; Padilla, Ruginski, & Creem-Regehr, 2017; Ruginski et al., 2016). For example, when viewing bar charts, people report that points within the bar are more likely to be a member of the distribution than points that are equidistant from the mean but outside the bar (Newman & Scholl, 2012). One possible cause of unintended visualization interpretations is the influence of the composition of marks (i.e., geometric primitives, such as dots and lines) in relation to visual encoding channels (i.e., the controls of the primitive's appearance, such as color and position; Munzner, 2014). For example, blurring the marks in a visualization may evoke a feeling of *out of focus*, which researchers propose intuitively communicates uncertainty in the data (Jiang, Ormeling, &

This article was published Online First September 26, 2019.

Lace M. K. Padilla, Department of Cognitive and Information Sciences, University of California, Merced; Sarah H. Creem-Regehr, Department of Psychology, University of Utah; William Thompson, School of Computing, University of Utah.

Correspondence concerning this article should be addressed to Lace M. K. Padilla, Department of Cognitive and Information Sciences, University of California, Merced, 5200 North Lake Road, Merced, CA 95343. E-mail: lace.padilla@ucmerced.edu

Kainz, 1995). Although many compositions of marks and encoding channels provide viewers with additional beneficial information, issues may arise when the possible unintended interpretations and subsequent decision-making have not been evaluated.

To examine cases in which seemingly useful visual compositions can distort interpretations, the current study draws on ensemble display visualizations of hurricane track forecasts, shown to be an effective technique when compared to more traditional summary displays (Liu, Padilla, Creem-Regehr, & House, 2019; Ruginski et al., 2016), but also susceptible to biases (Padilla et al., 2017; see Figure 1). In prior work, we compared hurricane path uncertainty visualizations, which revealed that people misinterpret summary displays of hurricane track forecasts (Figure 1A, D, and E). We found that people believe that the summary displays show the hurricane increasing in damage over time (Ruginski et al., 2016). In subsequent research, when participants were asked to explicitly judge the size and intensity of a predicted hurricane, responses were consistent with the previous damage ratings, suggesting that damage ratings incorporate perceptions of both size and intensity of the storm. Participants showed a greater increase in size and intensity ratings with time when viewing the cone display (Figure 1A) compared to ensemble display (Figure 1C; Padilla et al., 2017). In addition, with the ensemble display we found that damage ratings and intensity ratings closely align with the uncertainty in the storm path (Padilla et al., 2017; Ruginski et al., 2016). The results of these studies suggest that the ensemble display reduces misinterpretations about the storm path and is a promising alternative to the summarization visualization techniques that are currently used by the National Hurricane Center, which are similar to Figure 1A.

Before recommending adoption of the ensemble hurricane track visualization, we sought to examine whether the seemingly useful composition of marks and encodings tested in Ruginski et al. (2016) produced any unintended interpretations. Motivated by TV forecasts from numerous hurricanes in 2017, we wanted to understand how people reason with ensemble forecasts when one of the ensemble members or paths directly intersects their town. Padilla et al. (2017) tasked participants with comparing potential damage to two oil platforms—one platform was always collocated with a forecasted hurricane path and the other was not (see Figure 2). This study examined whether viewers believed that oil platforms closer to the center of the distribution of paths (i.e., the area with the most densely populated grouping of lines) would receive more damage, as reported in Ruginski et al. (2016), or if participants believed there would be greater risk associated with locations that were collocated with an ensemble member. Naïve viewers increased the proportion of trials in which they reported that the location farther from the center of the distribution of paths would receive more damage only when an ensemble member was *collocated* with the farther location (Figure 2A). In other words, people overemphasized the importance of a hurricane track when it overlaps an oil rig. We call this the *collocation effect*. In fact, in the ensemble hurricane forecasts presented, the lines are a subset of runs of the model, with perturbations to speed and bearing based on 5-year historical hurricane data (Liu et al., 2016, 2019; Padilla et al., 2017; Ruginski et al., 2016). As the lines are a randomly sampled subset of the model runs, each line is not a deterministic path, just one of many possible predicted paths.

Given the advantages of the ensemble hurricane visualizations over the current summary display method used by the National Hurricane Center and other proposed visualization techniques (Padilla et al., 2017; Ruginski et al., 2016), we sought to understand and reduce the collocation effect to ensure that ensemble displays are as effective to use as possible. Using the cognitive framework

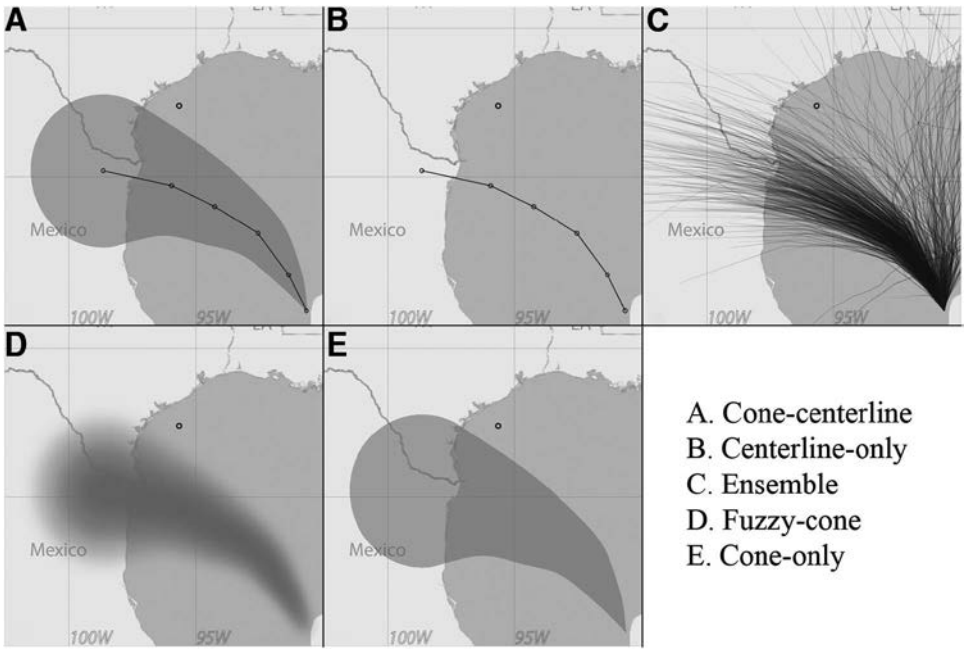

Figure 1. Example of the five hurricane track visualizations compared in Ruginski et al. (2016). Reprinted with permission.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

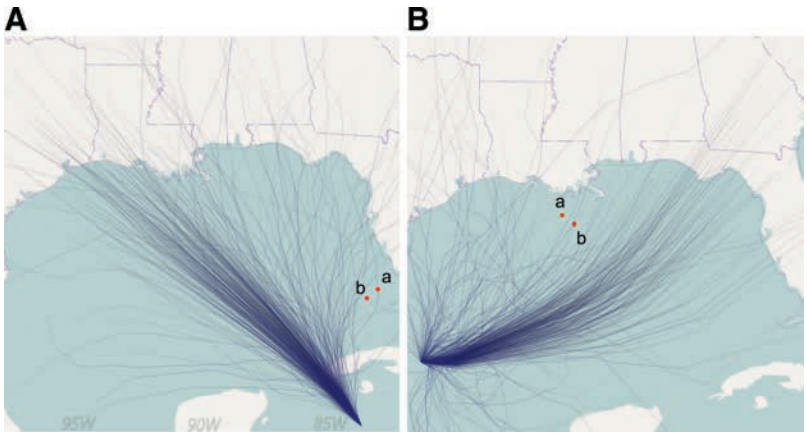

Figure 2. Hurricane forecast display stimuli used in Padilla et al. (2017), where the red dots indicate the location of offshore oil platforms. In A, location a is collocated. In B, location b is collocated. See the online article for the color version of this figure.

for visualization decision-making proposed by Padilla, Creem-Regehr, Hegarty, and Stefanucci (2018), we identified two key ways to reduce the collocation effect. The Padilla et al. (2018) framework suggests that both bottom-up and top-down processing can influence decisions with visualizations. Using bottom-up processing we sought to change the visual properties of the ensemble hurricane forecast, which would have downstream effects on all the subsequent decision-making processes. Drawing on top-down processes, our second approach encouraged viewers to use knowledge-driven processing via instructions to override the collocation effect.

Modifying the Properties of the Visualization

One concept that can help us identify the properties of the ensemble display that could be changed to reduce the collocation effect is that of *visual thought*, proposed by Tversky (2014). Visual thought suggests that in addition to communicating the relationships in the data, marks and encodings relay additional information that informs how we conceptualize the information. For example, numerous studies have documented a *containment* bias, where viewers interpret elements within a boundary as more similar than elements outside a boundary (Belia et al., 2005; Boone, Gunalp, & Hegarty, 2018; McKenzie, Hegarty, Barrett, & Goodchild, 2016; Newman & Scholl, 2012). In one study, McKenzie et al. (2016) showed that participants who viewed a geospatial uncertainty visualization with a hard boundary were more likely to use a containment heuristic than those who saw the same data but represented with a blurred edge created by a Gaussian fade. Freksa and Barkowsky (1996) suggest that sharp boundaries denote distinct concepts more than fuzzy boundaries. In hurricane ensemble track displays, the nature of the marks and encodings could be communicating additional information to the viewer that is producing the collocation effect. The properties of the ensemble display that can be changed to potentially reduce the collocation effect are the number of lines plotted (i.e., changing the marks) and color, line width, and line quality (i.e., changing the encodings of the lines). Prior research has successfully used the color of the hurricane tracks to communicate the category of the storm (Liu et al., 2019). As our long-term goal is to communicate the uncertainty in the category, size, and speed of the storm in addition to the path, here we save the encoding channels of color, line width, and line quality for other data parameters and focus on reducing the collocation effect by modifying the number of ensemble members shown.

The collocation effect may occur, in part, from the way in which the depiction of the ensemble lines leads people to believe that each line is a deterministic path that the hurricane could take rather than representing a sampling from a distribution of paths. Savelli and Joslyn (2013) have also documented cases where participants assume that probabilistic information is deterministic, entitled a *deterministic construal error*. When participants view upper and lower confidence intervals from a distribution of temperatures, they assume the intervals represent deterministic forecasted high and low temperatures (see also attribute substitution in Kahneman, 2011). A deterministic assumption would be appropriate for many visualizations, but it leads to misunderstanding in the case of uncertainty communication. If people assume that each of the hurricane paths represents individual paths the hurricane could take rather than a sampling from a distribution of possible paths, participants may associate a probability with each line. For example, see the cartoon figure of two fictitious hurricane ensemble track forecasts in Figure 3. In both fictitious forecasts, the ensemble tracks are sampled from the same distribution of paths. If a viewer believes that each route is a deterministic path the hurricane could take, he or she might conclude that in forecast A, New Orleans has a 33% chance of being hit by the storm, whereas, in forecast B, New Orleans has only a 10% chance.

The first goal of this work is to attempt to reduce the colocation effect by increasing the number of hurricane paths plotted. This approach takes advantage of the viewer's current conceptualization of the visual display but does not change it. If we can reduce the colocation effect by changing the number of lines, this would suggest that the visual depiction of paths is partially responsible for the previously observed bias (Padilla et al., 2017). Further, this work would offer clear recommendations for the number of ensemble members to plot for visualization practitioners.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

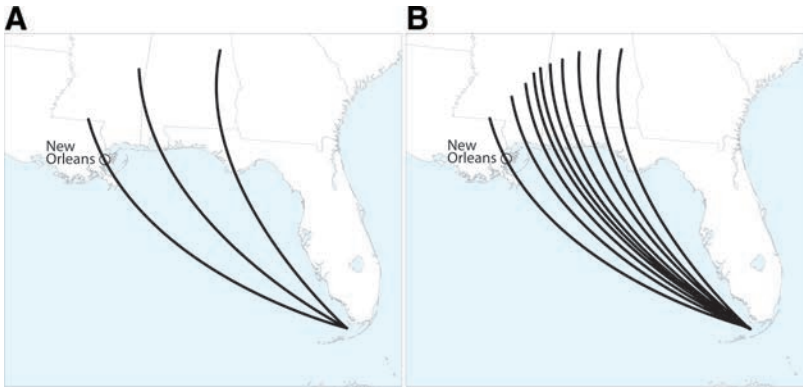

Figure 3. Illustration of two fictitious hurricane ensemble track forecasts. A depicts three ensemble members and B depicts 10 members. See the online article for the color version of this figure.

Activating Knowledge-Driven Processing via Instructions

The second approach highlighted in the Padilla et al. (2018) cognitive framework for visualization decision-making is to encourage viewers to use knowledge-driven processing to override the collocation effect. The interpretation that the paths in the ensemble hurricane forecast are a set of all deterministic forecasted paths is reasonable because viewers were given little information about how the ensemble forecasts were generated. For example, the task instructions used in Padilla et al. (2017) were,

> Throughout the study you will be presented with an image that represents a hurricane forecast, similar to the image shown above. An oil rig is located at each of the two red dots. Your task is to decide which oil rig will receive more damage based on the depicted forecast of the hurricane path.

To encourage viewers to adopt decisions more consistent with the modeling procedure used, at a minimum we need to provide viewers with instructions about how the ensemble visualizations are made.

Whereas providing viewers with detailed instructions about a visualization may seem like an obvious necessity, a growing body of research demonstrates inconsistent findings concerning how effectively viewers incorporate additional information, such as instructions or decision aids, into their decisions (Boone et al., 2018; Grounds, Joslyn, & Otsuka, 2017; Joslyn & LeClerc, 2013; Pugh, Wickens, Herdener, Clegg, & Smith, 2018; Savelli & Joslyn, 2013). Savelli and Joslyn (2013) found that participants maintained the incorrect belief that the temperature error bars represented high and low temperature forecasts despite a key that detailed the correct way to interpret the visualizations (see also Grounds et al., 2017). Pugh et al. (2018) found that training with a summary hurricane forecast path visualization (similar to Figure 1A) improved hurricane path trajectory judgments only when the visualization was present, and the training had no benefit when the visualization was not present. Further, Boone et al. (2018) attempted to improve participants' judgments with the summary hurricane path visualization by providing instructions about how the hurricane forecasts were generated. They found that using several types of instructions helped to reduce misconceptions about the size of the storm growing over time, but the instructions did not consistently influence participants' behavioral risk judgments about the storm.

Beyond viewers inconsistently incorporating additional information about a visualization into their decisions in the laboratory, in real-world scenarios people are provided with limited information about how forecasts are created. The majority of Americans receive information about hurricanes in TV broadcasts (Lindell et al., 2005), and newscasters rarely provide information about how they created the forecast visualizations. For example, one day before hurricane Irma made landfall on September 10th, 2017, of the 20 most viewed forecasts archived on Archive.org, none of the newscasters detailed how the storm path models or visualizations were created, and the average running time of a video clip was 1:52 min (data available in the online supplementary materials). Further, three of these forecasts included misleading information concerning how to interpret the summary hurricane path visualization, such as suggesting that areas inside the boundary are in the danger zone whereas areas outside are relatively safe. In addition, people evaluate many factors when considering evacuation, such as their peers evacuating, businesses closing, living in high risk areas, and official warnings (Huang, Lindell, & Prater, 2016). For these practical reasons, we sought to identify the minimum amount of instructional information needed to reduce the collocation effect, so as not to require undue amounts of time and energy from the viewer. To this aim, we tested two types of instruction manipulations. In one set of video instructions, participants learned about how the ensemble hurricane forecast visualization was created, similar to instructions provided in Boone et al. (2018). In the second, we tested more extensive instructions that also explained the collocation effect to participants and gave them practice overcoming it.

Overview of Experiments

Experiment 1 sought to reduce the colocation effect observed in Padilla et al. (2017) by changing the number of hurricane paths plotted. We hypothesized that increasing the number of lines would significantly reduce the collocation effect if viewers assume each line to be a deterministic forecasted path rather than a sampling from a distribution. To test this hypothesis, we varied

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

whether the target oil rig location intersected a line and the number of ensemble members represented. We predicted that damage ratings would be greater for target locations that are intersected by a hurricane path compared to locations that are not, and damage ratings would be greatest when fewer hurricane paths are shown.

In Experiment 2, we provided participants with several types of video instructions on how to overcome the collocation effect to test whether viewers can use knowledge- driven processing to override the influence of marks and encodings. The first type of instructions included information about how the visualization was created (*visualization instructions*). The goal of testing the visualization instructions was to determine if information about the modeling technique used and visualization generation procedures is sufficient to help people significantly reduce the colocation effect. In addition, based on pilot think-aloud trials that provided insight into participants' conscious strategies, we developed a more elaborate task-specific video tutorial with information about the collocation effect and instructions about refraining from increasing damage judgments when an ensemble member was collocated with the point of interest. Given that prior work is inconclusive as to whether viewers can incorporate additional information to interpret a visual display, we tested if instructions that directly explained both how the visualization was created and how to complete the task could help participants overcome the collocation effect (*task-specific instructions*).

Experiment 1

To test whether increasing the number of lines plotted reduces the colocation effect, we conducted an experiment in which the number of hurricane-simulated ensemble members (9, 17, 33, and 65) was manipulated. Participants viewed a hurricane track visualization with either 9, 17, 33, or 65 tracks and estimated the level of damage that off-shore oil rigs at specified locations would incur, which fell either on or off an ensemble member (see Figure 4; e.g., stimuli). We predicted that the difference in damage estimates between locations falling on versus off an ensemble member (collocation effect) would decrease as the total number of ensemble members increased.

Prior work has found that damage ratings, although indirect, are reflective of the individual's conceptualization of the trajectory of the storm (Padilla et al., 2017). In a study in which viewers made damage, size, and intensity ratings using hurricane path forecasts, researchers found that viewers integrate their understanding of the trajectory of the storm with their assumptions about its size and intensity (Padilla et al., 2017). In this way, damage is a complex judgment that provides information about how people make risk assessments about the path of a storm. Further, we sought to avoid requiring viewers to make probability judgments (i.e., What is the probability that the storm will hit the oil rig?), which are classically challenging and error-prone (Gigerenzer & Hoffrage, 1995). For example, Belia et al. (2005) demonstrated that even experts perform poorly at statistical inference tasks with simple uncertainty visualizations of error bars.

Method

Participants. Based on the effect size described in Padilla et al. (2017), a power analysis was conducted using G*Power (Faul,

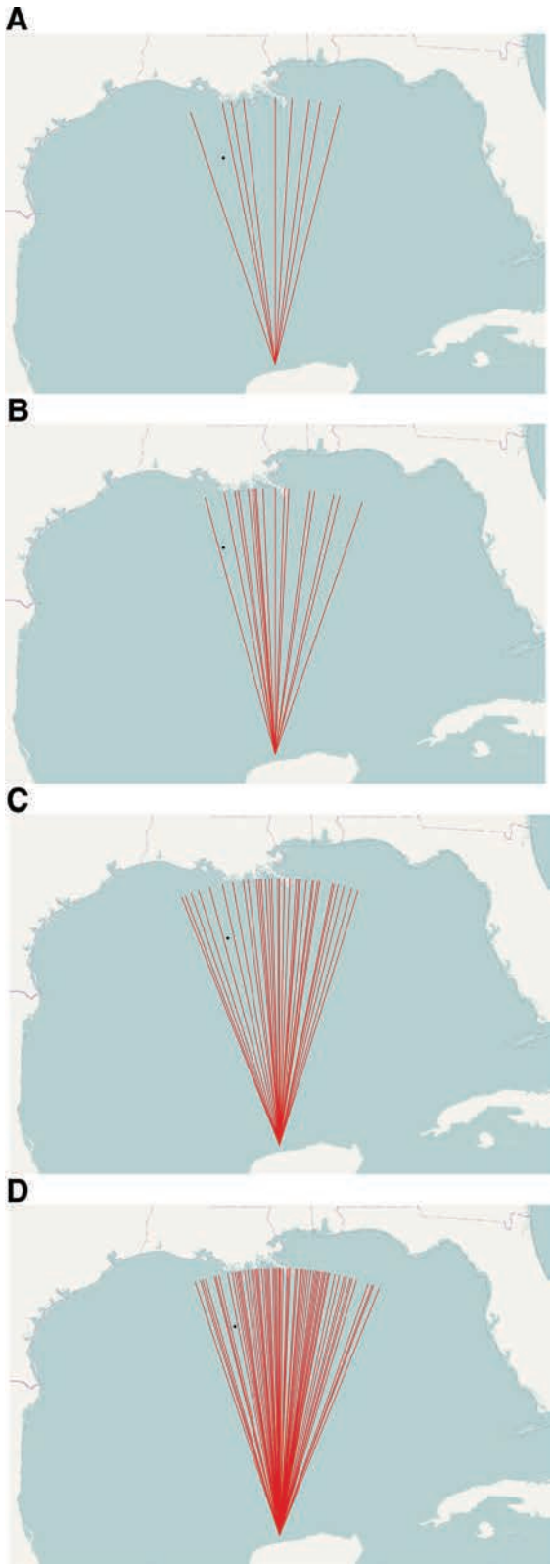

Figure 4. Example stimuli, showing the 9-track (A), 17-track (B), 33-track (C), and 65-track (D) displays. The black dot indicates the location of the offshore oil rig and none are collocated with a hurricane track. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Erdfelder, Buchner, & Lang, 2009) to determine an adequate sample size. At an alpha of level 0.05, power of 0.80, and an effect size of $f^2 = 0.11$, the minimum number of participants needed is 54 for two groups. Participants were 200 undergraduate students currently attending the University of Utah who completed this study for course credit, of which 73 were male and 127 were female, with a mean age of 22 ($SD = 5.58$). Each participant completed the task with visualizations that had one quantity of simulated ensemble members (9 tracks, $n = 52$; 17 tracks, $n = 50$; 33 tracks, $n = 50$; and 65 tracks, $n = 48$). Institutional review board (IRB) approval for this research was obtained from the University of Utah IRB.

Stimuli. Liu et al. (2019) proposed a path reconstruction procedure that more effectively represents a distribution of hurricane paths (Figure 5B) compared to random sampling (Figure 5A). By separating the paths, additional information can be encoded in the paths using color, line weight, and line quality. For example, Liu et al. (2019) added color to the paths to show the category of the hurricane and found that people demonstrate the same pattern of damage rating as with randomly sampled paths, which indicates they understand the uncertainty in similar ways in both techniques.

To examine the collocation effect, code was generated to create artificial hurricane forecast images that mimicked properties suggested by Liu et al. (2019) and distributions in such a way that one of the ensemble members passed through the center of a black dot used to depict the "oil rig" location (see Figure 6A). The stimuli used in the current study differed from the hurricane forecast tracks tested in Liu et al. (2019) in a few ways. Specifically, gaps were placed in the distributions and the tracks were straight lines, both of which were important for increasing the experimental control by creating the manipulations in the current study. The gap was specified in the distribution to ensure that only one line would be collocated with the oil rig location. To create trials where no line was collocated with the oil rig, the previously collocated line was moved to the other side of the distribution in such a way that it was equidistant from the center compared to its original location (see Figure 6C). This placement allowed for direct comparison of trials in which a line was and was not collocated with the oil rig, all other factors remaining constant. The transposition of the collocated line was also the reason the lines needed to be straight. Further, the straight lines allowed for the entire distribution of lines to be flipped over the distribution midline to counterbalance any skewing that may occur from moving the collocated line (see Figure 6B and 6D). In the code, a dot angle was specified that indicated the angle away from the midline at which the oil rig would be placed. N-2 hurricane track lines were sampled from a clipped normal distribution. No two lines could be oriented within 0.25° of one another. The midline was accidentally also plotted and did not adhere to the minimum angular distance constraint. Lines were excluded from two 1.5° gaps. The first gap was centered around the oil rig position. The second gap was equidistant from the center of the distribution, but in the opposite direction from the oil rig position, which allowed the distribution of paths to be flipped over the distribution midline. One additional line was then specified to intersect with the oil rig, producing one stimuli image (see Figure 6A). A second image was also generated in which the additional line was plotted through the gap that did not contain the oil rig (see Figure 6B). The two resulting images were then flipped over the midline to create a total of four mirrored images, two collocated referred to subsequently as "online" and two noncollocated were referred to as "offline" (see Figure 6C and 6D). Thin line widths for the hurricane tracks and a small diameter for the oil rig were selected to increase the precision of the dot overlap with the line.

We chose the following distances to place the oil platforms relative to the mean of the distributions: 14° and 12°. In our prior work, we used a wide range of distances from the center of the distribution and found correspondingly larger differences in damage ratings (Padilla et al., 2017; Ruginski et al., 2016). Our intention here was not to add to the distance finding, but to conceptually replicate it to ensure that the changes we made to the visualization technique did not result in unintended consequences for viewers' perception of the distribution. The distances were chosen to place the gaps in the tails of the distribution, thus reducing the noticeability of the gap by ensuring it was located in less densely populated regions of the distributions. Each simulated

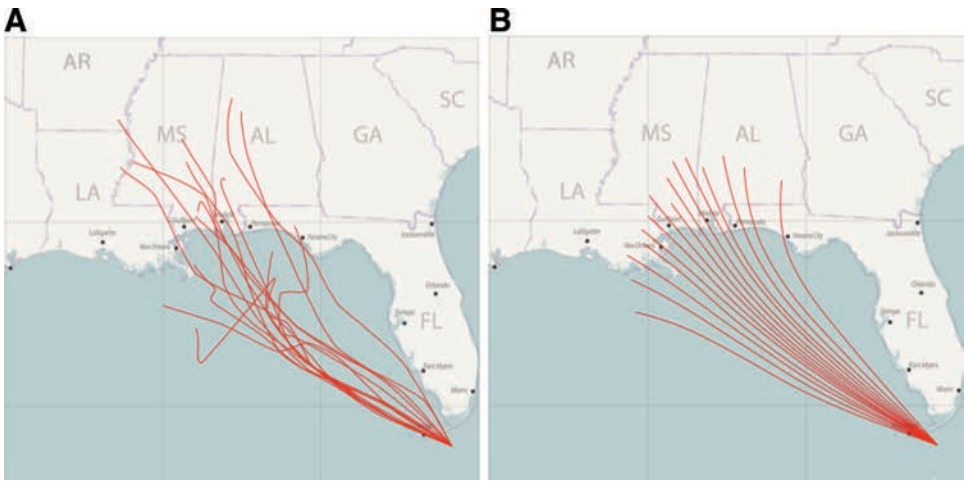

Figure 5. Comparison of randomly selected paths (A) and a path reconstruction procedure (B) proposed by Liu et al. (2019). Reprinted with permission. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

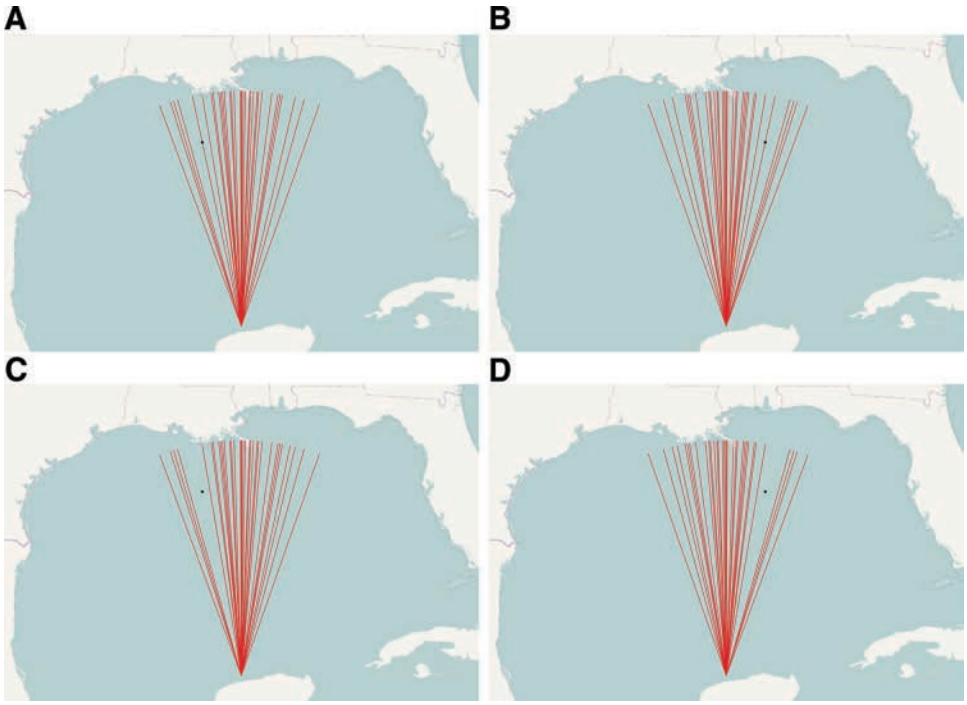

Figure 6. Example of the 33-track display, where one line was reflected over the mean line of the distribution of simulated ensemble members to make a collocation condition (A) and a noncollocation condition (C). B and D represent the mirror images of A and B, where the underlying map remained constant. See the online article for the color version of this figure.

ensemble member was a straight line of fixed length, characterized by its slope represented as an angle. Four quantities (9, 17, 33, and 65) of angles were randomly sampled from a clipped normal distribution with a maximum spread of 40°, a standard deviation of 5°, and a line thickness of 1 pixel. These quantities were selected to represent a wide range of values, which were created starting with a base of 8 and using a logarithmic scale to select 16, 32, and 64. Each quantity had an additional ensemble member, which was the transient ensemble member that was either collocated with the oil platform or moved to the other side of the distribution. Sixty-five simulated ensemble members were, subjectively, the upper bounds of reasonable ensemble members to represent, given the standard deviation, max-spread, and thickness of lines that we specified. In our case, more than 65 simulated ensemble members would have resulted in a distribution that would no longer be perceivable. Given that we aimed to test whether we could reduce the collocation effect by increasing the number of simulated ensemble members, we felt it was essential to test a wide range of quantities of ensemble members even if all the versions did not adhere to visualization design recommendations. Finally, to increase the number of trials, each of the permutations was seeded four times (i.e., randomly sampled four times), and each was then displayed with the midline oriented at three different angles (−30, 0, −30), resulting in a total of 96 trials. All simulated ensemble distributions were digitally composited over a map of the U.S. Gulf Coast that had been edited to minimize distracting labeling. These images were displayed to the subjects at a pixel resolution of 960 × 640 pixels. Underneath the forecast, a scale ranging from 1 (no damage) to 7 (severe damage) was displayed. For each trial, participants were shown one display depicting a hurricane path visualization. Stimuli were presented on the Qualtrics web application (Qualtrics, 2005).

Design. We used a 4 (number of simulated ensemble members: 9, 17, 33, and 65) × 2 (collocation: on- and off-line) × 2 (oil rig locations: 12° and 14°) × 2 (side of the distribution: left and right) × 3 (angle of storm: −30, 0, and 30) × 4 (seeds) mixed factorial design. Collocation, oil rig locations, side, angle of storm, and seeding were within-participant variables, resulting in a total of 96 trials per participant. Participants were randomly assigned to one of four visualization conditions (9, 17, 33, and 65 simulated ensemble members) as a between-participants factor.

Procedure. Participants completed these studies online for course credit on their personal machines, which were required to have screen sizes of larger than 9.4-in tall × 6.6-in wide. Individuals were first given the following instructions for the task and visualization:

> In the following experiment, you will view maps showing the forecast path of different hurricanes as they travel over the Gulf of Mexico toward land. The maps will also show the location of one offshore oil platform in the Gulf. Oil platforms are large structures on the surface of the water with components that extend to the ocean floor for drilling and storing oil.
>
> See the sample map below. A set of potential forecast paths of where the hurricane will move in the next three days is shown in red, and the location of the oil platform is shown by a small black circle. Your task is to estimate the level of damage that the platform will incur based on

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the depicted forecast of the hurricane path on a scale of 1 to 7 where 1 is no damage and 7 is severe.

You will make your judgments of potential damage to the oil platform using the damage scale provided below the map, which will be presented to you along with the forecast maps on each trial. To respond, you should check the box (1 through 7) associated with the level of damage that you believe will occur to the oil platform as a result of the forecasted hurricane. The hurricane forecasts and the locations of the oil platforms will vary across trials.

In addition, each trial included the following text as a reminder of the task: "What is the level of damage that the oil platform will incur?" Following the instructions, participants completed all the trials presented in a different random order for each participant and reported their confidence in their predictions on a Likert scale ranging from 1 (*not at all confident*) to 7 (*very confident*) for every trial. Lastly, at the end of the experiment, participants answered three questions related to comprehension of the hurricane forecasts.

Results

Multilevel models were fit to the data using the R lme package (Bates, Mächler, Bolker, & Walker, 2015) with restricted maximum likelihood estimation procedures (Raudenbush & Bryk, 2002). Multilevel modeling is a generalized form of linear regression that is used to analyze variance in experimental outcomes predicted by both individual (within-participant s) and group (between-participants s) variables. Visualization was dummy coded such that the 9-track visualization was the referent. Collocation was coded such that the coefficients indicated a change from off-line trials to online trials, meaning that a significant positive slope reveals a collocation effect. Collocation (off and on), visualization (9-track, 17-track, 33-track, and 65-track), distance (12° and 14°), and the interaction between collocation and visualization were entered as fixed effects. Participants were entered as random effects. Self-report measures of experience with hurricanes and hurricane prone regions were also collected. The results of this analysis can be seen in Table 1. The participants were students at the University of Utah, and few had experienced a hurricane (3%) or had lived in hurricane-affected regions (7%), so we did not include these measures as covariates.

Our primary hypothesis was that we would see less of a collocation effect for hurricane track visualizations with more simulated ensemble members. To start, there was a main effect of Collocation, meaning that for the 9-track display (the referent) and at the 12° distance (also the referent), damage ratings increased by 1.7 points (on the 7-point Likert scale) when the oil rig was intersected by one of the lines compared to when it was not.

Consistent with our predictions, we also found a significant interaction between collocation and each of the visualizations compared to the 9-track display. The negative coefficient for each of these interactions indicates that the difference between the online and off-line trials is significantly smaller for the 17-, 33-, and 65-track displays compared to the 9-track display at the closest distance. The 9-track online trials ($M = 4.55$, $SD = 1.56$) elicited 1.71 more damage than the off-line trials ($M = 2.84$, $SD = 1.47$). The difference between the online and off-line trials is 0.42 smaller for the 17-track display (online: $M = 4.27$, $SD = 1.56$, off-line: $M = 2.99$, $SD = 1.56$), 0.52 smaller for the 33-track display (online: $M = 3.98$, $SD = 1.36$, off-line: $M = 2.80$, $SD = 1.28$), and 0.15 smaller for the 65-track display (online: $M = 4.24$, $SD = 1.53$, off-line: $M = 2.68$, $SD = 1.47$) compared to the 9-track display.

To visualize the reduction in the collocation effect, we transformed the dependent variable by calculating the difference between the online damage ratings and off-line damage ratings at the same oil platform location, seed, and storm angle. The transformation produced a damage change score where zero indicates no collocation effect, positive values indicate an increase in reported damage for online trials compared to off-line trials, and negative values would indicate a decrease in reported damage for online trials compared to off-line trials. These data are displayed in Figure 7.

As illustrated in Figure 7, the 17-, 33-, and 65-track visualizations show significantly less of a collocation effect. However, unexpectedly, the 65-track visualization shows more of a collocation effect than the 17- and 33-track displays. A post hoc analysis confirmed that after setting the 65-track visualization as the referent and running the same model as previously described, collocation and each of the visualizations compared to the 65-track display had significant interactions. The negative coefficients for the interactions between collocation and the 17-track display ($b = -0.27$), $t(191) = -6.46$, $p < .00$, 95% confidence interval [CI: −0.35, −0.19] and the 33-track display ($b = -0.37$), $t(191) = -8.74$, $p < .000$, 95% CI [−0.45, −0.28] indicated that the collocation effect was significantly

Table 1
List of Fixed Effects With Coefficients, Standard Errors, t-Values, p Values, and 95% Confidence Intervals From the Statistical Model Predicting Damage Ratings

Fixed effects	Coefficients	*SE*	*t* value	*p* value	95% CI
(Intercept)	3.87	.17	21.83	.00	[3.52, 4.21]
Collocation	1.70	.02	58.40	.00	[1.65, 1.76]
17-track	.14	.21	.67	.50	[−.27, .55]
33-track	−.05	.21	−.23	.81	[−.46, .36]
65-track	−.16	.21	−.75	.45	[−.58, .25]
Distance	−.07	.007	−10.54	.00	[−.09, −.06]
Collocation × 17-track	−.42	.04	−10.21	.00	[−.50, −.34]
Collocation × 33-track	−.52	.04	−12.53	.00	[−.60, −.44]
Collocation × 65-track	−.15	.04	−3.58	.0003	[−.23, −.06]

Note. CI = confidence interval. Collocation was coded such that the effects indicate a change from offline to online.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

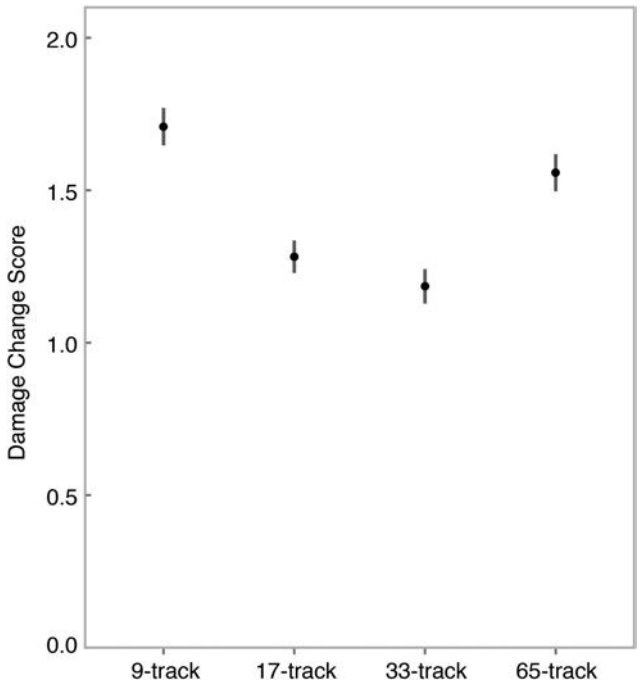

Figure 7. Damage change scores for the 9-, 17-, 33-, and 65-track displays. Bars represent 95% confidence intervals around the mean.

smaller for the 17- and 33-track displays compared to the 65-track display at the 12° distance.

A significant main effect of distance revealed that at the distance closer to the center of the distribution (12°, $M = 3.63$, $SD = 1.06$), participants believed that the oil rig would receive more damage compared to the farther oil rig location (14°, $M = 3.47$, $SD = 1.09$), which is in line with prior work and indicates that viewers perceived the uncertainty communicated in the distribution. Although significant, a change of .16 is small on a Likert scale from 1 to 7. In sum, this finding is in line with past work that suggests viewers effectively perceived the probability distribution that the hurricane track simulated ensemble visualization is intended to represent.

Participants also reported their confidence in their judgments for each trial using a Likert scale ranging from 1 (*not at all confident*) to 7 (*very confident*), along with follow-up questions about the visualizations. Using a multilevel model, we evaluated the impact of visualization and collocation (fixed effects) on confidence ratings with participants as random effects. This analysis revealed no significant change in confidence from the 9-track display ($M = 4.56$, $SD = 1.63$) compared to the 17- ($M = 4.65$, $SD = 1.51$), 33- ($M = 4.39$, $SD = 1.54$), and 63-track displays ($M = 4.87$, $SD = 1.44$). However, the main effect of collocation showed the participants were more confident about their judgments for the online ($M = 4.68$, $SD = 1.50$) trials than the off line trials ($M = 4.56$, $SD = 1.58$), $b = .117$, $t(198) = 9.03$, $p < .000$, 95% CI [0.09, 0.14]. However, the increased confidence for the online trials was quite small, 1.71%.

The results of the survey questions can be found in Table 2. Using a general linear model, visualization was used to predict question response with the 9-track display as the referent. Full output of the models can be found in the online supplementary materials.[1] For Q1, which references uncertainty in the visualization, no significant differences were found between the 9-track display and other visualization techniques, with participants at chance performance. For Q2, which references the collocation effect, participants viewing the 17- and 33-track displays showed fewer correct responses than those viewing the 9-track display, which is surprising because participants' behavioral judgments were in the opposite direction, where the 17- and 33-track displays show the least collocation effect. For Q3, viewers of the 65-track display were more likely to indicate that the hurricane forecast showed all possible paths the hurricane could take compared to the 9-track display. This result suggests that when too many ensemble members are plotted, one unintended effect may be that viewers believe that they represent all of the possible outcomes. To follow-up on this finding, a linear model was conducted with Q3 predicting the damage change score (collocation effect) of the 65-track display. This analysis revealed that participants who answered "no" to Q3 showed significantly less of a collocation effect ($M = 1.38$, $SD = 1.36$) compared to those who answered "yes" ($M = 1.68$, $SD = 1.56$), $b = -0.30$, $t(46) = -6.82$, $p < .00$, 95% CI [−0.38, −0.21].

Discussion

The results of this experiment showed that novice users are less influenced by the impact of a single simulated ensemble member when more ensemble members are represented. This finding supports our hypothesis that hurricane ensemble paths evoke an assumption that each line is a deterministic forecasted path rather than a sampling from a distribution and increasing the number of paths decreases the negative impacts of this effect. In addition, we found several unpredicted effects relating to visualizing the largest number of simulated ensemble members (65-track). The primary finding was that for the 65-track display, the collocation effect was greater than for the 17- and 33-track displays (although still less than the 9-track display). The postsurvey Q3 suggested that viewers were more likely to believe that the 65-track display represented all the possible paths the hurricane could take. A follow-up analysis provided evidence that, for the 65-track display, incorrect beliefs about the visualization representing all of the possible paths increased the collocation effect. It could also be the case that for the 65-track display, the gaps in the distribution of paths were more apparent, which could also influence the collocation effect. In sum, although increasing the number of simulated ensemble members can reduce the collocation effect, evidence suggests that when many ensemble members are represented, more viewers believe that all possible outcomes are shown.

Importantly, it should be noted that the collocation effect was never completely ameliorated. The 34-track visualization showed the largest (30%) reduction of the collocation effect compared to the 9-track display. Yet, participants still reported, using the Likert scale, that oil platforms that were directly hit by one of the simulated ensemble members would receive 1.28 units of more damage than oil rig locations that were not directly hit.

[1] Additional online materials can be found at: osf.io/j34g5/.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
Proportion of Responses for Each Visualization Condition

Questions	9-track	17-track	33-track	65-track
Q1. The display indicates that the forecasters are less certain about the path of the hurricane as time passes.	Yes: 50%	Yes: 46%	Yes: 58%	Yes: 38%
Q2. Locations that are touching a hurricane track are more likely to be hit by the storm than locations equidistant from the center of the forecast but not touching a hurricane track.	No: 35%	No: 16%*	No: 12%*	No: 25%
Q3. The hurricane forecast shows all possible paths the hurricane could take.	No: 63%	No: 54%	No: 54%	No: 42%*

* $p < .05$, with the 9-track display as the referent.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Experiment 2

In Experiment 2, we provided participants with several types of video instructions on how to overcome the collocation effect to test whether viewers can use knowledge-driven processing to override the influence of marks and encodings.

The first step in developing informative instructions was to identify what types of conscious decision-making strategies participants were aware of using, in order to determine how top-down knowledge may be able to influence the collocation effect. Based on prior work that examined the use of strategies in a mental rotation task that included visual information (Hegarty, 2017), we used a concurrent verbal protocol and a retrospective protocol to elicit participants' strategies while they completed 10 randomly sampled trials from Experiment 1. The objective of this pilot was to study the processes that participants were aware of, in case they may have been adopting deliberate cognitive strategies that could have contributed to the collocation effect. Twenty undergraduate and graduate students at the University of Utah participated in the pilot for $10 (male = eight, female = 12, and a *M* age of 25.75, *SD* = 4.3). Participants first received instructions on how to complete concurrent verbal protocols, which involved instructing them to verbalize their thoughts as they completed each stage of the study, including the practice trials (Ericsson & Simon, 1992). In line with recommendations from Ericsson and Simon (1992), three practice trials were used to help participants become comfortable with verbalizing their thoughts while completing the task. They were then given 10 think-aloud trials in which they were instructed to verbalize everything that came to mind as they completed all steps of the task. Following recommendations from (Ericsson & Simon, 1992), participants were encouraged to "keep talking," rather than a social communication request, such as "tell me what you think." Finally, participants completed retrospective protocols for which they reported what they thought while they completed the think-aloud protocols. The entire process was video recorded and transcribed.

Three distinct strategies and a combination of these strategies were observed, including (a) *distance strategy* for which participants reported determining their damage rating based on how far the oil rig was from the center of the distribution of simulated ensemble members, (b) *collocation* for which participants specifically commented on rating oil rig locations that are collocated with a simulated ensemble member as receiving more damage, and (c) *surrounding ensemble members* where participants reported making their damage judgments based on the distance of the oil rig to the surrounding simulated ensemble members. The results of this pilot provide evidence that some participants strategically increased their damage ratings when the oil rig was collocated with an ensemble member. Given that some participants were aware of the influence of collocation and the surrounding simulated ensemble members, it is possible that if they had been given instructions about how to overcome the collocation effect and interpret the visualizations correctly, they would have been able to incorporate this information into their decisions.

Using the findings of the think-aloud and retrospective protocols, we developed two types of video instructions to test whether participants can use top-down knowledge to overcome the collocation effect. The task-specific instructions included information about how to overcome the collocation effect, and the more general visualization instructions included only information about how simulated ensemble hurricane forecast tracks are generated. We predicted that the task-specific instructions would reduce the colocation effect significantly but, given the influence of visual features, not completely. This finding would suggest that the collocation effect could be influenced by top-down knowledge. In addition, we predicted that the visualization instructions would reduce the collocation effect but not to the degree of the task-specific instructions.

Method

Participants. Participants were 83 undergraduate students currently attending the University of Utah who received course credit for participation. Three participants were disqualified for not following instructions. Of the 80 (40 in each instructions group) who were included in the analysis, 23 participants were male, and 57 were female, with a *M* age of 21 (*SD* = 3.7).

Stimuli and design. The same 9-track display stimuli were utilized along with the same study design as in Experiment 1. However, before receiving the experiment instructions, participants viewed one of two videos. Both videos can be found in the online supplementary materials. The *task-specific video* included narrated instructions about the collocation effect and information about how the simulated ensemble hurricane forecasts were generated along with visual examples (length of 3.13 min). The full transcripts of the instructions are in the Appendix. The sequence of the task-specific instructions video was as follows:

1. Overview of the functions of hurricane forecasts
2. Description of how the type of hurricane forecast used in this study was generated
3. Information about uncertainty in hurricane forecasts

4. Instructions on how to identify the center of the distribution of paths and that the center represents the most likely path the hurricane will take

5. Illustration of how static simulated ensemble hurricane forecasts represent a subset of the many possible paths generated by the forecast models

6. Description of the collocation effect

7. Practice overcoming the collocation effect with example questions.

The *visualization-instructions* video was an edited version of the task-specific video (1.37 min) and included Elements 1–5 of the list above. Specific information about the collocation effect and practice overcoming the effect was not included.

Procedure. Participants were randomly assigned to one of two groups (task-specific instructions or visualization-instructions). After consent was obtained, participants viewed the relevant video and then completed the same procedure detailed in Experiment 1 but with only the 9-track visualization. As the 9-track visualization exhibited the largest collocation effect, we used it as a baseline to try to reduce the collocation effect with the instruction videos.

Results

As in Experiment 1, we used a multilevel logistic regression model to determine the influence of instructions on the damage ratings. We compared the 9-track display results from Experiment 1 to new data from participants who received the additional instructions. Instruction-type (none, task, and general), collocation (off and on), distance (12° and 14°), and the interaction between collocation and instructions were entered as fixed effects (see Table 3). Participants were entered as random effects. Collocation was coded such that effects indicate a change from off-line to online trials, and the no-instructions condition was specified as the referent.

As illustrated in Figure 8, the participants who viewed the general ($M = .96$, $SD = 1.22$) and task-specific instructions ($M = .66$, $SD = 1.01$) demonstrated significantly less of a collocation effect compared to those who received no instructions ($M = 1.70$, $SD = 1.56$). The coefficients for the interactions indicate that the task-specific instructions reduced the bias by 1.04 on the Likert scale, which corresponds to a 61% reduction of the collocation effect observed with the 9-track display. For the visualization instructions, the bias was reduced by .74 on the Likert scale or a 44% reduction of the collocation effect observed with the 9-track display. To test whether the task-specific instructions reduced the collocation effect more than the visualization instructions, the same analysis was conducted, but the visualization instructions were specified as the referent. This analysis revealed that the task-specific instructions reduced the collocation effect significantly more than the general instructions, $b = -0.30$, $t(125) = -6.12$, $p < .00$, 95% CI [−0.39, −0.20].

For confidence, a multilevel model was used to evaluate the impact of instructions and collocation (fixed effects) on confidence ratings with participants as random effects. This analysis revealed that participants who viewed both the visualization ($M = 4.6$, $SD = 1.4$), $b = .89$, $t(128) = 3.83$, $p < .000$, 95% CI [0.43, 1.35] and the task-specific instructions ($M = 4.93$, $SD = 1.48$), $b = 1.23$, $t(128) = 5.26$, $p < .000$, 95% CI [0.77, 1.69] had significantly more confidence in their damage ratings compared to those who received no instructions ($M = 3.7$, $SD = 1.74$). To test whether participants with task-specific instructions were more confident in their ratings compared to those who received the visualization instructions, the same analysis was conducted, but the visualization instructions were specified as the referent. This analysis revealed participants with the task-specific instructions were not more confident in their responses compared to those with visualization instructions, $b = 0.33$, $t(125) = 1.35$, $p = .17$, 95% CI [−.15, .82].

The results of the survey questions can be found in Table 4. Using a general linear model, instruction-type was used to predict question response with the no-instructions condition as the referent. Full output of the models can be found in the online supplementary materials. For Q1, no significant differences were found between no instructions and either of the instruction conditions. For Q2 (collocation), participants with the task-specific instructions were more likely to answer the question correctly compared to those who received no instructions. For Q3 (all possible paths), participants who received either instruction condition answered the question more correctly than those without instructions.

Discussion

In Experiment 2, we found that the collocation effect was significantly, but not entirely, reduced by instructions. The task-specific instructions attenuated the collocation effect to a greater degree than the visualization instructions, as evidenced by both the objective behavioral measures and the participants' self-report measures of their understanding. This finding illustrates both the effectiveness of instructions and the powerful influence of marks

Table 3
List of Fixed Effects With Coefficients, Standard Errors, p Values, and 95% Confidence Intervals (CIs) From the Statistical Model Predicting Damage Ratings

Fixed effects	Coefficient	*SE*	*t* value	*p* value	95% CI
(Intercept)	4.52	.19	23.87	.00	[4.15, 4.89]
Collocation	1.71	.03	56.07	.00	[1.64, 1.76]
Task-specific instruction	−.18	.22	−.82	.41	[−.60, .24]
Visualization instructions	.81	.22	3.74	.0001	[.38, 1.23]
Distance	−.13	.009	−13.46	.00	[−.14, −.10]
Collocation × Task-Specific Instruction	−1.04	.05	−22.54	.00	[−1.13, −.95]
Collocation × Visualization Instructions	−.74	.05	−16.02	.00	[−.83, −.64]

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

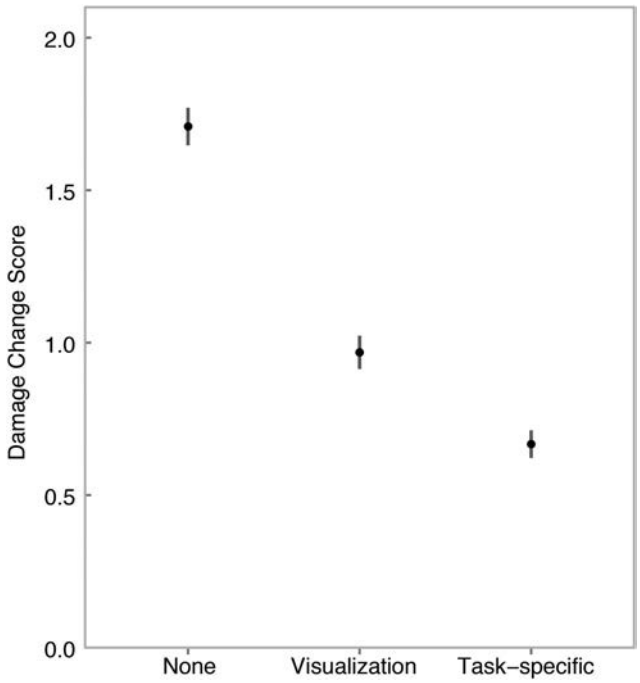

Figure 8. Damage change scores for the 9-track display conditions with no instructions, visualization instructions, and task-specific instructions. Bars represent 95% confidence intervals around the mean.

and channels in a visualization, as the collocation effect was never fully reduced by knowledge-driven processing.

General Discussion

Together, the current experiments explored the influence of visual properties and knowledge-driven processing on the previously observed collocation effect (Padilla et al., 2017) in hurricane forecast visualizations. We were able to demonstrate two approaches for reducing the collocation effect. In Experiment 1, we reduced the collocation effect by increasing the number of ensemble members plotted. When we increased the number of ensemble members shown, the importance of each ensemble member decreased, resulting in more similar damage judgments when a track directly hit the target compared to when a track just missed the target. Note that this approach does not assume a change in conceptualization from the viewer. Instead, we capitalized on the specifics of how people naturally reason with the information and then changed the visualization to elicit viewers' best possible judgments. In addition, both the postsurvey questions and the behavioral results suggest that with each of the visualizations viewers do understand the uncertainty communicated by the distributions. The results of this study suggest that ensembles can be used to effectively communicate uncertainty, but caution should be taken when using them for specific tasks for which viewers make judgments about a particular location.

With the second approach executed in Experiment 2, we also showed that the collocation effect could be significantly reduced when viewers used top-down knowledge-driven processing to incorporate instructions. In the second study, we found that viewers were consciously aware of the strategies they used to complete the task, including the influence of a simulated ensemble member when it was collocated with their point of interest. Although largely reduced with both approaches, the collocation effect was never completely eliminated. Notably, specific and admittedly frank instructions on how to overcome the collocation effect did not entirely reduce the bias. Our work proposes that the marks and channels in a visualization have a powerful influence on decision-making, which can be resilient to top-down knowledge-driven processing and should be taken seriously.

Other possible explanations for the collocation effect were not directly tested in the current study, such as the use of Gestalt principles of perceptual organization (Wertheimer, 1938) and object-based attention (Scholl, 2001). In the context of graph comprehension, Pinker (1990) has argued that we use Gestalt principles of perceptual grouping to constrain how visual features are linked together. Principles of grouping demonstrate that visual elements that are spatially proximal to each other (grouping by proximity), have smooth continuation with one another (grouping by good continuation), or are visually connected (grouping by connectedness) will be perceived as part of a single configuration. Several studies on graph understanding have shown that interpretations are affected by different Gestalt grouping principles. For example, the within-the-bar bias described earlier (Newman & Scholl, 2012; see also Okan, Garcia-Retamero, Cokely, & Maldonado, 2018) can be explained by the "visual chunking" of a bar as a single contained entity. Others have shown that modifying graphs so that certain Gestalt principles are in effect influences our ability to identify global trends in the data (Shah, Mayer, & Hegarty, 1999) or our accuracy in interpreting statistical interactions (Ali & Peebles, 2013). In the current study, the ensemble display may have evoked various grouping principles that influ-

Table 4
Proportion of Correct for Each Visualization Condition

Questions	None	Visualization instruction	Task-specific instruction
Q1. The display indicates that the forecasters are less certain about the path of the hurricane as time passes.	Yes: 50%	Yes: 35%	Yes: 42%
Q2. Locations that are touching a hurricane track are more likely to be hit by the storm than locations equidistant from the center of the forecast but not touching a hurricane track.	No: 34.6%	No: 42%	No: 90%***
Q3. The hurricane forecast shows all possible paths the hurricane could take.	No: 63%	No: 87%*	No: 92%*

* $p < .05$. *** $p = .000$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ence viewers' interpretations. A possible explanation based on perceptual grouping could be that viewers perceived the collocated dot-line configuration as a single object, enhanced by object-based attention mechanisms (Scholl, 2001). With additional lines added, less attention may have been directed at the individual ensemble member that contained the dot. Likewise, the amount of space between the lines changed when additional lines were added, possibly reducing attention to the distinction between on- and off-line trials. Future studies using eye-tracking methods could help to distinguish whether the lines and spaces attract less attention as the number of ensemble members increases. The displays could also be manipulated in terms of the thickness of the lines, the diameter of the dot, and the size of the spaces to test whether these features might affect processes of grouping.

This work also illustrated methods for developing both general and task-specific instructions and showed that instructions could change decisions with visualizations. Prior work demonstrated inconsistent findings as to whether people could utilize prior knowledge to change their judgments when viewing visualizations (Bailey, Carswell, Grant, & Basham, 2007; Boone et al., 2018; Joslyn & LeClerc, 2013; Shen, Carswell, Santhanam, & Bailey, 2012). Consistent with the recommendations from Zapata-Rivera, Zwick, and Vezzu (2016), this work finds that instructions can be used to reduce a specific error in reasoning with visualizations. We believe that the task-specific instructions were more successful than the general instructions because they targeted a specific distortion rather than trying to improve judgments broadly, which may be why other work did not find consistent improvements in visualization decision-making when providing viewers with more information (Boone et al., 2018; Savelli & Joslyn, 2013). Our intention for not providing participants with detailed instructions on how to interpret hurricane forecasts in Experiment 1 was, first, to understand what type of biases were naturally elicited purely by the visualization technique. Further, in real-world contexts, such as in hurricane forecasts on the news, it is rare that viewers are given a full description of how the forecast visualizations were generated and how to effectively interpret the forecasts. We sought to examine how people make judgments about storm damage with limited background information, to better understand what elements of the visualization technique are eliciting biases that would likely be observed in the real world.

The applied contributions of this work are to demonstrate that ensemble hurricane forecasts are effective for intuitively communicating uncertainty in hurricane paths, but also that it is important to consider unintended consequences of the ensemble display itself. We showed that when more simulated ensemble members are plotted, the adverse effects of this visualization technique are reduced, but not completely. Our stimuli showed the largest advantage when displaying about 30 ensemble members, but the maximum number could vary in other displays depending on numerous parameters such as spread and line thickness. We also demonstrated that providing instructions about how the hurricane forecast was generated and the collocation effect significantly reduced the adverse effects of this visualization technique. Additional work is needed to test how these findings generalize to other contexts where ensemble visualizations are used.

Although recent work reveals many merits to reconstructed ensemble paths compared to randomly sampled paths (Liu et al., 2019), one unconsidered component of randomly sampled paths is that they may convey more uncertainty compared to the smoother and more regularly distributed paths used in the current study. Liu et al. (2019) found that participants demonstrate a similar pattern of damage ratings between randomly sampled and reconstructed paths, where participants rate greater damage toward the center of the distribution and less damage as the distance from the center increases. However, damage ratings might not be sensitive enough to pick up on differences in the perception of uncertainty between the two techniques. Future work might consider the impact of the nature of the ensemble paths on the perception of uncertainty.

In addition, future research might consider how people with different levels of exposure to hurricane forecasts respond to various encodings of ensemble paths. The naive population used in the current study was selected as a baseline so that the influence of exposure to hurricane forecasts could be evaluated. Individuals who live in Florida, for example, might not demonstrate the collocation effect because they have first-hand experiences with the uncertainty of hurricane paths, or they might demonstrate a larger collocation effect, because they might have a greater emotional response to a line overlapping their town. Examining the influence of prior experiences with hazards on the perception of visual elements in a forecast is an open area of exploration.

Conclusions

Ensemble visualizations are an increasingly popular method for visualizing data, because emerging research demonstrates that ensemble visualizations can effectively and intuitively communicate traditionally complex statistical concepts to novice viewers. Ensemble visualizations are now being used to help local officials make large-scale decisions such as whether to evacuate a town before a hurricane strike. Given their widespread use and impact, we need to understand how ensemble visualizations influence our judgments and actions. We found that simulated ensemble visualizations that include a greater number of ensemble members (but not too many) can be an effective visualization technique for various types of decision-making tasks and that providing instructions about how the visualization is created can help people make more effective decisions. Further, this work demonstrates the importance of evaluating both the lower level perceptual and higher level cognitive processes at work when making decisions with visualizations. By understanding the cognitive processes associated with visualization reasoning, we can make more effective predictions about viewers' judgments and create increasingly targeted visualization improvements and decision aids.

References

Ali, N., & Peebles, D. (2013). The effect of Gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *Human Factors, 55,* 183–203. http://dx.doi.org/10.1177/0018720812452592

Bailey, K., Carswell, C. M., Grant, R., & Basham, L. (2007). Geospatial perspective-taking: How well do decision makers choose their views? *51st Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1246–1248). Los Angeles: SAGE Publications.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Advance online publication. http://dx.doi.org/10.18637/jss.v067.i01

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Boone, A. P., Gunalp, P., & Hegarty, M. (2018). Explicit versus actionable knowledge: The influence of explaining graphical conventions on interpretation of hurricane forecast visualizations. *Journal of Experimental Psychology: Applied, 24,* 275–295. http://dx.doi.org/10.1037/xap0000166

Driscoll, P., & Salwen, M. B. (1996). Riding out the storm: Public evaluations of news coverage of Hurricane Andrew. *International Journal of Mass Emergencies and Disasters, 14,* 293–303.

Ericsson, K. A., & Simon, H. A. (1992). *Protocol analysis: Verbal reports as data* (rev. ed). Cambridge, MA: MIT Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160.

Freksa, C., & Barkowsky, T. (1996). On the relation between spatial concepts and geographic objects. In P. Burrough & A. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 109–121). London, U. K.: Taylor & Francis.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704. http://dx.doi.org/10.1037/0033-295X.102.4.684

Grounds, M. A., Joslyn, S., & Otsuka, K. (2017). Probabilistic interval forecasts: An individual differences approach to understanding forecast communication. *Advances in Meteorology, 2017,* Article ID 3932565. http://dx.doi.org/10.1155/2017/3932565

Hegarty, M. (2017). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review, 25,* 1212–1219.

Huang, S.-K., Lindell, M. K., & Prater, C. S. (2016). Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environment and Behavior, 48,* 991–1029. http://dx.doi.org/10.1177/0013916515578485

Jiang, B., Ormeling, F., & Kainz, W. (1995, August). *Visualization support for fuzzy spatial analysis*. Paper presented at the American Congress of Survey Mapping/American Society of Photogrammetry and Remote Sensing Conference.

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

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Lindell, M. K., Lu, J.-C., & Prater, C. S. (2005). Household decision making and evacuation in response to Hurricane Lili. *Natural Hazards Review, 6,* 171–179.

Lindell, M. K., & Perry, R. W. (2004). *Communicating environmental risk in multiethnic communities*. Thousand Oaks, CA: Sage.

Liu, L., Boone, A., Ruginski, I., Padilla, L., Hegarty, M., Creem-Regehr, S. H., . . . House, D. H. (2016). Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization and Computer Graphics, 23,* 2165–2178.

Liu, L., Padilla, L., Creem-Regehr, S. H., & House, D. (2019). Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization Computer Graphics Forum, 25,* 882–891.

McKenzie, G., Hegarty, M., Barrett, T., & Goodchild, M. (2016). Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science, 30,* 221–239. http://dx.doi.org/10.1080/13658816.2015.1082566

Munzner, T. (2014). *Visualization analysis and design*. Cleveland, OH: CRC Press. http://dx.doi.org/10.1201/b17511

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review, 19,* 601–607. http://dx.doi.org/10.3758/s13423-012-0247-5

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. J. (2018). Biasing and debiasing health decisions with bar graphs: Costs and benefits of graph literacy, *71,* 2506–2519.

Padilla, L., Creem-Regehr, S., Hegarty, M., & Stefanucci, J. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications, 3,* 29. http://dx.doi.org/10.1186/s41235-018-0120-9

Padilla, L. M., Ruginski, I. T., & Creem-Regehr, S. H. (2017). Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications, 2,* 40. http://dx.doi.org/10.1186/s41235-017-0076-1

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pugh, A. J., Wickens, C. D., Herdener, N., Clegg, B. A., & Smith, C. A. P. (2018). Effect of visualization training on uncertain spatial trajectory predictions. *Human Factors, 60,* 324–339. http://dx.doi.org/10.1177/0018720818758770

Qualtrics. (2005). Qualtrics [Computer software]. Provo, Utah: Author.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Ruginski, I. T., Boone, A. P., Padilla, L., 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

Scholl, B. (2001). *Objects and attention: The state of the art. Cognition, 80*(1–2), 1–46.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91,* 690–702. http://dx.doi.org/10.1037/0022-0663.91.4.690

Shen, M., Carswell, M., Santhanam, R., & Bailey, K. (2012). Emergency management information systems: Could decision makers be supported in choosing display formats? *Decision Support Systems, 52,* 318–330. http://dx.doi.org/10.1016/j.dss.2011.08.008

Tversky, B. (2014). Visualizing thought. In W. Huang (Ed.), *Handbook of human centric visualization* (pp. 3–40). New York, NY: Springer.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). London, UK: Kegan Paul, Trench, Trubner & Company. http://dx.doi.org/10.1037/11496-005

Zapata-Rivera, D., Zwick, R., & Vezzu, M. (2016). Exploring the effectiveness of a measurement error tutorial in helping teachers understand score report results. *Educational Assessment, 21,* 215–229. http://dx.doi.org/10.1080/10627197.2016.1202110

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix

Full instruction videos are in additional online materials, which can be found at: https://osf.io/j34g5/.

Visualization Instructions Transcript

Hurricane forecasts can help you understand where a hurricane may go. Meteorologists create mathematical models to predict the path of a hurricane, and sometimes they use polylines to represent the hurricane's predicted path. One line represents one output of the model. While hurricane forecast models are improving, even with these models, meteorologists aren't 100% sure of the exact path that the hurricane will take. Many factors can change the path of the hurricane, such as unexpected weather changes. To show the range of possible paths that the hurricane could take, meteorologists use various techniques to plot multiple lines. The spread and distribution of lines is intended to show you the general direction that the hurricane will go. Here is an animation of a real hurricane forecast, where the scientists plotted many lines based on historical data. Notice where most of the lines are grouped. This is the most likely path that the hurricane will take. Also notice that each line is only one of hundreds of lines produced by the model. For this specific model, this means that any one line isn't very meaningful. Scientists cannot always show you animations, like the one you just saw. Sometimes they have to use a single image for print publications such as newspapers or reports. If you see an image like this, it doesn't show you all of the other possible lines, like the animation did. The lines that you see are only a small subset of all possible lines.

Task Specific Instructions Transcript

Hurricane forecasts can help you understand where a hurricane may go. Meteorologists create mathematical models to predict the path of a hurricane, and sometimes they use polylines to represent the hurricane's predicted path. One line represents one output of the model. While hurricane forecast models are improving, even with these models, meteorologists aren't 100% sure of the exact path that the hurricane will take. Many factors can change the path of the hurricane, such as unexpected weather changes. To show the range of possible paths that the hurricane could take, meteorologists use various techniques to plot multiple lines. The spread and distribution of lines is intended to show you the general direction that the hurricane will go. Here is an animation of a real hurricane forecast, where the scientists plotted many lines based on historical data. Notice where most of the lines are grouped. This is the most likely path that the hurricane will take. Also notice that each line is only one of hundreds of lines produced by the model. For this specific model, this means that any one line isn't very meaningful. Scientists cannot always show you animations, like the one you just saw. Sometimes they have to use a single image for print publications such as newspapers or reports. If you see an image like this, it doesn't show you all of the other possible lines, like the animation did. The lines that you see are only a small subset of all possible lines. This means that if you see one line overlapping your town or just missing it that is not meaningful because, as we learned, there are many lines that are not represented. Rather it is more important, to identify the center of the grouping of lines. Areas near the center of the grouping of lines have the highest likelihood of being hit by the storm. Let's try an example. Which location do you think would receive more damage? Keep in mind that these lines do not give you information about the size or intensity of the storm - just the path that it might take. Location A has the highest likelihood of being hit by the storm because it is closer to the center of the grouping of lines. Let's try one more example. Which location would receive the most damage? In this case, Location B has the highest likelihood of being hit by the storm. Remember to not base your judgment on an individual line. To interpret these correctly you must imagine where the center of the storm is, and locations closest to the center have the highest likelihood of being hit by the storm. To sum up what we've learned today, each line is only a sampling of the many possible lines. So it does not matter if one line overlaps your point of interest or not. Instead, you should focus on the center of the grouping of lines, which shows the most likely path that the hurricane will take.

Received October 14, 2018
Revision received July 25, 2019
Accepted July 30, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.