Expressing Flood Likelihood: Return Period versus Probability

MARGARET A. GROUNDS, JARED E. LECLERC, AND SUSAN JOSLYN

University of Washington, Seattle, Washington

ABSTRACT

The likelihood of floods and other potentially destructive natural phenomena is often expressed as a return period or recurrence interval, such as a 100-yr flood. However, the expression might give users the impression that the event will occur exactly once within the described period, obscuring the intended probabilistic meaning. If so, users may think a flood is less likely when one has just occurred or more likely when it has not, leading to a ''flood is due'' effect. This hypothesis was tested experimentally in two studies reported here. Participants were given either a return period or a probability expression and asked to rate flood likelihood and concern. Flood recency was also manipulated. The results from both studies support a flood is due effect when the return period expression is used. In the return period condition alone, participants rated floods as more likely and expressed greater concern when no flood had occurred recently. When no likelihood information was conveyed in the control condition, the opposite effect was observed. Participants rated flood likelihood as higher and expressed greater concern when a flood had occurred recently. Participants using the percent chance expression were least affected by flood recency. This adds to the growing body of research suggesting that nonexperts can benefit from probabilistic weather forecasts.

1. Introduction

The likelihood of floods and other hydrologic events, including storms and droughts, is often expressed as a return period, such as a ''100-yr flood.'' This information is used to make decisions about whether a project should go forward in a particular location and how to design structures to withstand an event with a certain magnitude as well as to inform the public of risk. Return periods are based on historic data from which the average recurrence interval over an extended period of time is derived. For instance, a 100-yr flood surpasses a level that is exceeded on average once in every 100 years. The likelihood in a given year is the inverse of the expected number of times that the level will be exceeded in any one year. For example, a 100-yr flood has a 0.01% or 1% chance of being exceed in any one year:

$$\text{Recurrence interval} = \frac{n}{m},$$

n is number of years on record, and

m is the number of recorded occurrences of the event being considered.

It is important to note, however, that in any given 100-yr period, a 100-yr event may occur once, twice, more, or not at all. A 100-yr flood does not mean that a flood of that magnitude will happen regularly every 100 years or only once in 100 years. Nonetheless, the expression almost invites this misinterpretation.

Effective communication of flood risk has emphasized two different factors: understanding of the risk and the persuasiveness of the communication [see Bell and Tobin (2007) for a review]. The effort to increase the public's understanding of the risk of floods is based largely on a deficit model, in which it is believed that better communication of information could reduce the public's deficit of knowledge about the risks, thereby leading to better decisions (Ramos et al. 2013). However, another gauge of the effectiveness of flood communication is whether it persuades the public to take precautionary action. The return period expression of flood likelihood has been recognized as problematic, as it has been suggested that it fails both to convey meaning and to motivate concern (Ludy and Kondolf 2012; Bell and Tobin 2007).

Interestingly, within an expansive literature on flood risk perception, past research on the effectiveness of

Corresponding author: Margaret A. Grounds, mgrounds@uw.edu

return period expression in particular and flood risk communication in general is limited (Kellens et al. 2013). There is, however, some evidence to suggest that the general public misinterprets the return period expression. For example, in the San Joaquin delta, an area protected by 100-yr flood levees, 31% of surveyed households claimed to understand the term, but only 3 out of 114 respondents correctly defined it (Ludy and Kondolf 2012). The most common erroneous definition was that a 100-yr flood occurs exactly once every 100 years. A similar misinterpretation was selected by 40% of respondents in Boulder Creek (Gruntfest et al. 2002).

This misinterpretation may create what we refer to as a "flood is due" effect. People may think that floods are more likely if a flood has not occurred in a span of time approaching the return period. Conversely, if a flood of that magnitude has just occurred, people may think the likelihood of another similar flood is less than what is intended by the expression. Thus, those with such misunderstandings may seriously mistake the risk they are facing. They may overestimate the likelihood of a serious flood in some situations and underestimate it in others, potentially exposing themselves to dangerous situations.

The same information expressed as percent chance or probability may lead to a more accurate understanding of the likelihood of a flood. Recent evidence suggests nighttime low temperature forecasts that include probabilistic information allowed participants to make better decisions than deterministic or single-value forecasts (Roulston et al. 2006; Joslyn and LeClerc 2012). There are similar advantages for forecasts presented as predictive intervals describing the boundaries within which the observation is expected with specified probability (Savelli and Joslyn 2013). In fact, adding numeric likelihood estimates has been shown to improve compliance with advice comparable to weather warnings, to a greater degree than does lowering the false alarm rate (LeClerc and Joslyn 2015).

The advantage for numeric likelihood estimates is in part because people understand that all forecasts involve uncertainty, even if it is not acknowledged (Joslyn and Savelli 2010; Morss et al. 2008). As a result they have greater trust in the information when an uncertainty estimate is included (LeClerc and Joslyn 2015). However, evidence also suggests that uncertainty information must be carefully expressed. For instance, there is considerable evidence that verbal expressions of uncertainty (e.g., very likely) are too vague to provide much benefit (Fischer and Jungermann 1996; Wallsten et al. 1986). Numeric expressions are more precise (Windschitl and Wells 1996), although they are better understood when expressed in a manner that matches users' expectations (Joslyn et al. 2008). However, somewhat surprisingly, there is new evidence that the benefit for numeric

likelihood expressions does not depend on the level of education (Grounds 2016). In sum, there is now considerable evidence for the benefit of probabilistic forecast (Joslyn and LeClerc 2013) to end users.

There is also evidence that a probabilistic format leads to a better understanding of flood likelihood (Bell and Tobin 2007; Keller et al. 2006), although in some cases concern is reduced with a percent chance expression (Bell and Tobin 2007). However, none of these experiments systematically varied the recency of target events. Thus, although it is clear that the return period expression is confusing to nonexpert end users, convincing evidence for the flood is due effect or the superiority of a percent chance expression in the face of the flood is due effect does not yet exist. The experiments reported here were designed to address these issues directly.

In the two experiments described below, we test these hypotheses by comparing the participants' estimate of the likelihood of floods when they are given either a return period expression or a percent chance expression. In addition, flood recency is systematically manipulated to determine its impact on likelihood expression understanding. If a flood is due effect prevails, participants with the return period expression will perceive higher flood likelihood when a flood has not occurred recently and lower likelihood when a flood has just occurred.

2. Experiment 1: Pilot study

First, a pilot study was conducted to determine whether a flood is due effect was observed with the return period expression (10-yr flood). A percent chance expression (10% chance) was also tested to determine whether it better conveyed the probabilistic nature of the information, reducing the flood is due effect. These conditions were compared to a control condition in which no flood likelihood was conveyed. In the control condition, flood level was labeled alphabetically (see Fig. 1). We systematically varied flood recency for all participants.

a. Method

1) PARTICIPANTS

As part of a course requirement, 243 university undergraduates (51.9% female, mean and median age = 19.00 years, range = 18–26) participated.

2) PROCEDURE

The study questionnaire was one of several pen-and-paper questionnaires administered in a mass testing setting. After providing informed consent, participants read the following paragraph: Bison City is a midwestern American town with a small river, Yellowtail Creek,
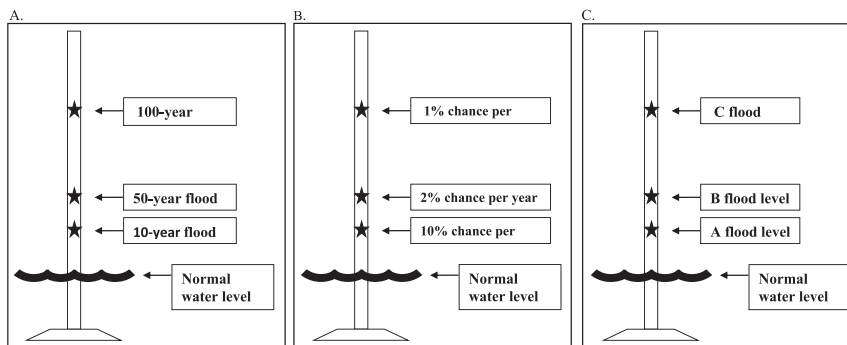
FIG. 1. The visualizations used in both experiments to present the likelihood for three flood levels. (a) Some participants saw the return period, (b) others the percent chance, and (c) others the control/no information.

running through it. Yellowtail Creek is normally quite calm, but sometimes, after heavy rain, it can flood. The figure below depicts a water level marker that sits in Yellowtail Creek. It shows flood levels for the river, with each mark representing a different likelihood of potential flood severity.

Below the paragraph, was a flood marker image with three flood levels labeled in one of three ways: 1) return period, 2) percent chance, or, in the control condition, with the letters 3) A, B, and C (see Fig. 1). Below the flood marker image was one of two sentences: 1) last year, Bison City experienced a flood at the 10-yr (10% chance, C) flood level (recent) or 2) Bison City has not experienced a flood at the 10-yr (10% chance, C) flood level in about 10 years (distant). Then participants indicated their concern for flooding in the coming year, by selecting from a 6-point Likert scale, ranging from not at all concerned to extremely concerned. They also indicated the likelihood of flooding in the coming year on a slider with a left anchor of extremely unlikely and a right anchor of extremely likely (see the appendix). Note that we did not use numeric probabilities as anchors because they would match the expression in the probabilistic condition, providing an unfair advantage to those participants and introducing a potential confound in the design.

3) DESIGN

The experiment had a 3 (expression: return period, percent, A–B–C) by 2 (flood recency: recent, distant) between-participants design, resulting in six conditions. The dependent variables were likelihood and concern ratings.

b. Results

To determine whether there was a flood is due effect that depended on flood expression, we conducted a pair of ANOVAs examining the impact of flood recency and

flood expression on participants' likelihood and concern ratings. A flood is due effect is indicated if participants rated likelihood and/or concern higher in the return period condition when no flood had occurred recently compared to when a flood had just occurred. This hypothesis was supported. In this experiment, the alpha level for each analysis was 0.05. Pairwise comparisons were made using Tukey's post hoc tests to correct for familywise error rates. Post hoc power analyses utilizing G*Power were also calculated.

We first analyzed participants' perception of the likelihood of flooding, which they expressed by marking a line somewhere between the end points labeled extremely unlikely and extremely likely. Each participant's mark was measured from the left anchor of the 140-mm line. The lengths were divided by 140 mm to create percentages. Then an ANOVA was conducted on mean percentage with two independent variables: likelihood expression (return period, percent, A–B–C) and flood recency (recent, distant). In support of our hypothesis, there was a significant interaction: $F(2, 236) = 4.10$, $p = 0.02$, $\eta_p^2 = 0.03$, power = 0.72. Those with the return period expression thought a flood was more likely in the distant ($M = 51.50\%$, SD = 17.21) as compared to the recent condition ($M = 43.15\%$, SD = 23.55; $p = 0.02$, Cohen's $d = 0.40$, power = 0.99). For those with the A–B–C expression, the pattern was reversed; participants thought a flood was more likely in the recent ($M = 59.31\%$, SD = 18.20) as compared to the distant condition ($M = 48.02\%$, SD = 21.25; $p = 0.03$, Cohen's $d = 0.57$, power = 0.99; see Fig. 2). Those with the percent chance expression thought a flood was approximately equally likely regardless of flood recency (recent: $M = 35.88\%$, SD = 19.37; distant: $M = 36.77\%$, SD = 27.43; $p = 0.86$, Cohen's $d = 0.04$, power = 0.06).

In addition, those in the A–B–C condition thought a flood was most likely ($M = 53.42\%$, SD = 20.51), those
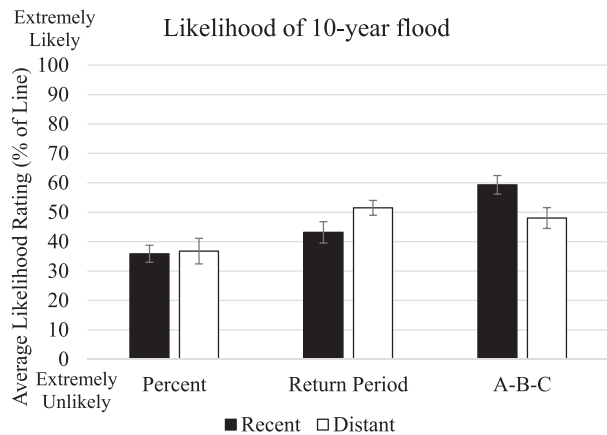
FIG. 2. Average likelihood ratings in experiment 1 for the occurrence of a 10-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.

FIG. 3. Average concern ratings in experiment 1 about the occurrence of a 10-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.

in the return period condition thought a flood was next most likely ($M = 47.56\%$, SD = 20.76), while those in the percent chance condition thought it was least likely ($M = 36.30\%$, SD = 23.41).[1]

Next, an ANOVA on concern ratings was conducted with the same two independent variables, likelihood expression (return period, percent, and A–B–C) and flood recency (recent, distant). There was a significant interaction, $F(2, 237) = 5.30$, $p < 0.01$, $\eta_p^2 = 0.04$, power = 0.83, that followed the pattern of likelihood ratings. Those with the return period expression were more concerned in the distant ($M = 3.51$, SD = 1.14) as compared to the recent condition ($M = 3.14$, SD = 1.13), $p = 0.01$, Cohen's $d = 0.33$, power = 0.23, whereas those with the A–B–C expression were more concerned in the recent ($M = 3.55$, SD = 0.91) as compared to distant condition ($M = 2.81$, SD = 1.04; $p = 0.004$, Cohen's $d = 0.76$, power = 0.73; see Fig. 3). Those with the percent expression were approximately equally concerned in the two recency conditions (recent: $M = 3.05$, SD = 1.08; distant: $M = 2.98$, SD = 1.05; $p = 0.76$, Cohen's $d = 0.07$, power = 0.06). In addition, those in the percent chance condition were least concerned ($M = 3.01$, SD = 1.06), while those in the A–B–C ($M = 3.16$, SD = 1.04), return period ($M = 3.33$, SD = 1.14) conditions were more concerned.

c. Discussion

These results support a flood is due effect when the return period expression is used. Although the term is intended to convey the likelihood of floods independent of flood recency, participants in the return period condition appeared to take recency into account when evaluating flood likelihood. Both likelihood and concern ratings were lower when a flood had just occurred in the return period condition. This suggests that participants thought that the likelihood was reduced because a flood occurs once and only once in the time frame mentioned. Interestingly, when no likelihood information was conveyed in the control condition, flood recency was also taken into account. Participants in the A–B–C condition thought a flood was more likely when one had just occurred, exactly the opposite bias. Perhaps they thought that the weather conditions that produced the first flood continued to exist: a "persistence effect." We return to this issue in the general discussion. Participants using the percent chance expression were least affected by flood recency. Moreover, those using the percent chance expression thought that the likelihood of flooding was lower overall and perhaps more similar to the intended likelihood. Notice that in all of the other conditions people thought that the likelihood was about midway between extremely unlikely and extremely likely. Although it is not possible to know how participants might have translated these anchors into percentage chance, it is arguable that 10% would represent something less than midway between the two anchors.

The flood is due effect could make people more vulnerable in real-life situations in which they might be reluctant to take precautionary action under specific flood recency conditions. The persistence effect may also result in a misunderstanding of flood likelihood in some situations. These results suggest that the best flood likelihood expression is percent chance, which was not influenced by the recency of flooding.

---

[1] The statistics that speak to the main effect of forecast format were omitted at the request of one reviewer because they had not been predicted beforehand but are available from the authors.

3. Experiment 2

Because the pilot study reported here had several limitations as well as some interesting unpredicted results, a second study was conducted. In experiment 1, the likelihood and concern questions did not specify a flood level. Some participants may have interpreted them to mean a flood at the level that had been specified (10 year, 10% chance), but others may have been thinking of another or a combination of flood levels. Thus, the flood is due effect may be specific to the flood level mentioned or more general. To address this issue, in experiment 2, flood levels were specified in the questions. In addition, the anchors on the likelihood scale were changed to impossible and certain to better compare with percent chance. In addition, in order to test a more representative sample than the university student participants in experiment 1, in experiment 2 a more diverse sample was collected on the Internet.

A number of additional control conditions were also included in experiment 2. In experiment 1, all conditions included a visual representation of flood likelihood. Visualizations themselves may play a role in participants' perception of flood risk, although the evidence is mixed. Some evidence suggests that visualizations give rise to greater perception of risk than other expressions (Stone et al. 1997). Other evidence fails to detect a difference between visualizations and other formats (Galesic et al. 2009; Weinstein et al. 1994). Still other evidence suggests that visualizations increase risk perception for high risk and decrease risk perception for low risk situations (Sandman 1998). Therefore, in order to determine the role of the flood marker visualization, it was compared to a tabular format in experiment 2. Moreover, although the pilot study included a control condition for flood expression, there was no control condition for flood recency. To determine participants' perception of flood likelihood in the absence of recency information, a control condition was added in which it was not mentioned.

Thus, experiment 2 was conducted to replicate and extend the findings in the pilot study. Likelihood and concern ratings were requested for three flood levels [10 year (10%), 50 year (2%), and 100 year (1%)] with more extreme anchor labels. Two additional control conditions, a tabular format and a condition that did not specify flood recency, were added. Here, based on the results of experiment 1, we also predicted a main effect for forecast format such that participants would think that a flood was most likely and be more concerned when no flood likelihood information was given and least likely when percent chance was given. We also predicted a main effect for recency such that participants would

TABLE 1. Table used in experiment 2 to present the likelihood for three flood levels. Some participants saw return period, other percent chance (in first bracket), and others control/no information (in second bracket). Bold text indicates water height above normal water level for participants.

Water height	
11 ft	**100-yr flood [1% chance per year]** **[C flood level]**
7 ft	**50-yr flood [2% chance per year]** **[B flood level]**
5 ft 6 in. **(5.5 ft)**	**10-yr flood [10% chance per year]** **[A flood level]**
3 ft	Normal water level
0 ft	Creek bed

think that floods were most likely and be more concerned when one had just occurred: a persistence effect.

a. Method

1) PARTICIPANTS

Participants were recruited from Amazon's Mechanical Turk (M-Turk), an online crowdsourcing service that hires workers for human intelligence tasks (HITs). Participants were compensated $0.10 for their responses. Only the 803 participants who were residents in the United States and had a 90% prior "approve rate" (percentage of prior HITs accepted by requester) were included. The median age (33 years old; $M = 36$, range $= 18–77$) was slightly younger than in the 2012 U.S. census (median $= 37.2$), and there were slightly more females (63%) than in the 2012 U.S. census (50.5%). Although education level was not collected, M-Turk samples tend to be slightly better educated than the general population (Ross et al. 2010).

2) PROCEDURE

After participants gave informed consent, they read the same description of a fictional town and the adjacent river that was used in experiment 1. Participants in the experimental conditions saw graphics identical to those used in experiment 1 (Fig. 1) or a table (Table 1) that represented the likelihood for three flood levels expressed either as a return period, percent chance, or A, B, and C. The table listed flood levels in feet that corresponded to the distance between the normal river level and flood levels in the graphic at a ratio of 1 ft to 1/16 in. In addition, as with experiment 1, participants in the experimental conditions were told when the most recent 10-yr (10% chance per year, C) flood had occurred, either last year (recent) or 10 years ago (distant). In the control condition, no flood recency information was provided.

After reading the instructions and examining the likelihood information, participants rated concern and

likelihood at each of the three levels (presented from lowest to highest). Participants indicated their concern for flooding in the coming year by selecting from a 6-point Likert scale, ranging from not at all concerned to extremely concerned. They indicated how likely they thought flooding was in the coming year on a slider with a left anchor of impossible and a right anchor of certain. Because of the greater diversity of this sample compared to the pilot study, we also asked if they had ever lived in an area that was regularly threatened by flooding. Prior flood experience may have played a role in likelihood perception.[2] To determine whether participants were aware of their level of comprehension of the likelihood expressions, they were asked to indicate their confidence in their own understanding of the flood term. Once all tasks were completed, participants received a completion code that they entered into the M-Turk website. Completion codes were verified and then payments were approved.

3) DESIGN

A $3 \times 3 \times 2$ between-groups full factorial design was used. Participants were randomly assigned to one of three flood-term conditions: (i) return period (10-yr flood), (ii) percent chance (10% chance per year), or (iii) control [no information (A, B, C)]. Within each of those groups, participants were randomly assigned to one of three flood recency conditions: (i) 10-yr (10% chance, C) flood happened 10 years ago, (ii) 10-yr (10% chance, C) flood happened last year, or (iii) control (no information). Finally, participants were randomly assigned to either the graphic or a table display (see Fig. 1; Table 1). This resulted in 18 separate conditions. The dependent variables were likelihood and concern ratings.

b. Results

To determine whether the flood is due effect was replicated, we first examined the effect of flood expression and flood recency on participants' likelihood and concern ratings for the 10-yr flood, for which the experimental conditions had flood recency information. Here, because we predicted an interaction as well as two main effects, the alpha level was reduced to 0.017 for each analysis. Then we conducted the same analyses on likelihood and concern ratings for the 50- and 100-yr

floods. Finally, we compared likelihood estimates for all three flood levels to one another to determine whether participants understood which events were rarer and whether this understanding depended on flood likelihood expression. All post hoc pairwise comparisons were made using Tukey's tests.

1) LIKELIHOOD 10-YR FLOOD

Indeed, the flood is due effect was replicated in experiment 2. A three (expression: return period, percent, and A–B–C) by three (flood recency: recent, distant, and control) by two (display: flood marker graphic and table) between-groups ANOVA was conducted on mean likelihood ratings for the 10-yr flood. As predicted, there was a flood term by recency interaction [$F(4, 785) = 11.36$, $p < 0.001$, $\eta_p^2 = 0.06$, power = 0.99; see Fig. 4]. For the return period, likelihood ratings were higher in the distant ($M = 61.63\%$, SD = 25.29) as compared to either the recent ($M = 50.85\%$, SD = 24.89; $p < 0.01$, Cohen's $d = 0.43$, power = 0.99) or control ($M = 47.94\%$, SD = 24.66; $p < 0.001$, Cohen's $d = 0.55$, power = 0.99) conditions. In the A–B–C condition, the pattern was reversed. The likelihood was higher in the recent ($M = 73.30\%$, SD = 18.42) compared to the control condition ($M = 53.98\%$, SD = 27.44; $p < 0.001$, Cohen's $d = 0.83$, power = 0.99), which was higher than distant condition ($M = 46.99\%$, SD = 24.83; $p < 0.05$, Cohen's $d = 0.27$, power = 0.99). For the percent chance expression, as with experiment 1, likelihood was rated similarly in the recent ($M = 42.88\%$, SD = 25.32) and distant ($M = 38.34\%$, SD = 25.60; $p = 0.45$, Cohen's $d = 0.18$, power = 0.95) conditions, although likelihood was higher in the recent than in the control ($M = 33.35\%$, SD = 22.89; $p < 0.01$, Cohen's $d = 0.39$, power = 0.99). There was a main effect for flood expression [$F(2, 785) = 45.26$, $p < 0.001$, $\eta_p^2 = 0.11$, power = 0.99]. Likelihood ratings were significantly higher in the A–B–C ($M = 57.58\%$, SD = 26.37; $p < 0.001$, power = 0.99) and the return period ($M = 53.42\%$, SD = 25.54; $p < 0.001$, power = 0.99) conditions as compared to the percent chance ($M = 37.88\%$, SD = 25.17) condition. In addition, there was a main effect for flood recency [$F(2, 785) = 12.39$, $p < 0.001$, $\eta_p^2 = 0.03$, power = 0.99]. Likelihood ratings were significantly higher in the recent ($M = 55.96\%$, SD = 26.36) as compared to distant ($M = 48.31\%$, SD = 27.23; $p < 0.001$, Cohen's $d = 0.29$, power = 0.99) or the control ($M = 45.10\%$, SD = 26.54; $p < 0.001$, Cohen's $d = 0.41$, power = 0.99) conditions. However, as noted above, this pattern clearly did not hold in the return period condition. Likelihood ratings in the graphic ($M = 49.33\%$, SD = 27.54) and table conditions ($M = 49.64\%$, SD = 26.66) were similar (Cohen's $d = 0.01$, power = 0.05).

---

[2] Approximately 60% of participants said they lived in an area that regularly flooded. Although they were on average more concerned about the 10-yr flood, prior experience did not interact with any of the variables of interest and will not be mentioned in the analyses below.
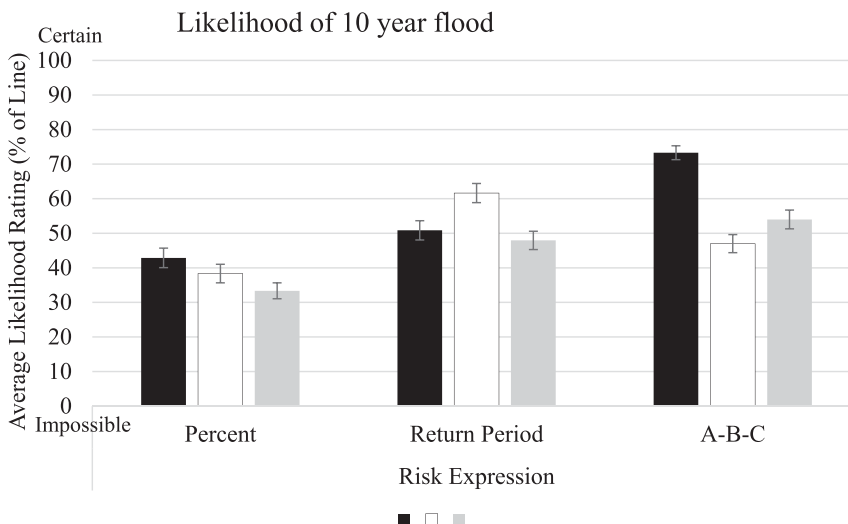
FIG. 4. Average likelihood ratings in experiment 2 for the occurrence of a 10-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.

2) CONCERN 10-YR FLOOD

A similar pattern was observed for concern ratings, again supporting the flood is due effect. A three (likelihood expression: return period, percent chance, A–B–C) by three (flood recency: recent, distant, control) by two (display: flood marker graphic and table) between-groups ANOVA was conducted on concern ratings. As predicted, there was a significant flood expression by recency interaction [F (4, 785) = 7.30, $p < 0.001$, $\eta_p^2 = 0.04$, power = 0.99]. Following the pattern of likelihood, for the return period expression, concern was higher in the distant ($M = 3.45$, SD = 1.22) than in the control condition ($M = 3.08$, SD = 1.04; $p < 0.05$, Cohen's $d = 0.33$, power = 0.99). In the other two conditions, the pattern was reversed. For the A–B–C expression, concern was higher in the recent ($M = 4.39$, SD = 1.32) as compared to the distant ($M = 3.21$, SD = 1.34; $p < 0.001$, Cohen's $d = 0.89$, power = 0.99) or control ($M = 3.26$, SD = 1.24; $p < 0.001$, Cohen's $d = 0.88$, power = 0.99) conditions (see Fig. 5). For the percent chance expression as well, concern was higher in the recent ($M = 3.49$, SD = 1.11) than in the distant ($M = 2.88$, SD = 1.12; $p < 0.01$, Cohen's $d = 0.55$, power = 0.99) or control ($M = 3.01$, SD = 1.14; $p < 0.01$, Cohen's $d = 0.43$, power = 0.99) conditions. There was a main effect for likelihood expression [F (2, 785) = 12.98, $p < 0.001$, $\eta_p^2 = 0.03$, power = 0.99]. Concern was higher in the A–B–C ($M = 3.59$, SD = 1.37) than in the return period ($M = 3.26$, SD = 1.14; $p < 0.01$, Cohen's $d = 0.26$, power = 0.99) or percent chance ($M = 3.10$, SD = 1.15; $p < 0.001$,

Cohen's $d = 0.39$, power = 0.99) conditions; however, there was no difference in concern between return period and percent chance conditions ($p = 0.29$, Cohen's $d = 0.14$, power = 0.80). There was also a main effect for flood recency [F (2, 785) = 18.30, $p < 0.001$, $\eta_p^2 = 0.05$, power = 0.99]. In general, participants expressed greater concern in the recent ($M = 3.72$, SD = 1.26) as compared to distant ($M = 3.16$, SD = 1.24; $p < 0.001$, Cohen's $d = 0.27$, power = 0.99) or control ($M = 3.12$, SD = 1.15; $p < 0.001$, Cohen's $d = 0.27$, power = 0.99) conditions. The mean concern rating in the graphic ($M = 3.27$, SD = 1.32) and table conditions ($M = 3.36$, SD = 1.17) was similar ($\eta_p^2 = 0.002$, power = 0.28).

Thus, the basic flood is due effect was replicated here in the return period condition. In addition, experiment 2 confirmed that absent flood likelihood information, people are inclined to think that a flood is more likely and to be more concerned when a flood has occurred recently, the opposite to the flood is due effect.

Next we examined likelihood estimates and concern ratings for 50-yr (2% chance per year) and 100-yr (1% chance per year) floods using identical analyses to those described above. Because the recency information did not address these flood levels directly we expected a reduced or absent effect of flood recency and no interaction with flood expression (flood is due effect). We retained the 0.017 alpha level used in the previous analyses. In a separate analysis, we compared likelihood estimates across flood levels to determine whether participants understood the relative rarity of the three flood levels and whether this distinction depended on flood expression.

FIG. 5. Average concern ratings in experiment 2 about the occurrence of a 10-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.

3) 50-YR FLOOD

The ANOVA on 50-yr (2%, B) flood likelihood ratings (Fig. 6) revealed no flood term by recency interaction [$F(4, 785) = 1.80$, $p = 0.13$, $\eta_p^2 = 0.01$, power = 0.55]. In other words, there was no evidence for the flood is due effect influencing the perception of the 50-yr (2% chance per year) flood. However, there was a main effect for likelihood expression [$F(2, 785) = 53.97$, $p < 0.001$, $\eta_p^2 = 0.12$, power = 0.99].

Likelihood ratings were higher in the A–B–C ($M = 43.53\%$, SD = 21.47) than in the return period ($M = 33.51\%$, SD = 22.45; $p < 0.001$) or percent chance ($M = 23.65\%$, SD = 23.86; $p < 0.001$) conditions. There was also a main effect for 10-yr flood recency [$F(2, 785) = 4.81$, $p < 0.01$, $\eta_p^2 = 0.01$, power = 0.80]. Likelihood ratings for a 50-yr flood were higher in the recent ($M = 37.52\%$, SD = 23.31) as compared to distant condition ($M = 30.58\%$, SD = 23.31; $p < 0.001$, Cohen's $d = 0.30$, power = 0.99).
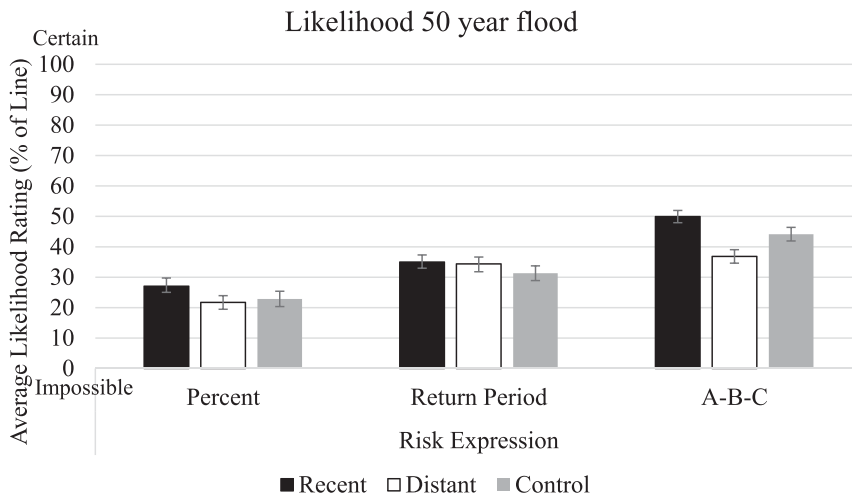
FIG. 6. Average likelihood ratings in experiment 2 of the occurrence of a 50-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.
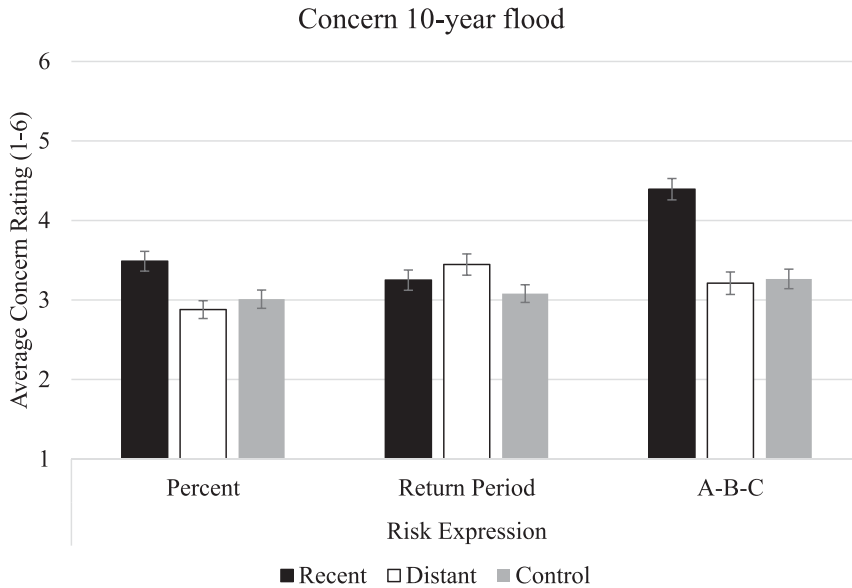
Concern 50 year flood this year
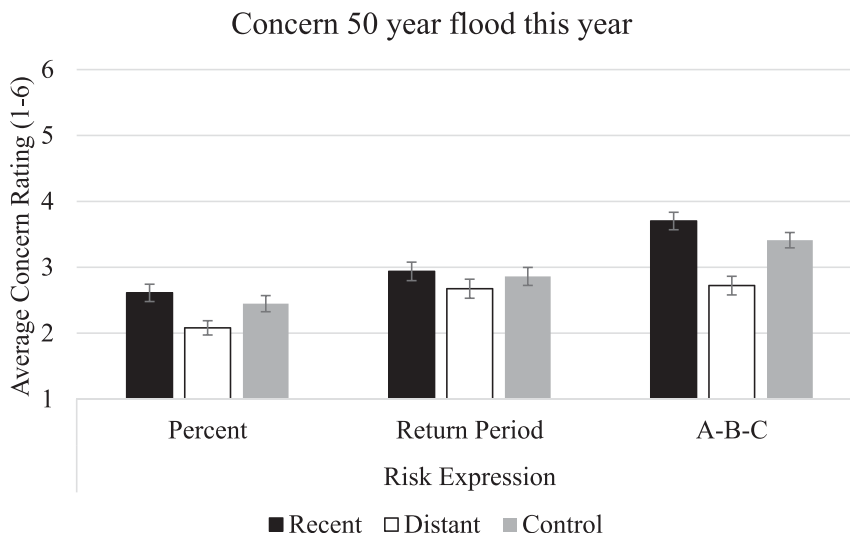
FIG. 7. Average concern ratings in experiment 2 of the occurrence of a 50-yr flood this year in the three likelihood expression and two flood recency conditions. Error bars represent standard error of the mean.

The pattern for concern followed that of likelihood ratings. There was no significant flood term by recency interaction [$F(4, 785) = 2.54$, $p = 0.15$, $\eta_p^2 = 0.01$, power $= 0.52$], again providing no support for a flood is due effect (see Fig. 7). There was, again, a main effect for likelihood expression [$F(2, 785) = 36.78$, $p < 0.001$, $\eta^2 = 0.09$, power $= 0.99$]; participants were more concerned about a 50-yr (2% chance) flood in the A–B–C ($M = 3.28$, SD $= 1.31$) as compared to the return period ($M = 2.82$, SD $= 1.28$; $p < 0.001$, Cohen's $d = 0.36$, power $= 0.99$) or percent chance ($M = 2.36$, SD $= 1.17$; $p < 0.001$, Cohen's $d = 0.74$, power $= 0.99$) conditions. There was also a main effect of 10-yr flood recency [$F(2, 785) = 15.65$, $p < 0.001$, $\eta_p^2 = 0.04$, power $= 0.99$]. Participants were more concerned about a 50-yr flood in the recent ($M = 3.09$, SD $= 1.29$) as compared to the distant ($M = 2.47$, SD $= 1.28$) conditions.

4) 100-YR FLOOD

The ANOVA on 100-yr (1% chance per year, A) flood likelihood ratings revealed no flood term by recency interaction [$F(4, 785) = 1.17$, $p = 0.32$, $\eta_p^2 < 0.01$, power $= 0.37$]. In other words, there was no evidence for the flood is due effect influencing perception of 100-yr (1% chance per year) flood likelihood. There was, however, a main effect for likelihood expression [$F(2, 785) = 30.53$, $p < 0.001$, $\eta^2 = 0.07$, power $= 0.99$; see Fig. 8]. Likelihood ratings were higher in the A–B–C ($M = 33.92\%$, SD $= 28.30$) as compared to the return period ($M = 22.12\%$, SD $= 25.98$; $p < 0.001$) and percent chance ($M = 17.36\%$, SD $= 25.37$; $p < 0.001$)

conditions. Surprisingly, likelihood ratings for the table display were higher ($M = 26.13\%$, SD $= 28.49$) than for the visualization ($M = 22.83\%$, SD $= 26.30$).[3]

The concern analysis was similar to likelihood ratings, with no significant flood term by recency interaction [$F(4, 785) = 1.32$, $p = 0.26$, $\eta_p^2 < 0.01$, power $= 0.41$]. There was, again, a main effect for likelihood expression [$F(2, 785) = 26.14$, $p < 0.001$, $\eta_p^2 = 0.06$, power $= 0.99$]. As with all previous analyses, participants were more concerned about a 100-yr (1% chance, A) flood in the A–B–C ($M = 3.16$, SD $= 1.74$) as compared to the return period ($M = 2.61$, SD $= 1.75$; $p < 0.001$, Cohen's $d = 0.32$, power $= 0.99$) condition. Those in the return period condition were more concerned than those with the percent chance condition ($M = 2.16$, SD $= 1.42$; $p < 0.01$). There was also a main effect for 100-yr flood recency [$F(2, 785) = 7.46$, $p < 0.001$, $\eta_p^2 = 0.02$, power $= 0.94$]. Participants were more concerned in the recent ($M = 2.70$, SD $= 1.54$) as compared to the distant condition ($M = 2.32$, SD $= 1.59$; see Fig. 9).

Thus, the flood is due effect, as evidenced by an interaction between flood expression and flood recency, was only observed for the flood level specified (10-yr flood) and not for the other two flood levels. Although it is possible that there may be a smaller effect size at the other two flood levels than we had the power to detect with this experiment.

---

[3] The statistics that speak to the main effect of visualization were omitted at the request of one reviewer because they had not been predicted beforehand but are available from the authors.

Likelihood 100-year flood
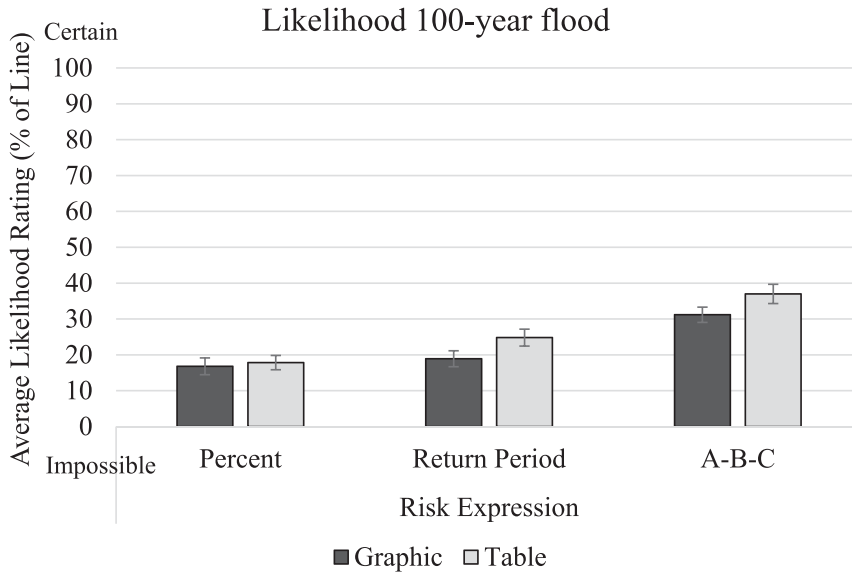
FIG. 8. Average likelihood ratings in experiment 2 of the occurrence of 100-yr flood this year in the two display conditions. Error bars represent standard error of the mean.

Nonetheless the other basic patterns were replicated among all flood levels. Somewhat surprisingly, both likelihood and concern were higher when no likelihood information was provided in the control condition (A–B–C). We return to this issue in the general discussion. In addition participants tended to subscribe to a persistence model of flood events. If a flood had just occurred they thought it was more likely that a flood would occur in the future.

Finally, we compared participants' likelihood ratings for the three flood levels to one another to determine whether participants understood the relative likelihood of the events described. We conducted a repeated measures ANOVA on likelihood ratings with flood level (10, 50, and 100 years) as the within-groups independent variable and likelihood expression (A–B–C, return period, and percent) and display (visualization or table) as the between-groups independent variables. It was clear that participants understood the relative likelihood of the three flood levels. There was a main effect for flood level [$F(2, 1570) = 427.36, p < 0.001, \eta_p^2 = 0.33, power = 0.99$]. Likelihood ratings for the 10-yr (10%, C) flood

Concern 100-year Flood
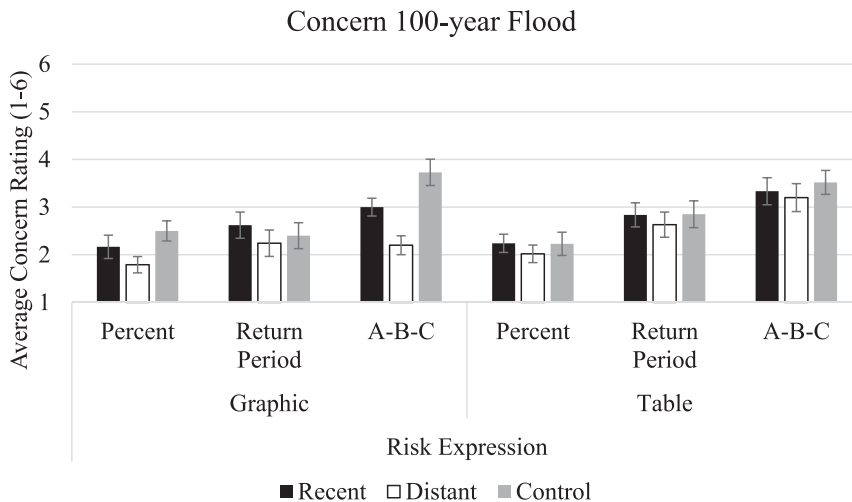
FIG. 9. Average concern ratings in experiment 2 for the occurrence of a 100-yr flood this year in the likelihood expression, return period, and display format condition. Error bars represent standard error of the mean.

Flood Likelihood Ratings
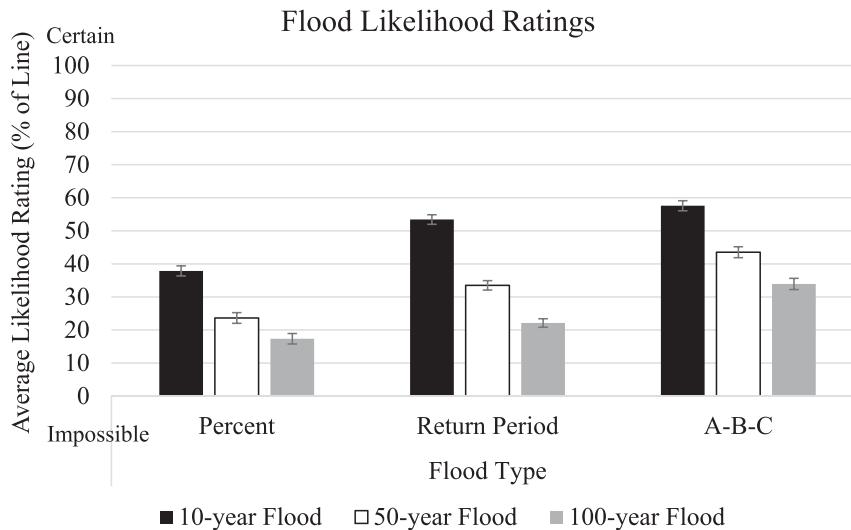
FIG. 10. Average likelihood ratings in experiment 2 for the occurrence of each flood type (10, 50, and 100 years) this year. Error bars represent standard error of the mean.

were higher ($M = 49.93\%$, SD $= 27.07$) than for 50-yr flood (2%, B; $M = 33.61\%$, SD $= 24.05$; $p < 0.001$, Cohen's $d = 0.64$, power $= 0.99$), which, in turn, were higher than for the 100-yr (1%, A) flood ($M = 24.53\%$, SD $= 27.48$; $p < 0.001$, Cohen's $d = 0.35$, power $= 0.99$). Thus, regardless of likelihood expression (even when no information was provided in the A–B–C condition), participants understood that moderate floods are more likely than extreme floods (see Fig. 10).

4. Discussion and conclusions

Importantly, a strong flood is due effect was observed in 10-yr flood likelihood estimates in experiment 2, providing further support for the conclusions of experiment 1. In the return period condition, participants thought a 10-yr flood was significantly more likely when a flood of that magnitude had not occurred for 10 years and less likely when it had occurred recently, suggesting that they thought the expression meant that a flood would occur exactly once every 10 years. It was clear that the flood is due effect was independent of the graphic because it did not interact with that variable. Moreover, experiment 2 clarified that the flood is due effect was specific to the flood level addressed by the term (10-yr flood) because there was no evidence, given the power of the present experiment, for a flood is due effect in either the 50- or 100-yr flood questions.

The power of the return period expression to influence expectations is even more dramatic when considering the fact that the flood is due effect was perhaps counterintuitive in the absence of flood likelihood information. When no information was given in the control condition, people tended to think a flood was more likely when a flood had just occurred, perhaps because they thought that the weather conditions that produced the first flood continued to exist: a persistence effect. This may reflect at least partially valid intuitions about flood events, which can occur in clusters (Gu et al. 2016; Merz and Blöschl 2008; Robinson and Sivapalan 1997; Villarini et al. 2013). Moreover participants appeared to have valid intuitions about the relative rarity of flood events. They rated the 100-yr (1% chance per year) flood as less likely than the 50-yr flood (2% chance per year), which was rated less likely than the 10-yr (10% chance per year) flood. Interestingly, they understood this based on flood level alone when no likelihood expression was provided (A, B, C), suggesting that they believed that extreme events tend to be rare. This pair of findings, the persistence assumption, and the assumed rarity of extreme events contribute to a growing literature documenting the largely accurate intuitions of nonexperts regarding the uncertainty that accompanies weather forecasts even when it is not specified (Joslyn and Savelli 2010; Morss et al. 2008; Savelli and Joslyn 2012).

The other major result of this pair of experiments concerns the probabilistic expression. It is interesting to note that those with the percent chance expression had lower likelihood ratings than did those with the return period or no information. This is particularly dramatic for the 10% chance flood because all other expressions led to estimates well over 50% of the distance on the rating scale between impossible and certain. Although it

is not clear that all participants would have understood that the anchor impossible was comparable to 0% chance and certain to 100% chance, it is arguable that 10% would represent something less than midway between the two anchors for most participants. This suggests that participants tended to overestimate the likelihood of floods, perhaps out of over cautiousness (Weber 1994) but to a smaller degree when provided percent chance expressions. Moreover, those using percent chance were least affected by the recency of similar floods, which may or may not be informative depending on the specific situation. Thus, although percent chance is often thought to be a confusing form of likelihood expression (e.g., Gigerenzer et al. 2005), the evidence reported here suggests that this format conveys the intended likelihood information, without a significant loss in concern, better than the return period or omitting likelihood information altogether.

An interesting but unexpected finding was that participants were often most concerned when they were not given any likelihood information. This was true for the 10-yr flood as well as for the two other flood levels (50, 100) that were not addressed by the likelihood information that was provided. This may suggest that people find the lack of likelihood information unsettling. There is now strong evidence that people understand that all forecasts involve some level of uncertainty, even when it is not specified (Joslyn and Savelli 2010; Morss et al. 2008). Perhaps it is reassuring to users to have the level of uncertainty made explicit.

Nonetheless, it is clear that the return period expression leads to serious misunderstandings about the likelihood involved, especially when a similar event was experienced recently or long ago. This misunderstanding could make users reluctant to protect themselves, exposing themselves unnecessarily to danger. Moreover, participants seemed to be aware of their lack of understanding, even if they were unable to correct it, providing significantly lower ''understanding'' ratings in the return period condition ($M = 4.03$, SD = 1.42) compared to the percent chance condition ($M = 4.37$, SD = 1.37; $p < 0.02$, Cohen's $d = 0.24$, power = 0.99) or even the control condition in which no likelihood information was provided ($M = 4.51$, SD = 1.36; $p < 0.001$, Cohen's $d = 0.35$, power = 0.99). Not only was the percent chance expression rated more understandable, but participants had arguably better objective understanding of the likelihood involved and were less susceptible to task bias effects when they used it.

Thus, the research reported here makes several important contributions to our understanding of how users interpret likelihood expressions. No previous research had systematically manipulated both likelihood

expression and recency to reveal a strong flood is due effect. Because it is now abundantly clear that this expression leads to serious misunderstanding of the risk users face, we recommend that it be avoided wherever possible. Because of the increased access on the Internet by everyday users, the application of this term in a wide range of contexts could be a problem. Moreover, we believe there is a better alternative. This research adds to the growing body of evidence supporting the notion the everyday users can make good use of numeric probabilities. In this case not only did probability provide an arguably better understanding of flood likelihood but also protected users from biases related to recency noted in both other conditions.

APPENDIX

Study Questions

1) If you lived in Bison City, how concerned would you be about Yellowtail Creek flooding this year, at any level?

Not at all   A little Somewhat   Quite a bit Very Extremely

2) On the line below, use your pen or pencil to put a mark indicating how likely Bison City is to experience a 10-year flood this year.

Extremely unlikely   Extremely likely

REFERENCES

Bell, H. M., and G. A. Tobin, 2007: Efficient and effective? The 100-year flood in the communication and perception of flood risk. *Environ. Hazards*, **7**, 302–311, doi:10.1016/j.envhaz.2007.08.004.

Fischer, K., and H. Jungermann, 1996: Rarely occurring headaches and rarely occurring blindness: Is rarely=rarely? The meaning of verbal frequentistic labels in specific medical contexts. *J. Behav. Decis. Making*, **9**, 153–172, doi:10.1002/(SICI)1099-0771(199609)9:3<153::AID-BDM222>3.0.CO;2-W.

Galesic, M., R. Garcia-Retamero, and G. Gigerenzer, 2009: Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychol.*, **28**, 210–216, doi:10.1037/a0014474.

Gigerenzer, G., R. Hertwig, E. Van Den Broek, B. Fasolo, and K. V. Katsikopoulos, 2005: ''A 30% chance of rain tomorrow'': How does the public understand probabilistic weather forecasts? *Risk Anal.*, **25**, 623–629, doi:10.1111/j.1539-6924.2005.00608.x.

Grounds, M. A., 2016: Communicating weather uncertainty: An individual differences approach. Ph.D. dissertation, University of Washington, 141 pp.

Gruntfest, E., K. Carsell, and T. Plush, 2002: An evaluation of the Boulder Creek local flood warning system. Department of Geography and Environmental Studies, University of Colorado at Colorado Springs, 110 pp.

Gu, X., Q. Zhang, V. P. Singh, Y. D. Chen, and P. Shi, 2016: Temporal clustering of floods and impacts of climate indices in

the Tarim River basin, China. *Global Planet. Change*, **147**, 12–24, doi:10.1016/j.gloplacha.2016.10.011.

Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteor. Appl.*, **17**, 180–195, doi:10.1002/met.190.

——, and J. E. LeClerc, 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol. Appl.*, **18**, 126–140, doi:10.1037/a0025185.

——, and ——, 2013: Decisions with uncertainty: The glass half full. *Curr. Dir. Psychol. Sci.*, **22**, 308–315, doi:10.1177/0963721413481473.

——, L. Nadav-Greenberg, M. U. Taing, and R. M. Nichols, 2009: The effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. *Appl. Cognit. Psychol.*, **23**, 55–72, doi:10.1002/acp.1449.

Kellens, W., T. Terpstra, and P. De Maeyer, 2013: Perception and communication of flood risks: A systematic review of empirical research. *Risk Anal.*, **33**, 24–49, doi:10.1111/j.1539-6924.2012.01844.x.

Keller, C., M. Siegrist, and H. Gutscher, 2006: The role of the affect and availability heuristics in risk communication. *Risk Anal.*, **26**, 631–639, doi:10.1111/j.1539-6924.2006.00773.x.

LeClerc, J., and S. Joslyn, 2015: The cry wolf effect and weather-related decision making. *Risk Anal.*, **35**, 385–395, doi:10.1111/risa.12336.

Ludy, J., and G. M. Kondolf, 2012: Flood risk perception in lands "protected" by 100-year levees. *Nat. Hazards*, **61**, 829–842, doi:10.1007/s11069-011-0072-6.

Merz, R., and G. Blöschl, 2008: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resour. Res.*, **44**, W08432, doi:10.1029/2007WR006744.

Morss, R. E., J. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting*, **23**, 974–991, doi:10.1175/2008WAF2007088.1.

Ramos, M. H., S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, **17**, 2219–2232, doi:10.5194/hess-17-2219-2013.

Robinson, J. S., and M. Sivapalan, 1997: Temporal scales and hydrological regimes: Implications for flood frequency scaling. *Water Resour. Res.*, **33**, 2981–2999, doi:10.1029/97WR01964.

Ross, J., I. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson, 2010: Who are the crowdworkers?: Shifting demographics in mechanical turk. *CHI EA '10: Extended Abstracts on Human Factors in Computing Systems*, Atlanta, GA, Association for Computing Machinery, 2863–2872, doi:10.1145/1753846.1753873.

Roulston, M. S., G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins, 2006: A laboratory study of the benefits of including uncertainty information in weather forecasts. *Wea. Forecasting*, **21**, 116–122, doi:10.1175/WAF887.1.

Sandman, P. M., 1998: Communications to reduce risk underestimation and overestimation. *Risk Decis. Policy*, **3**, 93–108.

Savelli, S., and S. Joslyn, 2012: Boater safety: Communicating weather forecast information to high stakes end users. *Wea. Climate Soc.*, **4**, 7–19, doi:10.1175/WCAS-D-11-00025.1.

——, and ——, 2013: The advantages of 80% predictive interval forecasts for non-experts and the impact of visualizations. *Appl. Cognit. Psychol.*, **27**, 527–541, doi:10.1002/acp.2932.

Stone, E. R., J. F. Yates, and A. M. Parker, 1997: Effects of numerical and graphical displays on professed risk-taking behavior. *J. Exp. Psychol.: Appl.*, **3**, 243–256, doi:10.1037/1076-898X.3.4.243.

Villarini, G., J. A. Smith, R. Vitolo, and D. B. Stephenson, 2013: On the temporal clustering of US floods and its relationship to climate teleconnection patterns. *Int. J. Climatol.*, **33**, 629–640, doi:10.1002/joc.3458.

Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth, 1986: Measuring the vague meanings of probability terms. *J. Exp. Psychol.: Gen.*, **115**, 348–365, doi:10.1037/0096-3445.115.4.348.

Weber, E. U., 1994: From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychol. Bull.*, **115**, 228–242, doi:10.1037/0033-2909.115.2.228.

Weinstein, N. D., P. M. Sandman, and W. K. Hallman, 1994: Testing a visual display to explain small probabilities. *Risk Anal.*, **14**, 895–895.

Windschitl, P. D., and G. L. Wells, 1996: Measuring psychological uncertainty: Verbal versus numerical methods. *J. Exp. Psychol. Appl.*, **2**, 343–364, doi:10.1037/1076-898X.2.4.343.