Geographic Scale and Probabilistic Forecasts: A Tradeoff for Protective Decisions?

Cassandra A. Shivers-Williams[1,2] and Kimberly E. Klockow-McClain[1,2]

[1]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma
[2]National Severe Storm Laboratory, National Oceanic and Atmospheric Administration
Norman, OK, USA

ORCID: Shivers-Williams, 0000-0002-1190-780X; Klockow-McClain, 0000-0002-3859-4831

Corresponding Author: Cassandra A. Shivers-Williams, cs-w@ou.edu

Acknowledgement: Funding was provided by the National Oceanic and Atmospheric Administration (NOAA) Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce.

ABSTRACT

This pilot study aimed to examine the impact of varying geographic scales, probabilities of tornado occurrence, and presentation formats within severe weather forecasts on individuals' protective decisions. This pilot was unique in its specific examination of the tradeoff between highly localized geography and higher valued probabilistic threat information in weather-related decision-making. This pilot utilized a 4 (geographic scale) x 12 (probability) x 3 (forecast presentation format) mixed, nested experimental design. Participants were 440 United States adults who completed electronic questionnaires containing experimentally manipulated severe weather forecasts. A linear mixed model analysis revealed several findings. First, participants who saw only categorical forecasts made similar preparatory decisions across geographic scales. Additionally, they were more willing to take preparatory action as categorical risk increased. Second, when probabilities were presented, the propensity to take protective action was greater at higher probabilities and at larger geographic scales, affirming the regional geographic reference class selected by the Storm Prediction Center in today's outlook system. Third, individuals' propensity for action generally increased as scale and probability increased but the pattern varied across presentation formats. Lastly, participants reported having a map to look at was moderately important to their decisions and having probabilistic and categorical risk information was highly important to their decisions. Taken together, the findings suggest a complex relationship between geographic scale and probability, which is further complicated by forecast presentation format.

KEY WORDS: Geographic scale; forecast uncertainty; protective decisions; forecast presentation format

Declarations

Funding: Funding was provided by the National Oceanic and Atmospheric Administration (NOAA) Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce.

Conflicts of interest/Competing interests: The authors declare that they have no conflicts or competing interests.

Ethics approval: This study was approved by the University of Oklahoma institutional review board.

Consent to participate: Participants provided electronic informed consent before completing the online survey used for data collection.

Consent for publication: Not applicable

Availability of data and material: Data and materials available upon request from authors

Code availability: Not applicable

Authors' contributions: Both authors contributed to the conception and design of the study, but that effort was led by Kimberly Klockow-McClain. Material preparation was primarily performed by Cassandra Shivers-Williams and then approved by Kimberly Klockow-McClain. Data collection was managed by Qualtrics, but that process was navigated by Cassandra Shivers-Williams. Analyses were also performed by Cassandra Shivers-Williams. The first draft of the manuscript was written by Cassandra Shivers-Williams, with Kimberly Klockow-McClain contributing heavily to the introduction, literature review, and discussion. Both authors commented on previous versions of this manuscript and have read and approved this final manuscript.

1. INTRODUCTION

Protective decision-making during severe weather is often complicated; many factors must be considered before action is taken. The first factor people often think about is whether an impending storm will personally impact them, but even that decision is multifaceted as timing, likelihood, location, and intensity, among other things, are considered (Johnson 2013). The present research hones in on two key factors, namely likelihood and location, to explore their relative impact on individuals' willingness to take preparatory action in response to a fictitious tornado. Each factor is described in more detail below.

At present, a research program called Forecasting a Continuum of Environmental Threats (FACETs) aims to expand the ways forecast information is created and ultimately delivered to the public as a means of facilitating improved decision-making; a key attribute of concern for this program is the creation and communication of forecast uncertainty products (Rothfusz et al. 2018). While the program aims to create new products, several uncertainty products already exist, and understanding how they were created and influence decisions can improve the concepts applied to new product creation. One of the signature uncertainty products currently in use in the National Weather Service (NWS) severe suite is the Storm Prediction Center (SPC) Convective Outlook, which contains probabilistic information about the likelihood of experiencing severe hazards like tornadoes, wind, and hail up to a day in advance. These products correspond to categorical risk levels, including High, Moderate, Enhanced, Slight, and Marginal categories.

The SPC products were created in the mid-1990s, when SPC Director Joe Shafer began to collaborate closely with noted statistician Alan Murphy (personal communication, H.E. Brooks 2019). They wanted to create a product that could offer an at-a-glance view of the relative severe weather risks (e.g., for tornadoes, hail, and wind) faced across the country on a given day beyond the categorical system that was in place at the time (High, Moderate, Slight Risk categories). At the time, forecasters created these risk categories using somewhat subjective judgments, and the SPC desired to formalize the system with objective criteria. They recognized that this could involve at least two components: the coverage of storm reports, defined as the number of storm hazards reported per unit area, and the potential severity levels for each hazard. They initially experimented with a system that offered expected coverage for severe reports at two different intensity levels: minimum NWS severe criteria, and "hatched" areas that could incur significant severe threats (EF2+ tornadoes, wind in excess of 65 kts, hail in excess of 2"). With those severity criteria defined, the next major issue was to decide the spatial resolution/coverage that would define the reference class of the probability.

The SPC could have made any choice; they explored what would happen if they offered larger-radius coverage probabilities, for example in an area of approximately 37 mi (120km^2 coverage area), a more moderate range of approximately 25 mi (80km^2 coverage area), or a smaller neighborhood of approximately 12 mi (40km^2 coverage area). Going into this analysis, Brooks noted a desire to have probability values that could escalate as an event unfolds. SPC wanted a probability value from the convective outlook that could be increased when a watch was issued, and then increased again when a warning was issued. Brooks noted in particular that SPC wanted numbers "large enough to feel," but small enough to still increase. When calculating the probability values that arose from these choices, Brooks found that the highest probability values achieved for High Risk days at the 37 mi radius could be 60% or greater, which did not leave much room for additional inflation of the probability value. The highest probability values achieved for High Risk days at a 25 mi radius were approximately 25%, leaving a lot of room for growth, but also offering values that "felt" meaningful. The highest probability values achieved for High Risk days at a 12 mi radius were approximately 6%, which left a lot of room for growth, but also "felt" less meaningful. Thus, they determined that the 37 mi radius was too large and the 12 mi radius was too small, and they decided to implement a 25 mi radius for the coverage area.

Implicit in these decisions were forecaster judgments about spatial resolutions that were meaningful. The resolution chosen offered probability values that forecasters found meaningful, but potentially at the expense of offering probabilities that were more local. A primary challenge facing the FACETs program is knowing how to best connect probabilistic information, including in all of its potential formats, with user decisions. And while probability value and format may play a role in this (Joslyn and LeClerc 2012, 2013), in the spatial context of decisions, so may localness/nearness (Klockow 2013; Nagele and Trainor 2012). Thus, research should be conducted to explore the potential tradeoffs between these effects. In other words, what is more important—a probability value that is relatively large, or a value that reflects an event that is likely to occur nearby?

While there is minimal research examining the potential psychometric tradeoffs between scale and numerical value in the weather realm, this kind of relationship has been studied in other domains. For example, psychologists have found that bystanders, in both emergency and non-emergency situations, are less likely to help a person in need as the number of other onlookers increases (e.g., Darley and Latané 1968; Latané and Darley 1968).

At the same time, psychologists have also found that individuals are more likely to help a single, identifiable person in need than a larger number of statistical victims (Jenni and Loewenstein 1997). In both of these cases, the felt effect of helping decreases as the scale of other involved persons increases. Examining the influence of geographic scale, Severtson and Burt (2002) found that concern over an environmental pollutant may increase with distance from the hazard. This was the result of motivated reasoning; those close to the hazard valued the economic benefits of industrial activity that gave rise to the pollution, while people living in urban areas farther away valued the pristine nature of the rural area. Thus, localness might not be the most significant determinant in personally connecting to a hazard. Gibson-Graham (2002) argued it is traditionally assumed that forces operating on global scales are more powerful because they subsume the local; however, processes occurring at local scales also have the power to bring global events to a halt. Thus, while larger scales may be assumed to have more power you can feel, the power of the smaller scale cannot be overlooked. Across these studies, the effect of scale appears mixed and subject to context. This research aims to bridge these studies that examine the power of various scales of influence with literature on weather decision-making under conditions of uncertainty.

1.1. Communicating and Understanding Forecast Uncertainty and the Impact on Protective Decisions

There has been a large vein of research examining how individuals use forecast uncertainty information to make various decisions. One of the most commonly provided forms of uncertainty information in modern US weather forecasts is the probability of precipitation. Previous findings (Gigerenzer et al. 2005; Murphy et al. 1980) have suggested that people do not properly understand probability of precipitation because the reference class—the class of event to which probabilistic forecast information refers—is often ambiguous. When the reference class is specified, however, recent studies have found that people can effectively use forecast uncertainty information to make better decisions (Grounds and Joslyn 2018; Joslyn and LeClerc 2012; Joslyn et al. 2007; LeClerc and Joslyn 2015) even though they may not always understand the technical or meteorological definitions of the uncertainty information provided (Morss et al. 2008). In sum, research has shown that providing forecast uncertainty information could have benefits, though it is unclear whether very specific attributes of the reference class are particularly important. In the context of this research, we introduce a new kind of reference class: a geographic reference class, or the scale attribute of the probabilistic information that arises when uncertainty information is mapped.

One argument for why receiving uncertainty information helps improve decisions is people are better able to calibrate to the risk posed by the situation. This often leads to correctly taking protective action during elevated risks and not taking protective action during lower risks. For example, Joslyn and LeClerc (2012) conducted a series of studies where college students engaged in cost-loss driven decision tasks to assess whether receiving forecast uncertainty information improved their decision-making. Participants were responsible for making decisions regarding whether to salt the roads in advance of icy conditions. In one study, one group of participants received only the nighttime low temperature (i.e., a typical deterministic forecast) while another group of participants received temperature along with the probability of freezing. Findings showed participants who received probability of freezing information were more likely to correctly salt when the probability of freezing was above the rationally correct decision threshold and correctly withhold salt when the probability fell below the threshold.

These results can vary based on the format of the probabilistic information, however. A format change would present the same reference class, e.g., the probability of reaching a freezing temperature threshold, with a mathematically equivalent but different numerical expression (e.g., percentage, odds ratio, frequency). Several studies have examined the effect of presenting forecast uncertainty with different expressions. Examples include presenting uncertainty information as odds ratios (LeClerc and Joslyn 2012; Morss et al. 2008), frequencies and percentages (Joslyn et al. 2009; Morss et al. 2008), intervals (Grounds et al. 2017; Morss et al. 2008), and verbal expressions (Grounds and Joslyn 2018; Wallsten et al. 1993). These studies revealed that each format has a particular influence on decision-making; for example, odds ratios increased the propensity to take protective action even when unwarranted, leading to a higher rate of false alarms (LeClerc and Joslyn 2012; Morss et al. 2008), and frequency representations improved decisions overall as compared to probability formats (Joslyn et al. 2009; Morss et al. 2008). Verbal formats have led to a wide array of interpretations, even though some studies have suggested these formats may be easier for non-experts to use (Grounds and Joslyn 2018; Wallsten et al. 1993). Additionally, the format could also interact with the decision task and produce particular effects on judgment. For example, if the decision a user must make relates to temperatures below freezing, but the uncertainty information is given as the probability of exceeding the freezing threshold, more errors in judgement may result (Joslyn et al. 2009).

Importantly for the present work, all of these studies examined the likelihood of reaching particular thresholds of temperature or wind speed—forecast attributes that are commonly experienced. LeClerc and Joslyn (2012) argued that presentation formats like odds ratios may be especially appropriate for rare, extreme weather, because they motivate more protective actions; however, from a calibration perspective, these behaviors can also result in a large incidence of false alarms and, over time, reduced trust in the forecast. Care must then be taken to consider the most appropriate expressions for forecast information in very rare events like severe weather. Relevant to the present study, the SPC currently employs a system with both a verbal expression (High, Moderate, Enhanced, Slight, and Marginal Risk) and numeric probabilities. Forecasters do not know whether categories enhance or in some other way interact with individuals' understanding of the probabilistic information. Further, scant research has examined how best to present forecast probabilities for severe convective storms on maps, specifically (Klockow-McClain et al. 2019).

Previous research has also examined laypersons' use and understanding of alternative forecast presentation formats such as visual forecasts for a variety of threats, including tornadoes and hurricanes (e.g., Boone et al. 2018; Lindell 2020; MacPherson-Krutsky et al. 2020; Millet et al. 2020; Padilla et al. 2017; Ruginski et al. 2016). Lindell's (2020) review of hazard map research showed individuals use heuristic shortcuts, in particular the proximity heuristic, to form risk judgments of tornadoes and hurricanes such that greater risk is perceived closer to storm tracks and comparatively lesser risk is perceived outside of a tornado warning polygon or hurricane uncertainty cone. Other studies (e.g., Liu et al. 2017; Padilla et al. 2017; Ruginski et al. 2016) have shown different hurricane forecast graphical visualization formats lead to different biases in laypersons' interpretations of the size, intensity, and potential damage of a hurricane. For instance, summary displays such as the National Hurricane Center's (NHC) "cone of uncertainty" have often led laypersons to erroneously assume the widening of the cone—which is meant to convey forecast uncertainty over time—conveys the hurricane growing in size or intensity (Padilla et al. 2017; Ruginski et al. 2016). However, by providing explicit instructions, Boone et al. (2018) found improved understanding of the graphic and reduced likelihood of endorsing size misconceptions. On the other hand, ensemble displays have been found to help users make more accurate risk estimates and reduce misconceptions of storm size (Lui et al. 2017; Padilla et al. 2017) but have also biased users' point-based judgments (Padilla et al. 2017). Taken together, researchers acknowledge it is important to consider the type(s) of task(s) users will complete while using visual forecasts because each visualization type has different inherent biases that affect decision-making. Consequently, Millet et al. (2020) argued different risk communication strategies may be needed to meet the needs of different user groups because their tasks vary widely.

1.2. The Role of Personal Geography in Protective Decisions

While successfully communicating uncertainty can involve probability format, much of the forecast information presented to audiences comes in a mapped format, where geography adds another layer of complexity. The format considerations relevant to spatial information have seen less work. Geographical formats for probabilistic information, including mapped attributes of color and probability distributed over space, can also affect judgment (Klockow-McClain et al. 2019; Miran et al. 2018). Klockow-McClain and colleagues (2019) examined the complex interplay between warning boundaries, physical location, probabilistic forecasts, and cartographic coloring schemes on individuals' protective decisions. This study found that protective decisions varied based on the length of the forecast guidance, the physical distance from the storm, the likelihood of the hazard, the expression of the probability (in percentage or verbal formats) and the format aspects of the map such as color. In response to deterministic warnings, greater risk was perceived at points closer to the storm than at points further away. Also, protective decisions at equivalent distances varied depending on whether the warning was short or long. Moreover, these relationships were moderated by providing verbal guidance; when respondents were told that the likelihood of a tornado was high, they responded more frequently than with deterministic warning information alone. Similarly, when responding to probabilistic information, participants responding from points closer to the storm were more likely to take protective action than at points farther away, even at the same objective probability level. In other words, 60% meant something different to respondents when they were close to the storm than when they were farther away; distance framed the probability. However, when the roles of deterministic and probabilistic information were compared, participants were more likely to take protective action when needed and forego protective action when it was unnecessary after receiving probabilistic information. This study showed the relationship between geographic properties and protective decisions is more complex than previous research has assumed. This study heavily informed the design of the present research. As noted above, outstanding questions remain about how best to display probabilistic information in a geospatial context, for example, a key attribute of mapped information is its resolution; the reference class in the Klockow et al. study (2019) was held constant, but varying spatio-temporal reference classes are possible.

Examining the influence of spatial relationships between individuals and hazards in real-world decision contexts, Teigen (2005) showed individuals' estimates of the likelihood of adverse outcomes were more strongly related to near-miss accidents compared to actual accidents. Across a series of studies, he showed that individuals based these estimates on a proximity heuristic, which is a mental shortcut in which a person judges the likelihood of a threat occurring using their physical proximity to the threat. Similarly, Aguirre and colleagues (1991) examined the effects of geographic specificity in Texas residents' responses to a tornado and their results showed the residents thought the broadly defined county risk information communicated in the warning messaging was hard to interpret. Another study by Nagele and Trainor (2012) suggested that people who are closer to a tornado threat are better able to experience the associated hazardous conditions (i.e., can see or feel the winds, hail, rain), which leads to increased belief that there is in fact a risk to safety and subsequent protective behavior. To test this proposition, they examined the impacts of warning polygon size and closeness to a tornado track—operationalizations of geographic specificity—on individuals' willingness to seek additional information and shelter under tornadic conditions. Their study produced mixed results; there was not a significant relationship between being located within a warning polygon and seeking shelter, however, in situations where a warning polygon was smaller than 50% of the county, people were more likely to seek shelter. Moreover, participants who were within five miles of the tornado track were more likely to seek shelter than do nothing in response to the threat. These researchers argued that, while storm-based warnings are helpful, larger polygons may not be specific enough to elicit protective action; there may be critical geographic information missing from these larger polygons that leads to inaction. A meta-synthesis review of tornado response literature (Johnson 2013) found that people did not decide how to act in the face of tornado threats solely on the warning and information contained therein; instead, the protective decision-making process included people attempting to confirm the existence of a threat and feeling some sense of danger from the threat before they decided to take protective action.

1.3. The Present Research

The present research drew on the literature briefly discussed here, namely the cognitive geography, meteorology and psychology research. This research constituted a pilot study aimed to fill gaps in existing literature by examining the impact of a potential tradeoff between highly localized geographic tornado threat information and higher valued probability information on individuals' protective decisions, a relationship not previously studied within a mapped weather risk context.

Several hypotheses guided this pilot study. First, individuals were hypothesized to be more likely to take protective action as the geographic scale, otherwise referred to as the geographic reference class, was more localized. It was expected that as a threat became more personally relevant, individuals would be more likely to respond. Second, individuals were hypothesized to be more likely to take protective action as the probability of a threat occurring increased. It was expected that as a threat became more likely to occur, individuals would be more likely to respond. Third, there were no a priori expectations of how these two factors would interact, because when applied together to create different reference classes, the effects are competing. Smaller, more localized reference classes will have correspondingly lower probability values. Lastly, there were no a priori expectations of how presentation format would interact with these factors as scant research has examined this interaction in the context of mapped information.

2. METHODOLOGY

2.1. Design

This pilot study employed a 4 (geographic reference class) x 12[1] (probability) x 3 (forecast presentation format) mixed, nested experimental design (see Fig. 1) where geographic reference class and probability were both within-subjects variables and forecast presentation format was a between-subjects variable. Further, probability was a nested factor within the geographic reference class factor; each geographic reference class contained a series of 12 severe weather forecasts containing unique probabilities of tornado occurrence. Geographic reference class consisted of four levels: city, county, region, and multi-region. Probability consisted of 12 levels, adapted from the SPC Threat Scales for Convective Hazards "Day 1 Outlook Probabilities" product (2017) and scaled appropriately for each geographic reference class (4 scales x 12 levels = 48 unique probability values). Presentation format consisted of three levels: probability forecasts, categorical forecasts, and combination probability/categorical forecasts. Each manipulation will be described in greater detail below.

2.2. Participants and Procedure

Data discussed in this paper were collected as part of a larger study conducted that examined several geographical, meteorological, and psychological variables posited to have influenced individuals' protective decisions. The overall study was approved by the university's institutional review board. Qualtrics sampled and managed online data collection for 440 participants during October and November 2018. Participants completed an electronic consent form and were administered questionnaires via Qualtrics. The median survey completion time was 25.5 minutes. Qualtrics screened for "fast responders"—participants who completed the entire survey in one-half the median time or less—and terminated them for not responding thoughtfully. After survey completion,

---

[1] The experimental design was not fully balanced for all combinations of geographic reference class, probability, and presentation format. This will be discussed in the "Practical Limitations" section.

participants read an electronic debriefing form and were paid for their time commensurate with the Qualtrics pay schedule.

For the purposes of this paper, we focused on a subset of the overall study's variables (the independent and dependent variables described below) and examined adults living across the United States (US) in the four continental US National Weather Service (NWS) Regions: Eastern, Southern, Central, and Western. These regions were chosen because they are the most severe weather-prone. Participants were sampled in representative proportions from each NWS region and similar to the 2018 US Census (United States Census Bureau 2019) (see Table 1).

2.3. Stimulus Materials

Participants responded to a series of decision scenarios consisting of severe weather probability forecasts and fictitious maps. Specifically, each scenario presented a severe weather forecast, modeled after the SPC Convective outlook, describing the day's chance of tornado risk at a given geographic reference class and a corresponding map of the geography (see Fig. 2a-d). Further, the scenarios contained the experimental manipulations of geographic reference class, forecast probability, and presentation format (explained in greater detail below) as well as the primary dependent measure—a question assessing participants' likelihood of taking preparatory action.

Scenarios were presented in sets, grouped together by geographic reference class, in a randomized order. Participants responded to every scenario for all geographic reference classes one reference class at a time. For example, a participant may have first been randomly assigned to view all forecasts for the city reference class. After completing all of these scenarios, this participant may have then been randomly assigned to view all of the forecasts for the region reference class, and then the county reference class, and finally the multi-region reference class. Again, the order in which participants viewed and responded to each reference class scenario set was randomized. Within each geographic reference class, the scenarios were presented in ascending probabilistic order such that the first scenario was the lowest chance of tornado occurrence and the last scenario was the highest chance of tornado occurrence; this was done to minimize confusion.

2.4. Independent Variables

2.4.1. Geographic Reference Class

To assess the impact of "localness" on decision-making, participants responded to sets of scenarios staged in four different geographic scales: a city (100 sq. mi; see Fig. 2a), county (625 sq. mi; see Fig. 2b), region (2500 sq. mi; see Fig. 2c), and "multi-region" (10,000 sq. mi; see Fig. 2d). This approach uses square grid cells to connote probabilities at each scale.

2.4.2. Probability

To assess the impact of probabilistic information on decision-making, participants responded to scenarios expressing different likelihoods of tornado occurrence. The probabilities and risk categories utilized were adapted from the current SPC Threat Scales for Convective Hazards "Day 1 Outlook Probabilities" (2017) product. Specifically, the SPC probabilities were adapted and converted to mathematical equivalents for each geographic reference class.

The SPC probabilities, as used today, refer to the likelihood an event will occur in a 24-hour period within 25 miles of any given point, which creates a region of 2500 sq. mi. Thus, the verbatim SPC probabilities were used for the region reference class scenarios. For the other reference class scenarios, mathematically equivalent SPC probabilities were computed by converting the probabilities based on scale differences. For example, the regional grid is 2500 sq. mi. whereas the city grid is only 100 sq. mi. To compute the city-scale mathematically equivalent SPC probabilities, each region-scale probability was divided by 25 (2500:100 reduces to 25:1). Similarly, the region-scale probabilities were divided by four to compute the county-scale probabilities (2500:625) and multiplied by four to compute the multi-region probabilities (2500:10000). See Table 2 for a complete breakdown of probability points, including corresponding severity levels and risk categories, used across geographic reference classes.

2.4.3. Presentation Format

To assess the impact of presentation format on decision-making, participants were randomly assigned to receive all scenarios in one of three presentation formats. In the probabilistic format condition, forecast uncertainty was presented as percent chance (e.g., there is a **2% chance** of a tornado today) and the accompanying maps were left uncolored (see Fig. 3a). In the categorical format condition, forecast uncertainty was expressed as a color-coded risk category (e.g., there is a **Marginal risk** of a tornado today) and the accompanying maps and category text were colored according to the SPC (2017) scale (see Fig. 3b). In the combined format condition, forecast uncertainty was expressed as both percent chance and risk category (e.g., there is a **Marginal risk** of a tornado today, which corresponds to a **2% chance**) and the accompanying maps and category text were colored (see Fig. 3c). Appendix A also outlines the risk categories that accompanied each probability level across reference classes.

2.4.4. Practical Limitations

Two practical limitations led to an imbalance of probability scenarios across geographic reference classes and presentation formats. This is worth noting because it led to an unequal number of stimuli presented to participants. In total, respondents viewed either 32 scenarios (categorical forecast condition) or 44 scenarios (probabilistic forecast and combined forecast conditions).

First, there were an unequal number of probability scenarios across the four geographic reference classes. The probability and combination forecast conditions maintained all 12 SPC probability levels within the city, county, and region scenarios. The multi-region reference class, however, only retained 8 of the SPC probabilities. The two highest probabilities—45% and 60%—and their "significant severe" complements converted to probabilities greater than 100%. Practically speaking, a tornado likelihood greater than 100% is unfeasible so these probabilities were dropped. Additionally, the conversion of 30% from the region to the multi-region was greater than 100%. However, if the conversion of 15% remained the highest probability used for the multi-region, the probabilities would have maxed out at 60%. Thus, the conversion of 30%, which was 120%, was truncated at 100%. The remaining lower probabilities were converted as previously described, which left 8 probability points for the multi-region reference class. These conversion issues were inconsequential for the category condition because participants did not receive the probabilities underlying the risk categories.

Second, there were an unequal number of scenarios across the three presentation formats. The probability and combination forecast conditions had 44 identical scenarios: 12 city scenarios, 12 county scenarios, 12 region scenarios, and 8 multi-region scenarios. However, the category condition only had 32 scenarios—8 per location. Practically speaking, it would have been redundant to ask participants in the category condition multiple identical questions, which would have happened if the categories were repeated for balance. Therefore, category condition participants only received the meaningfully different categories for every geographic reference class.

Both of these imbalances—the multi-region reference class having fewer scenarios than the other reference classes and the category presentation format having fewer scenarios than the other presentation formats—were addressed statistically. Specifically, missing data were estimated using a data imputation process in the Statistical Package for Social Sciences (IBM SPSS version 26). To correct the multi-region reference class imbalance, responses from the highest probability point used in the multi-region reference class (i.e., 100%) were imputed for the four probability points that were not administered in the probability and combination conditions. It was expected that judgments and decisions between probability point 8 (100% chance of occurrence) and probability points 9-12 (chances of occurrence greater than 100%) would not differ because respondents would have seen the same forecasts across these scenarios. Similarly, to correct the category presentation format imbalance, responses from parallel risk category scenarios were imputed for the four missing scenarios; it was expected that judgments and decisions between the risk categories participants saw and the parallel categories they did not see would not differ because they would have seen the same forecasts across these scenarios. As an example, participants responded to one scenario with a forecast for a high significant severe risk of a tornado, but the SPC Threat Scale has three levels that correspond to high significant severe risk. Participants' responses to the presented high significant severe risk scenario were imputed for the two high significant severe risk scenarios that were not presented. This process (using appropriately parallel categories) was used to impute data for the remaining two missing categorical forecast scenarios.

2.5. Measures

2.5.1. Demographic Items

At the beginning of the study, prior to receiving experimental materials, participants completed standard demographic items, some of which were used for sampling criteria. Sample items included age and gender.

2.5.2. Dependent Variables

2.5.2.1. Willingness to take preparatory action. After every scenario, participants responded to the question, "How likely are you to take preparatory action today in response to the potential tornado threat?" Responses were indicated on a 6-point Likert scale, with options ranging from *extremely unlikely* to *extremely likely*.

2.5.2.2. Message evaluation. After all decision trials for every geographic reference class were completed, several items assessed the extent to which different aspects of the forecast information impacted decisions. The message aspects evaluated were: (1) probabilistic forecast information in the text, (2) categorical/color-coded risk information in the text and on the map, (3) map coloring scheme, and, (4) having a map to look at while making decisions. Participants only rated the aspects to which they were exposed. For example, participants in the probability condition rated the importance of having "numerical probabilistic forecast information." Similarly, all participants rated the importance of "having a map to look at while making decisions."

Responses were indicated on 7-point Likert scales, with options ranging from *extremely unimportant/I did not consider this information at all when making decisions* to *extremely important/I heavily based my decisions on this information*. Higher scores indicated greater importance to decision-making.

2.6. Data Analysis

To investigate the research question, a linear mixed model (LMM) analysis and one-way analyses of variances (ANOVAs) were performed. Statistical Package for Social Sciences (IBM SPSS version 26) was used for all analyses. LMM was used because it allowed us to simultaneously account for variability across geographic reference classes and presentation formats as well as the nested effect of probability within geographic reference classes and the repeated measures of geographic reference classes and probabilities. Multiple responses from the same participant were expected to be more similar to each other than responses from other participants. Accounting for these fixed, random, and subject effects simultaneously was expected to reduce the error in the model. A series of ANOVAs were performed to examine the effect of presentation format on evaluations of the message features (e.g., having a map, color-coded risk information).

3. RESULTS

3.1. Impacts of Geographic Reference Class, Probability, and Presentation Format

The overarching goals of this experiment were to determine the effects of geographic reference class (i.e., scale), probability, and forecast presentation format on individuals' willingness to take preparatory action. To examine these relations, LMM estimated through maximum likelihood was computed. In the LMM, participants' willingness to take preparatory action served as the outcome variable and scale, presentation format, and probability were the predictors. Scale and probability were both modeled as repeated measures. The main effects of scale and presentation format as well as their interaction were modeled as fixed effects. Following guidance from statistical resources (e.g., Keppel and Wickens 2004, Chapter 25; Seltman 2009, Chapter 15; Starkweather and Harrington 2018), the nested effect of probability within scale was modeled as a random factor. The equation for the LMM is:

$$y_{ij} = \beta_1(scale_{ij}) + \beta_2(presentation_{ij}) + b_1(probability_{ij}) + \varepsilon_{ij}$$

where y_{ij} is the willingness to take preparatory action for a particular ij case, β_1 and β_2 are fixed effect coefficients for geographic scale and presentation format (respectively) for observations j in groups i, b_1 is the random effect coefficient for probability, and ε_{ij} is the error for case.

3.1.1. Fixed Effects

The main effects of both scale and presentation format were examined and both reached statistical significance. Specifically, the effect of scale on preparatory action, *F*(3, 47.89) = 4.85, *p* = .005, showed that participants' willingness to take preparatory action significantly varied as a result of the geographic reference class to which they were responding (see Fig. 4). Using a Bonferroni correction (to account for the number of simultaneous comparisons), pairwise comparisons were computed and examined for group differences; only two of the pairwise comparisons reached statistical significance. Namely, the mean preparatory action scores at the city scale (*M* = 4.07, □= -0.86, *SE* = .24, *p* = .005) and the county scale (*M* = 4.23,□= -0.70, *SE* = .24, *p* = .032) were both significantly lower than the mean multi-region scale score (*M* = 4.93). None of the region scale score (*M* = 4.49) comparisons reached statistical significance. Looking across scales, participants' willingness to take preparatory action was significantly lower in the smaller geographic scales (i.e., the city and county scales) compared to the largest geographic scale. On the other hand, scales that were similar to each other in size (i.e., city/county, county/region, region/multi-region) did not have preparatory action scores that varied significantly, which suggests participants were responding to these scales similarly.

The main effect of presentation format on preparatory action was also statistically significant, *F*(2, 19000.96) = 382.00, *p* = .000, showing that participants' willingness to take preparatory action significantly varied as a result of the forecast presentation format they viewed (see Fig. 5). After the Bonferroni correction, all pairwise comparisons reached statistical significance. First, the mean preparatory action score for those in the probability forecast condition (*M* = 4.12) was significantly lower than the mean score for those in the categorical forecast condition (*M* = 4.79, □ = -0.67, *SE* = .02, *p* = .000) and the combination forecast condition (*M* = 4.37, □ = -0.25, *SE* = .02, *p* = .000). Further, the mean score for those in the combination forecast condition was also significantly lower than that of the categorical forecast condition (□ = -0.42, *SE* = .02, *p* = .000). These results suggest categorical forecasts encouraged taking preparatory action, on average, while probabilistic forecast information, even when coupled with categorical forecast information, seemed to lessen the propensity to take preparatory action by comparison. These findings could have implications for how risk was interpreted as a function of presentation format.

The interactive effect of scale and presentation format on preparatory action was also statistically significant, *F*(6, 10605.39) = 137.17, *p* = .000, showing the effect of scale on preparatory action varied as a function of forecast presentation format (see Fig. 6). In examining the simple effects, a few trends emerged. First, the mean preparatory action score for those in the categorical forecast condition was consistent across scales. This makes sense because participants in this condition saw identical forecasts regardless of scale. Second, the mean preparatory action score for those in the probabilistic and combination forecast conditions both increased as scales increased in size. Interestingly, the mean score among combination forecast participants was higher than the mean score among probabilistic forecast participants for all scales except the multi-region; at this scale, those who saw probabilistic forecasts were more likely to take preparatory action. Moreover, the multi-region scale was the only scale for which the mean score among those who received categorical forecasts was lower than both other groups.

3.1.2. Random Effects

The nested effect of probability within scale was modeled as a random effect. While we purposefully chose the SPC probabilities for use in this study, these probabilities are only some of the possible probabilities across a probability distribution and thus were treated as a random factor. Figure 7 shows the preparatory action score trends across probability points for each scale. First, across all scales, preparatory action willingness increased as probability increased. Participants were generally most willing to take preparatory action when responding to a potential tornado impacting a multi-region area and less willing to take action when responding to a potential tornado threat impacting a city, even though they were asked to imagine they lived in that specific city. Also, the propensity to take preparatory action across probability points (i.e., the slope of each line) varied across scales; people were not only more willing to take action when responding to scenarios within the multi-region but they were also more willing to take action at a faster rate and lower probability point than when responding to any other scale. Propensity to take action at the other scales (city, county, and region) were fairly similar until the seventh probability point (1.20%, 7.50%, and 30%, respectively). At these probability points, the likelihood of taking action in the city decreased slightly and mostly plateaued while the likelihood of taking action slightly but steadily increased in the county and increased much more steeply in the region scales.

3.1.3. Examining Scale, Presentation Format, and Probability in Tandem

While the statistical analysis does not formally test for a significant three-way interaction, we can still examine the trend of preparatory action willingness across the three factors simultaneously (see Fig. 8a-d). The trend revealed the pattern of relations among scale, probability, and willingness to take action differed depending on forecast presentation format. Participants receiving the combined forecasts were more willing to take preparatory action in response to a tornado threat in a city and county at all probabilities of occurrence. This pattern also held for the region scale until its eleventh probability point (60% chance of occurrence) and beyond when participants who received only probabilistic forecasts became more willing to take preparatory action. The multi-region location, on the other hand, told a much different story. The only point at which participants who received the combination forecasts were more willing to take action was at the first probability point, or 8% chance of a tornado occurring. Beyond that probability point, participants who received only probabilistic forecasts were more willing to take action. Moreover, the propensity for taking preparatory action varied across locations. Participants were more likely to take action at a lower probability and at a much faster (steeper) rate for the region and multi-region locations than for the city and county locations.

3.2. The Value of Message Features

Participants were also asked to rate the importance of multiple aspects of the warning information to their decision-making process. On average, participants who received probabilistic forecast information rated it as highly important to their decisions, $M = 5.59$, $SD = 1.59$, $n = 279$. Similarly, participants who received color-coded categorical risk information also rated it as very important to their decisions, $M = 5.44$, $SD = 1.73$, $n = 270$. These same participants rated the map coloring scheme as moderately important to their decisions, $M = 5.02$, $SD = 1.96$, $n = 270$, but less important than the categorical information itself. Lastly, all participants reported having a map to look at while making decisions was moderately important, $M = 4.98$, $SD = 1.94$, $n = 407$; this aspect received the lowest importance rating. These means did not differ significantly as a function of presentation format, $F_{ProbabilityInfo} = 0.32$, $p = .58$; $F_{CategoryInfo} = 0.02$, $p = .89$; $F_{MapColoring} = 0.34$, $p = .56$; $F_{HavingMap} = 0.26$, $p = .77$. In other words, for example, having a map to look at while making decisions was equally important to all participants regardless of the type of information they saw.

4. DISCUSSION

This paper examines the interplay between geographic reference classes and probabilistic forecasts for individuals' protective decisions. The results partially supported the hypotheses; the first hypothesis, that information presented on local geographic scales might drive more protective actions, was not confirmed. However, the second hypothesis was confirmed—protective action was more likely as probabilities increased across any geographic scale. Further, when probabilities were presented, protective action was *less* likely as geographic scale decreased. Finally, individuals' willingness to take preparatory action across probability points and scales varied in complex ways depending on the format in which forecasts were presented. In particular, participants were more likely to take protective action when viewing the categorical forecasts—especially at higher risk categories—and least likely to take protective action when viewing probabilistic forecasts—across most probability points—with the likelihood of taking protective action in response to combination forecasts falling in-between the two. However, this pattern depended on geographic scale with less separation in tendencies at larger geographic scales (and even trend reversals). This suggests that probabilistic information alone was not enough to encourage protective action unless people were responding to larger geographic scales; probabilistic information may have been more helpful for helping respondents calibrate to the risk that was presented and was most helpful when *supplemented* with categories.

The novel contribution of this research was elucidating the interplay between geographic reference class and probabilistic forecasts in this specific context, which has interesting theoretical and practical implications. Previous literature examining the role of geography in protective decisions suggests that people are more likely to take protective action against personalized severe weather threats; that is, threats that are impacting their particular geographic location (e.g., Lindell 2020). Our results contradict this finding by showing that when probabilities are part of the risk information, participants were more likely to take protective action at larger geographic scales rather than more local scales. Our findings show tangible support for SPC's desire to create meaningful "probabilities one can feel." People were more likely to act at the regional and multi-regional scales and correspondingly higher valued probabilities as compared to the smaller, more local scales with correspondingly lower probabilities. These findings suggest, then, that there is in fact a tradeoff between personalized geographic scales and higher valued probabilistic forecasts. If the weather community wishes to provide forecasts that are very local, it would make sense to create them in such a way that the values could be relatively high (e.g., as in the storm-based, 1 km x 1 hr-scaled Probabilistic Hazard Information prototype; Rothfusz et al. 2018).

Similar to past research, presentation format also played a significant role in participants' decisions. Moreover, presentation format also impacted the effects of geographic scale and probability. Generally speaking, the combination of categorical and probabilistic forecast information encouraged participants to take preparatory action beyond the effect of either categorical or probabilistic forecast information alone. However, this pattern was reversed when participants were responding to the multi-region scale; in this case, only receiving probabilistic forecasts was more encouraging of preparatory action. It could have been that, at this scale, the categorical labels were not accurately capturing the risk perceived at the probability levels represented. In other words, maybe seeing 60% does not mentally equate to an “Enhanced” risk. Even though participants in the combined forecast condition saw both types of information, and as a whole the sample rated probabilistic information as more important to decision-making than categorical information, participants could have been anchoring on the category, or even the color coding of it, and placing less cognitive attention on the probabilistic information. Unfortunately, we did not include a measure to tease this possibility apart, but this would be good fodder for future research. There may also be individual differences in decision-making and numerical ability (e.g., Grounds and Joslyn 2018) as well as spatial ability (e.g., MacPherson-Krutsky et al. 2020) that may play a role in the extent to which participants used the probabilistic information and understood the map graphics. Future research should explore these possibilities as well. Lastly, there is a possibility that these findings are generalizable to other point hazards and these relations should be replicated in other contexts.

4.1. Limitations and Future Research

One major limitation of the present study was the design imbalance. Future work should aim to construct stimuli that vary in presentation format but maintain methodological balance so direct comparisons across presentation format can be made. Second, the stimuli in this study were hypothetical. While participants were primed to consider the presented geographic reference classes as places where they lived, a study examining the proposed relations between scale and probability may be more informative if there was a feedback loop that enabled a person’s actual geography to be inserted into the materials. Third, using an online questionnaire for experimental control to address the constructs studied also has an impact on the ecological validity of the findings, i.e., whether these results can be generalized to or expected to occur in real-life settings. Most significantly for this study, participants were all presented with locations that were unfamiliar and thus geography was a highly conceptual variable in ways it would not be for more naturalistic settings. As previously mentioned, future work should also assess the extent to which protective decisions, such as those made here, are impacted by individual psychological differences such as numerical ability and cognitive processing, account for more behavioral responses, and occur in response to actual severe weather events. Past work involving these types of factors has been minimal and produced mixed results. Also, as previously mentioned in section 2.3, the forecasts were presented in ascending order within each reference class, which may have led participants to anchor their decisions relatively (as the probabilities increased) instead of independently considering the forecast information in each scenario and then responding.

Future research should work to fully randomize all stimuli materials to ensure a complete test of these types of decisions independent of the order in which the materials were presented. In addition, our experiment held risk levels constant, exploring responses to geographic scales that increased alongside probability values. This limits our ability to determine what drives increasing responses where they are found, scale or probability. Future research could vary risk levels and hold probability values constant while increasing geographic scales, further isolating these effects. Lastly, this work studied individuals' responses to severe weather information using only one format for communicating probabilistic forecasts—using the existing SPC system. However, there are several ways to communicate probabilistic forecasts, some of which include climatological probabilities. One example would be to present odds ratios, which would account for climatology. This kind of format is beyond the scope of the present research but could be examined in future work.

4.2. Conclusion

Taken together, this work offers a novel approach to examining the tradeoff between geographic reference class and probabilistic forecast information in weather-related protective decisions. The relationship between these variables is complex and also impacted by the way in which forecast information is presented. These findings lend support for SPC's notion of creating "probabilities one can feel" at a scale that might seem large and impersonal but has practical forecast importance.

REFERENCES

Aguirre BE, Anderson WA, Balandran S, Peters BE, White MH (1991) Saragosa, Texas, Tornado May 22, 1987: An Evaluation of the Warning System. Washington, DC.

Boone AP, 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

Darley JM, Latané B (1968) Bystander intervention in emergencies: Diffusion of responsibility. Journal of Personality and Social Psychology 8:377-383. https://doi.org/10.1037/h0025589

Dziak JJ, Coffman DL, Lanza ST, Li R (2012, June 27) *Sensitivity and specificity of information criteria* (Technical Report Series #12-119). The Methodology Center. https://www.methodology.psu.edu/files/2019/03/12-119-2e90hc6.pdf

Gibson-Graham JK (2002) Beyond global vs. local: Economic politics outside the binary frame. In: Herod A, Wright MW (eds) Geographies of power: Placing scale. Blackwell, Oxford, pp 25-60

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

Grounds MA, Joslyn SL (2018) Communicating weather forecast uncertainty: Do individual differences matter? Journal of Experimental Psychology: Applied 24:18-33. http://dx.doi.org/10.1037/xap0000165

Grounds MA, Joslyn S, Otsuka K (2017) Probabilistic interval forecasts: An individual differences approach to understanding forecast communication. Advances in Meteorology 2017:1-18. https://doi.org/10.1155/2017/3932565

Jenni K, Loewenstein G (1997) Explaining the identifiable victim effect. Journal of Risk and Uncertainty 14:235–257. https://doi.org/10.1023/A:1007740225484

Johnson N (2013) How people respond: A meta-synthesis of post-tornado interviews and surveys. Master’s Thesis, North Carolina State University

Joslyn SL, LeClerc JE (2012) Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. Journal of Experimental Psychology: Applied 18:126-140. https://doi.org/10.1037/a0025185

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

Joslyn SL, Nadav-Greenberg L, Taing MU, Nichols RM (2009) The effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. Applied Cognitive Psychology 23:55-72. https://doi.org/10.1002/acp.1449

Joslyn S, Pak K, Jones D, Pyles J, Hunt E (2007) The effect of probabilistic information on threshold forecasts. Weather and Forecasting 22:804-812. https://doi.org/10.1175/WAF1020.1

Keppel G, Wickens TD (2004) *Design and analysis: A researcher's handbook* (4th ed). Pearson Education.

Klockow K (2013) Spatializing tornado warning lead-time: Risk perception and response in a spatio-temporal framework. Dissertation, University of Oklahoma

Klockow-McClain KE, McPherson RA, Thomas RP (2019) Cartographic design for improved decision making: Trade-offs in uncertainty visualization for tornado threats. Annals of the American Association of Geographers 110:314-333. https://doi.org/10.1080/24694452.2019.1602467

Latané B, Darley JM (1968) Group inhibition of bystander intervention in emergencies. Journal of Personality and Social Psychology 10:215-221. https://doi.org/10.1037/h0026570

LeClerc J, Joslyn S (2012) Odds ratio forecasts increase precautionary action for extreme weather events. Weather, Climate, and Society 4:263-270. https://doi.org/10.1175/WCAS-D-12-00013.1

LeClerc J, Joslyn S (2015) The cry wolf effect and weather-related decision making. Risk Analysis 35:385-395. https://doi.org/10.1111/risa.12336

Lindell MK (2020) Improving hazard map comprehension for protective action decision making. Frontiers in Computer Science 2:1-14. https://doi.org/10.3389/fcomp.2020.00027

Liu L, Boone AP, Ruginski IT, Padilla L, Hegarty M, Creem-Regehr SH, Thompson WB, Yuksel C, House DH (2017) Uncertainty visualization by representative sampling from prediction ensembles. IEEE Transactions on Visualization and Computer Graphics 23:2165-2178. https://doi.org/10.1109/TVCG.2016.2607204

MacPherson-Krutsky CC, Brand BD, Lindell MK (2020) Does updating natural hazard maps to reflect best practices increase viewer comprehension to risk? International Journal of Disaster Risk Reduction 46:101487. https://doi.org/10.1016/j.ijdrr.2020.101487

Millet B, Carter AP, Braod K, Cairo A, Evans SD, Majumdar SJ (2020) Hurricane risk communication: Visualization and behavioral science concepts. Weather, Climate, and Society 12:193-211. https://doi.org/10.1175/WCAS-D-19-0011.1

Miran SM, Ling C, Gerard A, Rothfusz L (2018) The effect of providing probabilistic information about a tornado threat on people's protective actions. Natural Hazards 94:743-758. https://doi.org/10.1007/s11069-018-3418-5

Morss RE, Demuth JL, Lazo JK (2008) Communicating uncertainty in weather forecasts: A survey of the U.S. public. Weather and Forecasting 23:974-991. https://doi.org/10.1175/2008WAF2007088.1

Murphy AH, Lichtenstein S, Fischhoff B, Winkler RL (1980) Misinterpretations of precipitation probability forecasts. Bulletin of the American Meteorological Society 61:695-701. https://doi.org/10.1175/1520-0477(1980)061<0695:MOPPF>2.0.CO;2

Nagele DE, Trainor JE (2012) Geographic specificity, tornadoes, and protective action. Weather, Climate, and Society 4:145-155. https://doi.org/10.1175/WCAS-D-11-00047.1

Padilla LM, Ruginski IT, Creem-Regehr SH (2017) Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. Cognitive Research: Principles and Implications 2:1-16. https://doi.org/10.1186/s41235-017-0076-1

Rothfusz L, Schneider R, Novak D, Klockow KE, Gerard A, Karstens C, Stumpf G, Smith T (2018) FACETs: A proposed next-generation paradigm for high-impact weather forecasting. Bulletin of the American Meteorological Society 99:2025–2043. https://doi.org/10.1175/BAMS-D-16-0100.1

Ruginski IT, Boone AP, Padilla LM, Liu L, Heydari N, Kramer HS, Hegarty M, Thompson WB, House DH, Creem-Regehr SH (2016) Non-expert interpretations of hurricane forecast uncertainty visualizations. Spatial Cognition & Computation 16:154-172. http://dx.doi.org/10.1080/13875868.2015.1137577

Seltman HJ (2018) *Experimental design and analysis* [eBook]. Carnegie Mellon University. https://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf

Severtson DJ, Burt JE (2012) The influence of mapped hazards on risk beliefs: A proximity-based modeling approach. Risk Analysis 32:259-280. https://doi.org/10.1111/j.1539-6924.2011.01700.x

Starkweather J, Herrington R (2018, November 6) *Linear mixed effects modeling* [SPSS Short Course]. Data Science & Analytics@UNT.

http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/SPSS_SC/Module9/M9_LMM/SPSS_M9_LMM.htm

Storm Prediction Center (2017) SPC products: Day 1 outlook probability to category conversion. https://www.spc.noaa.gov/misc/about.html

Teigen KH (2005) The proximity heuristic in judgments of accident probabilities. British Journal of Psychology 96:423-440. https://doi.org/10.1348/000712605X47431

United States Census Bureau (2019). 2018 population estimates by age, sex, race and Hispanic origin. https://www.census.gov/newsroom/press-kits/2019/detailed-estimates.html

Wallsten TS, Budescu DV, Zwick R, Kemp SM (1993) Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bulletin of the Psychonomic Society 31:135-138.

FIGURE LIST

Fig. 1 Pictorial representation of the experimental nested design: 3 (Presentation Format) x 4 (Geographic Reference Class) x 12 (Probability)

Fig. 2. Example of a scenario given to participants

Fig. 3. Panel of presentation format variations. Fig. 3a depicts a sample forecast with probability presentation format, Fig. 3b depicts a sample forecast with the category presentation format, and Fig. 3c depicts a sample forecast with the combined presentation format

Fig. 4. Effect of geographic reference class on preparatory decisions

Fig. 5. Effect of presentation format on preparatory decisions among participants receiving probabilistic and combination forecasts

Fig. 6. Interactive effect of geographic reference class and presentation format on preparatory decisions among participants receiving probabilistic and combination forecasts

Fig. 7. Nested effect of probability within geographic reference class on preparatory decisions

Fig. 8. Panel of three-way interaction trend. Graphs show the interaction between probability and presentation format separated by geographic reference class. Fig. 8a depicts the interaction at the city scale, Fig. 8b depicts the interaction at the county scale, Fig. 8c depicts the interaction at the region scale, and Fig. 8d depicts the interaction at the multi-region scale

TABLE LIST

Table 1. Demographic Information for Participants Compared with the 2018 US Census

Characteristic	Participants[a] (%)	US Adult Population[b] (%)
NWS Region		
Eastern Region	31.7%	30.7%
Southern Region	27.1%	27.2%
Central Region	20.7%	20.6%
Western Region	20.5%	20.6%
Age Group		
18 to 24	12.2%	12.0%
25 to 34	18.1%	18.0%
35 to 44	16.3%	16.3%
45 to 54	16.3%	16.4%
55 to 64	16.7%	16.7%
65 and up	20.4%	20.6%
Gender		
Female	51.4%	50.8%
Male	46.6%	49.2%

[a] Nine (9) participants did not respond to all demographic items.

[b] Population estimates were obtained from the U.S. Census 2018 Population Estimates by Age, Sex, Race and Hispanic Origin.

Table 2. Breakdown of Probability Conversions by Risk Category, Geographic Reference Class, and Severity

Probability	Risk Category	Geographic Reference Class				Severity
		City	County	Region	Multi-Region	
1	Marginal	0.08%	0.50%	2%	8%	a
2	Slight	0.20%	1.25%	5%	20%	a
3	Enhanced	0.40%	2.50%	10%	40%	a
4	Enhanced	0.40%	2.50%	10%	40%	Sig. Severe Risk
5	Enhanced	0.60%	3.75%	15%	60%	a
6	Moderate	0.60%	3.75%	15%	60%	Sig. Severe Risk
7	Moderate	1.20%	7.50%	30%	100% [b]	a
8	High	1.20%	7.50%	30%	100% [b]	Sig. Severe Risk
9	High	1.80%	11.25%	45%	c	a
10	High	1.80%	11.25%	45%	c	Sig. Severe Risk
11	High	2.40%	15.00%	60%	c	a
12	High	2.40%	15.00%	60%	c	Sig. Severe Risk

[a] Severity designation does not exist for this probability point.

[b] This probability was truncated at 100% for practicality; the conversion was greater than 100%.

[c] This probability converted to a number greater than 100% and thus was not shown to participants.

Fig. 1
(Prepared using Google Slides)

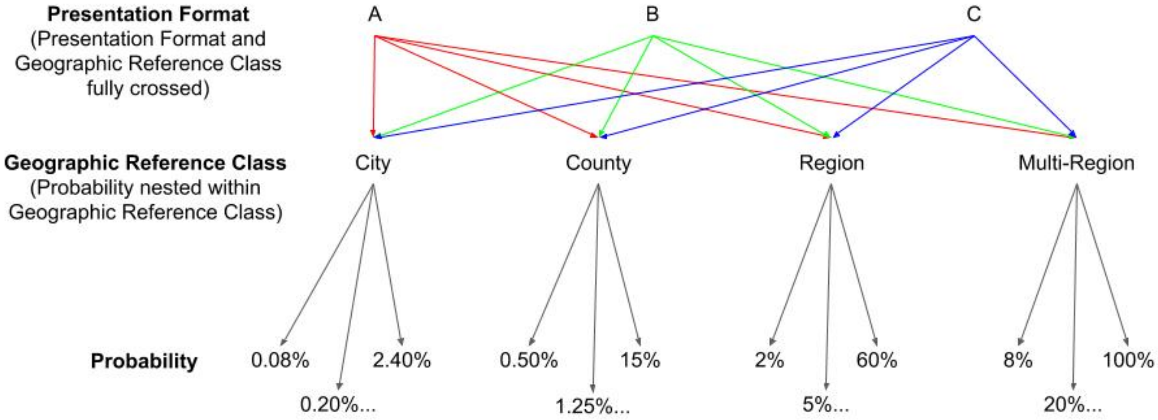
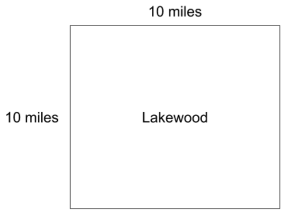


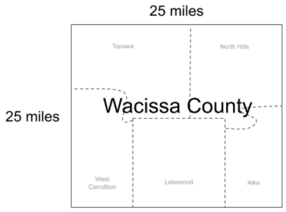
Fig. 2
(Prepared using Google Slides)

(a)



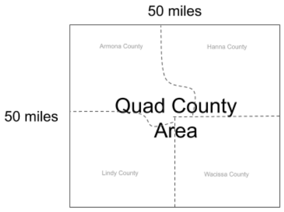
According to the Storm Prediction Center (SPC) Convective Outlook, there is a **0.08% chance** of a tornado **today in Lakewood.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

(b)



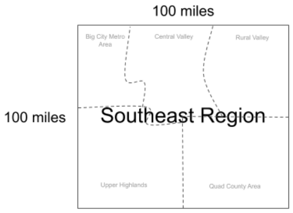
According to the Storm Prediction Center (SPC) Convective Outlook, there is a **0.50% chance** of a tornado **today in Wacissa County.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

(c)



According to the Storm Prediction Center (SPC) Convective Outlook, there is a **2% chance** of a tornado **today in the Quad County Area.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

(d)

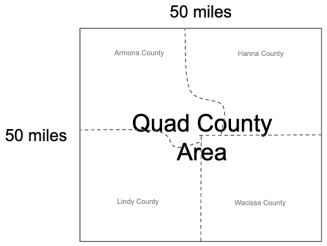


According to the Storm Prediction Center (SPC) Convective Outlook, there is an **8% chance** of a tornado **today in the Southeast Region.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

Fig. 3

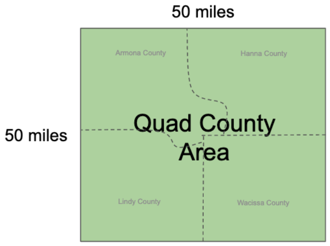
(Prepared using Google Slides)

(a)



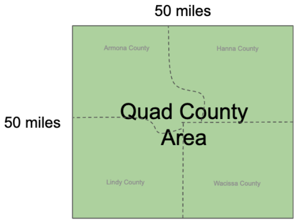
According to the Storm Prediction Center (SPC) Convective Outlook, there is a **2% chance** of a tornado **today in the Quad County Area.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

(b)



According to the Storm Prediction Center (SPC) Convective Outlook, there is a **Marginal Risk** of a tornado **today in the Quad County Area**. Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

(c)



According to the Storm Prediction Center (SPC) Convective Outlook, there is a **Marginal Risk** of a tornado **today in the Quad County Area**, which corresponds to a **2% chance** of a tornado **in the Quad County Area.** Based on the information provided, how likely are you to take preparatory action today in response to the potential tornado threat?

Fig. 4
(Prepared using SPSS 26)

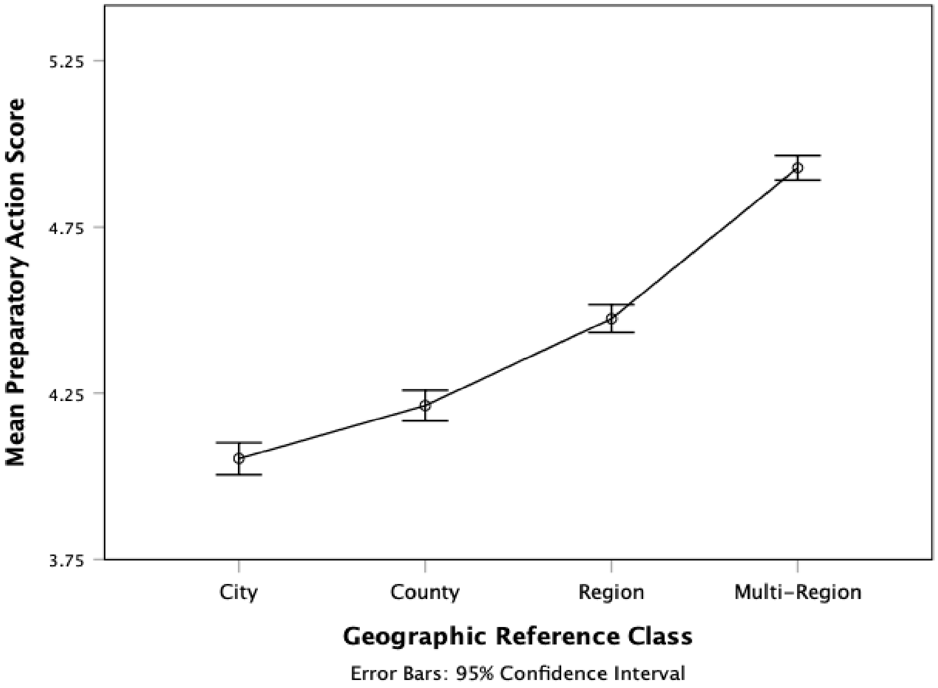


Fig. 5
(Prepared using SPSS 26)

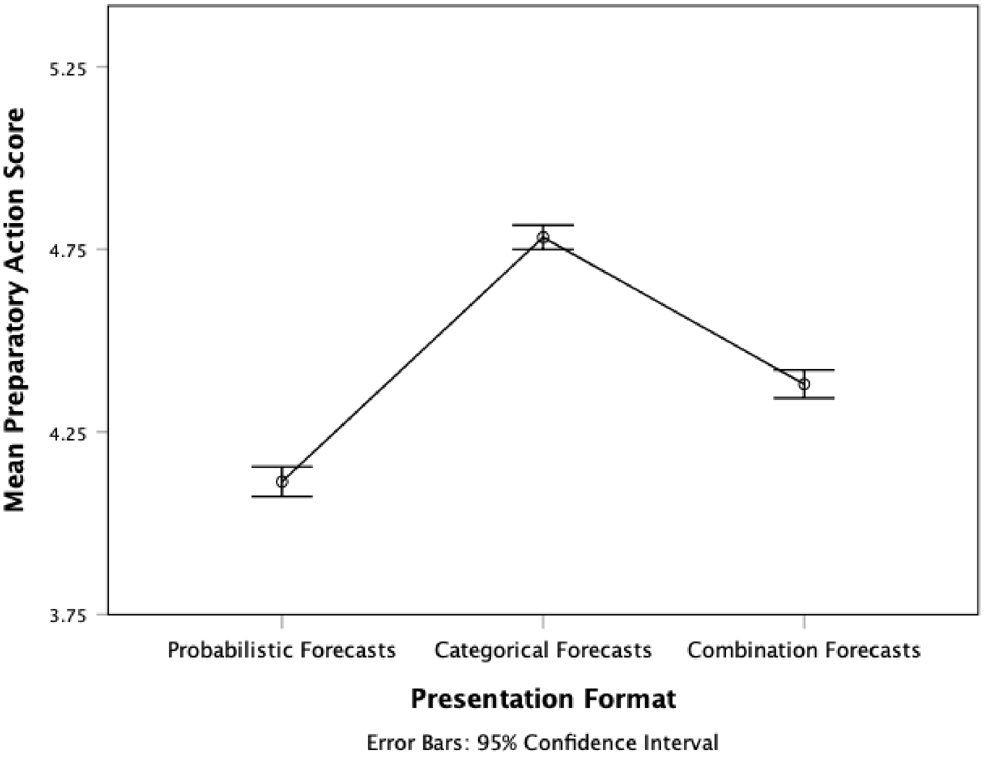


Fig. 6
(Prepared using SPSS 26)

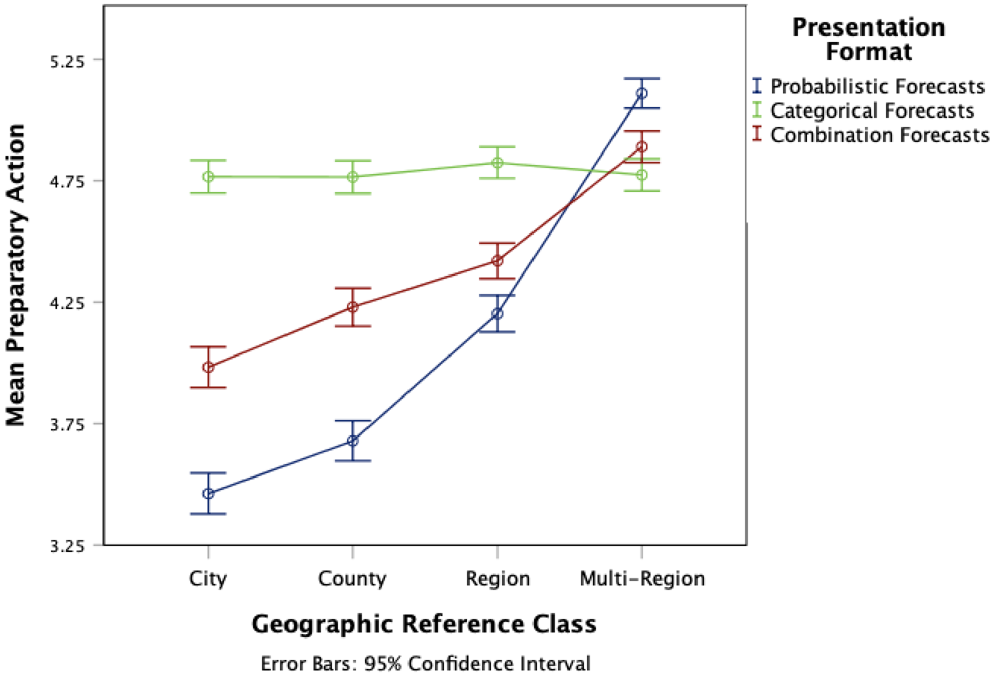


Fig. 7
(Prepared using SPSS 26)

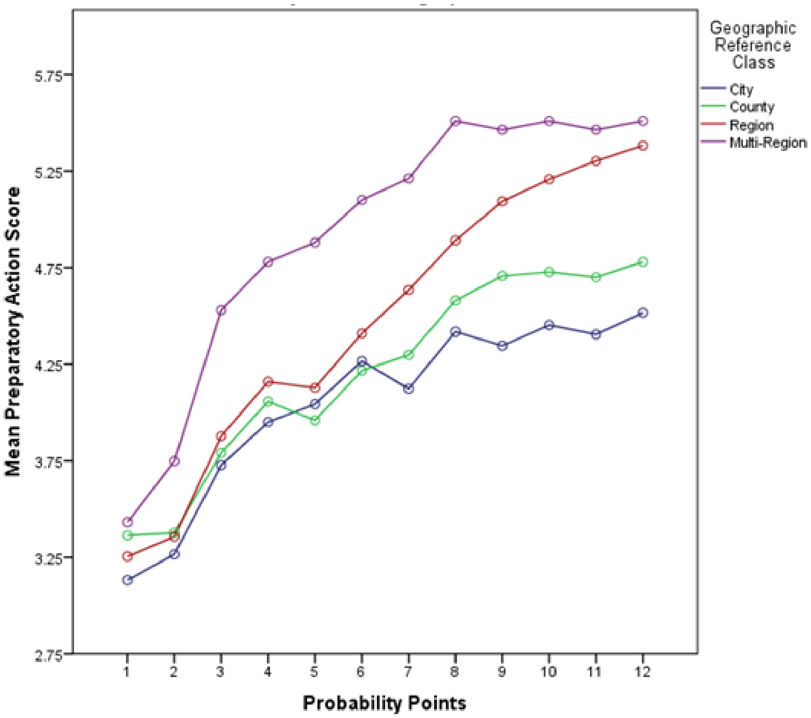


Fig. 8
(Prepared using SPSS 26)