Journal of Experimental Psychology: Applied

Dynamic Ensemble Visualizations to Support Understanding for Uncertain Trajectories

Jessica K. Witt and Benjamin A. Clegg

Online First Publication, June 17, 2021. http://dx.doi.org/10.1037/xap0000370

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000370

Dynamic Ensemble Visualizations to Support Understanding for Uncertain Trajectories

Jessica K. Witt and Benjamin A. Clegg
Department of Psychology, Colorado State University

When making decisions about uncertain spatial trajectories, such as storm forecasts, people rely on visualizations to support their understanding. Four experiments explored novel visualizations—dynamic ensembles. Nonexperts used visualizations to interpret probabilistic information about potential paths of a hurricane. Experiment 1 focused on global properties of the distribution, and showed dynamic ensembles imply a larger area at risk than traditional cones of uncertainty. Experiment 2 compared decisions with cones versus dynamic ensembles at specific individual locations. Dynamic ensembles offer more appreciation of risk outside the center of the distribution, and less abrupt in transitions from evacuation to nonevacuation choices. Experiment 3 compared decisions for dynamic ensembles versus static line ensembles. Similar evacuation rates across the two conditions suggest ensembles, rather than dynamics, are the more critical feature. Experiment 4 examined whether an additional dimension can be included in dynamic ensembles using color coding. Decisions reacted to this ancillary feature, with higher evacuation rates for locations threatened by more severe outcomes. Outcomes highlight the ability to systematically vary the level of risk communicated through the ensembles while also communicating the continuous nature of the risk. The overall findings show the viability of presenting uncertain spatial information using dynamic ensembles.

Public Significance Statement

Current approaches to showing the potential path of an incoming hurricane have well-known flaws, and this research investigates a new method to allow people to understand the areas that might be at risk. Findings from several experiments demonstrate that showing multiple, simultaneous, fast moving icons to illustrate the spread of possible tracks can provide a different sense of the possible threat, and specifically, these "zoomies" better convey risk that the storm will deviate from the most likely forecast path. Furthermore, these displays offer several design opportunities that could be used to communicate additional factors such as the magnitude of the storm and thus could afford a more complete communication of hurricane threats.

Keywords: visualizations, ensembles, decision making, hurricane forecasts

As a tropical cyclone such as a hurricane or typhoon approaches, people have to decide how to prepare. Preparations can range from doing nothing at all to packing as many items as possible and evacuating the area. Although an array of factors will influence their decisions (see Cox et al., 2013; Dombroski et al., 2006; Goldberg et al., 2020; Pham et al., 2020), one key element identified is risk perception (Thompson et al., 2017). In the case of an approaching storm, gauging risk depends critically on understanding the potential path (e.g., Broad et al., 2007). Greater lead time naturally facilitates evacuations and is especially important for the timely evacuation of large populations. But, there is a trade-off due to the nature of weather forecasting with larger look ahead time associated with more variability in the possible trajectory, which complicates decision making (Regnier, 2008; Regnier & Harr, 2006).

Reliable and appropriately presented information about uncertainty can improve decision making (e.g., Kirschenbaum & Arruda, 1994; MacEachren et al., 1998, 2012; Nadav-Greenberg & Joslyn, 2009). However, nonexperts are often poor at grasping spatial variability even when they directly experience a set of instances from a distribution (Herdener et al., 2016), including when the variability is their primary focus (Herdener et al., 2018). Wickens et al. (2020) point to the general tendency to underestimate spatial variability, but also highlight how a range of factors influence the extent to which that occurs. Importantly for the current context, according to the model (Wickens et al., 2020), rapid perceptual-based processing (e.g., Witt, 2019a) offers one potential mechanism to overcome a bias toward the underestimation of variability.

Jessica K. Witt https://orcid.org/0000-0003-1139-1599
Benjamin A. Clegg https://orcid.org/0000-0001-6026-5076
The authors have no conflicts of interest to disclose. This work was supported by the National Science Foundation (grant numbers BCS-1632222 and SES-2030059) and a fellowship from the CSU School of Global Environmental Sustainability to Jessica K. Witt and by Dr. Jeff Morrison and the Office of Naval Research (grant numbers N00014-17-1-2825 and N00014-20-1-2518 to Benjamin A. Clegg). Materials, data, and analysis scripts can be found on the Open Science Framework webpage (https://osf.io/fd682/).
Correspondence concerning this article should be addressed to Jessica K. Witt, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523-1876, United States. Email: Jessica.Witt@colostate.edu

The question then arises as to how best to convey the future risk to various locations from an inherently uncertain future path of a storm? This article presents initial evidence to support a novel method, dynamic ensembles, to enhance understanding of uncertain geospatial information such as forecasts for storm tracks. Although we ground this example in the context of hurricane paths, the same underlying methods can be applied to assist people in understanding other forms of spatiotemporal uncertainty, like the potential spread of wildfires, the search for a downed plane, or the hunt for a submarine.

Visualizations to Support Understanding

An effective way to improve understanding is through the use of visualizations (Card et al., 1999; Scaife & Rogers, 1996; Spence, 2001; Thomas & Cook, 2005; Tufte, 2001). Specifically, graphical depictions of relevant information can improve understanding and decision making (e.g., Hegarty et al., 2010; Sanfey & Hastie, 1998), which importantly extends to situations featuring spatial uncertainty (e.g., Cheong et al., 2016; Greis et al., 2018). However, there are challenges for presenting effective visualizations of geospatial information, and associated strengths and weaknesses for the different approaches that can be adopted (MacEachren & Kraak, 2000; Mason et al., 2017).

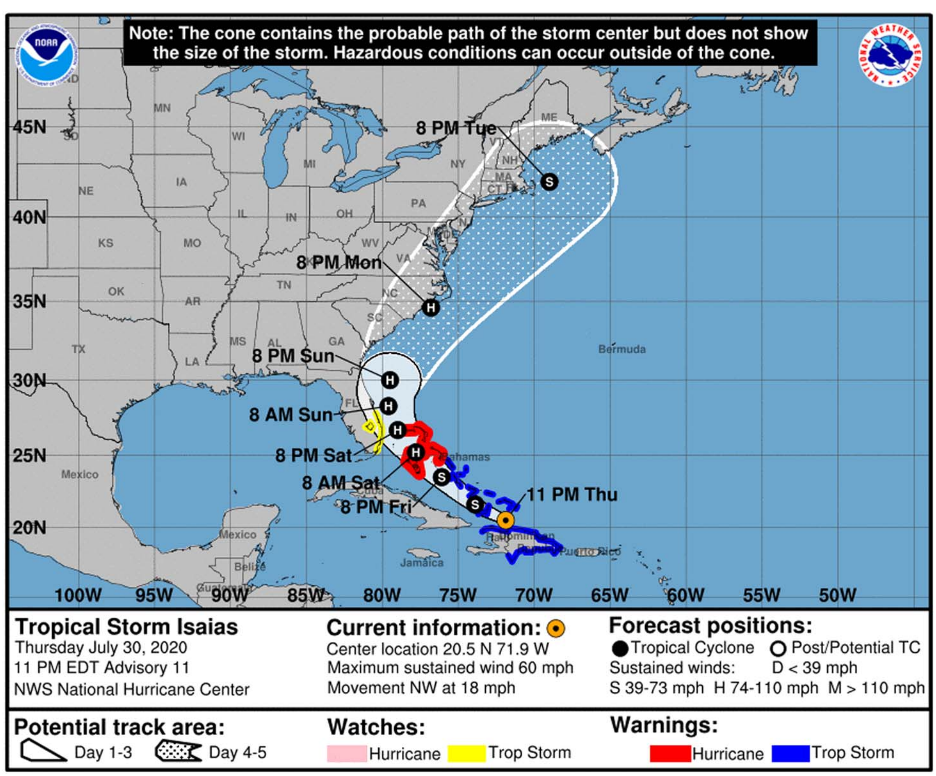
In the United States, one way the National Weather Service provides information about the potential path of storms is using National Hurricane Center Track Forecast Cones ("cones of uncertainty," e.g., see Figure 1). A cone of uncertainty represents deviation from most likely path, showing an area that captures two thirds of the forecast errors from a 5-year sample (National Oceanic and Atmospheric Administration, 2020) with a solid area showing the projection for 3 days, and a shaded area capturing out to 5 days. This same general approach of presenting a cone to represent potential variability of the path is commonly adopted by local weather forecasts presented to the general public (Broad et al., 2007), and these presentations are the primary source of information for the general public (Demuth et al., 2012).

Problems With the Cone of Uncertainty

Unfortunately, a number of problems have been identified with using cones of uncertainty and its attempt to convey the range of potential paths that a storm might take. At the most fundamental level, it is not always apparent to nonexperts what the cone represents. In looking at impressions in the general public, Broad et al. (2007) found that the people consistently misinterpreted or misapplied the features of cones, including an assumption that locations outside the cone would not be impacted, and example statements such as that the graphic represents past and future locations (rather than just current and forecast positions). The use of a single "most likely" path, to which the historically derived deviation around that is added as the cone, means the typical representation includes only one

Figure 1

National Oceanic and Atmospheric Administration Images of Cone of Uncertainty for Hurricane Isaias From the National Weather Service



Note. See the online article for the color version of this figure.

actual instance of a path. This creates a tendency to overly focus on that central forecast track of the storm.

Similarly, as measured in the context of lab-based experiments, people consistently confuse increased size of the cone of uncertainty, which represents an increase in uncertainty, with increased size or intensity of the storm instead (Boone et al., 2018; Padilla et al., 2017; Ruginski et al., 2016). Indeed one drawback of a standard cone representation is that, despite attempting to convey multiple dimensions of information, it only displays the most likely path of the storm and its historical potential for deviation from that path, and does not actually convey the important information on the predicted size of the storm and the area likely to be affected (although dots with letters are included to denote whether the predicted strength is that of a tropical storm or of a hurricane).

The boundaries of the cone also lend themselves to the *containment heuristic*, with the implication that locations just inside the boundary of the cone are incorrectly seen as disproportionately more probable than those just outside the boundary simply by virtue of being contained within the visualization (Padilla et al., 2018; see, e.g., McKenzie et al., 2016, and the related earlier example of the "within-the-bar bias," Newman & Scholl, 2012). For cones of uncertainty, this produces a large drop-off in perceived likelihood of paths outside the cone boundary compared to those inside, as if the cone contains all possible hurricane paths.

Another set of concerns result from probabilistic nature of the visualization. From the graphic alone (as shown in Figure 1), it is not self-evident that one would anticipate a 1/3 chance the storm's track will fall outside the cone. In addition, even with information about the probabilities, decision making could be improved through methods that expose individuals to concrete, natural frequencies of events rather than abstract, probabilistic information (see, e.g., Gigerenzer, 1994; Gigerenzer et al., 2005).

Experiments also raise questions about the basic efficacy of cone of uncertainty visualizations. For example, the findings from Ruginski et al. (2016) suggest that cone representations caused people to falsely assume that a storm is growing over time, and likely to inflict more damage, which would logically imply those same people do not understand that actual meaning: that the future path of the storm is increasingly uncertain. Moreover, in an abstract task that involved experiencing a set of instances from a spatial distribution, the use of cone visualizations did little to improve understanding of the variability (Pugh et al., 2018).

Other potential drawbacks with the use of cones of uncertainty in forecasts can be seen from real-world examples. As Hurricane Irma approached Florida in 2017, a number of forecast potential trajectories indicated it would pass up either the East or West coast of Florida, but the cone of uncertainty does nothing to convey the higher relative probabilities of those types of paths (see Figure 2 right panel). In 2004, Hurricane Jeanne was forecast to make a loop (which it did), but this behavior turned the cone of uncertainty into a circle (see Figure 2 left panel), and therefore failed to provide information to the viewer on how to best extrapolate its future path.

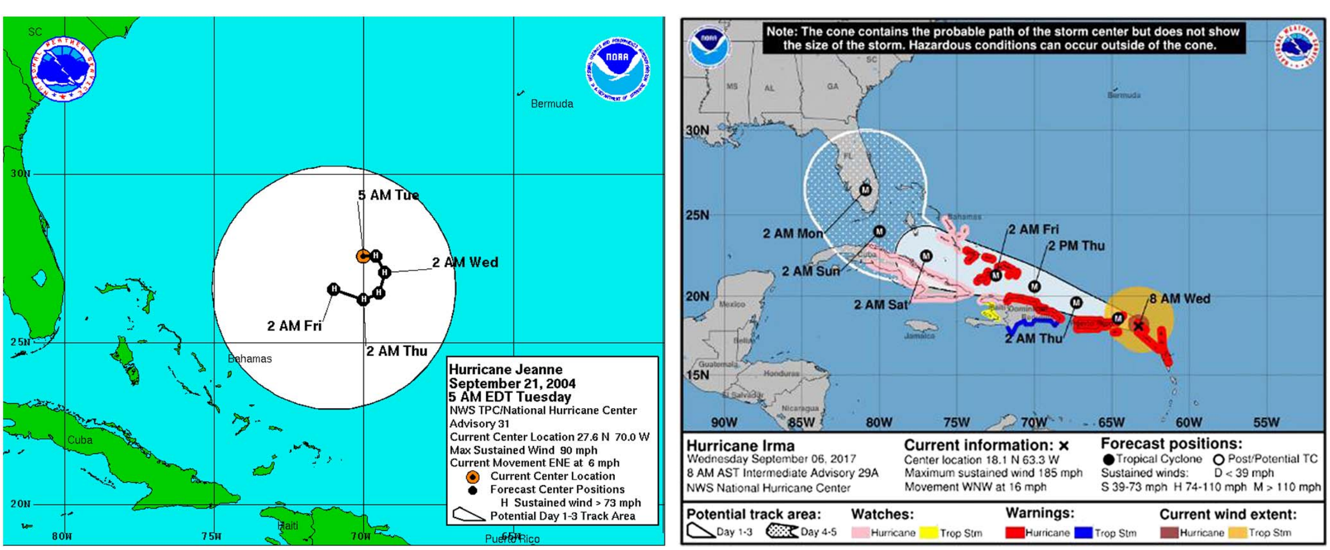
Alternative Visualizations

Although there are a variety of different types of visualizations employed in meteorology (for a review, see Rautenhaus et al., 2018), given emerging techniques and technologies that allow the production of multiple, parallel forecasts for a storm's trajectory, Hamill et al. (2012) proposed the potential value of ensemble techniques to convey uncertainty around tropical cyclones. An ensemble representation involves showing simultaneously multiple individual paths, with these instances illustrating divergence over time. A common class of these would be lagged ensemble track forecasts ("track ensembles") that show lines for several possible predicted paths.

Ensembles occur when collections of objects or features share common elements. The human visual system is adept at integrating across the individual items and extracting a summary of that

Figure 2

National Oceanic and Atmospheric Administration Images of Cones of Uncertainty for Hurricane Jeanne (Left Panel) and Hurricane Irma (Right Panel) From the National Weather Service



Note. See the online article for the color version of this figure.

information (see Whitney & Yamanashi Leib, 2018, for a comprehensive review). People can readily extract the average size, direction of motion, or length and orientation of a set of entities (e.g., Miller & Sheldon, 1969; Watamaniuk & McKee, 1998). Importantly for the current context, ensemble perception includes an ability to derive not just the average of the group, but also an accurate sense of the variability (e.g., Norman et al., 2015).

The inherent ability of the human visual system to process ensembles across multiple levels is leveraged in ensemble visualizations. Although one approach to visualization, as in the case of the cone of uncertainty, is to provide a graphic that attempts to summarize the average trajectory and in some form represent the possible variability; ensemble visualizations convey multiple instances that reflect a range of potential outcomes and let the visual system extract information from those. This approach can overcome some of the pitfalls, problems, and biases associated with the cone of uncertainty that are outlined above. As discussed above, the use of short time course perceptual information can also overcome a tendency to underestimate the variability present (see Wickens et al., 2020). In addition, because multiple ensemble features can be extracted simultaneously, this offers the potential to increase the amount of information that could be effectively communicated—adding, for example, dimensions to convey the predicted size or intensity of different storm tracks.

There has been research examining line, scatter, and heat map-based ensembles and their potential advantages versus various forms of cones of uncertainty (see, e.g., Cox et al., 2013; Ruginski et al., 2016). Cox et al. (2013) looked at a set of specific hurricane cases, comparing ensembles generated from overlaid tracks that faded out gradually to cones. Although this visualization changed behavior, the findings were somewhat mixed and failed to demonstrate a consistent benefit for ensembles. In addition, this work looked only at relative estimates of probabilities within eight sectors across 360°, and the authors note the importance of also including fine-grained decisions in future studies. Ruginski et al. (2016) employed damage judgments as measures, and although their findings suggest some important misunderstandings associated with cone visualizations, the results do not speak directly to evaluations of the likely paths of the storms.

One problem with the use of ensembles is a propensity for some participants to make judgments based on the individual instances presented (Padilla et al., 2017; Pappenberger et al., 2013; Ruginski et al., 2016). Rather than extracting general properties from the ensemble, at least some individuals focused on specific paths of the instances presented. For example, in one experiment, participants judged which of two oil rigs was more likely to be damaged by the impending hurricane. The rig closer to the center of the predicted path is always the correct answer. When this rig was located on a specific track, participants always selected this rig. However, they selected the farther rig 40% of the time when that rig was located on one of the specific paths and the closer rig was not. In other words, they were biased to believe rigs located *on* the paths were at higher risk than rigs located off the paths and biased to believe this information was nearly as relevant as the general properties of the ensemble (Padilla et al., 2017).

An alternative ensemble technique to showing predicted hurricane paths is to instead show an ensemble of possible hurricane positions at various time points (Liu et al., 2015, 2017). For example, rather than convey the paths, a display could convey anticipated risk and uncertainty at various locations using scalar fields with high saturation depicting areas at highest risk (Liu et al., 2015). Alternately, a display could also depict anticipated location and magnitude of the storm by positioning icons resembling hurricanes along the areas at risk and labeling each icon with a numeric code (e.g., 5 = category 5 hurricane). These displays encode anticipated storm severity using numbers and encode uncertainty using the density or quantity of icons (Liu et al., 2017). These displays led to increased rating of expected damage to locations at the center of the predicted path, but failed to communicate the increased risk beyond the center paths (Liu et al., 2017). It seems that the icons also led people to use the containment heuristic (Padilla et al., 2018).

Dynamic Ensemble Visualizations

So how should hurricane predictions be visualized? We propose the use of a dynamic ensemble that resembles hurricane trajectories. A selection of predicted hurricane paths is shown with marks that move, dynamically, along each of the predicted paths. With an increase in digital communication, dynamic or animated visualizations are a realistic option to convey information to large numbers of people. A dynamic ensemble leverages powerful and automatic ensemble processes in the visual system, has the advantage of presenting information using natural frequencies rather than probabilities, and requires less explanation for how the various marks map onto the underlying concepts (Hullman et al., 2015). Aligning the compatibility between the visual presentation and the underlying concepts is important for visualization comprehension (Witt, 2019b).

With respect to understanding hurricane forecasts, people understand the threat to areas located at the center of the forecasted path. What is lacking is a clear understanding that areas beyond the boundaries of the cone of uncertainty still have some risk, and that this risk declines with increased distance. Thus, to determine whether the dynamic ensemble improves understanding, we measured performance in two ways. One was to assess whether people would set the zone to be evacuated as greater after seeing the dynamic ensemble than after seeing the cone of uncertainty. The second method was to measure evacuation decisions for specific locations—towns located at the center, on the edge, and beyond the edge for both types of visualizations. To preview our results, the data showed increased understanding for risk at the edges after seeing the dynamic ensemble than after seeing the cone of uncertainty.

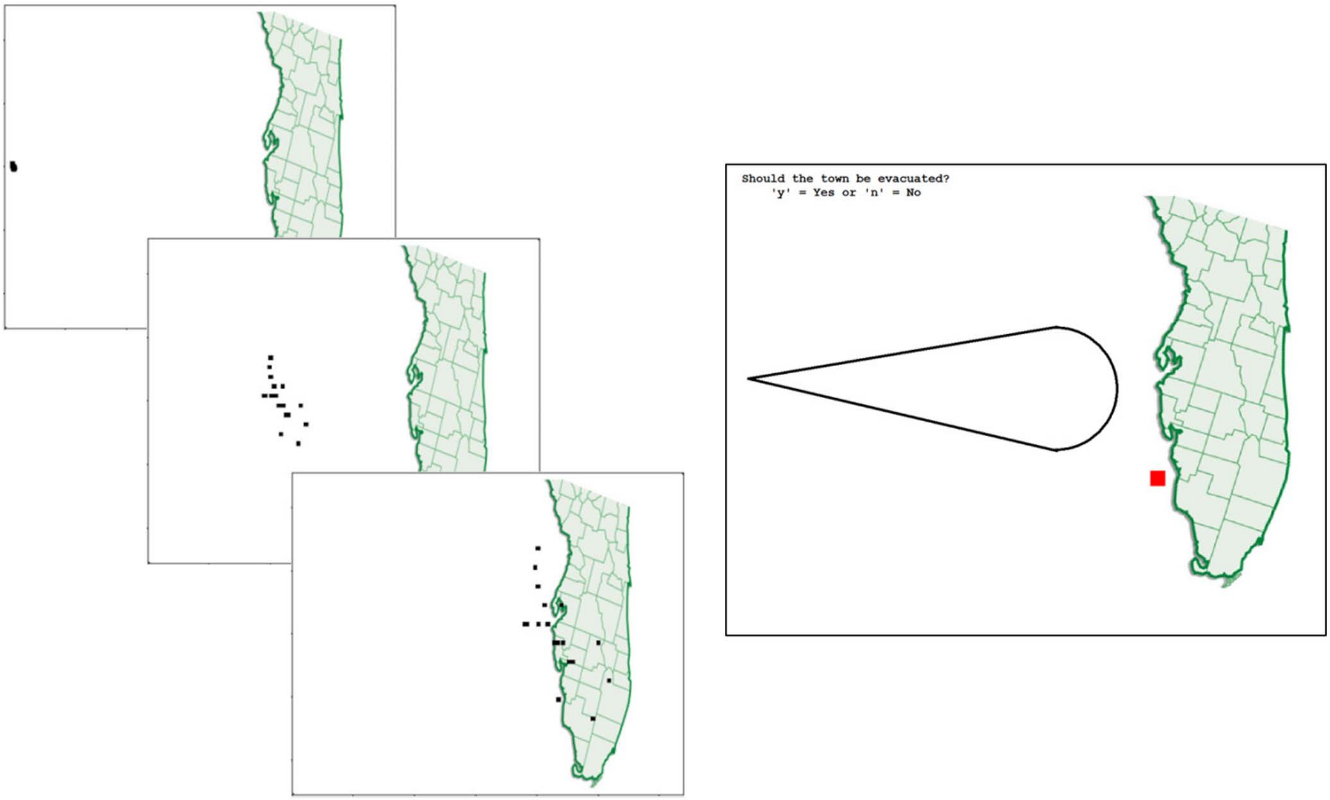
Experiment 1: Setting an Evacuation Zone

The first experiment compared decisions made when viewing a cone of uncertainty to convey the extent of the area under threat versus a dynamic ensemble that demonstrated an array of possible paths with a probabilistic distribution (see Figure 3). Given that, as discussed above, people view paths outside of the cone of uncertainty to be unlikely, the key question was whether extracting information from dynamic ensembles offers a sense that a larger area could be under threat. In this initial experiment, to best assess whether ensembles can convey effective information about the potential variability of a storm's path, participants were asked to judge the total area that needed to be evacuated.

Participants viewed a hurricane forecast prediction then adjusted the size of the evacuation zone. Forecast predictions were presented as a cone of uncertainty or with a dynamic display of potential paths, hereby referred to as zoomies. We predicted larger evacuation zones for the dynamic ensemble displays than the cone of uncertainty displays.

Figure 3

Illustration of the Zoomies (Left) and the Simplified Cone of Uncertainty (Right)



Note. The zoomies moved from left to right. Here, successive screenshots are shown to illustrate their paths. All zoomies traveled in straight linear paths with smooth and continuous motion from their origination point but varied in angle and speed to create the scattering of potential paths for the incoming storm. An animation of the zoomies can be found at https://osf.io/fd682/. See the online article for the color version of this figure.

Method

Design

The study measured the evacuation zone size set by participants across 2 types of visualization (cones vs. zoomies) × 3 levels of prediction uncertainty (low, medium, and high). To maximize statistical power to detect effects and minimize the role of individual differences in interpretation, the studies employed a within-subjects design. The first visualization presented was counterbalanced across participants. Participants completed 72 trials for each visualization condition for a total of 2,160 trials in the analysis.

Participants

Nineteen participants volunteered in exchange for course credit. Pilot studies were conducted with 7–15 participants per experiment (Witt et al., 2020). With so many trials per participant and such large effects, not many participants were needed to detect significant effects. Participants were recruited from the pool of students taking introductory psychology courses. We did not ask whether they had ever lived in a geographic location for which hurricanes were a risk. As our governor pointed out, Colorado is not at risk for hurricanes.[1] Over 70% of the undergraduates in our college are from Colorado, and only around 5% of undergraduates throughout the university are from states or territories that are the most hurricane-prone areas in the country. In this case, however, there is an advantage to assessing the basic properties of these visualizations using naive participants whose judgments are not likely to be confounded by frequent previous exposure to hurricane forecasts and outcomes (we return to this topic in the Discussion). All experiments were approved by the Colorado State University Institutional Review Board.

Stimuli and Apparatus

The stimuli were viewed on a 19” computer monitor with 1,280 × 1,024 resolution. On the right side of the screen, a map of a coastline was presented (see Figure 3). There were two types of trials: cones and zoomies. On cone trials, a cone (created as two lines and an arc connecting them) was presented to the left of the coastline. The left–right length of the cone was 18 cm. The width of the cone was set to 1 of 3 widths depending on the level of spread. For the narrow spread, the width was 4.7 cm. For the medium and wide spreads, the widths were 7.1 and 9.4 cm, respectively. These three levels of spread corresponded to three levels of *prediction uncertainty* (low, medium, and high, respectively). A narrow spread meant there

[1] https://www.facebook.com/PolisForColorado/posts/even-though-colorado-is-not-at-risk-for-hurricanes-scientists-at-colorado-state-/10157383273703921/

was less uncertainty about the predicted hurricane path. The center of the cone was either at the center of the screen, 0.6 cm above the center, or 0.6 cm below the center of the screen. This corresponded to the manipulation of the angle of the storm (up, center, or down). The cone was shown for 500 ms before the evacuation zone appeared.

On the zoomie trials, 18 small squares (0.44 cm^2) originated from the center of the left side of the screen (1.5 cm from the left edge) and moved toward the coastline. Their distribution approximated a normal distribution with 4 zoomies just above the center path, 4 just below, 2 each on either side of that, and 3 evenly spaced outside each of those (see Figure 3). The 12 in the center (67%) would all have been inside the cone had the cone been present (cones and zoomies were never presented at the same time). As the spread increased, the spatial paths of the zoomies were more spread apart. The spatial paths were also manipulated to be up, center, or down depending on the angle of the storm. Each zoomie moved at a constant speed for which some random noise was added at the beginning of each trial. The speeds ranged from 1.2 to 1.9 m/s. It took approximately 144 ms for them to make their movements.

Procedure

Participants gave informed consent and then the following instructions were provided on self-paced introductory screens: "You are in charge of evacuating a coastal region when hurricanes approach. If you choose not to evacuate a zone and a hurricane hits, damage will be extensive and costly. If you choose to evacuate a zone and the hurricane does not hit there, money will be spent on the evacuation for nothing. Thus, there are benefits and costs to selecting smaller and larger evacuation zones. Regions must be evacuated 12 hr in advance of when the hurricane will hit. For each decision, a hurricane is hovering and is approximately 12 hr away, so it will be time to make your decision. You will see a model that shows the best predictions of the hurricane's anticipated path. *The model shows a region that contains 67% of the predicted paths*. Hurricanes are unpredictable, so make your best guess of how much of the coast to evacuate." None of the text was italicized but it is here because this is the part that was different depending on the condition. For the zoomies condition, the italicized text was replaced with "The model shows animations of the most likely potential hurricane paths." Everything else about the instructions was the same.

Each trial began with a fixation screen for 1,000 ms. For cone trials, the coast line and the cone of uncertainty appeared. After 500 ms, a red rectangle signaling the evacuation zone appeared. The rectangle was either 1.5 cm or 5.9 cm tall. Participants used the arrows on the keyboard to move the evacuation zone up and down. Each key press corresponded to a move of 0.44 cm. Participants used the numbers 1, 2, 4, and 5 on a keypad to make the evacuation zone bigger or smaller by small increments (0.1 cm) or big increments (0.3 cm). Their task was to set the evacuation zone as they saw fit before continuing on to the next trial. Participants completed 72 trials in the cone condition (3 spreads × 3 angles × 2 start sizes × 4 repetitions) and the same number of trials in the zoomies condition. Order of visualization condition was randomized across participants. Trial order within block was randomized.

Results and Discussion

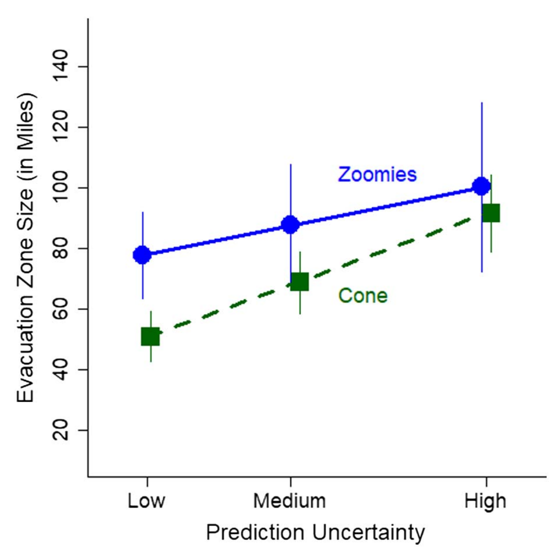
Five participants did not complete the experiment, although one completed all but 4 trials. The other four were excluded from analyses. Evacuation zones and prediction uncertainty spreads were transformed into approximate miles based on the assumption that the visible coastline was 800 miles. In this metric, the width of the cone for the 3 levels of prediction uncertainty were 70, 105, and 140 miles, respectively. The sizes of the evacuation zones were analyzed with a linear mixed model. The dependent variable was evacuation zone size. The fixed effects were visualization type, prediction uncertainty, and their interaction. Visualization type was entered as a factor with the zoomies as the reference. Prediction uncertainty was minimum centered by subtracting 70 miles. The random effects were intercepts and slopes for each fixed effect by participant.

The main effect of visualization type was significant, $t = -4.13$, $p = .001$, estimate $= -26.87$ miles, $SE = 6.51$ miles. Participants made the evacuation zone nearly 27 miles wider after viewing the zoomies than after viewing the cone of uncertainty. The main effect of prediction uncertainty was significant, $t = 3.22$, $p = .006$, estimate $=$ 0.29 miles, $SE = 0.09$ miles. For each increase of 40 miles in prediction uncertainty, participants set the evacuation zone to be 11 miles wider. The interaction between visualization type and prediction uncertainty was significant, $t = 2.61$, $p = .020$, estimate $= 0.23$ miles, $SE = 0.09$ miles. As prediction uncertainty increased, participants increased the size of the evacuation zone at a greater rate when viewing the cone of uncertainty than when viewing the zoomies (see Figure 4).

In both conditions, participants set the size of the evacuation zone to cover the most likely storm paths, while excluding predicted storm paths located at the edge and beyond the cone or predicted storm paths where the density of the zoomies was low. However, the size of the evacuation zone was set to be larger with the zoomies than with the cone of uncertainty. This suggests that participants consider larger areas to be at risk with the zoomies. The findings from Experiment 1 imply potential real-world value from employing dynamic ensembles to convey the possible future path of an incoming storm, specifically through benefits

Figure 4

Size of the Evacuation Zone as a Function of Prediction Uncertainty and Visualization Type for Experiment 1



Note. Lines represent linear regressions and error bars correspond to 95% CIs calculated from the model. See the online article for the color version of this figure.

from increased appreciation of uncertainty being represented. In anticipating an actual storm, this suggests dynamic ensembles could lead to earlier preparations for potential evacuation by more people.

Experiment 2: Decision to Evacuate a Town

Although Experiment 1 provides evidence that dynamic ensembles when viewed globally have the potential to be an effective visualization, the nature of the judgment in the first study may have encouraged people to think about the general properties of the ensemble. Although emergency planners may have to make decisions about the size of an area to be evacuated, most individuals are making decisions about whether to evacuate their specific location. This shift in focus from the area potentially impacted to the need to evacuate a single location could change the effectiveness of dynamic ensemble visualizations. Instead of assessing the properties of the ensemble, single-location judgments might make participants prone to focus on the individual instances (as has been observed with static ensemble displays) overriding the benefits from the ensemble information. Thus, Experiment 2 was the same as Experiment 1 except instead of setting an evacuation zone, participants were shown a single town and had to decide whether to evacuate it.

We hypothesized that participants would have similar understanding of risk for locations at the center of the storm's predicted path but that participants would perceive locations beyond the center as being low risk for the cone of uncertainty (thereby demonstrating the aforementioned containment effect; Padilla et al., 2018) and at a relative higher risk for the zoomies. In other words, we predicted that people would not be prone to the containment effect with the zoomies, but it would occur with the cone of uncertainty.

Method

Design

The experiment employed a fully within-subjects 2 (types of visualization: cones vs. zoomies) × 3 (levels of prediction uncertainty) × 6 (town location; labeled as town zone) design. Visualization condition was blocked; starting order was counterbalanced across participants. Prediction uncertainty and town zone were randomized within block.

Participants

Twelve students were recruited in the same way as in Experiment 1 and participated in exchange for course credit.

Procedure

Everything was the same as in Experiment 1 except the display showed a single town, and participants decided whether to evacuate the town. The following instructions were given prior to the zoomies block:

> You are in charge of evacuating a town when hurricanes approach. The town will be marked with a red square. If you choose not to evacuate the town and a hurricane hits, damage will be extensive and costly. If you choose to evacuate the town and the hurricane does not hit there, money will be spent on the evacuation for nothing. Thus, there are benefits and costs to evacuating the town. Towns must be evacuated 12 hr in advance of when the hurricane will hit. For each decision, a hurricane is hovering and is approximately 12 hr away, so it will be time to make your decision. *The forecast shows animations that illustrate some of the potential paths the hurricane might take.* However, hurricanes are unpredictable, so make your best guess of whether to evacuate the town.

The text shown to participants was not italicized. The text in italics was the only part that differed for the instructions with the cone block, which instead stated "The model shows a region that contains 67% of the predicted paths."

On each trial, a blank screen was shown for 20 ms. Then, the coastline appeared. The town was a red square that was 0.9 cm^2 and appeared along the coast. With equal frequency, the town could be in a location corresponding to above and outside the cone boundary, within the cone boundary, or below and outside the cone boundary. Within each area, the specific placement was randomized using a uniform randomizer. When the town was above or below the cone, the town was limited to be at least 1.5 cm from the top or bottom of the display. The same placement rules were used on the zoomie trials, so the town could be above, within, or below the center 12 zoomies.

Participants completed 108 trials with each visualization (3 levels of prediction uncertainty × 3 storm angles × 3 town positions × 4 repetitions). Order within condition was randomized, and order of condition was randomized across participants.

Results

Four participants did not complete the experiment and were excluded from the analyses. The data that were analyzed constituted 6,912 trials (8 participants × 2 blocks × 108 trials). We analyzed four performance metrics. One performance metric was the proportion of evacuations for towns located within the areas most likely to be hit by the hurricane. These were towns that were or would have been located within the cone of uncertainty. Another performance metric was the proportion of evacuations for towns located in areas less likely to be hit by the hurricane. These were towns that were or would have been located beyond the boundaries of the cone of uncertainty. We were also interested in how behavior transitioned from high-likelihood regions to low-likelihood regions. We discuss our two novel measures of transition behavior below. For each performance metric, we assessed how visualization type and prediction uncertainty affected evacuation decisions.

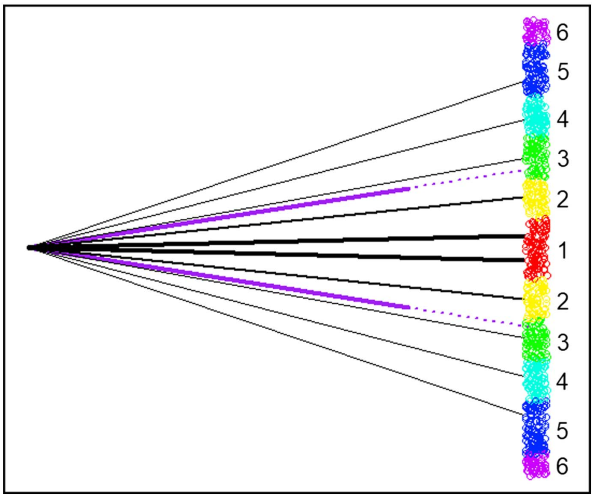
We recoded town position into zone such that towns located at the center of the predicted path were coded as Zone 1, towns located beyond zone 1 but within the boundaries of the cone were coded as Zone 2, towns located at the cone's edge as Zone 3, and towns located beyond the cone as Zones 4–6 as distance from center increased. Zone 6 was beyond the boundary of the most extreme zoomies (see Figure 5). How evacuation decision varied across zones is shown in Figure 6.

Highest Risk Town Decisions

To assess decisions for in-cone towns, we conducted a general linear mixed model with data for which town were or would have been located well within the boundary of the cone. These included Zones 1 and 2. The dependent variable was the binary decision to evacuate or not (coded as 1 and 0). The within-subjects fixed effects were visualization type (cone, zoomies), prediction uncertainty (low, medium, high coded as −1, 0, 1, respectively), and their interaction.

Figure 5

The Black Lines Show the Mean Path Traveled by Each Set of Zoomies for the Medium Level of Prediction Uncertainty and the Middle Angle



Note. The thickness of the line indicates the number of zoomies traveling along that path (4 zoomies for thickest lines, 2 zoomies for medium lines, and 1 zoomie for thinnest lines). The solid purple lines show the outer edges of the cone of uncertainty, and the dotted purple lines show how these lines would extend to the town location. All town locations for this condition across all participants are shown as points. The points are color coded according to zone, and each zone is labeled. See the online article for the color version of this figure.

Random effects were included for participant. The model was singular or did not converge when we included random slopes for both trial range and prediction uncertainty (including or excluding the interaction between them). The model converged with either random slope included and with all random slopes excluded. The three models produced nearly identical outcomes with respect to the interaction. The models differed for the main effects such that the p values were less when a factor was not included as a random slope. Thus, we present the model with no random slopes to show the minimum p values for the main effects, none of which were significant.

Participants were highly likely to evacuate towns located in the areas most likely to get hit, $M = 0.95$, $SEM = 0.01$. Evacuation decisions did not vary across visualization type, estimate = 0.17, $SE = 0.09$, $z = 1.83$, $p = .068$. Evacuation decisions also did not vary across prediction uncertainty, estimate = −0.16, $SE = 0.11$, $z = -1.49$, $p = .137$. However, as prediction uncertainty increased, evacuation decisions differed between the two visualization types, estimate = −0.577, $SE = 0.11$, $z = -5.26$, $p < .001$. For the cone condition, evacuation decisions increased as the prediction uncertainty increased, $p = .004$ (Ms = 0.93, 0.95, 0.97), whereas for the zoomies, evacuation decisions decreased as prediction uncertainty increased, $p < .001$ (Ms = 0.98, 0.96, 0.93).

These results (see also Witt et al., 2020) suggest that at least some people derive a sense that changes in the uncertainty of the path of the tracks imply different need to evacuate from what would be high-probability areas along the center of the distribution. One possibility is that, as in the case of track ensembles (Padilla et al., 2017), people are using instances to determine the risk—that is, making judgments based on one of the tracks intersecting the location or passing close by. We return to this issue in the General Discussion.

Lower Risk Town Decisions

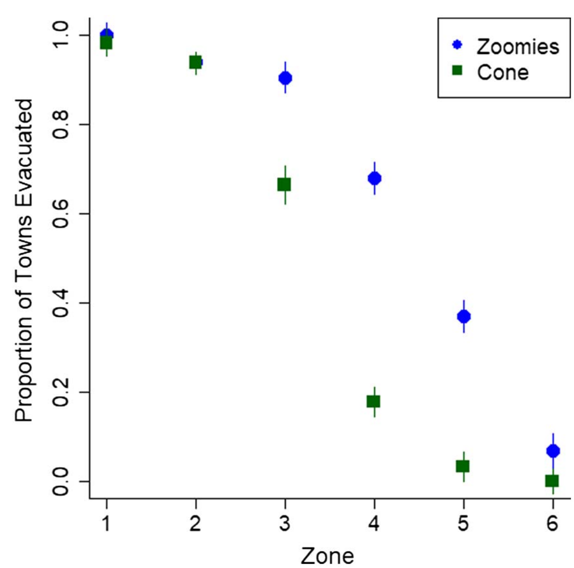
To assess decisions for towns located beyond the boundary of cone (or the equivalent location for the zoomie condition), we conducted a general linear mixed model on trials for towns located in this range. This included towns located in Zones 4–6. The dependent variable was the binary decision to evacuate or not (coded as 1 and 0). The within-subjects fixed effects were visualization type (cone, zoomies), prediction uncertainty (low, medium, high coded as −1, 0, 1, respectively), and their interaction. Random effects were included for participant, including random slopes for both fixed effects and their interaction.

Decisions to evacuate these towns that were beyond the central zones of the predicted hurricane path varied across visualization type, $z = 8.50$, $p < .001$, estimate = 1.99, $SE = 0.23$. Participants were far more likely to evacuate these towns when the visualization was zoomies ($M = 0.39$, $SEM = 0.07$) than when the visualization was the cone of uncertainty ($M = 0.08$, $SEM = 0.02$).

Participants were more likely to evacuate these towns as prediction uncertainty increased, $z = 5.17$, $p < .001$, estimate = 0.48, $SE = 0.09$. Thus, less certain predictions made participants evacuate a greater region. The interaction between visualization type and prediction uncertainty approached significance, estimate = −0.20, $SE = 0.11$, $z = -1.77$, $p = .078$. As prediction uncertainty increased, the increase in evacuations increased for both the cone of uncertainty, $p = .015$, and the zoomies, $p < .001$, and the increase in evacuation rate was almost double for zoomies as for the cone (Ms = 0.48, 0.28, respectively).

Figure 6

Median Proportion of Towns Evacuated by Eccentricity of Location for Cone and Zoomies Conditions for Experiment 2



Note. Medians were calculated after computing a mean for each participant. Error bars are 1 *SEM* calculated within-subjects. See the online article for the color version of this figure.

Transitions in Evacuation Decisions

To understand how participants' responses transitioned from being more likely to evacuate to being less likely to evacuate, we analyzed their data with a general logistic mixed regression. The dependent measure was whether the town was evacuated (coded as 1) or not (coded as 0). The fixed effects were the town's zone, the visualization type, and the prediction uncertainty. All two-way and three-way interactions were also included. Random effects for each participant were included (intercepts and slopes for each fixed effect). Zone was centered by subtracting 3, so all main effects and interactions that did not include zone correspond to responses when towns were at the boundary of the cone. Visualization type was entered as a factor with the cone condition serving as the reference factor. Prediction uncertainty was centered and entered as −1, 0, and 1 from less to more uncertain, respectively.

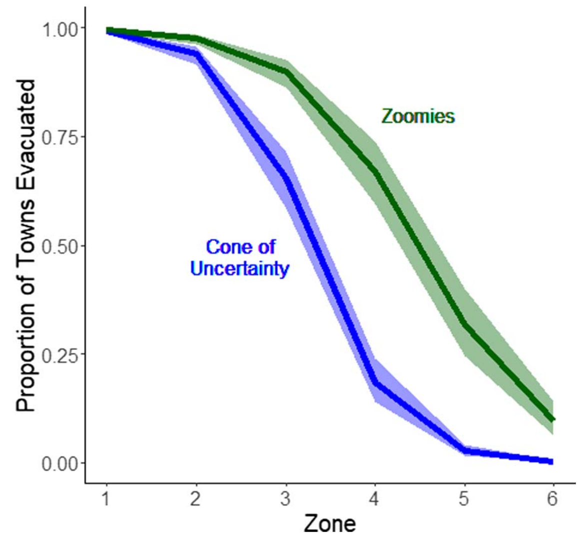
There are two performance metrics of interest. One is the strictness by which evacuation decisions transition from being more likely to evacuate to being less likely to evacuate. For example, if people only evacuate towns located in the cone and not beyond the cone, the decision transition would be extremely strict, thereby showing the containment effect (Padilla et al., 2018). In contrast, decisions might be more probabilistic with a gradual decline in evacuation rates. The statistical test of decision transition strictness is the interaction between visualization type and zone, with steeper slopes corresponding to stricter decisions. This interaction was significant, $z = 3.89$, $p < .001$, estimate = 0.64, $SE = 0.16$. As shown in Figure 7, decision transitions were stricter (steeper) for the cone of uncertainty than for the zoomies.

Decision transition strictness was affected by trial range, and this effect differed by visualization type as revealed by the significant three-way interaction, $z = 5.48$, $p < .001$, estimate = 0.42, $SE = 0.08$. For predictions with the lowest uncertainty, slopes were similar between the cone and the zoomies, $p = .23$, but for predictions with medium or highest uncertainty, the slopes for the cone were steeper than for the zoomies, $ps < .001$ (see Figure 8).

Another performance metric is the location at which decisions switch from being more likely to evacuate to being less likely to evacuate. This threshold was calculated using the coefficients from

Figure 7

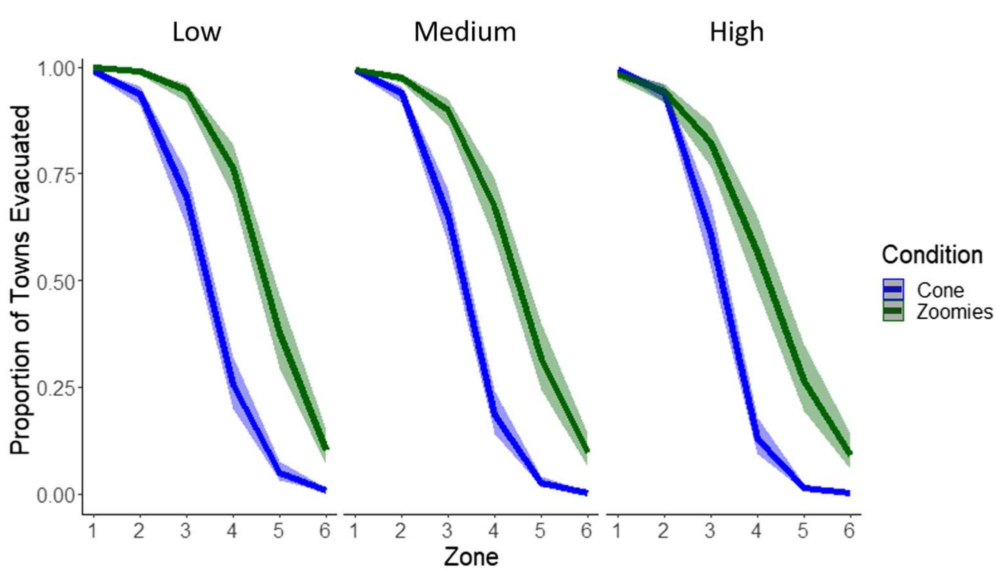
The Proportion of Towns That Were Evacuated is Plotted as a Function of Zone and Visualization Type From Experiment 2



Note. Shading corresponds to 1 *SEM* calculated from the model. See the online article for the color version of this figure.

Figure 8

The Proportion of Towns That Were Evacuated is Plotted as a Function of Zone, Visualization Type, and Prediction Uncertainty for Experiment 2



Note. Shading corresponds to 1 *SEM* calculated from the model. See the online article for the color version of this figure.

the model as the zone for which evacuation decisions were 0.50. For the cone of uncertainty, this point was 3.30, which means this point was approximately just beyond the edge of the cone. This value is what was expected based on claims that the cone of uncertainty lead to a containment effect (cf., Padilla et al., 2018). For the zoomies, the point was 4.48. Given that 6 was beyond the range of all zoomies, a value of 4.48 corresponds to the locations for which the zoomies were sparse but still present. Thus, the type of visualization altered the threshold at which decisions switched from more likely to less likely to evacuate.

Discussion

As in Experiment 1, these results suggest that dynamic ensembles imply to individuals a larger area at potential risk. Once again, the real-world value of employing this type of visualization might be seen in people recognizing that they may be at risk even if they are located some distance away from the single, most likely forecast path.

The outcomes show striking differences between the two visualization types. The cone of uncertainty lead to strict decision transitions with thresholds that were located just beyond the edge of the cone. In contrast, when presented with zoomies, a decision transitions were less strict, these transitions were more affected by prediction uncertainty, and the threshold was where zoomies were sparse. Thus, visualization type affected both transition strictness and the threshold for decisions.

Overall the findings of the second experiment suggest several positive prospective aspects for using dynamic ensembles in real-world contexts, including an increased sense of risk for peripheral locations, avoiding containment effects, and decisions that react to the level of uncertainty being portrayed.

Experiment 3: Zoomies Versus Track Ensembles

In the previous experiment, zoomies led to different kinds of behavior than the cone of uncertainty. Behavior consisted of less strict decision transitions, particularly as prediction uncertainty increased. In addition, towns that were at risk but further from high-probability paths were more likely to be evacuated with the zoomies visualization than with the cone of uncertainty. This raises the issue of whether these differences in behavior were due to the animated nature of zoomies or to the fact that individual prediction paths were presented rather than aggregated predictions as with the cone of uncertainty. Another format for presenting individual prediction paths is track ensembles (sometimes colloquially referred to as "spaghetti plots"). Track ensembles show the predicted hurricane paths as lines. Like zoomies, they show individual predicted paths, but they are not dynamic and thus do not have the same visual-conceptual compatibility as zoomies, which show the paths as a visual analog to storms moving toward the coast. We examined whether behavior would be equivalent when storm path predictions were presented via dynamic ensembles versus static track ensembles.

Method

Participants

Thirty-nine participants were recruited as before. We initially recruited fewer participants. Once it was clear there were little-to-no differences between evacuation rates for the dynamic ensembles and the static track ensembles, we recruited more participants to be able to differentiate between a small effect and a null effect. This type of sequential analysis requires adjusting α to avoid inflating type I errors. Based on simulations using the phackRM function (Sherman, 2014), α was adjusted to 0.018.

Design and Procedure

The procedure was the same as in Experiment 2 except instead of the cone of uncertainty, we compared zoomies to track ensembles. The track ensembles were 18 straight lines that followed the same parameters as the zoomies (as in Figure 5). Participants completed one block with the zoomies and one with the track ensembles; starting order was randomized across participants. A total of 16,772 trials were included in the analysis.

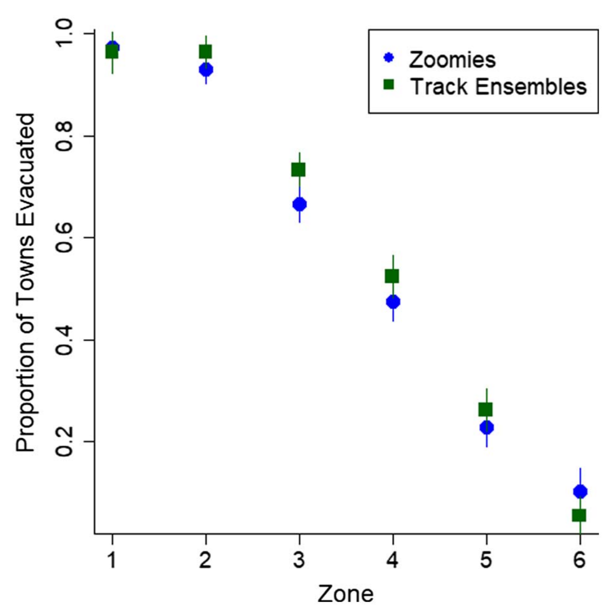
Results and Discussion

As before, we analyzed data for the towns most likely to get hit separately from the towns less likely to get hit, then we analyzed the transitions. An overview of the evacuation decisions is shown in Figure 9.

Evacuation decisions for towns in Zones 1 and 2 were analyzed with a general linear mixed model. The fixed effects were visualization type, prediction uncertainty, and their interaction. The random effects for participant included intercepts and both main effects. The main effect of visualization type was not significant, $z = -0.30$, $p = .76$, estimate $= -0.08$, $SE = 0.28$. The main effect of prediction uncertainty was significant, $z = -6.52$, $p < .001$, estimate $= -0.94$,

Figure 9

Median Proportion of Towns Evacuated by Eccentricity of Location for Zoomies and Track Ensemble Conditions for Experiment 3



Note. Medians were calculated after computing a mean for each participant. Error bars are 1 *SEM* calculated within-subjects. See the online article for the color version of this figure.

$SE = 0.14$. The interaction between visualization type and prediction uncertainty was significant, $z = 2.34$, $p = .019$, estimate $= 0.42$, $SE = 0.18$. Participants were more likely to evacuate when prediction uncertainty was lower than when it was higher, and this difference was greater for the zoomies condition (Ms $= 0.98, 0.96, 0.90$) than the track ensembles condition (Ms $= 0.97, 0.95, 0.93$, respectively).

Evacuation decisions for towns in zones 4–6 were similarly analyzed. There was no difference between visualization types, $z = 0.30$, $p = .76$, estimate $= 0.04$, $SE = 0.14$. As prediction uncertainty increased, evacuation rates increased, $z = 3.11$, $p = .002$, estimate $= 0.30$, $SE = 0.10$. This increase was slightly greater for zoomies than for track ensembles but was not significant, $z = 1.87$, $p = .062$, estimate $= 0.12$, $SE = 0.07$.

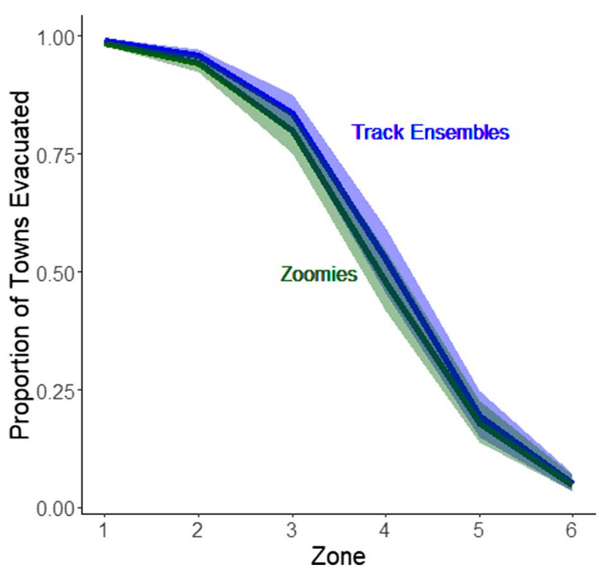
Decision Transitions: Strictness and Thresholds

To analyze decision transition strictness and thresholds, we ran a general binary linear model as in Experiment 2. The main effect of zone was significant, $z = -11.39$, $p < .001$, estimate $= -1.53$, $SE = 0.13$. As zone increased, evacuation decision decreased. The interaction between zone and visualization type was not significant, $z = 0.93$, $p = .35$, estimate $= 0.08$, $SE = 0.09$. The three-way interaction was also not significant, $z = -0.73$, $p = .47$, estimate $= -0.03$, $SE = 0.04$. These nonsignificant results suggest that the strictness of the decision transition was similar for zoomies and track ensembles (see Figure 10).

The two visualization types differed in how prediction uncertainty affected the threshold at which decisions switched from more likely to less likely, as shown by the significant interaction between visualization type and prediction uncertainty, $z = -2.55$, $p = .011$, estimate $= -0.17$, $SE = 0.07$. We calculated the threshold at which decisions were 0.50 to evacuate. For the track ensembles, these thresholds were 4.30, 4.07, and 3.80 as decision uncertainty increased. For the zoomies, these thresholds were 4.20, 3.95, and 3.66. Prediction uncertainty had a greater impact on thresholds for the zoomies than the track ensembles.

Figure 10

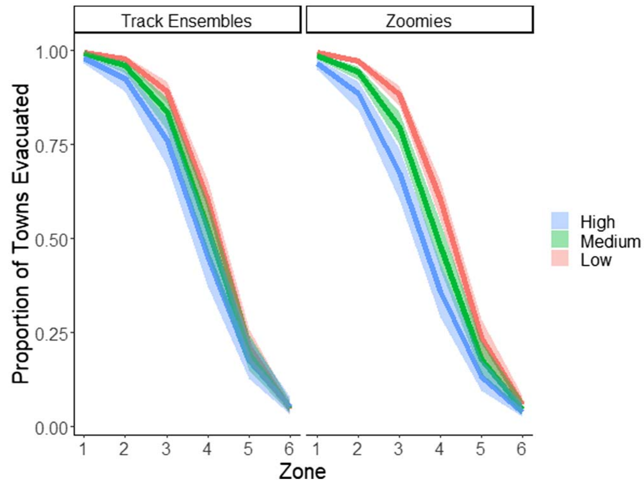
The Proportion of Towns That Were Evacuated is Plotted as a Function of Zone and Visualization Type From Experiment 3



Note. Shading corresponds to 1 *SEM* calculated from the model. See the online article for the color version of this figure.

Figure 11

The Proportion of Towns That Were Evacuated is Plotted as a Function of Zone, Visualization Type, and Prediction Uncertainty for Experiment 3



Note. Shading corresponds to 1 *SEM* calculated from the model. See the online article for the color version of this figure.

Overall, performance was similar for both the dynamic ensemble display (zoomies) and the static ensemble display (track ensembles). The only difference that emerged was when prediction uncertainty was high, the overall rate of evacuations was lower for the zoomies than the track ensembles (blue lines in Figure 11).

One known issue with track ensembles is that people overestimate the risk of towns located on one of the paths (Padilla et al., 2017). We did not code whether a line or a zoomie crossed the town in this experiment; future studies should explore whether this proximity bias exists for zoomies as it does for track ensembles. If zoomies do not lead to the same bias as track ensembles that would be an important advantage for zoomies over track ensembles. In general, the results suggest that ensemble displays, rather than animate displays, largely drive behavior following the presentation of zoomies.

We should also note that it is possible that dynamic ensembles have advantages over track ensembles that were not captured in the current experiment. For example, relative to a static display, a dynamic display may be more effective at communicating information about predicted timing of when the storm will progress to a certain location or differences in predicted storm speeds across potential paths. Future studies are needed to address these questions.

Experiment 4: Severity of the Storm

Although Experiment 3 suggests a high degree of similarity in the decisions from static track versus dynamic ensembles, a potential advantage of a visualization like zoomies is how other dimensions can be readily incorporated into the representation (for a related example see, Liu et al., 2018). For example, specific predictions about the severity of the storm across locations can be included as an additional feature. In real-world situations, the severity of a storm might vary depending on the trajectory it follows, with potential storm paths moving over shallower, warmer water leading to more severe outcomes. Experiment 4 tested whether people could effectively interpret such additional dimensions incorporated into the ensemble displays. Here, we represented severity by the color of the storm, with red indicating more severe and orange indicating less severe. We measured the effect of representing predicted storm severity on evacuation decisions. We hypothesized that participants would be able to take into account both the density of the zoomies and their color when making evacuation decisions.

Method

Design

The experiment employed a within-subjects 2 (storm intensity denoted: black vs. red/orange) × 3 (level of prediction uncertainty) × 6 (town zone) design. Whether storm intensity was denoted (using red and orange zoomies) or not (using all black zoomies) was blocked. The block that was presented first was counterbalanced across participants. A total of 6,676 trials were included in the analysis.

Participants

Participants were recruited as before. A total of 17 students completed the experiment.

Procedure

Participants completed two blocks of trials. One block was the same as the zoomie blocks in Experiments 2 and 3 for which the zoomies were black. In the other block, some of the zoomies were red and some were orange (see Figure 12). The red zoomies were either at the top or bottom of the 18 zoomies in which case 5 of the 18 zoomies were red and the rest were orange. Or the red zoomies were in the center in which case 8 of the 18 zoomies were red and the rest were orange. Prior to this block of trials, the instructions stated as follows:

> The task is the same except now you will see predictions that also tell you about the severity of the incoming storm. Orange means the storm is predicted to be moderate whereas red means the storm is predicted to be severe. The forecast shows animations that illustrate some of the potential paths the hurricane might take. However, hurricanes are unpredictable, so make your best guess of whether to evacuate the town. You decide whether or not to evacuate each town.

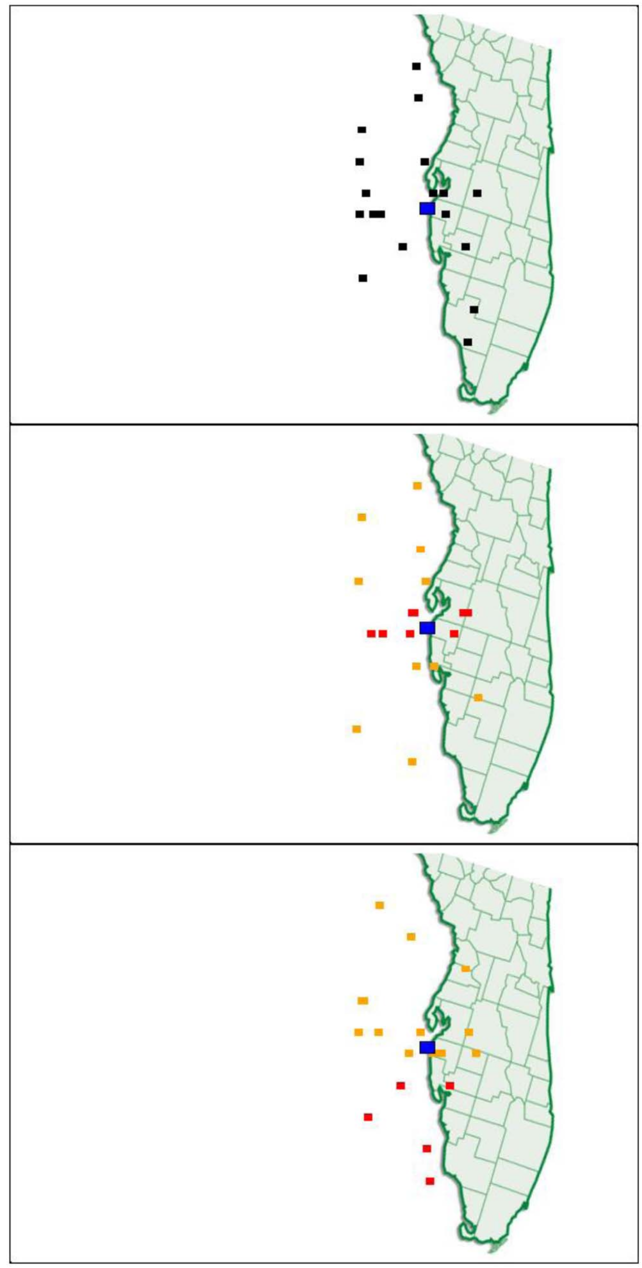
Results and Discussion

The mean values are shown in Figure 13. We analyzed the decision to evacuate (1 = evacuate; 0 = do not evacuate) with a binary general linear mixed model. The fixed effects were zone, the color of the zoomie that was closest to the town (black, red, and orange), prediction uncertainty, and all two-way and three-way interactions. Because no zoomies went into zone 6 by definition, we only included data for Zones 1–5 in this analysis. Zoomie color was entered as a factor with black as the reference. Zone and prediction uncertainty were centered as before. Random effects for participant included intercepts and slopes for each main effect and for the zone by color interaction. Model outcomes are shown in Figure 14.

The main effect of zone was significant, $z = -9.51$, $p < .001$, estimate $= -1.24$, $SE = 0.13$. Towns further from the predicted

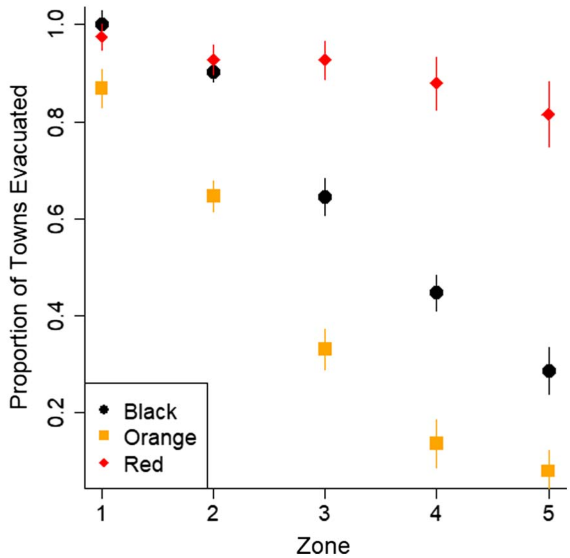
Figure 12

Illustrations of Differences Between Stimuli With All Black Zoomies (Top Panel) and Red and Orange Zoomies



Note. Middle panel shows a sample stimulus with red zoomies in the middle, and bottom panel shows a sample stimulus with red zoomies on the bottom. The town (blue square) is located in the middle in each display. See the online article for the color version of this figure.

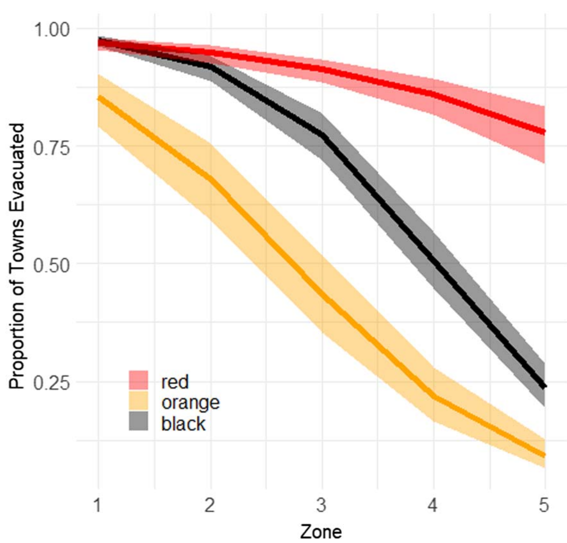
Figure 13
Median Proportion of Towns Evacuated by Zone and Zoomie Color Conditions for Experiment 4



Note. Medians were calculated after computing a mean for each participant. Error bars are 1 *SEM* calculated within-subjects. See the online article for the color version of this figure. See the online article for the color version of this figure.

center of the storm were evacuated at lower rates. Zoomie color affected evacuation rates. Towns were evacuated at a higher rate for red zoomies than black zoomies, $z = 4.69$, $p < .001$, estimate $= 1.11$, $SE = 0.24$. Towns were evacuated at a lower rate for orange zoomies than for black zoomies, $z = -5.76$, $p < .001$, estimate $= -1.65$, $SE = 0.29$.

Figure 14
Proportion of Towns Evacuated is Plotted as a Function of Zone and Zoomie Color (Black, Orange, or Red) for Experiment 4



Note. Shading represents 1 *SEM* from the model. See the online article for the color version of this figure.

Because black was the reference, the model did not give a direct estimate of red versus orange, but post hoc analysis showed evacuation rates differed between the two colors, $p < .001$.

The significant interactions between zone and zoomie color showed that evacuation rates decreased with distance to town at different rates across the colors (see Figure 14). The interaction between zone and black versus red zoomies was significant, $z = 4.19$, $p < .001$, estimate $= 0.67$, $SE = 0.16$. The evacuation rates decreased more rapidly for the black zoomies than for the red zoomies as distance to the town increased. The interaction between zone and black versus orange zoomies was significant, $z = 2.37$, $p = .018$, estimate $= 0.26$, $SE = 0.11$. Again, the evacuation rates decreased more rapidly for the black zoomies than for the orange zoomies as distance to the town increased. Post hoc analysis also showed a significant difference between red and orange zoomies, $p = .013$.

The outcomes of the analyses with prediction uncertainty are shown in Figure 15. The main effect of prediction uncertainty and two-way and three-way interactions with prediction uncertainty were all significant, *p*s $< .019$. To better interpret these effects, we reran the model separately for each zoomie color with zone, prediction uncertainty, and their interaction as fixed effects. For the black zoomies, both the main effect of prediction uncertainty and the interaction between prediction uncertainty and zone were significant, *p*s $< .001$. With lower levels of prediction uncertainty, towns located farther from the center of the predicted path were evacuated at a higher rate and the rate of evacuation decreased more dramatically than with higher levels of prediction uncertainty. For the orange zoomies, there was a main effect of prediction uncertainty, $p < .001$, but not a significant interaction between prediction uncertainty and zone, $p = .24$. The threshold at which storms were evacuated shifted with prediction uncertainty for the orange zoomies, but the rates of evacuation did not. For the red zoomies, evacuation rates did not significantly differ across levels of prediction uncertainty, $p = .25$, and the interaction between zone and prediction uncertainty was not significant, $p = .94$. Evacuation rates were similar regardless of prediction uncertainty for the red zoomies.

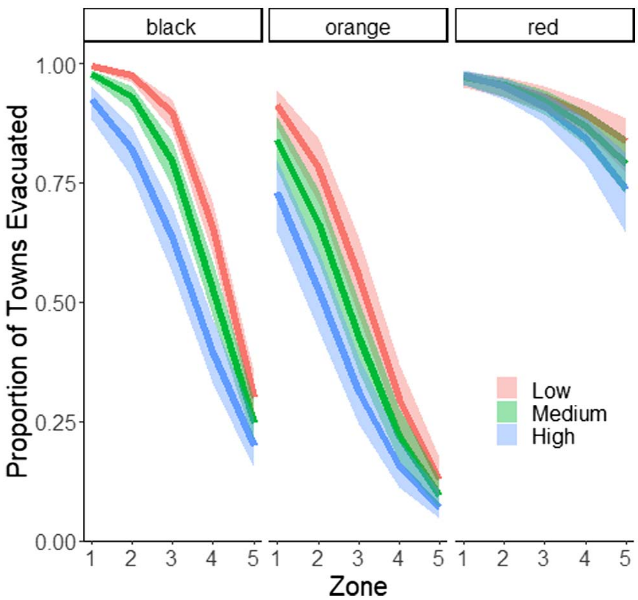
The key thing this experiment demonstrates is the potential to include other dimensions into dynamic ensembles, with such information factored into decisions in appropriate ways—in this case with higher severity storms showing greatly increased rates of evacuation in those areas. As predicted, people could extract multiple dimensions of information from the dynamic ensembles. This benefit of dynamic ensembles could be leveraged to include other types of information within a forecast beyond the current generic example of storm severity. Other potential elements might include changing the size of the zoomies to represent, for example, storm surge or changing the rate of rotation of an icon to convey wind speed threat. Although exploring such variants is beyond the scope of the present article, the value of representations that can successfully convey additional valuable information further enhances their utility.

General Discussion

Hurricanes do enormous damage, taking lives, destroying homes and businesses, disrupting infrastructure, and costing millions of dollars in repairs and lost tourism. Two important steps to reducing the damage are to better predict when and where hurricanes will occur and communicating these predictions to the public. Because

Figure 15

Proportion of Towns Evacuated Plotted as a Function of Zone, Zoomie Color (Red, Black, Orange), and Prediction Uncertainty (Low, Medium, High) From Experiment 4



Note. Shading represents 1 *SEM* estimated from the model. See the online article for the color version of this figure.

hurricane predictions have a large amount of uncertainty, communication to the public must make both the information and the uncertainty clear. However, people struggle to reason with uncertainty (e.g., Herdener et al., 2018), and have specific difficulties understanding current visualizations used to communicate hurricane predictions (e.g., Broad et al., 2007).

We propose a new way to visualize spatial uncertainty using dynamic ensembles and show how these "zoomies" can be utilized to convey the uncertain future path of hurricane predictions. Although the mechanisms to generate the instances for real storm ensembles are well beyond the scope of the present article, this could be accomplished given the presence of multiple different forecasts for storms, and the ability to rerun models with minor variations in parameters (see, e.g., Cox et al., 2013). Emerging cases where static track ensembles are already used for storms suggest such information can be generated. This new method does require animation. Animation would not have been feasible when news was primarily communicated through static media such as newspapers but is increasingly more feasible with digital media being more universal. Thus, the types of visualizations examined here could in the future plausibly be generated and could be presented to large numbers of nonexperts with no specialized equipment required.

Although there were no major differences between these novel dynamic ensembles and the static, track ensembles under the conditions employed here, the findings illustrated that ensemble visualizations avoid many of the problems associated with the current method of displaying the cone of uncertainty. The cone of uncertainty shows an increase in the size of the cone in an attempt to communicate the increase in uncertainty, but people misinterpret the visualization as showing the prediction that the storm will increase in size (e.g., Broad et al., 2007; Ruginski et al., 2016). Empirically, the cone of uncertainty leads to a containment effect (Padilla et al., 2018). People interpret the visualization as showing the only areas of risk being those contained within the cone, so they misunderstand that areas beyond the cone are also at risk. The cone of uncertainty can also be misleading if the predicted hurricane paths generally fall along two clusters of paths directions. Although the central path through the cone is at the highest risk (cones are generated around the most likely trajectory), risk may not necessarily be even along both sides of the cone depending on the cluster positions, and indeed the cone may not even necessarily encapsulate both clusters if one set of predictions diverges radically from the most likely path.

In contrast, dynamic ensembles do not have a clear outline. Our findings here showed that with the zoomies, people's decisions about what towns to evacuate did not show the same containment effect as with the cone of uncertainty. The zoomies eliminated the sense of a container and communicated the presence of risk in the surrounding areas. The potential to recognize greater uncertainty of the path might be vitally important for real world storms, where, for example, with Hurricane Katrina, the city of New Orleans did not fall inside the National Weather Service (NWS) cone of uncertainty until just a couple of days before it struck close to the city. In addition, the zoomies do not have features of the cone of uncertainty that are likely to be misunderstood, such as that it represents the size of the storm, rather than uncertainty that increases with time.

Although there can be important advantages from using a visualization like dynamic ensembles that offers heightened awareness of the possible threat from an incoming storm for those outside the current forecast most likely path, in real world situations these benefits need to be weighed against potential costs. For example, if larger areas see themselves at potential risk then more preparatory resources might be demanded, and if people move beyond preparations to actually begin to attempt to evacuate, then increased strain on infrastructure like roads could result. Critically, features of the zoomies can be used to convey threat such as the color of the zoomies (as shown in Experiment 4). Although not directly tested here, presumably the number or density of the zoomies also conveys the threat as shown by the decrease in evacuation rates as the number or density of the zoomies decreased. Thus, the overall level of risk and corresponding consequences for thresholds for decisions might be manipulated by changing aspects of the zoomies display.

It is unclear how people will come to view repeated instances when an ensemble display gives an indication of some risk that is not then followed by the storm impacting their local area. Whether these scenarios are seen as a natural function of the uncertainty or come to be viewed as generating more false alarms might have important implications for future reactions. In addition, future work will be needed to assess how these ensemble displays including the zoomies impact trust over time.

Two other aspects of the dynamic ensembles provide advantages for better communicating predictions and uncertainty. One is that the zoomies, like the track ensembles, show natural frequencies: Each zoomie shows a potential hurricane path, rather than a summary of the probabilities as is shown in the cone of uncertainty. The other is that there is intuitive mapping between the visual features in the display and their underlying concepts. The zoomies move along the same paths the hurricanes are predicted to move.

This kind of conceptual compatibility increases comprehension and decreases bias in other visualization contexts (Witt, 2019b).

Zoomies also have the potentially incredible advantage of being able to communicate more information. With hurricanes, there are many risk factors at play. Although wind speed is typically the focus, storm surge, rainfall, and flash flooding are often the major hazards, particularly in some areas. Communicating all the different risk factors as well as the storm's anticipated timing leads to a cluttered display, particularly with the cone of uncertainty. The dynamics of the ensemble visualizations can inherently convey a sense of temporal uncertainty, illustrating how uncertainty in speed might impact the time of arrival, or how the risk of stalled forward movement of the weather system might create particularly hazardous conditions in a certain location. Dynamic ensembles can also accommodate the use of different visual features to communicate all the risk factors. In Experiment 4, we showed that one visual feature, color, could effectively communicate the severity of the storm. The next step is to determine which visual features are most effective and how many features can be effectively communicated. Drawing from human factors, it would also be important that any continuous scale also did a good job of communicating when high risk is predicted, regardless of the source of that risk.

Moving forward, there are many other questions that need to be answered before the dynamic ensembles should be implemented for hurricane predictions. One is whether there are biases to perceive areas that are directly in the path of a zoomie as being at higher risk than areas not along one of the paths. This pattern was found for track ensembles (e.g., Padilla et al., 2017) and reveals a lack of understanding the displays as showing a *sample* of possible hurricane paths. One way to alleviate this confusion, should it also occur with the zoomies, is to continually repeat the display and show different samples for each presentation. Another possibility would be to substantially increase the number of zoomies in the displays so that more central locations are in the path of at least one instance. Yet, another alternative is to increase the size of each element shown, with fewer items then able to encompass a larger area for each prospective path.

These issues should be addressed so that there are clear recommendations on how to present the zoomies and thus the presentation is standardized. In observing that track ensembles are rarely shared on social media as information about hurricanes (chosen only 3% of the time), Bica et al. (2019) point to the potential role of unfamiliarity and lack of standardization with these line ensemble visualizations. To avoid a similar issue with zoomies, the displays should first be empirically evaluated and standardized. These recommendations should determine the number of zoomies that should be displayed and the speed by which they travel. Too many zoomies may lead to visual crowding and confusion; too few zoomies may lead to decreased risk perception in areas with fewer zoomies. Zoomies that move fast allow for repeated presentations but might miscommunicate the speed of the storm. It is possible that faster presentation will also be harder to comprehend, although we are less worried about this possibility given that ensemble processing is amazingly accurate even with brief presentations (see Whitney & Yamanashi Leib, 2018).

We also note that our initial exploration of the viability of these types of visualizations only examined straight line paths, and the question of their relative benefits for more complex paths remains one for future research. When visualizing hurricane predictions, forecasters can choose to display specific paths outputted from the model, or they can display a set of paths that was constructed from the model predictions but has properties to assist the visual processing of the spatial distribution (Liu et al., 2018). These same decisions can be made with dynamic presentations. For both static and dynamic displays, more research is necessary to determine the benefits and costs of displaying specific model paths versus a set constructed to represent overall predictions.

The principles underlying the use of dynamic ensembles to communicate predictions and their uncertainty are not specific to hurricanes. Indeed, they could also be applied to other kinds of natural disasters such as wildfires or other kinds of spatial–temporal trajectories with uncertainty such as tracking other boats or drones. Empirical evaluations need to confirm their generalizability, but the theoretical underpinnings support the idea of principles that should apply to many different contexts.

Summary

In conclusion, displaying hurricane predictions as a dynamic ensemble, or zoomies, can influence understanding of risk, particularly for areas on the edges of the set of the predicted hurricane paths. This visualization may have particular promise for early notification of potential areas under threat where lack of appreciation of the uncertainty may lead those far from the central path failing to begin to prepare. The zoomies have an intuitive mapping, show natural frequencies rather than probabilities, leverage the immense ensemble processing capabilities of the visual system, and can be further designed to communicate additional risk factors. With the ubiquity of digital media, visualizations that are animated are increasingly feasible. We recommend pursuing the use of visualizing hurricane predictions using ensembles and highlight some potential advantages of using dynamic ensembles.

References

Bica, M., Demuth, J. L., Dykes, J. E., & Palen, L. (2019). Communicating hurricane risks: Multi-method examination of risk imagery diffusion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* [Paper 315] (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300545

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*(3), 275–295. https://doi.org/10.1037/xap0000166

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*(5), 651–668. https://doi.org/10.1175/BAMS-88-5-651

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann Publishers.

Cheong, L., Bleisch, S., Kealy, A., Tolhurst, K., Wilkening, T., & Duckham, M. (2016). Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science*, *30*(7), 1377–1404. https://doi.org/10.1080/13658816.2015.1131829

Cox, J., House, D., & Lindell, M. (2013). Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, *3*(2), 143–156. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2012003966

Demuth, J. L., Morss, R. E., Morrow, B. H., & Lazo, J. K. (2012). Creation and communication of Hurricane Risk Information. *Bulletin of the American*

Meteorological Society, *93*(8), 1133–1145. https://doi.org/10.1175/BAMS-D-11-00150.1

Dombroski, M., Fischhoff, B., & Fischbeck, P. (2006). Predicting emergency evacuation and sheltering behavior: A structured analytical approach. *Risk Analysis*, *26*(6), 1675–1688. https://doi.org/10.1111/j.1539-6924.2006.00833.x

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). Wiley.

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623–629. https://doi.org/10.1111/j.1539-6924.2005.00608.x

Goldberg, M. H., Marlon, J. R., Rosenthal, S. A., & Leiserowitz, A. (2020). A meta-cognitive approach to predicting hurricane evacuation behavior. *Environmental Communication*, *14*(1), 6–12. https://doi.org/10.1080/17524032.2019.1687100

Greis, M., Joshi, A., Singer, K., Schmidt, A., & Machulla, T. (2018). Uncertainty visualization influences how humans aggregate discrepant information. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* [Paper 505] (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174079

Hamill, T. M., Brennan, M. J., Brown, B., DeMaria, M., Rappaport, E. N., & Toth, Z. (2012). NOAA's future ensemble-based hurricane forecast products. *Bulletin of the American Meteorological Society*, *93*, 209–220. https://doi.org/10.1175/2011BAMS3106.1

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 37–53. https://doi.org/10.1037/a0017683

Herdener, N., Wickens, C. D., Clegg, B. A., & Smith, C. A. P. (2016). Overconfidence in projecting uncertain spatial trajectories. *Human Factors*, *58*(6), 899–914. https://doi.org/10.1177/0018720816645259

Herdener, N., Wickens, C. D., Clegg, B. A., & Smith, C. A. P. (2018). *Attention does not improve impaired understanding of variability in spatial prediction*. Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting (pp. 232–236). https://doi.org/10.1177/1541931218621054

Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS ONE*, *10*(11), Article e0142444. https://doi.org/10.1371/journal.pone.0142444

Kirschenbaum, S. S., & Arruda, J. E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors*, *36*(3), 406–418. https://doi.org/10.1177/001872089403600302

Liu, L., Boone, A. P., Ruginski, I. T., Padilla, L., Hegarty, M., Creem-Regehr, S. H., Thompson, W. B., Yuksel, C., & House, D. H. (2017). Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization and Computer Graphics*, *23*(9), 2165–2178. https://doi.org/10.1109/TVCG.2016.2607204

Liu, L., Mirzargar, M., Kirby, R., Whitaker, R., & House, D. (2015). Visualizing time-specific hurricane predictions, with uncertainty, from storm path ensembles. *Computer Graphics Forum (Proceedings of EUROVIS '15)*, *34*(3), 371–380.

Liu, L., Padilla, L., Creem-Regehr, S. H., & House, D. H. (2018). Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 882–891. https://doi.org/10.1109/TVCG.2018.2865193

MacEachren, A. M., Brewer, C. A., & Pickle, L. W. (1998). Visualizing georeferenced data: Representing reliability of health statistics. *Environment & Planning A*, *30*, 1547–1561. https://doi.org/10.1068/a301547

MacEachren, A. M., & Kraak, M.-J. (2000). Research challenges in geovisualization. *Cartography and Geographic Information Science*, *28*, 3–12. https://doi.org/10.1559/152304001782173970

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics and uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2496–2505. https://doi.org/10.1109/TVCG.2012.279

Mason, J. S., Retchless, D., & Klippel, A. (2017). Domains of uncertainty visualization research: A visual summary approach. *Cartography and Geographic Information Science*, *44*(4), 296–309. https://doi.org/10.1080/15230406.2016.1154804

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*(2), 221–239. https://doi.org/10.1080/13658816.2015.1082566

Miller, A. L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology*, *81*(1), 16–21. https://doi.org/10.1037/h0027430

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty Forecasts Improve Decision Making Among Nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209–227. https://doi.org/10.1518/155534309X474460

National Oceanic and Atmospheric Administration. (2020). *National Hurricane Center and Central Pacific Hurricane Center: Definition of the NHC Track Forecast Cone*. https://www.nhc.noaa.gov/aboutcone.shtml

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, *19*(4), 601–607. https://doi.org/10.3758/s13423-012-0247-5

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision (Charlottesville, Va.)*, *15*(4), Article 3. https://doi.org/10.1167/15.4.3

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Impacts*, *3*(29), Article 29. https://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*(1), Article 40. https://doi.org/10.1186/s41235-017-0076-1

Pappenberger, I., Stephens, E., Thielen, J., Salamon, P., Demeritt, D., van Andel, S., Wetterhall, F., & Alfieri, L. (2013). Visualizing probabilistic flood forecast information: Expert preferences and perceptions of best practice in uncertainty communication. *Hydrological Processes*, *27*(1), 132–146. https://doi.org/10.1002/hyp.9253

Pham, E., Li, Z., Emrich, C. T., Mitchem, J., & Cutter, S. L. (2020). Evacuation departure timing during Hurricane Matthew. *Weather, Climate, and Society*, *12*(2), 235–248. https://doi.org/10.1175/WCAS-D-19-0030.1

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*(3), 324–339. https://doi.org/10.1177/0018720818758770

Rautenhaus, M., Böttinger, M., Siemen, S., Hoffman, R., Kirby, R. M., Mirzargar, M., Röber, N., & Westermann, R. (2018). Visualization in meteorology-a survey of techniques and tools for data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, *24*(12), 3268–3296. https://doi.org/10.1109/TVCG.2017.2779501

Regnier, E. (2008). Public evacuation decisions and hurricane track uncertainty. *Management Science*, *54*(1), 16–28. https://doi.org/10.1287/mnsc.1070.0764

Regnier, E., & Harr, P. A. (2006). A dynamic decision model applied to hurricane landfall. *Weather and Forecasting*, *21*(5), 764–780. https://doi.org/10.1175/WAF958.1

Ruginski, I. T., Boone, A. P., Padilla, L. M., Liu, L., Heydari, N., Kramer, H. S., Hegarty, M., Thompson, W. B., House, D. H., & Creem-Regehr, S. H. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition and Computation*, *16*(2), 154–172. https://doi.org/10.1080/13875868.2015.1137577

Sanfey, A., & Hastie, R. (1998). Does evidence presentation format affect judgment? An experimental evaluation of displays of data for judgments. *Psychological Science*, *9*, 99–103. https://doi.org/10.1111/1467-9280.00018

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*, 115–143. https://doi.org/10.1006/ijhc.1996.0048

Sherman. (2014, June 27). *phack: An R function for examining the effects of p-hacking*. Sherman's Head: About Psychology, Statistics, and R. http://rynesherman.com/blog/phack-an-r-function-for-examining-the-effects-of-p-hacking/

Spence, R. (2001). *Information visualization*. ACM Press.

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: Research and development agenda for visual analytics*. IEEE Press.

Thompson, R. R., Garfin, D. R., Silver, R. C. (2017). Evacuation from natural disasters: A systematic review of the literature. *Risk Analysis*, *37*(4), 812–839. https://doi.org/10.1111/risa.12654

Tufte, E. R. (2001). *The Visual display of quantitative information* (2nd ed.). Graphics Press.

Watamaniuk, S. N. J., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics*, *60*(2), 191–200. https://doi.org/10.3758/BF03206028

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Wickens, C. D., Clegg, B. A., Witt, J. K., Smith, C. A. P., Herdener, N., & Spahr, K. S. (2020). Model of variability estimation (MOVE): Factors influencing human prediction and estimation of variance in continuous information. *Theoretical Issues in Ergonomics Science*, *21*(2), 220–238. https://doi.org/10.1080/1463922X.2019.1679907

Witt, J. K. (2019a). The perceptual experience of variability in line orientation is greatly exaggerated. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(8), 1083–1103. https://doi.org/10.1037/xhp0000648

Witt, J. K. (2019b). Graph construction: An empirical investigation on setting the range of the y-axis. *Meta-Psychology*, *3*, Article MP.2018.895.

Witt, J. K., Clegg, B. A., Wickens, C. D., Smith, C. A. P., Laitin, E. L., & Warden, A. C. (2020). Dynamic ensembles versus cones of uncertainty: Visualizations to support understanding of uncertainty in hurricane forecasts. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1644–1648. https://doi.org/10.1177/1071181320641399

Received December 9, 2020
Revision received April 22, 2021
Accepted May 1, 2021 ■