The Impact of Forecast Inconsistency and Probabilistic Forecasts on Users' Trust and Decision-Making

Jessica N. Burgeno and Susan L. Joslyn

University of Washington, Seattle, WA

Corresponding author: Susan L. Joslyn, susanj@uw.edu

Note: This paper was also submitted in partial fulfillment of a dissertation.

Early Online Release: This preliminary version has been accepted for publication in *Weather, Climate, and Society*, may be fully cited, and has been assigned DOI 10.1175/WCAS-D-22-0064.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

When forecasts for a major weather event begin days in advance, updates may be more accurate but inconsistent with the original forecast. Evidence suggests that resulting inconsistency may reduce user trust. However, adding an uncertainty estimate to the forecast may attenuate any loss of trust due to forecast inconsistency as has been shown with forecast inaccuracy. To evaluate this hypothesis, the experiment reported here, tested the impact on trust of adding probabilistic snow accumulation forecasts to single value forecasts in a series of original and revised forecast pairs (based on historical records) that varied in both consistency and accuracy. Participants rated their trust in the forecasts and used them to make school closure decisions. Half of participants received single-value forecasts and half also received the probability of 6 or more inches (decision threshold in the assigned task). As with previous research, forecast inaccuracy was detrimental to trust although probabilistic forecasts attenuated the effect. Moreover, the inclusion of probabilistic forecasts allowed participants to make economically better decisions. Surprisingly, in this study, inconsistency increased, rather than decreased trust, perhaps because it alerted participants to uncertainty and led them to make more cautious decisions. Furthermore, the positive effect of inconsistency on trust was enhanced by the inclusion of probabilistic forecast. This work has important implications for practical settings, suggesting that both probabilistic forecasts and forecast inconsistency provide useful information to decision makers. Therefore, members of the public may well benefit from well-calibrated uncertainty estimates and newer, more reliable information.

SIGNIFICANCE STATEMENT

The purpose of this study was to clarify how explicit uncertainty information and forecast inconsistency impact trust and decision-making in the context of sequential forecasts from the same source. This is important because trust is critical for effective risk communication. In the absence of trust, people may not use available information, and subsequently may put themselves and others at greater than necessary risk. Our results suggest that updating forecasts when newer, more reliable information is available, and providing reliable uncertainty estimates can support user trust and decision making.

1. Introduction

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Forecasts for major weather events often begin days in advance. The weather models upon which forecasts are based update frequently and generally grow more accurate as lead times decrease (Lazo et al. 2009; Wilson and Giles 2013). However, meteorologists are sometimes reluctant to update the forecasts provided to members of the public out of fear that inconsistency in subsequent forecasts will be confusing and negatively affect user trust. "Inconsistency" in this context means that the most recent forecast (e.g., 10 inches snow accumulation) differs from the previous forecast (e.g., 2 inches of snow accumulation) for the same target date (next Saturday). In fact, maintaining consistency in forecasts is considered best practice by some institutions, like the National Oceanic and Atmospheric Administration (NOAA 2016). Yet, because forecasts tend to grow more accurate as lead time decreases, the choice to maintain consistency can be at a loss to accuracy (how closely the forecast matches the outcome).

There is strong evidence that forecast inaccuracy reduces trust. For instance, in a study in which participants used overnight low temperature forecasts to make road salting decisions, they rated trust significantly higher and took protective action more often with low-compared to high-error forecasts (Joslyn and LeClerc 2012). Similarly, in a study in which participants used reports from financial analysts to make investment decisions, participants rated competence, trust, and likelihood of buying future reports higher for accurate compared to inaccurate financial analysts (Kadous, Mercer, and Thayer 2009). Also, compared to patients who imagined receiving accurate initial mammogram test results, those who imagined receiving false positive breast cancer test results, reported reduced trust and being more likely to delay future mammography (Kahn and Luce 2003). Even preschoolers show reduced trust in inaccurate relative to accurate informants (Pasquini et al. 2007; Ronfard and Lane 2018).

By contrast, evidence on the effect of inconsistency in forecasts is sparce and comes largely from non-weather domains. For instance, consumers believe that consistency between two estimates from the same source is a signal of skill (Falk and Zimmermann, 2017). There is also evidence that information about an event from multiple sources is preferred when it is in agreement as opposed to conflicting, all else being equal (Smithson, 1999). Moreover, confidence in one's own decision is higher when based on information from financial advisors who agree with one another as opposed to those who do not agree (Budescu et al., 2003).

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

There is also recent evidence, from our own lab, that speaks to the effect of inconsistency in predictions on trust and decision making. For example, one study that manipulated forecast consistency in sequential thunderstorm and snow forecasts from a single source found that consistent (relative to inconsistent) forecasts led to greater trust (Losee and Joslyn 2018). There is also research comparing the impact of inconsistency to that of inaccuracy, suggesting that sequential forecast inconsistency reduces user trust, but that inaccuracy has a larger negative effect on trust (Burgeno and Joslyn 2020). In these experiments, participants based their school closure decisions on snow accumulation forecasts (e.g., Monday forecast: 4 inches of snow on Wednesday) from a single source, 2-days and 1-day in advance of an anticipated storm. Not only was inaccuracy more detrimental to trust in the forecast but inconsistency appeared to provide useful information. It increased participants' uncertainty expectations, reflected in a wider range of expected outcomes, and led to more conservative closure decisions. An inaccuracy by inconsistency interaction effect suggested that differences in trust due to inconsistency shrank when forecasts were inaccurate. In other words, the reduction in trust due to inaccuracy was substantial to the extent that inconsistency had little additional impact.

At least part of the reason that inconsistency is less detrimental to trust in sequential forecasts may be the fact that, when forecasts are inconsistent, people understand that the most recent forecasts is more likely to be more accurate and regard it as a replacement for an earlier forecast. Indeed, prior research on sequential forecasts suggests that participants' best estimates were more heavily influenced by recent forecasts (1-day in advance) than initial forecasts (2-days in advance), suggesting that participants expected the most recent forecast to be more reliable and were weighting it more heavily (Burgeno and Joslyn 2020).

The relative effects of inconsistency and inaccuracy have also been compared in an experiment based on snow forecasts from 2 different sources both provided at the same time, one day in advance of an anticipated storm. It revealed that while inaccuracy significantly reduced trust, inconsistency between the two sources did not (Su, Burgeno, and Joslyn 2021). In fact, participants incorporated information from both sources equally in their outcome estimates and appeared to glean useful information from inconsistencies. As with inconsistent sequential forecasts (Burgeno and Joslyn 2020), inconsistencies led participants to infer greater uncertainty and to make more cautious decisions. Therefore, inconsistency appears to be less problematic for trust than inaccuracy in both sequential forecasts coming

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

from the same source and simultaneous forecasts from different sources, and it may provide useful information.

It is important to note that a particular kind of trust was measured in this line of work, referred to as "calculative trust". There are at least two kinds of trust that could be affected, 1) relational trust, representing the social bond between the trustor and the trustee, which is based on factors such as the forecast providers intentions, attitudes or goals and 2) calculative trust, sometimes called "confidence," which is based on factors directly related to the quality of the forecast and derived from factors such as past performance (Siegrist, Gutscher and Earle 2005; Twyman, Harvey, and Harries 2008; Earle 2010). The kind of trust tested in the work reported below was also "calculative" trust.

Although the research on inconsistency reviewed above is both important and foundational, it is crucial to note that in order to isolate the effects of inaccuracy and inconsistency, all of the experiments cited above from our own lab used highly controlled forecast stimuli, limiting the range of forecasts values and closely matching the degrees of inaccuracy and inconsistency at small amounts (about 2 inches). In other words, both inconsistency and the inaccuracy were essentially categorical variables (either inconsistent or consistent, either inaccurate or accurate). Moreover, exactly half of forecast pairs were inconsistent and the other half consistent. Similarly, exactly half of forecasts in each consistency category were inaccurate, and half were exactly accurate. That begs the question, will the same effects be observed in more realistic forecast situations in which forecasts vary naturally and take on a wider range of values? Indeed, the degree of inconsistency may be crucial. For instance, the impact on users may be greater if the snow accumulation forecast decreases from 7 to 1 inch compared to from 3 to 1 inch in the subsequent forecast. This may, in turn, translate into a greater impact on trust. Indeed, for some users, small inconsistencies may not be regarded as inconsistency all, but rather as an informative update. Larger inconsistencies, however, may have a qualitatively different impact. In addition, there may be relationships between forecast values, inconsistency and inaccuracy in actual forecasts that may also be relevant. The experiment reported here was designed to evaluate the relative impact on trust of forecast inconsistencies and inaccuracies that vary naturally and take on a wide range of values.

The other question this work was designed to answer is whether there is a benefit to adding an uncertainty estimate to inconsistent forecasts. By "uncertainty estimate" in this

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

context we mean a probabilistic forecast (e.g., 30% chance) indicating the likelihood of a particular outcome. Although forecast inconsistency may reduce trust in some situations, it may be possible to preserve trust in the face of inconsistency by adding a probabilistic forecast as has been shown with inaccuracy. For example, in the road salting study (Joslyn and LeClerc 2012) mentioned above, probabilistic forecasts reduced the negative effects of forecast inaccuracy on both trust and decision making. When provided the probability of observing temperatures at or below the decision threshold (in addition to single-value forecasts) participants rated trust higher than those who received single-value low temperature forecasts alone. There are likely two main reasons for this effect. First, the acknowledgment of uncertainty may make the forecast seem "less wrong" when it fails to verify, preserving trust in the face of forecast error. In addition, people have an intuitive understanding of the uncertainty inherent in weather forecasts, even when it is not specified (Savelli and Joslyn 2010). Therefore, a forecast that makes the uncertainty explicit may seem more honest in the first place. In addition, in these experiments (Joslyn and LeClerc 2012), participants made better decisions from an economic standpoint when they were provided with probabilistic forecasts. In another study, probabilistic forecasts preserved trust to a greater degree than did lowering false alarm rates. In addition, probabilistic forecasts increased compliance with weather warnings (LeClerc and Joslyn 2015). In yet another set of studies, probabilistic forecasts added to flood warnings enhanced subjective understanding of flood likelihood and reduced recency biases compared to a return period expression (e.g., 10-year flood) and to a no-information control group (Grounds, LeClerc, and Joslyn 2018). Thus, a growing body of evidence suggests that laypeople can use explicit probabilistic information and that it may offer several benefits in the decision-making process, not the least of which is preserving trust. Therefore, uncertainty estimates may attenuate the loss of trust due to forecast inconsistency. However, for these benefits to be observed, it may be necessary for probabilistic forecasts to be reliable. In one study (Experiment 1, Burgeno and Joslyn 2020) when targeted outcomes were observed 50% of the time regardless of the probability predicted, no effect of including probabilistic forecasts (compared to single value forecasts) was observed. Thus, the experiment reported here was designed to test whether including *reliable*, probabilistic forecasts preserves trust in the face of forecast inconsistency.

In sum the experiment reported here was designed to test whether the reduction in trust due to forecast inconsistency extends to inconsistency values that vary naturally and if so, whether the reduction in trust is attenuated by including uncertainty estimates in the

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

forecast. It also tested the impact of these factors on participants own outcome estimates and decision quality. This experiment employs the school closure paradigm described above (Burgeno and Joslyn 2020). However, the new experiment reported below used entirely different, realistic forecast stimuli and was conducted two years later with a different group of participants than the previous studies. Participants' goal was to decide, based on a sequence of snow forecasts taken from historical records, whether it was appropriate to close schools due to a snowstorm based on a 6-inch or more accumulation rule. Half of participants received probabilistic forecasts in addition to the single-value snow accumulation amount to determine the impact of probabilistic forecasts on trust and decision quality.

We hypothesized that probabilistic forecasts would enhance trust and that inconsistency and inaccuracy would reduce trust. Furthermore, we hypothesized that probabilistic forecasts would attenuate the negative effects of forecast inconsistency and inaccuracy on trust and enhance decision quality. We predicted that inconsistency would be interpreted as indicating greater uncertainty in the forecast, reflected in a wider range of expected outcomes, and tested whether this would be affected by probabilistic forecasts. Finally, we hypothesized that more recent forecasts, in inconsistent pairs, would have a greater impact on participants' accumulation estimates. Hypotheses were preregistered on Open Science Framework and can be viewed at https://osf.io/dv6j8.

2. Method

a. Participants

A total of 419 University of Washington psychology students participated for course credit and the opportunity to earn a cash bonus. After executing data cleaning procedures (described below), data from 398 participants (62% female, mean age=19.5) remained and were included in the analyses below.

b. Procedure

Participants first gave informed consent and provided their age and gender. Next, they read and listened to, instructions spoken by the experimenter that explained the computer-

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

based task[1] (See Appendix A). Participants were asked to advise schools on whether to close due to an anticipated snowstorm based on weather forecasts provided by "a private weather service that specializes in local predictions". Although several factors are considered when actual closure decisions are made, in this simplified task, the decision was to be based upon snow accumulation alone. Participants were instructed to advise closing if they expected six or more inches of snow accumulation. Participants provided school closure advice for 65 schools across the region for each of two hypothetical winter periods, for a total of 130 trials. Each week was described as involving a different school district to encourage participants to regard the trials as independent of one another.

To better simulate actual weather-related decisions that have real consequences, a point system was used. Participants ending point balance was converted to cash at the conclusion of the experiment to encourage them to put forth their best effort. Participants began with a virtual budget of 332 points. Their goal was to retain as many points as possible. Closure recommendations cost 2 points to reflect the cost of makeup days. There was no cost for recommending that a school stay open; however, if participants advised staying open and six or more inches of snow was observed, they incurred a 6-point penalty to reflect the risk of accidents and injuries due to dangerous road conditions. Notice that, as with many real-world weather-related decisions, the cost of protection is less than the potential cost of the adverse weather event.

Participants earned a cash bonus for the ending point balance at the rate of $1 for every 32 points over 72 (final balance) points. A 72-point threshold was selected to discourage the simplistic and unrealistic strategy of recommending closure for every trial, which would result in a final balance at the payment threshold of 72 points[2]. In addition to

[1] The experiment was programed in Excel Visual Basic and conducted on standard desktop computers.

[2] The endowment was calculated by multiplying the number of trials (130) by the cost of closing (2 points), and adding that product to the payment threshold, (130*2) + 72=332. This was done to create a cushion of points so maintain engagement with the task.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

providing real consequences for the decisions made, this point system held constant the cost and the penalty across participants. In other words, *unlike* many real-life weather threats, for which the cost of protection or the vulnerability to consequences may be greater for some—in this context it was the same across participants reducing statistical noise and allowing us to better detect differences due to the forecasts alone.

For every trial, participants based their school closure decision on two snow forecasts for Wednesday, one issued on Monday (two days prior to the event) and one on Tuesday (one day prior to the event). Forecasts were presented sequentially, centered on separate screens with the current weekday in the top left-hand corner in bold font. To determine how the forecasts influenced participants own estimates, participants were asked to report the number of inches of snow they expected for Wednesday as well as the least (minimum estimate, "as little as") and greatest (maximum estimate, "as much as") number of inches that they would not be surprised by. Then, participants rated their trust in the forecast, "to help them make their [school closure] decision" on a 6-point drop-down menu, from "Not at all" to "Completely". Notice that this question asks participants to focus on the quality of the information itself, rather than the source of the information. See Appendix B for the exact wording of each question. The current point balance was displayed in the bottom left-hand corner of each screen. When participants completed all four questions, they clicked a "next" button in the bottom right-hand corner of the screen to progress to the next screen and could not go back and change responses on the previous screen. Then, the second forecast was shown and participants answered the same four questions with respect the second forecast. Next, the decision screen appeared. The current day (Tuesday) was displayed in bold font in the top left-hand corner and two buttons in the middle of the screen labeled "close" and "stay open." Below each respective button was a reminder of the associated point cost and that "close" meant "I think snow accumulation will be 6 inches or more" whereas "stay open" meant "I think snow accumulation will be less than 6 inches".

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

After submitting their school closure decision, a fourth screen appeared saying that the school followed their advice and either stayed open or closed. The observed snow accumulation on Wednesday was displayed and the resulting cost or penalty was shown (unless neither occurred). Participants' point balance and, if applicable, the penalty incurred, was displayed in the bottom left corner of the screen. Participants again rated their trust in the forecasts using the same pull-down menu. In sum, each trial consisted of 4 screens: 1) Monday forecast for Wednesday, 2) Tuesday forecast for Wednesday, 3) Tuesday night school closure decision, and 4) Wednesday outcome. Then, the next trial began with a new set of forecasts and outcome that pertained to a school in a different district. Participants completed four practice trials before the test trials began (See Appendix A).

c. Forecast Stimuli

The data upon which the snow accumulation forecasts (48 and 24 hours in advance), probabilities of 6 or more inches accumulation and observed 24-hour snow accumulation outcomes were based, were obtained from the Eastern Region Headquarters of the National Oceanographic and Atmospheric Administration (NOAA). The original set of 160 forecasts pertained to a snowstorm that occurred in several locations over the eastern United States on February 9, 2017[3]. In the experiment we used 130 of these, treating each pair of forecasts and outcome as a separate event. All single value forecasts and observed accumulation amounts were rounded to the nearest inch. Although some other small changes were made (described below), the vast majority of trials[4] included original forecast values, and all outcome values were identical to the original historical forecast set. As a result, forecasts varied naturally in terms of snow accumulation totals, accuracy, and consistency. Moreover, the critical characteristics of the original forecast data set were maintained (See Appendix C).

1) Consistency

[3] Special thanks go to David B. Radell at NOAA and the National Weather Service for providing us with the forecast data.

[4] 68% were presented in the same order as in the historical forecast set

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Consistency was defined as an exact match between the Monday and Tuesday forecasts. All inconsistent trials were inconsistent by 1 or more inches with a range of -8 to 8 inches and a mean of 2.86 inches. Because the original data set had few exactly consistent pairs, they were increased by making slight changes to the second forecast in 9 pairs (7%) in which the initial differences were small. The original cases pertained to a single weather event in which the expected accumulation increased over time, so there were very few descending forecast pairs (4%, N=7). This was problematic because some anecdotal evidence[5] suggests that downgraded forecasts (descending in this case) are more likely to be altered to maintain consistency by forecasters. Therefore, we increased the proportion of descending forecasts by flipping the order of 19 inconsistent forecast pairs. As a result, in the forecast set used here, 40 (61%) of the inconsistent trials were ascending (values increased from first to second forecast) and the remaining 26 (39%) were descending (values decreased from first to second forecast).

2) Accuracy

Accuracy was gauged relative to the second (Tuesday) forecast. By this standard, the proportion of exactly accurate forecasts was similar to that of the historical forecast set (See Appendix C). All inaccurate trials were inaccurate by 1 inch or more. Inaccuracies ranged from -6 to 10 inches and had a mean of -1.02 inches. Thus, like the historical forecast set, inaccurate forecasts were biased high by about an inch and less than 20% crossed the 6-inch decision threshold (e.g., a second forecast of 7 inches and an observed snow accumulation of 5 inches).

3) Probabilistic Forecasts

For half of participants, the forecast also included the probability of 6 or more inches of snow. The probabilities were also based on those provided in the historical data set. However, it was important to first test the impact of well-calibrated probabilistic forecasts, otherwise,

[5] Unpublished interviews with operational forecasters at National Weather Service Western Region, Seattle, WA.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

null effects could be due to either the genuine lack of an effect or simply the lack of an effect for uncalibrated probabilities. This was especially true of the most recent forecast used as the standard for accuracy. Therefore, some second forecast probabilities were altered slightly so that forecasted probabilities for 6 or more inches of snow accumulation roughly matched the frequency of observing 6 or more inches of snow. See Appendix D for the calibration procedures and Appendix C for forecast characteristics.

e. Design

A single factor (forecast format) between-participants design was used. Half of participants received a single value forecast, while the other half received the same single value and the probability of six or more inches of snow accumulation (e.g., "…4 inches of snow … however, there's a 30% chance of 6 or more inches of snow"). We refer to the former as deterministic in that they imply an exact outcome (e.g., "…4 inches of snow") and the latter as probabilistic. Thus, other than the additional probability of observing 6 or more inches of snow, the forecasts and outcomes seen by both groups of participants were identical. Forecasts were presented in one of four fixed orders[6].

Participants were randomly assigned to one of the two forecast format conditions and one of the four forecast orders. Forecast values, the magnitude of inconsistency and inaccuracy, and the economically optimal decision (see closure decision analysis below) were also included as predictor variables. The outcome variables were trust rating, participants' decisions about whether to close schools or not (closure decisions) and snow accumulation estimates.

3. Results

Prior to conducting the main analyses, we eliminated participants who did not understand the task, were not paying attention or not taking the task seriously. To this end, participant data were excluded if a) they provided a lower estimate for maximum than for

[6] Order was a control variable to ensure that any observed effects would not be tied to a particular order. All dependent variables were summarized across order.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

minimum snow accumulation estimate, or if b) their average best estimate, or c) highest day 2 maximum or minimum estimates were unreasonably large, i.e. greater than the national record accumulation amount for lowland (200m or less above sealevel) snowfall (49 inches, NOAA's National Climatic Data Center 2019). Twenty-two participants were excluded in this procedure, leaving 398 participants in the following analyses.

a. Trust

1) Hypotheses

The primary hypotheses for this research concerned whether trust was impacted by access to probabilistic forecasts, inconsistency between two consecutive forecasts for the same event, inaccuracy of the most recent forecast (when compared to the outcome), or interactions among these variables[7]. We hypothesized that:

H1. Probabilistic forecasts would increase trust in forecasts compared to deterministic forecasts.

H2. Inconsistency would reduce trust in forecasts.

H3. Inaccuracy would reduce trust in forecasts.

H4. The negative effect of forecast inconsistency on trust would be attenuated by the inclusion of a probabilistic forecast.

H5. The negative effect of forecast inaccuracy on trust would be attenuated by the inclusion of a probabilistic forecast.

2) Data Analysis Plan

Because of our interest in the effects of inaccuracy on trust (as well as inconsistency and probabilistic forecasts), we analyzed the post outcome trust measure, at which point

[7] The numbering of hypotheses as reported here is slightly different than those registered although the content is the same.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

forecast accuracy was known to participants (Question 6, Appendix B).[8] Trust was an ordinal variable. Therefore, it was analyzed with a series of Generalized Estimating Equations (GEEs) using cumulative link proportional odds regression models (see Appendix E for model details and regression tables) which are designed to model ordinal data and population-averaged (between-group) effects. To conduct these analyses, we used the 'multgee' package (Touloumis 2015) for R. We specified an 'independence' working correlation structure[9] and robust standard errors to build in resistance to possible misspecifications of the working correlation structure.[10] For this and all subsequent analyses, an alpha level of .05 was used to determine statistical significance.

3) Trust Ratings

As hypothesized, probabilistic forecasts increased trust (See Appendix E, Table E1). The estimated association between forecast format and trust was significant such that when trials included probabilistic forecasts, compared to equivalent trials with deterministic forecasts (inaccuracy and inconsistency held constant), the odds of reduced trust decreased (trust increased) by approximately 20%, *estimated odds ratio= 0.80, 95%CI = (0.65,0.99), p =.04*.

Contrary to our predictions, inconsistency (mismatch between Forecast 1 and Forecast 2) appeared to slightly increase (rather than decrease) trust (Appendix E, Table E1). The estimated association between inconsistency and trust ratings was significantly positive such

[8] The trust measure taken earlier (Question 4), was prior to making the decision or learning the outcome at which point accuracy was unknown to the participant. Nonetheless it yielded similar results with probabilistic forecasts increasing trust by approximately 35%, *estimated odds ratio=.65, 95%CI = (0.51,0.82), p<.001*. Inconsistency increased trust by approximately 5%, *estimated odds ratio = 0.95, 95%CI = (0.94, 0.97), p<.001*).

[9] This is a simplifying assumption that responses nested within a participant are independent of one another.

[10] A working correlation structure does not need to be specified correctly because robust standard errors, with wider confidence intervals than naïve standard errors, are agnostic to the structure specified. Therefore, even if the working correlation structure is mis-specified, the model will still generate appropriate estimates.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

that a 1-inch increase in the difference between Forecast 1 and 2, compared to otherwise equivalent trials (inaccuracy and format held constant), *decreased* the odds of trust reduction (increased trust) by approximately 8%, *estimated odds ratio = 0.93, 95%CI = (0.92, 0.94), p<.001*.

Meanwhile, inaccuracy, the degree of mismatch between Forecast 2 and the outcome appeared to decrease trust as predicted. See Appendix E, Table E1. The estimated association between forecast inaccuracy and trust was significantly negative such that a 1-inch difference between Forecast 2 and the observed accumulation, compared to otherwise equivalent trials (inconsistency and format held constant), *increased* the odds of trust reduction (reduced trust) by approximately 15%, *estimated odds ratio = 1.15, 95%CI = (1.14, 1.17), p<.001*. To reiterate, although the effect of inaccuracy on trust confirmed our hypothesis, the effect of inconsistency did not. Inaccuracy had a negative association with trust (decreased trust) whereas inconsistency had a slight positive association with trust (increased trust).

Previous research suggested that inconsistent forecasts had a smaller effect on trust when forecasts were inaccurate (Burgeno and Joslyn 2020). To better understand this relationship with naturalistic forecasts incorporating a wider range of inconsistencies and inaccuracies, we conducted exploratory analyses with inaccuracy dichotomized at 3 inches (roughly the mean of inaccuracies). See Appendix E, Table E4 and E4a. In these data, the strength of the positive association between inconsistency and trust differed significantly across levels of forecast accuracy, *p<.001* such that it was stronger (interaction odds ratio farther from 1) for trials with greater inaccuracy (more than 3 inches from the outcome), *estimated odds ratio=0.80, 95%CI= (0.78,0.82)*, compared to equivalent trials (forecast format held constant) with less inaccuracy (less than 3 inches), *estimated odds ratio=0.93, 95%CI= (0.92,0.94)*. This suggests that as with previous research, at low forecast inaccuracy there was an association between inconsistency and trust. However, as the inaccuracies increased (greater than 3-inches, not tested in previous research) the effect on trust was greater. In contrast with previous research, here, the association between trust and inconsistency was positive. Therefore, the positive effect of inconsistency on trust was greater when inaccuracy was greater.

The association between inconsistency and trust also differed significantly across forecast format, *p<.001*. The positive association between inconsistency and trust was

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

stronger (farther from odds ratio=1) for trials that included probabilistic forecasts, *estimated odds ratio=0.90, 95% CI= (0.88,0.91)*, compared to equivalent trials (inaccuracy held constant) with deterministic forecasts, *estimated odds ratio=0.96, 95%CI= (0.94,0.97)*. See Appendix E, Table E2 and E2a. In other words, probabilistic forecasts were associated with a stronger increase in trust due to inconsistency compared to equivalent trials with deterministic forecasts (for the general pattern, see Figure 1, Panel A).

Similarly, in support of our hypothesis, the association between inaccuracy and trust differed significantly by format, *p=.04*. The negative association was weaker for trials that included probabilistic forecasts, *estimated odds ratio=1.13, 95%CI= (1.12,1.15)*, compared to equivalent trials (consistency held constant) with deterministic forecasts, *estimated odds ratio=1.17, 95% CI = (1.14, 1.19)*. See Appendix E, Table E3 and E3a. In other words, as hypothesized, probabilistic forecasts attenuated the negative effect of inaccuracy on trust, compared to equivalent trials with deterministic forecasts (for the general pattern, see Figure 1, Panel B).

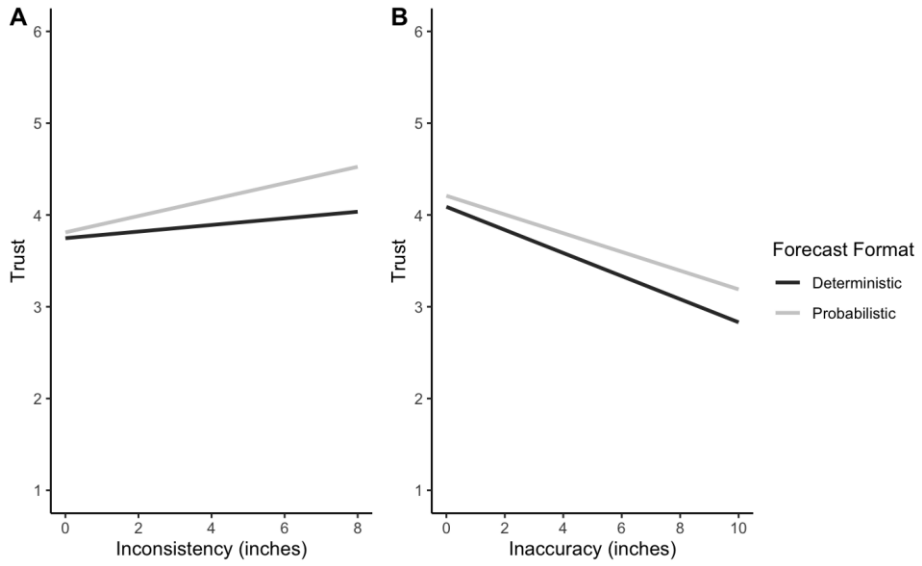


Figure 1. Panel A. Regression Lines for Trust Rating by Inconsistency and Forecast Format. Panel B. Regression Lines for Trust Rating by Inaccuracy and Forecast Format. Note that unlike the analyses reported above, these figures do not control for the effect of other variables. They merely provide an illustration of the general pattern.

Taken together, these results suggest that here, with more realistic forecasts, unlike previous experiments, inconsistency increased rather than decreased trust and the impact was

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

greater with greater inaccuracies. However, the rest our predictions were confirmed. Probabilistic forecasts increased trust and interacted with the effects on trust due to both inconsistency and inaccuracy. Probabilistic forecasts enhanced the positive association between inconsistency and trust. At the same time probabilistic forecasts attenuated the reduction in trust due to inaccurate forecasts.

Table 1 Descriptive Statistics:
Participants' Mean Response on Key Dependent Variables Within Each Forecast Format Condition (Deterministic, Probabilistic)

	Deterministic		Probabilistic	
	Mean	SD	Mean	SD
Post-Decision Trust (rating scale from 1-6)	3.82	1.41	3.99	1.42
Expected Value Difference (smaller indicates better decisions)	-1.47	0.95	-1.43	0.90
All below are inches:				
Best Estimate, F1	4.90	3.77	4.87	3.78
Best Estimate, F2	5.81	4.56	5.77	4.63
Minimum Estimate, F1	2.73	3.15	2.63	2.93
Minimum Estimate, F2	3.47	3.77	3.24	3.51
Maximum Estimate, F1	7.18	4.47	6.79	4.56
Maximum Estimate, F2	8.02	5.18	7.72	5.21
Range Estimate	4.55	3.51	4.48	3.51

Note: F1 refers to forecast 1 and F2 refers to forecast 2.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

b. Accumulation Estimates, Ranges and Closure Decisions

Next we examined participants' closure decisions and snow accumulation estimates. We hypothesized that:

H6. Probabilistic forecasts would enhance decision quality, defined here as the expected value of the decision (see calculation below) and greater differentiation of closure decisions across the decision threshold value (see below).

H7. Inconsistency would increase uncertainty expectations, defined here as the range of anticipated outcomes ("as much as", "as little as"), and increase decision quality. We also asked what further impact forecast format (deterministic, probabilistic) would have on uncertainty expectations.

H8. The most recent forecast (Forecast 2) would have a greater impact on participants' outcome estimates requested after Forecast 2 was shown, than would the initial forecast, (Forecast 1) suggesting that participants understood that the most recent forecast was more accurate.

1) Data Analysis Plan

The continuous variables, decision quality, uncertainty expectations and snow accumulation estimates were analyzed using linear mixed model regressions[11]. A t-statistic (coefficient divided by its standard error) and alpha levels of .05 were used to determine whether the coefficient of each predictor variable (See Appendix G for regression tables) was significantly different from 0, i.e., whether the contribution that predictor was significant. School closure decisions were analyzed as a binary variable, modeled with a series of binary logistic GEEs (see Appendix F for model details and regression tables). To conduct these analyses, we used the 'geepack' package for R (Højsgaard, Halekoh, and Yan 2006), with robust standard errors. We specified an 'independence' working correlation structure and binomial family. See Table 1 for descriptive statistics.

1) Decision Quality

[11] Linear mixed model regression analyses are also capable for accounting for clustered responses.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

First, we examined whether probabilistic forecasts improved decision quality. The quality of the participant's decision was defined as its value, prior to knowing the outcome, referred as the "expected value" (Bernoulli 1954). We describe it here as the "expected cost" because only losses (cost of closure or penalty) were possible in this task. For each trial the optimal choice was the one with the least expected cost (Murphy 1977). There were two possible options on every trial, to advise 1) keeping the school open or 2) closing. The expected cost of keeping a school open (there was no actual cost) was the product of the 6-point penalty and the chance of receiving it (the percent chance of 6 or more inches of snow for the second forecast on that trial). The cost of closing a school was the 2-points that participants paid when they selected that option. A 33% chance of 6 or more inches of snow was the breakeven point at which the expected cost of staying open (.33 x 6 = 2) was equal to the cost of closing (2 points). Therefore, whenever the chance of 6 or more inches was greater than 33% it was optimal to advise closing because the cost of closing was less than the expected cost of staying open. Whenever the chance of 6 or more inches was *less* than 33%, it was optimal to advise staying open. A difference score was calculated on each trial by subtracting the expected (or actual) cost of the participant's choice from the optimal choice on that trial (henceforth referred to as expected cost difference). A "0" difference indicates that the participant made the optimal choice. Otherwise, the value is negative. Then, a linear mixed model regression analysis was conducted on the expected cost difference (See Appendix G), with forecast format (probabilistic/deterministic), inconsistency, and the inconsistency by forecast format interaction entered simultaneously as predictors[12].

Confirming our hypothesis, the expected cost difference was smaller (decision quality was better) for probabilistic compared to the deterministic forecasts (See Table 1). In particular, shifting from the deterministic to the probabilistic format predicted a .06 unit decrease in the expected cost difference, $t(51736)=10.10$, $p<.001$.

[12] Inaccuracy was not included as a predictor because participants had not learned the outcome at the point at which they made a decision.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

There was also an unpredicted increase in decision quality due to inconsistency, although it was smaller than the effect of forecast format. For every 1 unit increase in inconsistency, there was a .02 unit decrease in expected cost difference (decision quality was better), *t(51736)=21.19, p<.001*. Additionally, the inconsistency by forecast format interaction was significant such that the probabilistic forecast reduced the expected cost difference (increased decision quality) for smaller inconsistencies, but less so for larger inconsistencies, where decision quality was already higher, *t(51736)=6.73, p<.001, B=.01*. We will return to this issue in the discussion.

To better understand the decision errors participants made we next examined the difference in participants decisions to close schools above and below the optimal decision threshold. As mentioned above, according to expected value theory, it was optimal to close schools whenever the probability of 6 or more inches of snow was 33% or higher, and to keep schools open otherwise. By this standard, as is common with decisions that involve only losses (Tversky and Kahneman 1979)[13]. most decision errors (65%) were risk-seeking (participants kept schools open when they should have closed) as opposed to risk-averse (closing schools when they should stay open). Binary logistic GEE models were used to examine the associations between closure decisions (open or close) and forecast format, inconsistency, and a categorical variable that indicated whether the optimal decision was to stay open or close on that trial. Two interactions were also tested, forecast format by inconsistency and forecast format by optimal decision. Thus, there were three models: one with the main effects entered simultaneously, and one for each of the two interaction effects (controlling for all main effects; see Appendix F).

Indeed, participants tended to follow the optimal strategy. The estimated association between optimal decision and actual closure decisions was significantly positive such that a day 2 forecast probability at or above 33% increased the odds of deciding to close by approximately 3000%, *estimated odds ratio=30.39, 95% CI= (28.30, 32.60), p<.001*. See Appendix F, Table F1.

[13] There are some exceptions to this at very small likelihoods (Tversky and Fox 1995)

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Importantly, as reflected in the expected value analysis, participants made fewer errors with probabilistic forecasts. The estimated association between optimal decision and closure decisions varied significantly across forecast format, $p<.001$. Probabilistic forecasts supported greater differentiation across the decision threshold, *estimated odds ratio = 44.2, 95% CI = (39.4, 49.5),* compared to deterministic forecasts (inconsistency held constant), *estimated odds ratio = 22.7, 95% CI = (21.2,24.4).* See Appendix F, Table F2 and F4. In other words, probabilistic forecasts decreased the odds of deciding to close when it was optimal to keep a school open and increased the odds of deciding to close when it was optimal to close compared to those who received deterministic forecasts.

In contrast, participants closed more often overall as the inconsistency in forecasts increased. The estimated association between inconsistency and closure decision was significantly positive, such that a 1-inch difference between Forecast 1 and 2, compared to otherwise equivalent trials (forecast format and threshold orientation held constant), increased the odds of deciding to close by approximately 44%, *estimated odds ratio=1.44, 95%CI =(1.42, 1.46), p<.001*. See Appendix F, Table F1. In addition, the association between optimal decision and actual closure decisions was stronger for larger inconsistencies compared to smaller inconsistencies, *estimated odds ratio =1.68, 95% CI = (1.59,1.77), p<.001*.[14] See Appendix F, Table F3.

Thus, examination of closure decisions above and below the optimal threshold (33% chance of 6 or more inches) aligned with the expected value analysis. Probabilistic forecasts allowed participants to make better decisions than did deterministic forecasts in both analyses. Participants also made better decisions when forecasts were inconsistent. This was due in part to the fact that inconsistency encouraged them to close the schools more often, an advantage in this task in which people tend to be risk seeking (majority of errors were *not* closing when closing was optimal).

[14] This may be explained by the fact that magnitude of inconsistency was positively correlated with the probability of greater than 6 inches (r=.67, p<.001), making it generally optimal to close in trials in which there were large inconsistencies.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

3) Range Estimates

The above analysis suggests that participants made better decisions (in this case more conservative, closing schools more often) both when they were provided with explicit uncertainty estimates, the precent chance of 6 or more inches of snow, and when there was greater inconsistency between the Day 1 and 2 forecasts. This latter result could be due in part to the fact that participants interpreted inconsistency as an indication of uncertainty in the forecast. To determine whether this was the case, a range of anticipated outcomes was calculated. This was done by subtracting participants' minimum (Question 2, Appendix B) from their maximum (Question 3, Appendix B) estimate of the number of inches that would not surprise them, taken after the second forecast. A wider range of anticipated outcomes suggests greater perceived uncertainty. Then, a linear mixed model regression was conducted on range of outcomes, with inconsistency, forecast format, and the inconsistency x forecast format interaction entered simultaneously as predictors. See Appendix G for regression tables. Confirming our hypothesis, forecast inconsistency tended to increase the range of anticipated outcomes. More specifically, every 1-inch increase in inconsistency predicted a .62-inch increase in range, $t(51340)=88.23$, $p<.001$. The main effect of forecast format did not reach significance, $t(404)=1.39$, $p=.17$. However, the inconsistency by forecast format interaction was significant such that participants who received probabilistic forecasts expected a smaller range of values for lower magnitude inconsistencies, and a larger range of values for higher magnitude inconsistencies, compared to participants who received deterministic forecasts, $t(51340)=10.37$, $p<.001$, $B=.09$ (see Figure 2). Thus, as predicted, participants expected greater uncertainty with greater inconsistency. In addition, probabilistic forecasts amplified the difference in uncertainty expectations across the range of inconsistency.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

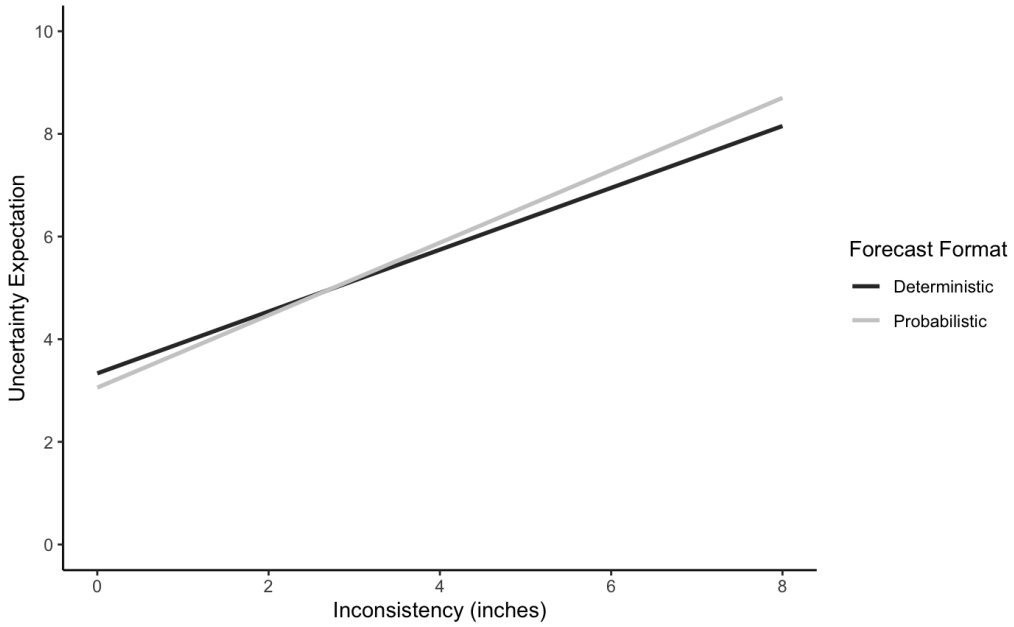


Figure 2. Regression Lines for Uncertainty Expectations by Inconsistency and Forecast Format.

3) Snow Accumulation Estimates

To determine how the two forecasts influenced a participant's own expectations of the outcome, we next examined snow accumulation estimates for Wednesday made after Forecast 2 (Question 1, Appendix B). A linear mixed model regression was conducted on snow accumulation estimates, with three continuous predictor variables (Forecast 1 value, Forecast 2 value, inconsistency) and the categorial predictor, forecast format (deterministic, probabilistic) entered simultaneously with the inconsistency by forecast format interaction.[15] See Appendix G for regression tables.

As hypothesized, the second forecast was a much better predictor of snow accumulation estimates than was the first forecast. For every 1 unit increase in the second

[15] Inaccuracy was not included as a predictor because participants had not yet learned the outcome at the point at which they made an estimate.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

forecast, there was a .87 unit increase in estimated snow accumulation, *t(51330) = 322.72, p<.001*.[16] In contrast, for every 1 unit increase in the first forecast, there was only a .08 unit increase in estimated snow accumulation, *t(51330) = 30.11, p<.001*. In addition, there was an unpredicted effect of inconsistency. Inconsistency slightly but significantly reduced estimates. More specifically, every 1 unit increase in inconsistency predicted a .03 unit decrease in estimated snow accumulation, *t(51330) =5.91, p<.001*. The main effect of forecast format failed to reach significance, *t(411) = .72, p=.47*. However, the inconsistency by forecast format interaction was marginally significant, *t(51330)=1.97, p=.05*, such that the reduction in estimates due to inconsistency was stronger for deterministic forecasts than for probabilistic forecasts. In sum, as predicted, these results suggest that participants weighted the most recent forecast ten times more heavily than the earlier forecast, in their own estimate.

4. Discussion/Conclusion

The experiment reported here is the first to demonstrate the benefits of probabilistic forecasts to enhance both trust in the forecast and decision quality in the face of forecast inconsistency. Participants made better decisions, in terms of both increased expected value and fewer decision errors with probabilistic than deterministic forecasts. A closer inspection of decision errors clarified the benefits of the probabilistic forecast. Because only costs and losses were possible in this task, participants made more risk seeking (failing to close schools when it was economically optimal) than risk averse errors (closing schools when it was NOT economically optimal). In cost/loss situations such as this, people tend to prefer to take a risk than to pay a small cost up front to protect against that risk, even when it is *not* economically optimal to do so (Tversky and Kahneman 1979). However, the error analysis revealed that those with probabilistic forecasts were less prone to this strategy. They differentiated to a greater degree across the optimal decision threshold. In other words, when provided with the

[16] Note that, due to the inclusion of random effects, R^2 is uninterpretable for mixed model regressions.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

probability of six or more inches of snow, participants closed schools more often when it was economically optimal to do so (probability of 6 or more inches was 33% or more), and kept schools open more often when it was economically optimal to do so (probability of 6 or more inches was less than 33%) compared to participants using the deterministic forecast alone.

This experiment is also the first to demonstrate the impact of forecast inconsistency on trust and decision-making using naturalistic forecast stimuli. The basic conclusions from these results align remarkably well with those reported in previous highly controlled experiments (Burgeno and Joslyn 2020; Su, Burgeno, and Joslyn 2021) suggesting that forecast inconsistency, as it is defined here, may not be as detrimental to trust as is often assumed.

However, here, in contrast to the highly controlled studies cited above, the results suggest that naturalistic forecast inconsistency may have a positive impact on trust. One potential explanation resides in the set of historical forecasts used here in which the inconsistent forecasts were predominantly ascending (the second forecast was for greater accumulation than the first). Moreover, the increasing forecast trend tended to be confirmed by the outcome in those trials. For 72% of the ascending trials the observed accumulation was higher than the most recent forecast. People may have expected the trend to continue, as has been shown in previous research (Hohle and Teigen 2015, 2018; Maglio and Polman 2016), and confirmation of those expectations may have increased trust. Another factor that may have increased trust slightly is that fewer of the inconsistencies between forecasts (31%) crossed the 6-inch decision threshold in this experiment, compared to the highly controlled studies (50% at minimum). Because participants decisions depended on whether 6 or more inches was expected, an inconsistency may be less trustworthy when the two forecasts point toward different choices (close/open). Therefore, the slight positive effect of inconsistency on trust found in this experiment may be specific to situations in which there is an ascending trend or the trend in forecasts is confirmed by the result, or the inconsistency is less consequential to the decision. Resolving these issues might be a fruitful line of future research, in which such variables could be systematically manipulated to determine their individual impacts on trust.

An alternative more general explanation is that inconsistency increases trust because it acts as an estimate of uncertainty. As with the prior research (Burgeno and Joslyn 2020), the results reported here demonstrated that participants expected a larger range of outcomes

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

with greater inconsistency, suggesting that they perceived greater uncertainty in these forecasts. However here, unlike the previous highly controlled studies in which inconsistency was held constant at a few inches, some of the inconsistencies were much larger. This may have enhanced the positive effect of perceived uncertainty on trust. It is clear that an explicit expression of uncertainty increases trust. As with numerous previous experiments (Joslyn and LeClerc 2012; LeClerc and Joslyn 2015; Grounds, LeClerc, and Joslyn 2018), the inclusion of the probabilistic forecast increased trust over the single value forecast. It may be that when uncertainty is acknowledged in some way, either with an explicit uncertainty estimate or implied by the inconsistency in forecasts, the forecast seems less "wrong" when the single value forecast does not match the observed snow accumulation.

Somewhat surprisingly, forecast inconsistency also increased decision quality slightly, perhaps because it was interpreted as a sign of uncertainty. Forecast inconsistency appeared to encourage greater cautiousness, closing schools more often overall, as was seen in previous research (Burgeno and Joslyn 2020). Greater cautiousness tended to increase decision quality in this task because the majority of errors were risk-seeking (failing to close when it was optimal). The increase in cautiousness with inconsistent forecasts seen here may have been due in part to the fact that with these forecast data, greater inconsistency tended to be correlated with higher forecasted snow accumulation totals in Forecast 2 ($r=.39$, $p<.001$). However, this could not have been the explanation in the previous research in which inconsistency also increased cautiousness (Burgeno and Joslyn 2020) because forecast values in those experiments were held constant across conditions. Thus, an explanation that accounts best for all of these results, is that the increase in decision quality is due to the fact that inconsistency acts to signal uncertainty which promotes cautiousness. Regardless of the reason, it is important to note that the positive effect of inconsistency on closure decisions differed qualitatively from that of probabilistic forecasts which was more precise. Probabilistic forecasts, because they specified the percent chance of snow accumulation surpassing the decision threshold (6 or more inches) increased closure decision mainly when it was optimal to do so and not otherwise.

We were also interested in how people integrate information from differing forecasts to form their own estimates. In line with the previous research on sequential forecasts (Burgeno and Joslyn, 2020), participants' snow accumulation estimates were influenced more strongly by the second than by the first forecast values. In other words, although participants

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

did not completely disregard the first forecast, they appeared to understand that the most recent forecast should take precedence. There are at least two possible explanations for this. It may be due to extensive extra-experimental experience with real weather forecasts, leading to many, oftentimes correct, intuitions about forecasts (Morss, Demuth, Lazo, 2008; Joslyn and Savelli, 2010; Savelli and Joslyn, 2012). However, it's important to note that our forecast stimuli were realistic in the sense that second forecasts (Mean inaccuracy= 2.14, SD = 1.81) were on average closer to accurate than first forecasts (Mean inaccuracy = 2.42, SD = 2.03). Participants might have learned (explicitly or implicitly) to discount first forecasts within the context of the experimental experience.

The main limitation of the research presented here is related to one of the primary goals: to evaluate the effects of forecast accuracy and consistency on trust and decision making in the context of naturalistic forecasts. Allowing forecasts and outcomes to vary naturally led to a loss in internal validity. In other words, some of the effects observed here may be limited to similar forecast sets. For instance, here (and perhaps in most naturalistic situations), inconsistency led to a slight increase in trust. This could have been due to the perception of greater uncertainty per se (perhaps due to larger inconsistencies than in the previous highly controlled studies), or to the predominance of ascending and confirmed trends in this forecast set. Similarly, participants' increased cautiousness with inconsistent forecasts may have been due to the perception of greater uncertainty per se, or to the fact that inconsistent forecasts often included slightly higher snow total values. Thus, future work should test these effects with different naturalistic forecast data as well as manipulate them systematically in controlled studies, to verify these particular effects. Another issue that could be resolved in future research is whether the source of inconsistency matters. For instance, inconsistency could be due to capricious weather situations or to lack of expertise among forecasters which may impact some form of trust. Finally, it is important to note that this was a student sample. It is possible that greater experience or differences in education level might lead to slightly different results. However, recent evidence suggests that the ability to use probabilistic forecasts to make better decisions is similar among college students to a broader population (Grounds 2016; Grounds and Joslyn 2018). It is also important to note, that decisions in a controlled experimental environment such as this, differ in many respects to those made in real world situations in which other factors play a role and the decision consequences can be very serious.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Importantly, the main results reported here align with a growing body of highly controlled experimental research. We have shown here that, in line with previous research (Joslyn and LeClerc 2012; LeClerc and Joslyn 2015; Grounds, LeClerc, and Joslyn 2018; etc.), explicit numeric uncertainty estimates preserved trust in the context of naturalistic forecasts and outcomes, especially as inaccuracy increased. Probabilistic forecasts also allowed users to make better decisions from an economic perspective. In addition, the research reported here provides converging evidence that the effect of forecast inconsistency is not as problematic as once thought and may also confer some benefits upon forecast users. It is important to consider the impact of forecast inconsistency in the context of forecast inaccuracy as we have done here because there can be a tradeoff between them. Weather models tend to grow more accurate as lead times decrease. Therefore, the artificial maintenance of forecast consistency can be at a cost to accuracy. As shown previously in studies with highly controlled forecast stimuli (Burgeno and Joslyn 2020, Su, Burgeno, and Joslyn 2021) and here with naturalistic forecast data, inaccuracy is much more detrimental to trust than is inconsistency. This is true whether inconsistency is based on a single source (presented experiment, Burgeno and Joslyn 2020) or resides in multiple sources (Su, Burgeno, and Joslyn 2021). It is true whether forecasts are encountered sequentially (present experiment, Burgeno and Joslyn 2020) or simultaneous (Su, Burgeno, and Joslyn 2021). All of this evidence points in the same direction: Inaccuracy is far more detrimental to user trust than is inconsistency. In fact, much of this research suggests that inconsistency may be beneficial in that it provides useful information to decision makers. Based on this converging evidence, we recommend that forecast providers avoid artificially preserving consistency at a potential loss to accuracy. Updating forecasts and including well calibrated uncertainty estimates, can preserve trust in the information source as well provide users with decision-relevant information.

Acknowledgments.

This research was supported by the National Science Foundation under Grant 1559126. Special thanks go to David B. Radell at NOAA and the National Weather Service for providing us with the forecast data, and to Mengying Xu, Keiko Shannon, Justin Takeuchi, Yuan (Eva) Yin, and Brandy Steed for conducting data collection sessions.

Data Availability Statement.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

All data used and collected for this study is available at https://osf.io/dv6j8.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

APPENDIXES

Appendix A: Task Instructions and Training Trials

Scenario

You have been hired to work for a decision consultancy. Your project this winter is to consult with school districts faced with widespread snowstorms. Your job is to provide decision advice regarding whether they should close school for the day or stay open for class.

Schools are closed when driving conditions are unsafe to prevent accidents and injuries. However, school closures are expensive to the district because days must be made up at the end of the school year.

You will be provided with forecasts for each school area 2-days and 1-day in advance of a storm to help you make your decision. Due to microclimates across the regions, snow accumulation can differ from location to location; therefore, you will receive weather forecast information from a private weather service that specializes in local predictions.

There will be two periods for which winter storms are anticipated across two regions. For each storm, you will provide school closure advice for 65 schools located throughout the region. You will see a screen indicating the new period after school 65.

If you think the school area will receive 6 inches of snow or more, advise closing. If you think the school area will receive less than 6 inches, advise staying open.

Your boss gives cash bonuses to the members of the decision consultancy staff who offer the best advice. You will begin with **332 points**. It will cost 2 points every time you advise closing. It will cost 0 points if you advise to stay open. However, if you advise to stay open and 6 inches of snow or more is observed, then you will be **penalized 6 points**. Your goal is to give the best advice possible and retain as many points as you can. **You will receive a**

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

cash bonus if your ending balance is above 72 points.

Summary

When you expect **6 inches or more** of snow, you should advise the school to close.
When you expect **less than 6 inches** of snow, you should advise the school to stay open.

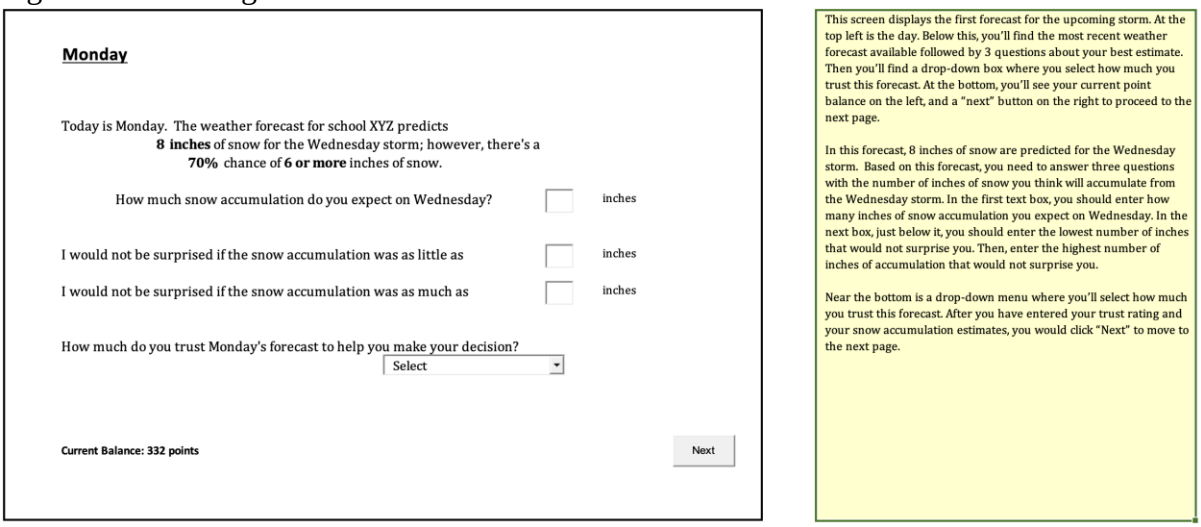
Cost to close schools: **2 points** to compensate for makeup days.

Penalty for staying open when 6 inches or more are observed: **6 points** to compensate for traffic accidents and injuries.

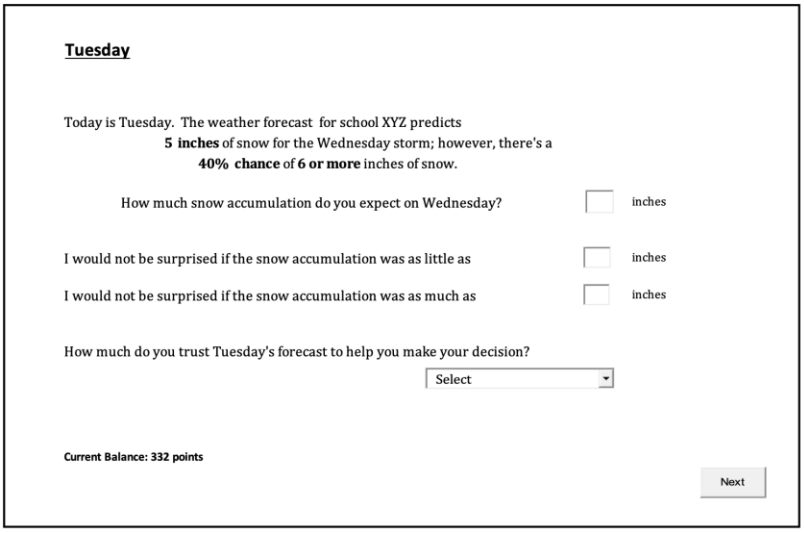
You will receive one dollar for every 32 points above 72 points at the end of the session.

You will now see several demonstration trials to help you understand your task. After those trials, you will begin making your own decisions. Your goal is to end up with the highest number of points possible.

Figure A1. Training Trials

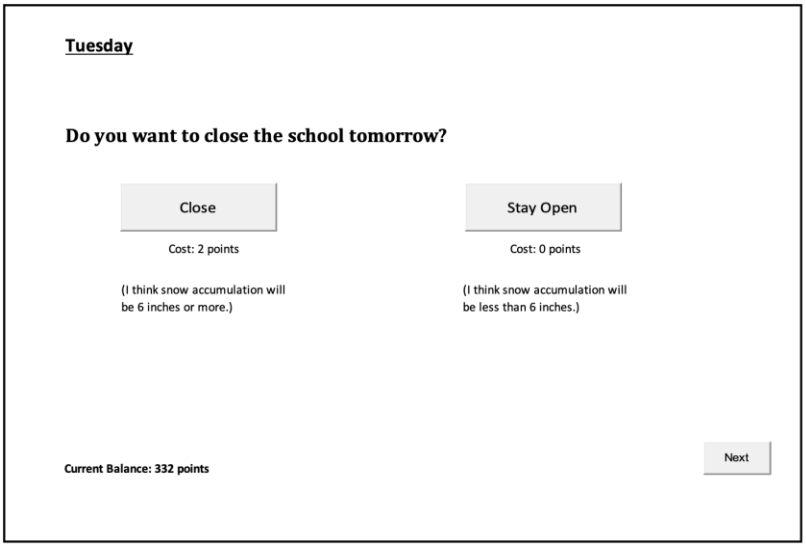


Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.



Notice that the day has changed from Monday to Tuesday. You will find the second forecast has updated to predict 5 inches for the Wednesday storm. Again, you would answer each of these questions.

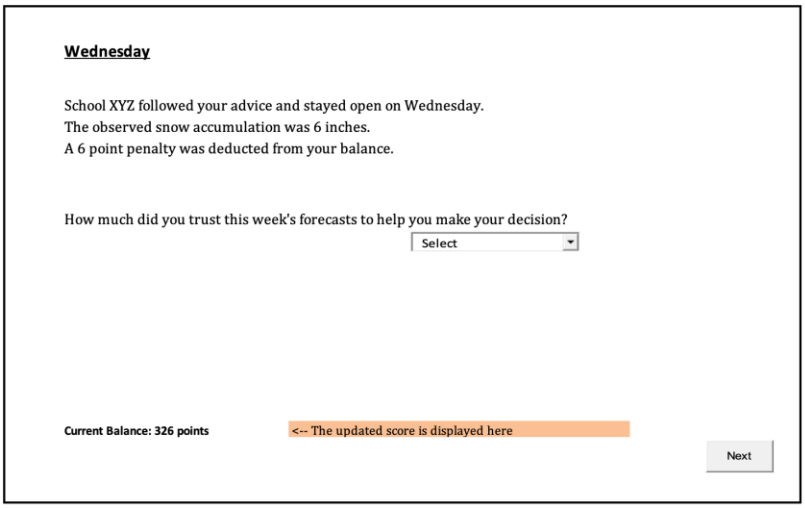
Notice that the previous forecast and this forecast are both for Wednesday.



It is now the night before the snowstorm. Now you must make a decision to close the school or stay open for classes based on the two previous weather forecasts that you've just seen. Remember that closing the school costs 2 points to cover the cost of makeup days, and staying open costs no points. However, if you choose to stay open and 6 inches or more of snow accumulates tomorrow, you will be penalized 6 points.

If you expect that the accumulation for Wednesday will be less than 6 inches, you should click the "Stay Open" button. If you expect 6 inches or more of snow, you should click the "Close" button.

In this example, the person performing these demonstration trials clicked the "Stay Open" button.



Notice it is now Wednesday and the snowfall has occurred.

As you can see to the left, the total snow accumulation for today was 6 inches. The person performing these example trials chose to keep the school open, but 6 inches of snow or more was observed and the person was penalized 6 points to cover the cost of equipment damages and injuries.

Notice the updated score in the bottom left reflects the 6 points the person lost.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Appendix B: Questions Asked on Each Trial

1. How much snow accumulation do you expect on Wednesday?____
2. I would not be surprised if the snow accumulation was as little as ____ inches
3. I would not be surprised if the snow accumulation was as much as ____ inches
4. How much do you trust Monday's *[or Tuesday's]* forecast to help you make your decision? [Response: 6-point scale from "not at all" to "completely"]
5. Do you want to close the school tomorrow?

   Response options:

   Close: cost 2 points (I think snow accumulation will be 6 inches or more.)

   Stay Open: cost 0 point (I think snow accumulation will be less than 6 inches.)

6. Trust rating taken after outcome is shown:
   How much did you trust this week's forecasts to help you make your decision?
   Response: 6-point scale from "not at all" to "completely"

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Appendix C: Features Preserved in the Forecast Set Compared to Historical Forecast Set

	Original Historical Forecast Data (N=160)				Final Experimental Forecast Data (N=130)			
	Min	Max	**Mean**	SD	Min	Max	**Mean**	SD
Observed Snow Accumulation	0	20	**4.46**	4.70	0	20	**5.08**	4.62
Deterministic Forecast 1 Values	0	11	**3.79**	3.26	0	17	**5.75**	4.50
Deterministic Forecast 2 Values	0	18	**5.83**	5.15	0	17	**6.11**	4.77
Forecast 1 Probabilities of 6"+	0	78	**26.06**	23.85	0	100	**40.13**	31.23
Forecast 2 Probabilities of 6"+	0	100	**37.78**	34.91	0	100	**40.13**	31.23
Inaccuracy of Deterministic Forecast 2	-8	9	**-1.37**	2.44	-6	10	**-1.02**	2.47
Absolute Value of (Forecast 2) Inaccuracy for Inconsistent* Deterministic Forecasts	0	9	**2.49**	1.75	0	10	**2.24**	1.89
Absolute Value of (Forecast 2) Inaccuracy for Exactly Consistent** Deterministic Forecasts	0	6	**1.18**	1.66	0	6	**1.89**	1.44
Proportion of All Cases:								
Exactly Accurate	21.9%				20.0%			
Threshold Crossing Inaccuracies	15.6%				19.2%			
Threshold Crossing Inconsistencies	17.5%				16.2%			

*Original inconsistent N=115; Final inconsistent N=66.

**Original exactly consistent N=45; Final exactly consistent N=64.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Appendix D: Probabilistic Forecast Calibration Procedure

A binning technique was used to examine the reliability of the probabilistic forecasts because there were very few cases at the same probability, precluding more conventional measures such as the Brier score (Brier, 1950). A bin was considered calibrated if the proportion of observed events with 6 or more inches fell within the probability range for that bin. For instance, Bin 2 ranged between 5-14% and contained 1 out of 10 trials (10%) in which 6 or more inches of snow accumulation was observed.

In the historical forecast data set, the proportion of outcomes at or above the 6-inch threshold was within a few percentage points of the bin boundaries in most cases. However, in the higher probability bins (65-74%,75-84%, and 85-94%), the proportions were as many as 25 percentage points higher than the upper bound of the bin suggesting a low bias in the forecasted probabilities for that day. Therefore, slight changes were made (cases were removed, duplicated, and/or the probabilities were modified) to perfect probabilistic forecast reliability while maintaining the basic characteristics of the historical forecast set. See Table D below.

Table D. Day 2 Forecasted and Observed Probabilities

Probabilistic Forecasts	Experienced Prob of 6"+	Original Data	Observed over 6"	Total Events	Binned real probabilities
0%	3.33%	2.27%	1	30	<5%
10%	10.00%	0.00%	1	10	5 to 14
20%	20.00%	7.69%	2	10	15 to 24
30%	30.00%	36.36%	3	10	25 to 34
40%	40.00%	40.00%	4	10	35 to 44
50%	50.00%	36.84%	5	10	45 to 54
60%	60%	64%	6	10	55 to 64
70%	70%	100%	7	10	65 to 74
80%	80%	100%	8	10	75 to 84
90%	90.00%	100.00 %	9	10	85 to 94
100%	100.00%	100.00 %	10	10	95 to 100
Total			56	130	

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Appendix E: Trust Analyses: GEE model Descriptions and Regression Tables by Hypotheses

Hypotheses

Hypothesis 1: Is Forecast Format associated with Trust?

Hypothesis 2: Is Inconsistency associated with Trust?

Hypothesis 3: Is Inaccuracy associated with Trust?

Hypothesis 4: Does the association between Inconsistency and Trust differ across Forecast Formats?

Hypothesis 5: Does the association between Inaccuracy and Trust differ across Forecast Formats?

Exploratory: Relationship between Inconsistency, Inaccuracy and Trust

Models

Table E1. Hypotheses 1, 2, and 3 are addressed by Model I with predictors Inaccuracy, Inconsistency, and Forecast Format.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inaccuracy	1.15	1.14	1.17	<.001
Inconsistency	.93	.92	.94	<.001
Forecast Format	.80	.65	.99	.04

Table E2. Hypothesis 4 is addressed by Model II with predictors Inaccuracy, Inconsistency, Forecast Format, and the Inconsistency by Forecast Format interaction.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inaccuracy	1.15	1.14	1.17	<.001
Inconsistency	0.96	0.94	0.97	<.001
Forecast Format	0.92	0.74	1.14	0.47
Inconsistency x Forecast Format Interaction	0.93	0.91	0.95	<.001

Table E2a. Model III with predictors Inaccuracy, Forecast Format, Inconsistency by Deterministic, and Inconsistency by Probabilistic interactions, was conducted to explore how the association between Inconsistency and Trust differed across Forecast Format.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inaccuracy	1.15	1.14	1.17	<.001
Forecast Format	0.92	0.75	1.15	0.47

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Inconsistency x Deterministic	0.96	0.94	0.97	<.001
Inconsistency x Probabilistic	0.90	0.88	0.91	<.001

Table E3. Hypothesis 5 is addressed by Model IV with predictors Inaccuracy, Inconsistency, Forecast Format, and the Inaccuracy by Forecast Format Interaction.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inaccuracy	1.17	1.14	1.19	<.001
Inconsistency	0.93	0.92	0.94	<.001
Forecast Format	0.85	0.68	1.06	0.16
Inaccuracy x Forecast Format Interaction	0.97	0.95	1.00	0.037

Table E3a. Model V with predictors Inconsistency, Forecast Format, Inaccuracy by Deterministic, and Inconsistency by Probabilistic interactions, was conducted to explore how the association between Inaccuracy and Trust differed across Forecast Format.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inconsistency	0.93	0.92	0.94	<.001
Forecast Format	0.85	0.68	1.07	0.16
Inaccuracy x Deterministic	1.17	1.14	1.19	<.001
Inaccuracy x Probabilistic	1.13	1.12	1.15	<.001

Table E4. Exploratory Model VI with predictors Inaccuracy, Forecast Format, and Inconsistency by Forecast Format interaction, was conducted to test whether the association between Inconsistency and Trust differs across Accuracy.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Inaccuracy	1.31	1.28	1.34	<.001
Inconsistency	1.06	1.04	1.07	<.001
Forecast Format	.80	.64	.99	0.04
Inaccuracy x Inconsistency Interaction	0.94	0.93	0.94	<.001

Table E4a. Exploratory Model VII with predictors Inaccuracy, Forecast Format, Inconsistency by High Accuracy and Inconsistency by Low Accuracy interactions, was conducted to explore how the association between Inconsistency and Trust differed across Accuracy.

Effect	Odds Ratio	95% CI	p

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

		LL	UL	
Inaccuracy	1.18	1.16	1.20	<.001
Forecast Format	0.80	0.65	0.99	.04
Inconsistency x High Accuracy	0.93	0.92	0.94	<.001
Inconsistency x Low Accuracy	0.80	0.78	0.82	<.001

Appendix F. Closure Decision Analyses: Binary GEE Model Descriptions and Regression Tables by Hypotheses

Hypotheses

Hypothesis 6a: Is Forecast Format associated with Closure Decisions?

Question 6b: Is Optimal Decision associated with Closure Decisions?

Question 6c: Does the association between Optimal Decision and Closure Decisions differ across Forecast Formats?

Question 6d: Does the association between Optimal Decision and Closure Decisions differ across Inconsistency?

Question 6e: Is Inconsistency associated with Closure Decisions?

Exploratory: Relationship between Optimal Decision, Forecast Format, and Closure Decisions

Models

Table F1. Hypothesis 6a and Questions 6b and 6e are addressed by Model I with predictors Optimal Decision, Inconsistency, and Forecast Format.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Forecast Format	0.95	0.85	1.08	0.43
Inconsistency	1.44	1.42	1.46	<.001
Optimal Decision	30.39	28.34	32.59	<.001

Table F2. Question 6c is addressed by Model II with predictors Optimal Decision, Inconsistency, Forecast Format, and the Optimal Decision by Forecast Format interaction.

Effect	Odds Ratio	95% CI		p
		LL	UL	

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Forecast Format	0.64	0.54	0.75	<.001
Inconsistency	1.44	1.42	1.46	<.001
Optimal Decision	22.74	21.22	24.36	<.001
Optimal Decision x Forecast Format	1.94	1.70	2.22	<.001

Table F3. Question 6d is addressed by Model III with predictors Optimal Decision, Inconsistency, Forecast Format, and Optimal Decision by Inconsistency

Effect	Odds Ratio	95% CI		p
		LL	UL	
Forecast Format	0.95	0.82	1.09	.43
Inconsistency	0.91	0.86	0.96	.001
Optimal Decision	18.2	16.83	19.68	<.001
Optimal Decision x Inconsistency	1.68	1.59	1.77	<.001

Table F4. Exploratory Model IV with predictors Forecast Format, Inconsistency, Optimal Decision by Deterministic Forecast Format, and Optimal Decision by Probabilistic Forecast Format, was conducted to test whether the association between Optimal Decision and Closure Decisions differed across Forecast Format.

Effect	Odds Ratio	95% CI		p
		LL	UL	
Forecast Format	0.64	0.54	0.75	<.001
Inconsistency	1.44	1.42	1.46	<.001
Optimal Decision x Deterministic	22.70	21.20	24.40	<.001
Optimal Decision x Probabilistic	44.20	39.40	49.50	<.001

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Appendix G: Linear Mixed Model Regression Tables for Expected Value, Uncertainty Range, and Best Estimate

Table G1. Regression table for Expected Value

Predictor	B	95% CI (B)	t	p
(Intercept)	3.33	[3.11,3.65]	24.25	<.001
Forecast Format	-0.28	[-0.67,0.11]	1.39	0.17
Inconsistency	0.62	[0.61,0.63]	88.23	<.001
Inconsistency x Forecast Format	0.09	[0.07,0.11]	10.37	<.001

Table G2. Regression table for Best Estimate

Predictor	B	95% CI (B)	t	p
(Intercept)	0.16	[0.05,0.27]	2.86	<.01
Forecast Format	-0.06	[-0.21,0.10]	0.72	.47
Forecast 2 Value	0.87	[0.86,0.87]	322.72	<.001
Inconsistency	-0.03	[-0.04,-0.02]	5.91	<.001
Forecast 1 Value	-0.08	[0.07,0.08]	30.11	<.001
Inconsistency x Forecast Format	0.01	[0.00,0.02]	1.97	.049

Table G3. Regression table for Uncertainty Range

Predictor	B	95% CI (B)	t	p
(Intercept)	3.33	[3.11,3.65]	24.25	<.001

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Forecast Format	-0.28	[-0.67,0.11]	1.39	0.17
Inconsistency	0.62	[0.61,0.63]	88.23	<.001
Inconsistency x Forecast Format	0.09	[0.07,0.11]	10.37	<.001

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

REFERENCES

Bernoulli, D., 1954: Exposition of a new theory on the measurement. *Econometrica*, **22**(1), 23-36.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**(1), 1-3.

Budescu, D.V., A.K. Rantilla, H.T. Yu, and T.M. Karaletiz, 2003: The effect of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior Human Decision Processes*, **90**(1), 178-194.

Burgeno, J. N., and Joslyn, S. L., 2020: The impact of weather forecast inconsistency on user trust. *Weather, Climate, and Society*, **12**(4), 679-694.

Earle, T.C., 2010: Trust in risk management: A model-based review of empirical research. *Risk Analysis*, 30(4), 541-574.

Falk, A., and F. Zimmermann, 2017: Consistency as a signal of skills. *Management Science*, **63**(7), 2197-2210.

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

Grounds, M., and Joslyn, S., 2018: Communicating weather forecast uncertainty: Do individual differences matter? *Journal of Experimental Psychology: Applied*, **24**(1), 18-33.

Grounds, M.A., LeClerc, J.E., and Joslyn, S., 2018: Expressing flood likelihood: Return period versus probability. *Weather, Climate and Society*, **10**(1), 5-17.

Hohle, S. M. and Teigen, K. H., 2015: Forecasting forecasts: The trend effect. *Judgment and Decision Making*, *10*(5), 416-428.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Hohle, S., and Teigen, K., 2018: When probabilities change: Perceptions and implications of trends in uncertain climate forecasts. *Journal of Risk Research,* 1-15.

Højsgaard, S., Halekoh, U., and Yan, J., 2006: The R package geepack for generalized estimating equations. *Journal of Statistical Software*, **15**, 1-11.

Joslyn, S. L., and LeClerc, J. E., 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied*, **18**(1), 126.

Joslyn, S., and Savelli, S., 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*, **17**(2), 180-195.

Kadous, K., Mercer, M., and Thayer, J., 2009: Is there safety in numbers? The effects of forecast accuracy and forecast boldness on financial analysts' credibility with investors. *Contemporary Accounting Research*, **26**(3), 933-968.

Kahn, B. E., and Luce, M. F., 2003: Understanding high-stakes consumer decisions: mammography adherence following false-alarm test results. *Marketing Science*, **22**(3), 393-410.

Lazo, J. K., Morss, R. E., and Demuth, J. L., 2009: 300 billion served: Sources, perceptions, uses, and values of weather forecasts. *Bulletin of the American Meteorological Society*, **90**(6), 785-798.

LeClerc, J., and Joslyn, S., 2015: The cry wolf effect and weather- related decision making. *Risk Analysis*, **35**(3), 385-395.

Losee, J. E., and Joslyn, S., 2018: The need to trust: How features of the forecasted weather influence forecast trust. *International Journal of Disaster Risk Reduction*, **30**, 95-104.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Maglio, S. J., and Polman, E., 2016: Revising probability estimates: Why increasing likelihood means increasing impact. *Journal of Personality and Social Psychology*, *111*(2), 141-158.

Morss, R. E., Demuth, J. L., and Lazo, J. K., 2008: Communicating uncertainty in weather forecasts: A survey of the US public. *Weather and Forecasting*, **23**(5), 974-991.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, *105*(7), 803-816.

NOAA, 2016: Risk communication and behavior: Best practices and research findings. NOAA Tech. Rep., 60 pp., https://www.performance.noaa.gov/wp-content/uploads/Risk- Communication-and-Behavior-Best-Practices-and-Research-Findings-July-2016.pdf.

NOAA's NCDC, 2019: *30 Years of Seattle Snow Accumulation Data (1989–2019)* [Dataset retrieved December 2nd, 2019]. NOAA's National Climatic Data Center. https://www.ncdc.noaa.gov/cdo-web

Pasquini, E. S., Corriveau, K. H., Koenig, M., and Harris, P. L., 2007: Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, **43**(5), 1216-1226.

Ronfard, S., and Lane, J. D., 2018: Preschoolers continually adjust their epistemic trust based on an informant's ongoing accuracy. *Child Development*, **89**(2), 414-429.

Savelli, S., and Joslyn, S., 2012: Boater safety: Communicating weather forecast information to high-stakes end users. *Weather, Climate, and Society*, **4**(1), 7-19.

Siegrist, M., Gutscher, H., and Earle, T.C., 2005: Perception of risk: The influence of general trust, and general confidence. *Journal of Risk Research*, **8**(2), 145-156.

Smithson, M., 1999: Conflict aversion: Preference for ambiguity vs conflict in sources and evidence. *Organizational Behavior and Human Decision Processes*, **79**(3), 179-198.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.

Su, C., Burgeno, J. N., and Joslyn, S., 2021: The effects of consistency among simultaneous forecasts on weather-related decisions. *Weather, Climate, and Society*, 1-30.

Touloumis, 2015: R Package multgee: A generalized estimating equations solver for multinomial responses. *Journal of Statistical Software*, **64**(8), 1-14. https://www.jstatsoft.org/v64/i08/

Tversky, A., and Fox, C.R., 1995: Weighing risk and uncertainty. *Psychological Review*, **102**(2), 269-283.

Tversky, A., and Kahneman, D., 1979: Prospect theory: An analysis of decision under risk. *Econometrica*, **47**(2), 263-291.

Twyman, M., Harvey, N., and Harries, C., 2008: Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. *Judgement and Decision Making*, 3, 111-120.

Wilson, L. J., and Giles, A., 2013: A new index for the verification of accuracy and timeliness of weather warnings. *Meteorological Applications*, **20**(2), 206-216.

Accepted for publication in *Weather, Climate, and Society*. DOI 10.1175/WCAS-D-22-0064.1.