Effect of Providing the Uncertainty Information about a Tornado Occurrence on the Weather Recipients' Cognition and Protective Action: Probabilistic Hazard Information versus Deterministic Warnings

Seyed M. Miran,[1],* Chen Ling,[1] Alan Gerard,[2] and Lans Rothfusz[2]

[1]Department of Mechanical Engineering, The University of Akron, Akron, OH, USA.

[2]NOAA National Severe Storms Laboratory, Norman, OK, USA.

Abstract

Currently, a binary alarm system is used in the United States to issue deterministic warning polygons in case of tornado events. To enhance the effectiveness of the weather information, a likelihood alarm system, which uses a tool called Probabilistic Hazard Information (PHI), is being developed at National Severe Storms Laboratory to issue probabilistic information about the threat. This study aims to investigate the effects of providing the uncertainty information about a tornado occurrence through the PHI's graphical swath on laypeople's concern, fear, and protective action, as compared with providing the warning information with the deterministic polygon. The displays of color-coded swaths and deterministic polygons were shown to subjects. Some displays had a blue background denoting the probability of any tornado formation in the general area. Participants were asked to report their levels of concern, fear, and protective action at randomly chosen locations within each of seven designated levels on each display. Analysis of a three-stage nested design showed that providing the uncertainty information via the PHI would appropriately increase recipients' levels of concern, fear, and protective action in highly dangerous scenarios, with a more than 60% chance of being affected by the threat, as compared with deterministic polygons. The blue background and the color-coding type did not have a significant effect on the people's cognition of the threat and reaction to it. This study shows that using a likelihood alarm system leads to more conscious decision making by the weather information recipients and enhances the system safety.

KEY WORDS: Likelihood alarm systems; probabilistic hazard information; probability matching; risk communication; WarnGen

Social Media Summary:

Using likelihood alarm systems leads to more conscious decision-making by the weather information recipients, and enhances system safety. In scenarios with a high level of urgency, the effect of probability matching on the people's behavior during a hypothetical tornado event is increased. In other words, there is an interaction between the effect of the urgency matching and the probability matching for probabilistic hazard information.

1. INTRODUCTION

Severe storms, including tornadoes, have killed more than 1,500 people and have caused more than $260 billion in damages in the United States since 1980 (National Oceanic and Atmospheric Administration, 2017). Tornadoes are one of the most devastating natural disasters that pose a grave risk to residents of this country (Hammer and Schmidlin, 2001).

Taking protective action is an effective way to mitigate negative consequences of such a disastrous natural hazard (e.g., Grothman and Reusswig, 2006). According to the Protective Action Decision Model by Lindell and Perry (2012), different types of weather warnings can lead to different protective action by people. The reason has been attributed to different levels of fear, as primary motivation for protective action, that different warnings pose to the people (Lundgren, 1994). To lower the adverse effect of tornadoes, people should be provided with effective weather information and ensure that they can comprehend the information easily and take protective behavior appropriately.

1.1. The Current Binary Alarm System

Ideally, a warning should be issued only if a severe tornado occurs. Because the atmosphere is dynamic and inherently unpredictable, any weather forecast has uncertainty (Murphy, 1998). Therefore, any binary warning system, which has two modes of alarm or no-alarm, results in a significant number of false alarms even when warnings are scientifically accurate.

False alarms can cause tremendous adverse effects for the public (Simmons and Sutter, 2009), such as business closure and unnecessary evacuation of people (Durage, Wirasinghe, and Ruwanpura, 2015). In addition, false warnings can cause people to ignore future warnings because of the "cry wolf" effect, and to avoid taking protective action in a future threat occurrence (Edwards and Lemon, 2002). As a result, the warning's credibility is reduced (Madhavan, Wiegmann, and Lacson, 2006) and the public safety decreases.

Currently, the National Weather Service (NWS) uses a system that issues binary warnings for severe hazardous events, such as tornadoes, with a tool called "WarnGen" (Coleman, Knupp, Spann, Elliott, and Peters, 2011). Forecasters use this tool to draw polygons to indicate the areas that are expected to be affected by the threat in the near future. These warnings are deterministic, and a given location is either in or out of the risk areas, and there is no differentiation for the chance of a threat occurrence in areas inside the polygon. In reality, there is an uncertainty involved in the threat occurrence in different locations (Murphy, 1998). It means that there is a chance that the threat would not occur in some locations inside the polygon (Sutter and Erickson, 2010). WarnGen polygons, however, cannot provide the weather information recipients with uncertainty information about the tornado occurrence in different locations. As a result, the chance of issuing false alarms increases and the credibility of the warnings decreases.

1.2. The New Likelihood Alarm System

Another type of alarm system is called a *likelihood alarm system*. In this type of system, there are several stages corresponding to different likelihoods that a critical event happens. These stages can be represented by different colors or with different wordings and characteristics (Sorkin, Kantowitz, and Kantowitz, 1988).

The likelihood alarm system is based upon two human-automation interaction ideas: probability matching and urgency matching (Bustamante, 2008). *Probability matching* means that recipients of the alerts match their responses to an alert with the likelihood of the event occurrence (Bliss, Gilson, and Deaton, 1995). *Urgency matching* means that the recipients of alerts coordinate their behaviors with the perceived urgency of the event occurrence (Haas and Casali, 1995). Research studies have found that the likelihood alarm system improves the accuracy of users' decisions (e.g., Bliss, Gilson, and Deaton, 1995;

Bustamante, 2008; Clark, Ingebritsen, and Bustamante, 2010; Wiczorek and Manzey, 2011; Balaud and Manzey, 2014).

Scientists at National Severe Storms Laboratory are working on a program called *Forecasting a Continuum of Environmental Threats* (FACETs; Rothfusz, Karstens, and Hilderbrand, 2014), which provides a continuous flow of information about weather updates. Probabilistic hazard information (PHI) is a core concept of FACET that is being developed at the National Severe Storms Laboratory to potentially replace WarnGen (Karstens et al., 2014). This concept incorporates the idea of probability matching by providing the likelihood of the threat occurrence at the location of an information recipient. The idea of urgency matching, which is assumed to be triggered by the perceived urgency of the threat occurrence owing to nearness to the threat, is also incorporated by informing the recipient about the time of the threat arrival.

2. Background

Bliss, Gilson, and Deaton (1995) investigated the effect of the probability matching on the people's response to alarms with varying reliability: 25%, 50%, and 75% true alarms. The researchers used a combination of visual and auditory alarms to resemble the real-world alarm conditions (e.g., civilian aircraft cockpits). The subjects were involved with a complex psychomotor activity and at the same time were presented with the alarm. The researchers found that 90% of the subjects matched their responses to the alarm with the probability of true alarms (probability matching).

Wiczorek and Manzey (2011) conducted an experiment and studied the possible advantages of a likelihood alarm system over a binary system in a multitask environment. The authors concluded that the likelihood system was more effective than the binary one, and it improved the user's decision-making accuracy. In another study, Wiczorek, Manzey, and Zirk (2014) compared a binary alarm system with a likelihood alarm system to investigate how the number of stages in a likelihood alarm system can affect people's responses to the alerts. The researchers found that the differentiated information provided by the likelihood alarm system improved the accuracy of the participants' responses compared with a binary system, and as the number of stages in a likelihood alarm system increased, the participants made fewer wrong decisions.

There are a few previous research studies investigating human probability-matching behavior in response to the tornado threats. Klockow (2013) conducted an experiment using a deterministic visual information about the threat occurrence and a probabilistic one. The researcher did not find any significant difference between people's responses to deterministic and binary warnings and the responses to probabilistic warnings.

Ash, Schumann III, and Bowser (2014) conducted an experimental study to investigate the effect of visual probabilistic tornado information on the people's interpretation of the threat and their reactions to it. The participants of that study were presented with the visual information in which they could identify their hypothetical locations relative to a tornado threat and the chance of being affected by that tornado. The authors concluded that the probability of the threat occurrence affected people's level of concern and the participants adjusted their protective action with that probability. The same conclusion was drawn in a similar experiment by Balaud and Manzey (2014).

2.1. How to Convey the Likelihood of Threat

Most people in the United States are aware that the weather information involves uncertainty (Morss, Demuth, and Lazo, 2008) and are prepared to be exposed to that uncertainty (Joslyn and Savelli, 2010). It is believed that providing explicit uncertainty information about the weather forecast benefits the recipient of the information (Joslyn and Savelli, 2010; Kox, Gerhold, and Ulbrich, 2015). It needs to be determined, however, how this uncertainty information should be presented to the public.

In the case of the probabilistic hazard information (PHI) swath, the likelihood of the threat occurrence at different levels of the mentioned swaths could be conveyed with verbal terms or numerical probability. Verbal expressions of the uncertainty (e.g., slight chance), however, can be interpreted differently by people, and each person can have a distinct perception of it (Saviers and Van Bussum, 1999). Besides, it is a challenge to standardize the meanings of the verbal probability to the public, and there is an unneglectable variability in interpreting the probability phrases (Budescu, Por, and Broomell, 2012).

Many researchers and meteorologists believe that quantitative probabilities are clearer and more precise, preferring to use them over the qualitative statements about the probability (Murphy and Winkler, 1971; Monahan and Steadman, 1996; Budescu, Broomel, and Por, 2009). However, the numerical probability cannot eliminate the ambiguity (Handmer and Proudley, 2007), and color-coding methods have been suggested to replace the numeric methods for conveying the uncertainty information (Christner and Ray, 1961). Previous research studies show that the color-coding can improve cognitive processing (Kopala, 1979), and can be a sensible method to convey the uncertainty information during a hazardous event (Hoffman, Detweiler, Conway, and Lipton, 1993; Ash et al., 2014).

Choosing the most appropriate colors to convey the appropriate level of risk has been a matter of research. According to the traditional way of color-coding, red should be used to indicate the highest level of danger, followed by orange, yellow, green, and blue (Bresnahan and Bryk, 1975).

Color saturation methods, which use different intensities of a hue, and color hue methods, which use different colors to denote different levels of the uncertainty, have been proposed to be used as possible visualization methods to convey the uncertainty (Kunz, Grêt-Regamey, and Hurni, 2011).

2.2. Probabilistic Hazard Information

With PHI, a swath with multiple color-coded probability levels is used to convey the likelihood that an area being impacted by a threat. Each swath's level represents different likelihood of the threat occurrence (Ling et al., 2015). There is an area at the beginning of the swath showing the location of the threat at the moment, and the end of the swath indicates the expected location of the threat within a specific time into the future, such as 50 minutes. The recipients could compare their locations with the current location of the threat at the beginning of the PHI swath to know how much time they have before the possible threat arrival.

It has been shown that an appropriate way of conveying the uncertainty information using the PHI swath is through color-coding of its levels along with the quantitative probability associated with each color on a separate reference bar (Miran, Ling, James, and Rothfusz, 2016; Miran, Ling, James, Gerard, and Rothfusz, 2017a; Miran, Ling, James, Gerard, and Rothfusz, 2017b; Miran, Ling, Gerard, and Rothfusz, 2018a). As a result, the PHI enables the probability matching through conveying the likelihood of the threat occurrence at the recipients' location via the color-coding and the urgency matching by providing information about recipient's relative location to the threat and the approximate lead time.

3. Aims of the research and the hypotheses

The main goal of the current research is to investigate how providing the uncertainty information through the PHI swath affects people's concern, fear, and protective action in case of a hypothetical tornado. In other words, how does providing the probability matching in the PHI swath affect people's cognition and protective action compared with not providing any probabilistic information in a WarnGen polygon?

Although the PHI swaths or polygons present information on a particular tornado threat, the atmospheric environment in the general area may be prone to produce more tornadoes. The secondary research goal is to study how providing the uncertainty information on formation of any tornado in the general surrounding area affects weather information recipients' levels of concern, fear, and protective action. In addition, we are interested in learning whether different ways of color-coding for PHI have different effects on people's responses.

It is hypothesized that conveying the uncertainty information about the threat occurrence through the PHI poses a more appropriate level of concern and fear, and it elicits more appropriate protective action, compared with a WarnGen polygon. In addition, it is hypothesized that providing the background probability information will render more appropriate responses compared with having no background information. We also hypothesize that people understand a conventional method for color-coding better than other possible methods.

4. METHODS

Fifty random people were hired from public premises in Akron, Ohio, to participate in this experiment. All the participants were the U.S. citizens with no color blindness. The Institutional Review Board at the University of Akron approved this study, and it was conducted in accordance with the university's code of ethics. Each participant signed an informed consent form and received a $25 gift card as remuneration.

4.1. Design of the Hazard Displays

One design for the PHI swath had a five-color color-coding scheme (Bresnahan and Bryk, 1975). AutoCAD 2017 was used to create designs and displays. The swath was divided into five levels denoting different probability levels of the threat occurrence. These five levels on the PHI swath, starting from the beginning of the swath, were denoted as recipient's level 1 through 5, respectively.

The color fuchsia was used to denote the first level, the innermost level of the swath, corresponding to the probability of 80%–100%. Two shades of red denoted the next two probability levels, corresponding to the probability of 60%–80% for the first shade and 40%–60% for the second one. For the next levels, one shade of orange, the probability of 20%–40%, and one shade of yellow, the probability of 0%–20%, was used (Fig. 1a, 1b).

To compare the effectiveness of the five-color design with another way of color coding, a red-scale design with five shades of red (Fig. 1c, 1d) showing the different probability levels was used (Miran, et al., 2017a). The color information can be seen in Table I.

The other design type was polygon. To evaluate the added value of presenting the uncertainty information for PHI, the experiment was designed so that the deterministic polygon could provide some cues for urgency matching. Five imaginary levels on the polygon were determined, from the narrower side to the wider side of the polygon (Fig. 1e, 1f). These levels, starting from the narrower side of the polygon, are referred to as *recipient's level* 1 through 5, respectively.

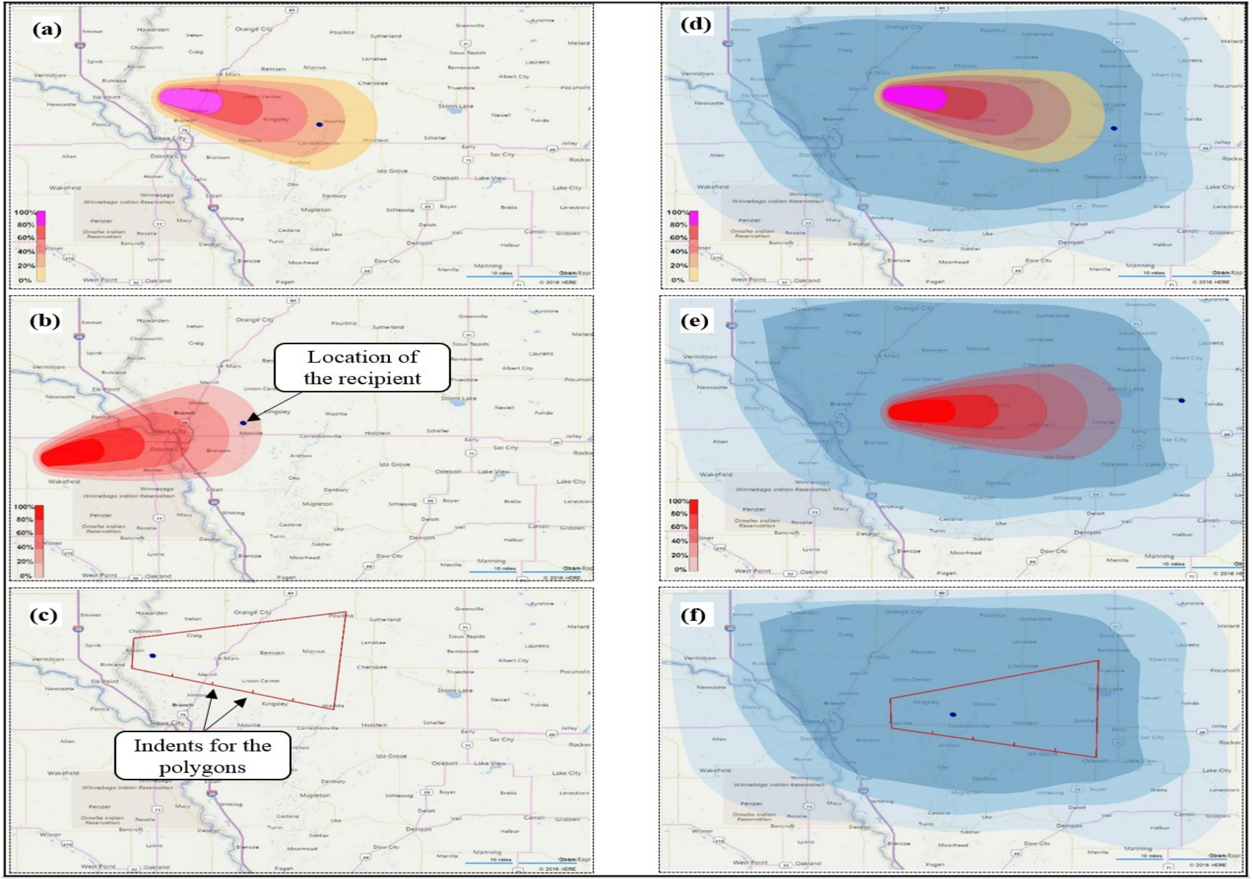

Fig. 1. Different designs used in the experiment: (a) five-color scheme without background; (b) five-color scheme with background; (c) red-scale scheme without background; (d) red-scale scheme with background; (e) polygon without background; (f) polygon with background

Three shades of blue were used for the background of the designs to convey the likelihood of any tornado formation in the general surrounding area. The darkest shade, the innermost level, corresponds to the highest likelihood and the lightest shade, the outermost level, corresponds to the lowest likelihood (Fig. 1b, 1d). The three levels on the background, starting from the innermost level (the darkest one), are referred to as *recipient levels* B1, B2 and B3.

For each swath's level or polygon's level, two dots were randomly selected by coding in Python's random module. Each dot represents the location of a recipient. For the WarnGen polygon, indenting was used to help participants distinguish different levels (Fig. 1c).

Table I. Color Information of the Color-Coded Objects

Color-Coded Objects	Level	RGB Values	Transparency
Five-Color	1	225, 0, 225	10
	2	245, 0, 0	40
	3	246, 141, 138	50
	4	244, 123, 62	50
	5	254, 204, 102	45
Red-Scale	1	255, 0 ,0	10
	2	255, 0, 0	30
	3	255, 0, 0	50
	4	255, 0, 0	65
	5	255, 0, 0	80
Background	B1	224, 239, 249	1
	B2	195, 224, 242	1
	B3	131, 187, 220	1

Design types were randomly located on a map, and different displays were created. For red-scale and five-color designs, a reference bar was shown on the bottom left of displays indicating the numeric probability of the threat occurrence at each level, and the participants received the necessary instructions before beginning the experiment.

For “with background” (WB) displays, two random dots were selected at the first two levels of the blue background (B1 and B2). For “without background” (WOB) displays, two imaginary levels, corresponding to the two blue levels of WB displays, were considered and two random dots on each of them were selected. The selection resulted in seven levels for each display. For each design type, there were 28 different displays: 7 (levels) × 2 (dots) × 2 (backgrounds). Eighty-four scenarios (28 × 3 design types) were created in the experiment.

4.2. Design of Questions

The following multiple-choice questions were asked for each of the 84 scenarios. The participants were instructed to answer the questions as quickly and accurately as possible without any prioritization for speed versus accuracy. Questions were shown below the displays.

1. How concerned are you about the threat?

A. Completely not concerned

B. Not concerned

C. Not much concerned

D. Neutral

E. Slightly concerned

F. Concerned

G. Completely concerned

2. At your current location in relation to the threat, how afraid are you for your life and propertyproperty?

A. Completely not afraid

B. Not afraid

C. Not much afraid

D. Neutral

E. Slightly afraid

F. Afraid

G. Completely afraid

3. At your current location in relation to the threat, what protective actions would you take?

A. I do not care at all.

B. I would vigilantly seek out weather reports and prepare the home for tornado.

C. I would take shelter immediately.

4.3. Design of the Experiment

There were three independent variables in the experiment: background, design, and recipient's level. The presence (WB) and absence of the blue background (WOB) were used as two levels of the background independent variable. The five-color and red-scale, two different types of the PHI swath, along with the polygon of WarnGen were three levels of design-independent variables. There were seven levels for the "recipient's level" variable, with levels 1 through 5 on the warning design and levels B1 and B2 on the background. The three dependent variables were the participant's responses to each question, which were about the levels of concern, fear, and protective action.

In the current study, a three-stage nested design was used. The first factor was the fixed-effect background. The second factor was design-nested in background, called "design (background)". The nested relationship is because the appearance of designs (both PHI and polygon) are different in WB and WOB displays (Fig. 1). The third factor was "recipient's level" nested in "design" and "background," called "level (background, design)." It is nested because the levels were presented differently in the different combination of designs and backgrounds. The PHI swath was color-coded, and the polygon was not. In PHI designs, red-scale and five-color designs had different colors for each specific level.

4.4. Procedure

Before the experiment, the participants were instructed that the color of each level on the PHI swath indicates the likelihood of the threat occurrence. They were also told that the tornado's location at the moment is at the beginning of the swath, and it takes 50 minutes for it to move from the beginning to the end of the swath. In reality, the WarnGen polygon cannot show the exact location of the tornado to the warning recipients. For the experimental purpose to evaluate the added value of having probability matching versus having only urgency matching information, we provided the information about urgency matching in the polygon design. Participants were instructed to assume that the tornado's location at the moment is at the first level near the narrower side of the polygon, and it takes 50 minutes to move to the wider side. It should also be assumed that it takes 10 minutes for the threat to pass each level.

E-Prime software was used to display images randomly to the participants, who used a computer keyboard to respond to the questions. At the end of the experiment, the participants were asked to select their most preferred display, list advantages and disadvantages of each of six display types—3 (design types) × 2 (with or without background)—and explain the effect of the presence of the background on their decisions.

4.5. Data Analysis

One way to analyze the ordinal response variable is to assign ordered scores to the responses and conduct an analysis of variance (Agresti, 2002). Numbers 1 through 7 were assigned to seven choices of questions 1 and 2. Among the response choices, "Completely not concerned" was mapped to 1, "completely concerned" was mapped to 7, and the choices between these two were mapped to 2 through 6. Regarding question 3, the first response choice was mapped to 1, the second one was mapped to 2, and the third one was mapped to 3.

Using Minitab statistical package, initially, a model with all three independent variables was built for each question. Three models were built for "concern," "fear," or "protective action" as the dependent variable in each model. Next, to investigate the effect of providing probability matching on the participants' concern, fear, or protective action at each recipient's level, a model with "design (background)" and "background" as independent variables and "concern," "fear," or "protective action" as the dependent variable was run for each recipient's level, with a total of 21 models (7 levels × 3 dependent variables). In cases that the "design (background)" was statistically significant, Tukey tests were used to compare design types. In case of discrepancy with a previous analysis of variance that used more lenient standards, the stricter Tukey results were used.

5. RESULTS

5.1. Analysis of the Level of Concern

In the model with “background,” “design (background),” and “level (background, design)” as independent variables and “concern” as the dependent variable, all three dependent variables are statistically significant. The results are shown in Table II.

Table II. Results of the Analysis of the Models with All Three Independent Factors

Dependent Variables	Level	Independent Variables					
		Background		Design (Background)		Level (Background, Design)	
		F Value (df)	*p* Value	*F* Value (df)	*p* Value	*F* Value (df)	*p* Value
Concern	Full	6.52 (1, 4158)	0.01*	3.38 (4, 4158)	0.01*	169.98 (36, 4158)	0.00*
Fear	Full	3.61 (1, 4158)	0.06	2.47 (4, 4158)	0.04*	277.86 (36, 4158)	0.00*
Protective action	Full	3.03 (1, 4158)	0.08	2.72 (4, 4158)	0.03*	159.05 (36, 4158)	0.00*

*The *p* value is less than α = 0.05.

Regarding results of the models for each recipient’s level, for the recipient’s levels 1 through 5, the “background” is not a significant factor but “design (background)” is statistically significant (except for model 3). For level B1, both “background” and “design (background)” are significant, but none of them are significant for level B2 (see Table III).

Tukey tests are performed to compare design types (Table IV). When the recipient’s location is within levels 1 and 2, where the lead time is less than 20 minutes, PHI swaths posed a higher level of concern than the polygons did for both WOB and WB displays. There is no significant difference between the PHI designs, and the background does not have a significant effect. Although significant differences are reported in Table III for level 4 and 5, according to Tukey’s result, there is no difference for level 4, but there is a significant difference for levels 5, where PHI resulted in a lower level of concern than the polygon. When the location is at level B1, although the recipient’s location is on the background, the design type plays a role in the participants’ level of concern; however, the color coding of the design does not cause an increase in the level of concern. The blue background is associated with lower concern in the participants at level B1.

Table III. Results of Analysis of Models at Each Recipient's Level

Dependent Variables	Level	Independent Variables			
		Background		Design (Background)	
		F Value (df)	*p* Value	*F* Value (df)	*p* Value
Concern	1	0.01 (1, 594)	0.93	9.27 (4, 594)	0.00*
	2	0.01 (1, 594)	0.91	10.52 (4, 594)	0.00*
	3	1.16 (1, 594)	0.28	0.83 (4, 594)	0.50
	4	0.28 (1, 594)	0.6	2.99 (4, 594)	0.02*
	5	0.62 (1, 594)	0.43	6.33 (4, 594)	0.00*
	B1	9.88 (1, 594)	0.00*	5.38 (4, 594)	0.00*
	B2	0.50 (1, 594)	0.49	1.99 (4, 594)	0.09
Fear	1	1.75 (1, 594)	0.19	22.17 (4, 594)	0.00*
	2	0.00 (1, 594)	0.96	9.08 (4, 594)	0.00*
	3	0.77 (1, 594)	0.38	1.67 (4, 594)	0.15
	4	0.63 (1, 594)	0.43	2.13 (4, 594)	0.07
	5	0.39 (1, 594)	0.53	2.27 (4, 594)	0.06
	B1	2.30 (1, 594)	0.13	5.36 (4, 594)	0.00*
	B2	0.17 (1, 594)	0.68	2.19 (4, 594)	0.07
Protective action	1	1.21 (1, 594)	0.27	12.15 (4, 594)	0.00*
	2	0.24 (1, 594)	0.63	11.97 (4, 594)	0.00*
	3	1.76 (1, 594)	0.18	0.11 (4, 594)	0.98
	4	0.55 (1, 594)	0.46	2.18 (4, 594)	0.07
	5	0.03 (1, 594)	0.86	1.95 (4, 594)	0.1

	B1	1.35 (1, 594)	0.24	7.80 (4, 594)	0.00*
	B2	0.50 (1, 594)	0.48	1.90 (4, 594)	0.11

*The *p* value is less than α = 0.05

5.2. Analysis of the Level of Fear and Protective Action

Because the results of the analysis for the level of fear and protective action have similar patterns, they are presented together in this section. In the model with “background,” “design (background),” and “level (background, design)” as independent variables and “fear” as the dependent variable. In the model with the same independent variables but the “protective action” as the dependent variable, “design (background)” and “level (background, design)” are statistically significant, but “background” is not (see Table II).

In the models for each recipient's level with “background” and “design (background)” as independent variables and “fear” or “protective action” as dependent variables, “design (background)” is a significant factor in models that correspond to recipient’s levels 1 and 2 and to level B1. In the models corresponding to other recipients’ level, none of the independent variables are statistically significant (Table III).

Tukey tests show that when the recipient’s location is within the first two recipient’s levels, with lead time of less than 20 minutes, the means of level of fear and protective action are significantly lower for “polygon” designs than those of PHI swaths for both WOB and WB displays (Table IV). The presence of a blue background does not have any significant effect on the responses, and there is no significant difference between the five-color and red-scale designs.

Table IV. Results of Tukey Tests for Different Models at Different Recipients’ Level: Mean (Standard Deviation)

Response	Level	Design					
		Five-Color	Red-Scale	Polygon	Five-Color	Red-Scale	Polygon

		(WOB)	(WOB)	(WOB)	(WB)	(WB)	(WB)
Concern	1	6.42 (0.80) A	6.40 (1.22) A	5.85 (1.33) B	6.43 (0.92) A	6.35 (0.90) A	5.91 (1.32) B
	2	6.05 (0.74) A	6.16 (1.07) A	5.51 (1.24) B	6.07 (1.02) A	6.12 (1.54) A	5.56 (1.55) B
	4	4.63 (0.81) A	4.88 (1.10) A	5.06 (1.05) A	4.68 (0.84) A	4.68 (1.58) A	5.05 (1.34) A
	5	3.99 (0.91) B C	3.97 (1.02) A B C	4.38 (1.10) A B	3.65 (0.98) A C	3.80 (0.92) B C	4.59 (1.21) A
	B1	2.87 (0.86) A B	2.93 (0.98) A B	3.37 (1.47) A	2.22 (0.96) C	2.71 (0.84) B C	3.06 (1.57) A B
Fear	1	6.94 (0.83) A	6.91 (1.09) A	6.22 (1.10) B	6.90 (0.93) A	6.91 (0.93) A	6.49 (1.11) B
	2	6.16 (1.13) A	6.10 (1.29) A	5.71 (1.27) B	6.10 (1.15) A	6.22 (1.30) A	5.66 (1.01) B
	B1	2.88 (0.88) A B	3.03 (0.94) A B	3.99 (1.24) A	2.24 (0.69) C	2.70 (0.81) B C	3.07 (1.16) A B
Protective action	1	2.90 (0.51) A	2.97 (0.71) A	2.71 (0.79) B	2.97 (0.57) A	2.96 (0.64) A	2.75 (0.85) B
	2	2.72 (0.53) A	2.75 (0.59) A	2.45 (0.78) B	2.75 (0.61) A	2.78 (0.57) A	2.45 (0.80) B

	B1	1.71 (0.98) A	1.69 (1.09) A	1.72 (1.22) A	1.27 (0.89) B	1.37 (0.98) B	1.68 (1.16) A

WB, with background; WOB, without background.

Note: Means that do not share a letter at the same level are significantly different.

For the model corresponding to level B1 with “protective action” as the dependent variable, the mean of responses is not significantly different among WOB displays; however, among WB displays, the mean for polygon designs is significantly lower than that of PHI swaths. For the similar model but with “fear” as the dependent variable, there is not a meaningful pattern for the difference between different designs.

Figures 2, 3, and 4 show the means of level of concern, fear, and protective action at each of the recipient’s level and for each design type regardless of the presence or absence of the blue background. It can be seen that in all three figures, the means of responses for polygon design is less than that of the PHI designs until recipient’s level 3. After level 3, the means becomes higher for the polygon design than for the two PHI designs.

The percentage of people who would immediately take shelter at recipient’s level 1 and 2 can be seen in Figure 5. It is clear that the PHI designs could elicit more immediate action from the recipients than the polygon design would.

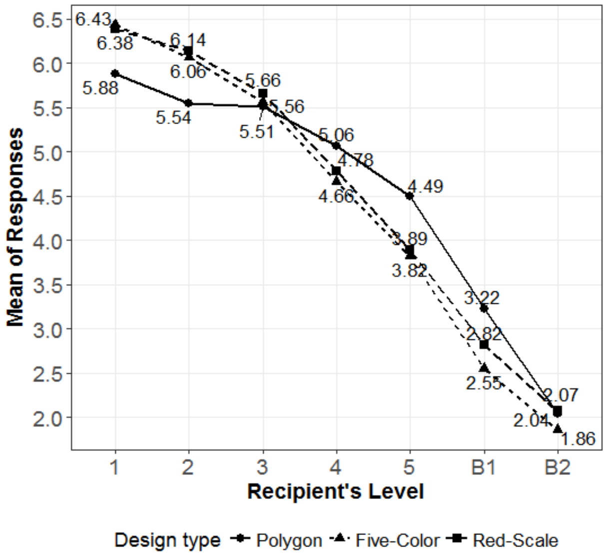

Figure 1. The means of level of concern at each level for different design types.

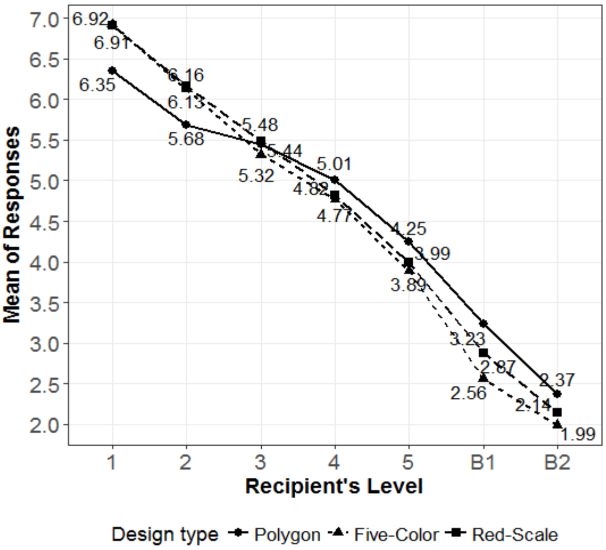

Figure 3. The means of level of fear at each level for different design types.

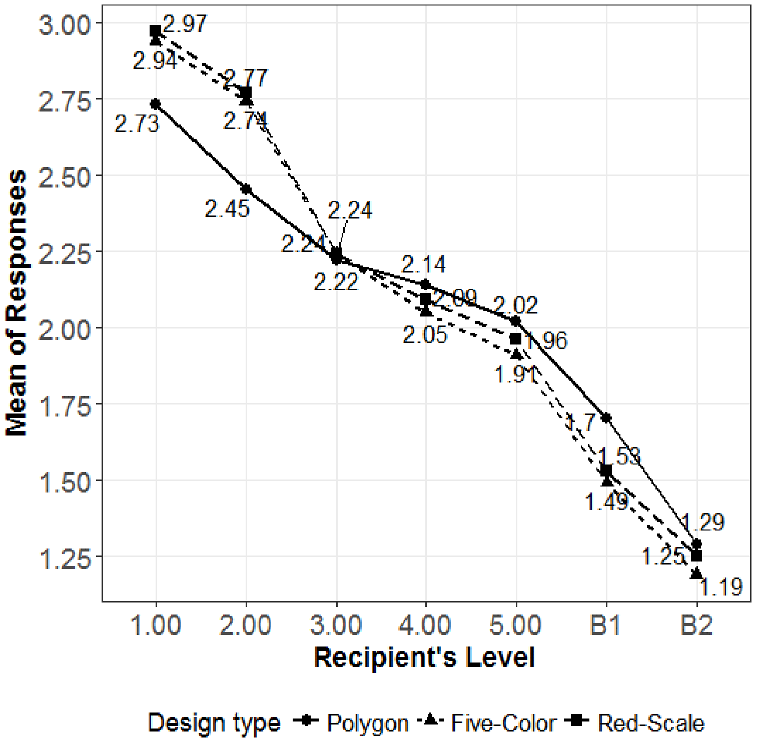

Figure 4. The means of level of protective action at each level for different design types.

At recipient's level 1, where it is crucial for the information recipients to immediately take shelter, almost 96% of the people reported that they would do so when they used PHI swaths, compared with around 76% on average when they used WarnGen polygons. Probabilistic designs elicited about 20% more immediate protective action at recipient's level 1. At recipient's level 2, where it is still appropriate for people to take immediate shelter because the threat is arriving within 20 minutes, almost 77% of people reported that they would take immediate shelter when they used PHI swaths, compared with around 47% on average when they used WarnGen polygons. Probabilistic designs elicited about 30% more immediate protective action at recipient's level 2.

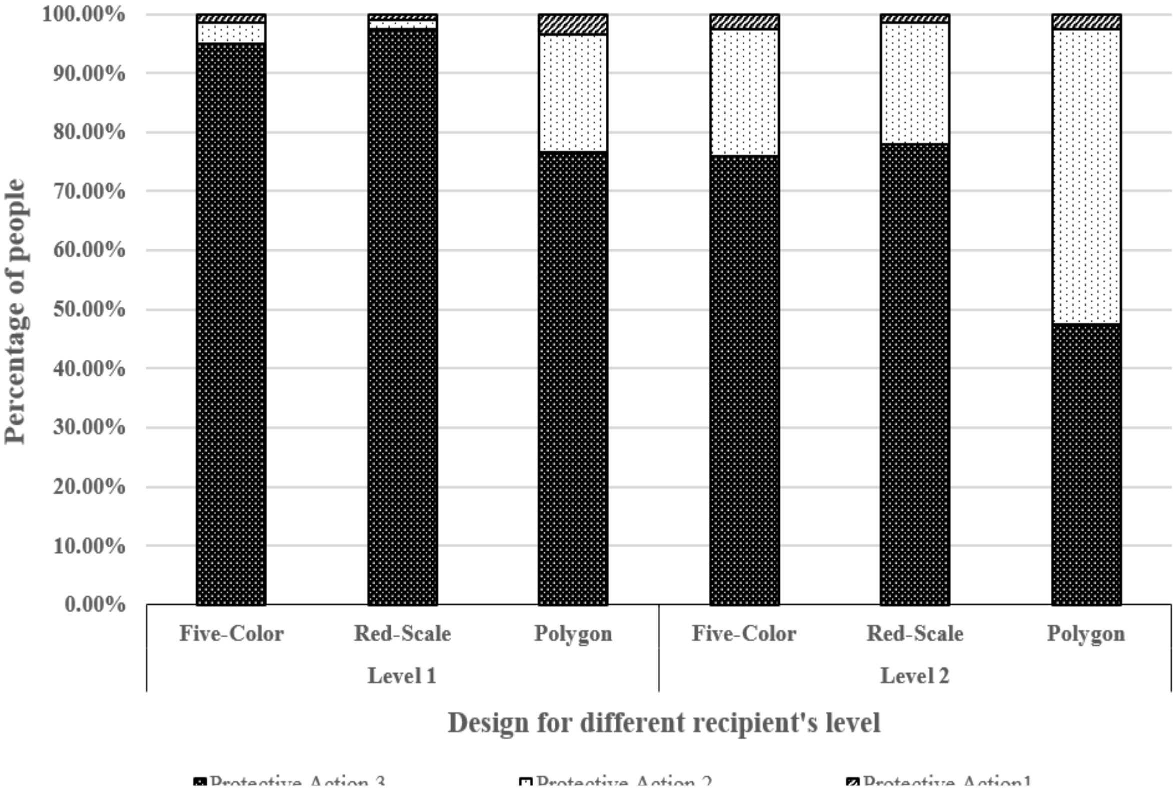

Figure 5. Percentage of people who immediately took protective action at each level.

5.3. The Most Preferred Display

Almost 84% of participants chose the five-color design with WOB display as their preferred design, followed by the five-color design with WB displays (10%). Only 4% of participant chose red-scale with WOB displays, and 2% chose red-scale with WB displays. No one preferred the polygon designs.

6. DISCUSSION

The main goal of this study was to investigate how providing the uncertainty information about the tornado occurrence affects the recipients' cognition of the threat and reaction to it. We measured the expected levels of concern, fear, and protective action in different scenarios using deterministic and probabilistic weather information.

As it was the case in a previous study by Ash et al. (2014), the mean of concern, fear, and protective action in the nearest areas to the tornado is lower for deterministic warnings than for probabilistic ones. As the proximity to the tornado decreases, the mean of the mentioned measures for the deterministic warnings surpasses the probabilistic ones. It suggests that there is an interaction between the probability matching and the urgency matching. In areas that the level of the urgency is high, the effect of the probability matching is higher and vice versa.

In highly dangerous scenarios in which the lead time is less than 20 minutes and the likelihood of the threat occurrence is greater than 60%, corresponding to recipient's levels 1 and 2, and the information recipients are expected to take immediate shelter, the use of PHI elicited more appropriate protective action from more people (20% more for recipient's level 1 and 30% more for recipient's level 2) compared with using the WarnGen polygon. In other words, in line with Ash et al. (2014) and contrary to Klochow (2013), probability matching played a significant role in increasing people's levels of concern, fear, and protective action in reaction to the threat and contributed to more accurate decision making. However, when the lead time is greater than 20 minutes and the likelihood is less than 60%, the levels of concern, fear and protective action are higher when the PHI is used (although not significant in all cases except one). This finding implies that when the recipients use the PHI, they make more differentiated responses based on the probability level that they perceive. They could translate the differences in threat probability into different levels of concern, fear, and protective action.

Regarding the effect of providing the likelihood of any tornado formation in the area through three shades of the blue background, this factor did not have a significant effect on the levels of fear and protective action. It had a significant effect only on the level of concern within level B1. At this level, this color coding significantly decreases the levels of concern in the weather information recipients. The original hypothesis is that by including the blue background in the area surrounding a particular threat to denote an unstable environment that could produce more tornadoes, people should be more concerned. It is likely that when people are presented with the threat information without any colored background, they just look at their relative locations to the PHI swath or the WarnGen polygon and think they might be in danger. When recipients are at the second level of blue background and farther from the swath or the polygon, they feel safer in general, and the blue background does not make any difference. From participants' evaluations of displays, the majority of them did not want to see the blue background behind the swaths or polygons. Although the background would help the information recipients to acquire more information about the surrounding area and be more vigilant in advance about a possible threat, it makes viewing the geographical map difficult and clashes with the PHI swath and the location indicator, participants argued.

Our analysis showed that the type of color-coding does not have any significant effect on the recipients' responses. This finding is in line with those of Klochow (2013) and Ash et al. (2014) that the color-coding type does not affect the perception of the uncertainty information. However, the majority of participants in this study preferred five-color PHI designs. Their main reasons for choosing "five-color" design are the use of the conventional way of color coding, its usefulness in distinguishing the probability levels on the swaths, and its good contrast with the background. Confirming the result of Miran et al. (2017a), participants noted that they were not able to easily distinguish different probability levels of a red-scale design. The reasons that none of the participants picked polygon design was their inability to grasp its contents quickly. It is suggested that the five-color design, which has been preferred by majority of the participants, be used for the PHI graphical interface.

7. CONCLUSION

This study supports findings of previous studies on the people probability matching behavior in reaction to different alarms (e.g., Bliss et al. 1995; Wiczorek and Manzey, 2011; Wiczorek et al. 2014; Balaud and Manzey, 2014) that the additional information that likelihood alarm systems provide contribute to more conscious decision making by the end users and ultimately increases the system's safety.

The PHI is considered more effective than the deterministic polygon, and it is recommended that the uncertainty information about the threat occurrence in different areas be included in the hazard information. It needs to be noted that in this study, to evaluate the added value of proving probabilistic information, we assumed that people could get urgency estimation from the WarnGen polygon. However, in the real world, WarnGen polygon might not offer such cues to users, making the advantage of PHI over deterministic polygon even more pronounced.

The main limitation of this study is its relatively small number of participants who mainly are residents of north east Ohio, where not many of tornadoes have occurred in the previous years. Because previous studies have considered the past tornado experience as a potential significant factor for people taking shelter against a tornado threat (Silver and Andrey, 2014; Miran, Ling, and Rothfusz, 2018b), the focus of a future study should be on investigating people's probability-matching behavior using PHI in areas that are more prone to tornado occurrence and where the residents have experienced a tornado previously.

ACKNOWLEDGMENTS

This study is funded by National Oceanic and Atmospheric Administration project: OAR-USWRP-R2O "FACETs Probability of What? Understanding and Conveying Uncertainty through Probabilistic Hazard Services."

REFERENCES

Agresti, A. (2002). Categorical Data Analysis. Second edition, Hoboken, NJ: Wiley.

Ash, K. D., Schumann III. R., & Bowser, G. C. (2014). Tornado Warning Trade-Offs: Evaluating Choices for Visually Communicating Risk. *Weather, Climate, and Society, 6,* 104–118.

Balaud, D., & Manzey, D. (2014). The more the better? The impact of number of stages of likelihood alarm systems on human performance. Available: https://www.hfes-europe.org/wp-content/uploads/2014/11/Balaud.pdf.

Bliss, J. P., Gilson R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38,* 2000-2012.

Bresnahan, T. F., & Bryk, J. (1975). The hazard association values of accident-prevention signs. *Professional Safety,* 17-25.

Budescu, D. V., Broomel, S., & Por, H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science, 20,* 299 –308.

Budescu, D. V., Por, H. H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change, 113,* 181–200.

Bustamante, E. A. (2008). Implementing likelihood alarm technology in integrated aviation displays for enhancing decision making: a two-stage signal detection modeling approach. *International Journal of Applied Aviation Studies, 8,* 241-262.

Christner, C. A., & Ray, H. W. (1961). An evaluation of the effect of selected combinations of target and background coding on map-reading performance: Experiment V. *Human Factors, 3,* 131-146.

Clark, R. M., Ingebritsen, A., & Bustamante, E. A. (2010). Differential effects of likelihood alarm technology and false-alarm vs. miss-prone automation on decision-making sensitivity and bias. *Proceedings of the Human Factors and Ergonomics Society,* 1508-1512.

Coleman, T. A., Knupp, K. R., Spann, K., Elliott, J. B., & Peters, B. E. (2011). The history (and future) of tornado warning dissemination in the United States. *Bulletin of the American Meteorological Society, 92,* 567–582.

Durage, S. W., Wirasinghe, S. C., & Ruwanpura, J. Y. (2015). Decision analysis for Tornado warning and evacuation. Nat Hazards Rev [On-line]. Available: https://doi.org/10.1061/(ASCE)NH.1527-6996.0000195.

Edwards, R., & Lemon, L. R. (2002). Proactive or reactive? The severe storm threat to large event venues. 21st Conf. Severe Local Storms, American Meteorological Society, San Antonio, 232–235.

Grothman, T., & Reusswig, F. (2006). People at risk of flooding: Why some residents take precautionary action while others do not. *Nature Hazards, 38*, 101–120.

Haas, E. C., & Casali, J. G. (1995). Perceived urgency of and response time to multi-tone and frequency modulated warning signals in broadband noise. *Ergonomics, 38*, 2313-2326.

Hammer, B. O., & Schmidlin, T. W. (2001). Vehicle-occupant deaths caused by tornadoes in the United States. *Environmental Hazards, 2*, 105-118.

Handmer, J., & Proudley, B. (2007). Communicating uncertainty via probabilities: The case of weather forecasts. *Environmental Hazards, 7*, 79–87.

Hoffman, R. R., Detweiler, M., Conway, J. A., & Lipton, K. (1993). Some considerations in using color in meteorological displays. *Weather and Forecasting, 8*, 505–518.

Joslyn, S., & Savelli, S. (2010). Communicating forecast uncertainty: public perception of weather forecast uncertainty. *Meteorological Application, 17*, 180–195.

Karstens, C. D., Smith, T. M., Calhoun, K. M., Clark, A. J., Ling, C., Stumpf, G. J., & Rothfusz, L. P. (2014). Prototype tool development for creating probabilistic hazard information for severe convective phenomena. 94th American Meteorological Society, Annual Meeting, Atlanta, GA, P2.2.

Klockow, K. E. (2013). Spatializing tornado warning lead-time: risk perception and response in a spatiotemporal framework. Norman: The University of Oklahoma.

Kopala, C. J. (1979). The Use of Color-Coded Symbols in a Highly Dense Situation Display. *Proceeding of the Human Factors Society*, 397-401.

Kox, T., Gerhold, L., Ulbrich, U. (2015). Perception and use of uncertainty in severe weather warnings by emergency services in Germany. *Atmospheric Research* [On-line]. Available: https://doi.org/10.1016/j.atmosres.2014.02.024.

Kunz, M., Grêt-Regamey, A., & Hurni, L. (2011). Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Natural Hazards, 59*, 1735–1751.

Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: theoretical modifications and additional evidence, *Risk Anal, 32*, 616–632.

Ling, C., Hua, L., Karstens, C. D., Stumpf, G. J., Smith, T. V., Kuhlman, K. M., & Rothfusz, L. (2015). A comparison between WarnGen System and probabilistic hazard information System. *Proceedings of the Human Factors and Ergonomics Society, 59*, 1791–1795. doi:10.1177/1541931215591387.

Lundgren, R. E. (1994). Risk Communication: A Handbook for Communicating Environmental, Safety and Health Risks. Columbus, OH: Battelle Press.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors*, 48, 241-256.

Miran, S. M., Ling, C., James, J. J., & Rothfusz, L. (2016). Comparing Effectiveness of Four Graphical Designs for Probabilistic Hazard Information for Tornado Threat. *Proceedings of the Human Factors and Ergonomics Society, 60*, 2029–2033. Available: https://doi.org/10.1177/1541931213601461

Miran, S. M., Ling, C., James, J. J., Gerard, A., & Rothfusz, L. (2017a). User perception and interpretation of tornado probabilistic hazard information: Comparison of four graphical designs. *Applied Ergonomics, 65*, 227-285. Available: https://doi.org/10.1016/j.apergo.2017.06.016.

Miran, S. M., Chen, L., James, J. J., Rothfusz, L. (2017b). Effective Method to Convey Threat Information for Tornado: Probabilistic Hazard Information vs. Deterministic Hazard Information. *Proceedings of the Human Factors and Ergonomics Society, 61*(1), 292–296. Available: https://doi.org/10.1177/1541931213601554.

Miran, S. M., Ling, C., Gerard, A., Rothfusz, L. (2018a). The effect of providing probabilistic information about a tornado threat on people's protective action. *Natural Hazards, 94*(2), 743–758. Available: https://doi.org/10.1007/s11069-018-3418-5.

Miran, S. M., Chen, C., Rothfusz, L. (2018b). Factors influencing people's decision making during three consecutive tornado events. *International Journal of Disaster Risk Reduction, 65*, 150–157. Available: https://doi.org/10.1016/j.ijdrr.2018.02.034.

Morss, R. E., Demuth, J., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: a survey of the U.S. public. *Weather and Forecasting, 23*, 974–991.

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people. *American Psychologist, 51*, 931–938.

Murphy, A. H., & Winkler, R. L. (1971). Forecasters and probability forecasts: Some current problems. *Bulletin American Meteorological Society, 52*, 239–247.

Murphy, A. H. (1998). The early history of probability forecasts: Some extensions and clarifications. *Weather and Forecasting, 13*, 5–15.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. Human Factors, 48, 241-256.

National Oceanic and Atmospheric Administration (2017). Billion-Dollar Weather and Climate Disasters: Table of Events. Available: https://www.ncdc.noaa.gov/billions/events/US/1980-2017.

Rothfusz, L., Karstens, C. D., & Hilderbrand, D. (2014). Forecasting a continuum of environmental threats: Exploring next-generation forecasting of high impact weather. *EOS Transactions, American Geophysical Union, 95*, 325–326.

Saviers, A., & Van Bussum, L. (1999). Juneau Public Questionnaire: Results, analyses, and conclusions. Western Region Technical Attachment. National Oceanic and Atmospheric Association, USA.

Silver, A., & Andrey, J. (2014). The Influence of previous disaster experience and socio-demographics on protective behaviors during two successive tornado events. *Weather Clim Soc, 6*(1), 91–103.

Simmons, K. M., Sutter, D. (2009). False alarms, tornado warnings, and tornado casualties. Weather Clim Soc., 1, 38–53

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*, 445-459.

Sutter, D., & Erickson, S. (2010). The time cost of tornado warnings and the savings with storm-based warnings. *Weather, Climate, and Society, 2*, 103–112.

Wiczorek, R. & Manzey, D. (2011). Evaluating Likelihood Alarm Systems as an Alternative to Binary Alarm Systems. In D. Waard, N. Gérard, L. Onnasch, R. Wiczorek & D. Manzey (Eds.), *Human Centred Automation* (pp. 69–83). Maastricht: Shaker Publishing.

Wiczorek, R., Manzey, D., Zirk, A. (2014). Benefits of decision-support by likelihood versus binary alarm systems: Does the number of stages make a difference? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting: Vol. 58* (pp. 380–384). Santa Monica: Human Factors and Ergonomics Society.