This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Risk Research on 02/05/2018, available online: http://www.tandfonline.com/ 10.1080/13669877.2018.1459801

https://www.tandfonline.com/eprint/wwdbsZMk4nGmuxIMMRzg/full

When probabilities change:

Perceptions and implications of trends in uncertain climate forecasts

Sigrid Møyner Hohle[1] and Karl Halvor Teigen[1, 2]

[1]Simula Research Laboratory, Lysaker, Norway

[2]Department of Psychology, University of Oslo, Norway

Acknowledgement:

This work was supported by The Research Council of Norway under Grant No. 235585/E10.

Corresponding author:

Sigrid Møyner Hohle, Simula Research Laboratory, P.O. Box 134, 1325 Lysaker, Norway.

E-mail: sigrid@simula.no, Telephone: + 4795733060

Abstract

Past research has revealed a *trend effect* when people are faced with a revised probabilistic forecast: A forecasted event that has become more (vs. less) certain is taken to signal a trend towards even stronger (weaker) certainty in future revisions of the forecast. The present paper expands this finding by exploring the boundary conditions of the trend effect and how it affects judgments of the forecaster. In Study 1, the trend effect was shown to persist when receivers process the forecast more deliberately, by considering reasons for the revision. In Study 2, trend continuation was predicted even when the two forecasts were made by different experts at different points in time. Study 3 demonstrated that the effect disappears when receivers are given an earlier forecast disrupting the linearity of the trend (e.g., a 60%-70% sequence preceded by a 70% forecast). In Study 4, two forecasters were perceived as more in agreement when revising divergent probabilities in the same rather than in opposite directions. If the event occurs, a forecast with downgraded probability (e.g. from 50% to 40%) was judged to be less accurate than an equally uncertain single forecast (40%). These results demonstrate the robustness of the trend effect based on two forecasts, affecting not only receivers' expectations of what comes next, but also their perceptions of the forecaster and of forecast accuracy. The findings have implications for how people communicate and understand risks and other uncertain events in areas such as climate science, weather prediction, political science and medicine.

When probabilities change: Perceptions and implications of trends in uncertain climate forecasts

What people decide upon today is heavily dependent on their forecasts for tomorrow. Forecasts guide daily life planning like taking an umbrella to work, as well as political decisions such as the amount of money the government should spend on flood control. But as the future is inherently uncertain, risks and other forecasted events, such as those dependent on climate change, can only be expressed in terms of probabilities. A weather forecaster can estimate the probability of rain in Oslo this weekend to be 40%. Likewise, rather than stating categorically that tigers will or will not become extinct, the IUCN Red List of Threatened Species (2015) indicates a 20% probability that they will be extinct in 20 to 100 years ahead. Such probabilistic estimates may change over time, due to new information or new models for analyzing data. In fact, probabilities *should* not stay the same for long, as a characteristic of successful forecasters is their frequent revisions of forecasts (Tetlock & Gardner, 2016). Revisions can make the target event emerge as more or as less likely than before (Dessai & Hulme, 2004). Also, different experts may produce probability estimates approaching towards or diverging from each other.

Because such changes often occur, it is important to know how the public will make sense of them. What does it mean that a probability has increased or decreased? What do changes tell about the future as well as of the expert who produced them? For instance, whose forecast is more accurate when the rain starts pouring: the weather forecaster who first said there was a 40% chance of rain, and then revised it up to 60%, or one who adjusted an initial 80% estimate downwards to a 60% chance? In the present paper, we explore how people perceive such probability revisions, coming from a single or from different experts. The aim is to understand how revisions affect (a) receivers' expectations about future forecasts, and (b) their perceptions of the forecaster.

The trend effect

Previous studies have demonstrated the existence a *trend effect* in people's reception of forecasts that are revised from one point in time to another (Hohle & Teigen, 2015). After having read about an expert who has become more certain (vs. more uncertain) about a target outcome, people expected a future forecast to be even more certain (or uncertain). When they were told that the likelihood of a 3 ˚C rise in global temperatures by the end of the century had increased from 60% in a previous forecast to 70% in a more recent report, a majority of participants expected that probabilities would rise above 70% in the next forecast.

Conversely, when the forecast had decreased from 80% to 70%, future probabilities below 70% were predicted. The effect was also found for point forecasts, such that a prediction of 40 cm sea level rise was expected to increase if preceded by a lower prediction, but decrease if followed by a higher prediction. A recent study has replicated the trend effect for interval forecasts. A forecast where the interval for expected outcomes had widened was predicted to continue widening, while a narrowed interval was believed to become still narrower in the future (Løhre, 2017). Even single forecasts can suggest a trend, depending upon the way they are described. Probabilities described as being "now over X%" or "now almost Y%" were believed to have been upgraded from a lower level than before (Hohle & Teigen, 2017; Study 2). "More than X% chance" reflected a belief that the target event would occur, whereas "less than Y%" suggested less belief than warranted by the numerical probabilities involved (Study 3).

The trend effect is in line with numerous studies demonstrating that non-random sequences are expected to continue their current path, at least by participants drawn from a Western population (Ji, Nisbett, & Su, 2001). Trend continuation dominates whether sequences represent temperature changes or share prices (Lewandowsky, 2011), company sales (Harvey & Reimers, 2013) or growth rate of the world economy (Ji et al., 2001). In these cases, the trends may correctly reflect an underlying factor of growth or decline that can be projected into the future. It is more difficult to justify predictions of a forecaster's future certainty based on her past revisions. Participants who believe that the forecaster is on her way to become even more certain (or uncertain) imply that they know more about the forecaster's future estimates than the forecaster herself.

Trends created by revised probabilities may affect actual behavior. People were willing to donate more money to research on a cancer type whose death risk has increased rather than decreased (Erlandsson, Hohle, Løhre, & Västfjäll, 2017), and preferred a product whose risk of a defect was adjusted downwards compared to upwards (Maglio & Polman, 2016). Importantly, in all these studies the current forecasts were the same, only the previous forecasts differed by being higher or lower than the most recent one. The past risk level thus shapes the urgency with which the present risk is perceived.

Revisions can also affect judgments of the forecaster. In a study of interval forecasts, Løhre (2017) found that upward revisions of the magnitudes of expected outcomes increased trust in the forecaster. However, we are not aware of studies examining the effects of revised probability estimates. How are forecasters perceived when they have become more or less

certain? Change may by itself be interpreted as a sign of uncertainty, lack of confidence or ignorance, but could also indicate an increased understanding of the processes involved.

Understanding the effects of changed probabilities on trust in the expert is central in the area of climate science. Trust is a key component of environmental risk perceptions and use of climate information (Peters, Covello, & McCallum, 1997; Weber, 2010), and an important determinant for public acceptance or rejection of scientists' risk assessments (Slovic, 1999).

The present research

The present studies aimed to explore the robustness and the boundary conditions of the trend effect. We examined 1) whether the effect holds in presence of additional information and more elaborate processing, and 2) its effects on evaluations of the forecast and the forecaster.

When an uncertain forecast is revised from one point in time (T1) to another (T2), a receiver may adopt one of three main strategies in predicting future forecasts. One might take the last forecast as the current "best guess", and accordingly rely on the T2 probability as a default value for future estimates (no further change). Alternatively, the difference between the estimates might be (partly) due to variability, leading to a future estimate in the range between T1 and T2 (averaging, or trend reversal). Finally, the change from T1 to T2 can be viewed as evidence of a dynamic trend, with future forecasts expected to be revised further in the same direction, as evidenced by the trend effect (trend continuation). In the present paper, we ask whether task differences (in number of forecasts, forecasters, and processing requirements) affect which of these three main strategies is being used.

In four studies, participants were given one or more probabilistic forecasts about climate-related events (agricultural productivity, the onset of El Niño, and the opening of a new field for oil drilling). Study 1 examined the role of deliberation on prediction strategy, by making respondents provide reasons for their estimates and thus process the forecasts more thoroughly before predicting future forecasts. The role played by the source of forecast was investigated in Study 2, where previous forecasts coming from the same forecaster was compared to forecasts from a different forecaster. Both studies showed trend continuation to be the dominant strategy with only two forecasts. Study 3 explored the effects of additional, even earlier forecasts disrupting a monotonic trend.

We further asked, in all studies, participants to evaluate the forecaster. In Study 4, they judged the perceived agreement between expert pairs who have revised their forecast in

either similar or opposite directions. Participants in this study were also assessed the accuracy of a revised forecast in retrospect, when the actual outcomes were known. The hypothesis to be tested was whether probabilities that had been revised “in the right direction” were considered more accurate than those that had been changed in the opposite direction or not revised.

Study 1: Reasons for forecast change

In absence of other information, the most recent forecast should be regarded as the forecaster’s updated best guess, thus the most rational strategy would be to expect similar forecasts in the future. Instead, people use the change as indication of a growing or declining trend (Hohle & Teigen, 2015), or an event becoming less or more remote (Maglio & Polman, 2016). This trend effect suggests as a dynamic view of changes in certainty, as “on its way” to be continued beyond the most recent forecast. However, it could also be a result of superficial processing of a task, where participants see two consecutive numbers (e.g., 60% and 70%) and use them simply to predict the third, without further thoughts about the meaning of probabilities and possible reasons for a probability change. This might be regarded as an instance of so-called Type 1 processing, which is fast, automatic and effortless, in contrast to the deliberate and slow Type 2 processing (Kahneman, 2003; Stanovich & Toplak, 2012). Techniques that make a task more difficult or disfluent are found to reduce reliance on simple heuristics and enhance analytic reasoning (Alter, Oppenheimer, Epley, & Eyre, 2007).

Study 1 was designed to test whether more elaborate processing would make people less disposed to see a forecast revision as a trend. The study was designed as a modified replication of Hohle and Teigen’s (2015) Study 3, which described an expert’s changed certainty in estimates of future grain productivity in Norway as a consequence on climate change. To encourage a more elaborate and analytic approach, they were this time required to produce potential reasons for the original estimates before suggesting the most likely future estimate by the same expert. This was to discourage spontaneous and superficial responses, and promote a more reflective mindset. Providing reasons is one aspect of accountability, which has been shown to attenuate people's tendency to biased judgments (Lerner & Tetlock, 1999). To explore how revisions would affect perceptions of the expert, we also asked participants to rate the expert on several measures. Personal beliefs about climate change were in the original study found to affect expectations of future forecasts, such that belief in climate change was associated with endorsing trends consistent with more future climate

change (Hohle & Teigen, 2015). A measure of individual belief in climate change was therefore included also in Study 1, to test whether belief would affect predicted forecast and also possibly perceptions of the forecaster.

Method and material

Participants were 111 students at Lillehammer University college (82 women and 29 men, median age 20 years) attending a lecture in introductory psychology. Data collection took place exactly one year after the one reported by Hohle and Teigen (2015), with a new cohort of students taking the same class as on the previous occasion. They were randomly allocated to two conditions (increasing versus decreasing probability) by receiving two versions of the same basic questionnaire.

Participants were told that the expert Randi Rugstad had made projections of future grain productivity in Norway by the year 2100, based on extant models of climate change. This scenario was chosen because of its ambiguity. Unlike many other effects of climate change, agricultural productivity might increase, due to a rise of temperatures, or decrease, as a consequence of extreme weather conditions (drought or floods). Rugstad had written two different reports, the first issued in 2003 and the second in 2013. In the increasing condition, she estimated in her first report a 50% probability that the future productivity would equal or exceed 500 kg/decare, and in the second she estimated this level of productivity would be reached with a 60% probability. In the decreasing condition, her estimates went down from 70% probability in 2003 to 60% in 2013. Participants were asked to write up to three possible reasons for her revised estimates. They were then asked about what they thought her next estimate would be, in a future report to be due in 2023.

The questionnaire was identical to the one used by Hohle and Teigen (2015), except for the question about reasons. Participants were also asked to assess how much the expert had changed her opinions, her degree of uncertainty, perceived competence and trust in her future prognoses, all on 7-point rating scales. To test a possible effect of personal belief in climate change, participants rated their agreement with two items on scales ranging from 1: strongly disagree to 7: strongly agree: “I am certain that climate change occurs” and “Statements about human activity changing the climate are exaggerated” (reverse coded).

Results

Next forecast. Participants produced on average two reasons for Rugstad’s estimates (2.19 in the increasing vs 1.83 in the decreasing condition, $t(109) = 2.057$, $p = .042$). The

search for potential reasons did not change the trend effect. Mean estimates of the next forecast were 64.6% (*SD* = 13.0) in the increasing condition vs. 52.0% (*SD* = 8.6) in the decreasing condition, $t(92.81) = 5.794$, $p < .001$, $d = 1.20$. These estimates are similar to those in Hohle and Teigen's (2015) original study, with a large majority continuing the trend in both conditions, as shown in the two upper panels of Table 1.

Expert evaluation. Participants in the decreasing condition perceived Rugstad as more uncertain than in the increasing condition, $M_{decreasing} = 4.68$ vs. $M_{inceasing} = 3.91$; $t(106) = 2.37$, $p = .019$, $d = 0.46$. There were no significant differences in perceived opinion change, competence or trust in the expert (all *ps* > .27).

Climate change belief. Perhaps surprisingly, climate change beliefs were higher in the decreasing condition, as indicated by the average score of the two climate belief items, $M_{decreasing} = 5.69$ vs. $M_{increasing} = 5.21$; $t(106) = 2.79$, $p = .037$, $d = 0.41$. In the downward condition, belief in climate change was positively associated with predicted future forecast ($r = .464$, $p = .001$). Beliefs were not significantly correlated with evaluations of the expert.

< Insert Table 1 about here >

Discussion

We sought in this study to encourage a more deliberate processing of the revised forecasts by making participants think about reasons for the revisions. Yet, a majority of participants answered in line with trend continuation, and results were very similar to the original study (Hohle & Teigen, 2015), where no reasons had been provided. Hence, deliberation did not remove the trend effect, and the effect cannot simply be understood as a mindless continuation of a number sequence. Unexpectedly, climate change beliefs were stronger in the decreasing condition. This effect did not appear in the original study (Hohle & Teigen, 2015), so it is possible that the search for reasons created an effect of revision direction on belief, by making participants focus on reasons suggesting more detrimental impacts of climate change. This topic will not be further explored in the present paper.

Participants were not instructed to produce a specific type of reasons. It is possible that some types of reasons could have changed prediction strategy. Maglio and Polman (2016) found that when participants were explicitly told that a forecast would not be revised again, events with increasing chances did no longer feel closer than events with decreasing

chances. If participants were prompted to provide reasons why a revision may be unique, they might not have expected future changes to a similar degree.

The expert was explicitly rated as being more certain after upward compared to downward revisions, even though the final probability of productivity was the same (60%) in both conditions. Thus, certainty was judged relative to the expert's past estimates, rather than to her current beliefs, as expressed in probabilistic terms.

Study 2: Diverging experts

Study 1 focused on how perceptions of an expert's forecast were affected by information about the same expert's previous forecast(s). But because expert disagreement is common within many scientific domains, people will sometimes receive diverging forecasts coming from different sources rather than from the same expert. Will such forecasts made at different times, by different experts, still be interpreted as a trend? Or will respondents in this case be more prone to perceive the difference as expressions of disagreement or variability, and expect future estimates to be less extreme than the most recent forecast?

All participants in Study 2 received a Scientist A's most recent probability estimate about a future event. They were told in one condition (A1) what the same expert said about the issue one month ago, and in two other conditions what another Scientist B said either one month ago (B1), or what Scientist B says now (B2). They then were asked to predict Scientist A's next forecast (one month ahead).

We expected that in Condition A1, trend continuation would dominate, replicating the trend effect. The diverging forecasts made in Condition B2 would rather be perceived as disagreement or variability, and lead participants to expect a future estimate between the two (averaging A's and B's estimates). In Condition B1, participants might have a choice between seeing the difference as a trend (expecting trend continuation) or as variability (leading to an expectation of trend reversal).

Method and material

After excluding one participant who failed a simple attention check, 250 US residents recruited from Amazon Mechanical Turk remained for analysis; 133 males, 115 females and 2 others. The mean age was 36.1 (*SD* = 19.7), and 80.8% had at least some college education.

Participants were told that two scientists, A and B, give regular forecasts about the chance of El Niño developing during the upcoming season. Because they use slightly different models and their interpretations of the model output may vary, these experts may

give different forecasts. Scientist A stated in May 2017 that: “It is 64% likely that El Niño will develop during the fall”. Participants were told they could access one additional forecast: Scientist A's forecast from one month ago (April 2017), Scientist B's forecast from one month ago (April 2017), or B's most recent forecast (May 2017), and were asked to rank them according to relevance by indicating which forecast they would prefer the most and the least to see.

Participants were then randomly assigned to three conditions, receiving either A’s previous forecast (Condition A1), B’s previous forecast (Condition B1) or B’s most recent forecast (Condition B2), all estimating a 52% chance of an El Niño this fall. Participants were then asked to predict Scientist A's next estimate. On the next page, they answered “How certain do you feel that [probability given in answer to last question] will be Scientist A's next forecast?”, on a scale from 1: very uncertain to 7: very certain.

All participants rated their agreement with statements about the perceived uncertainty, confidence and competence of Scientist A, trust in Scientist A, the degree to which scientists have improved in predicting El Niño, and the degree to which politicians should take their forecasts into account. Participants in Conditions B1 and B2 also rated whether “Scientists A and B seem to disagree”. All items were rated from 1: strongly disagree to 5: strongly agree.

After answering some unrelated questions, participants’ beliefs in human caused climate change were assessed with two statements: “I am quite sure that climate change is occurring now” and “Climate change is merely a natural fluctuation, not caused by human activity”, rated on a scale from 1: strongly disagree to 5: strongly agree. After reverse coding the second statement ratings were positively correlated ($r = .64$, $p < .001$) and were averaged into one belief in climate change score.

Results

Forecast preference. A majority of participants rated the most recent forecast from the other expert, Scientist B, as most relevant (79.6%), whereas B’s previous forecast was the least relevant to know (51.6%). When forecasts were ranked from 1-3, participants preferred on average B’s most recent forecast, $M = 1.29$ ($SD = 0.59$), followed by A’s previous forecast, $M = 2.24$ ($SD = 0.70$) and B’s previous forecast, $M = 2.49$ ($SD = 0.58$). A repeated ANOVA indicated a difference in preference, $F(1.86, 444.35) = 165.14$, $p < .001$, $\eta^2_p = .41$, with all conditions differing from each other ($p < .004$).

Next forecast. Trend continuations (guesses above 64%) dominated when A’s first forecast had been revealed, and also, to a slightly less extent, when B’s first forecast was

revealed, as summarized in Table 1. Guesses above 64% were less popular when B's most recent forecast was revealed, suggesting mostly averaging in this condition. Mean estimates of future prediction were highest in Condition A1 (M = 67.1%) and B1 (M = 65.8%) and lowest in Condition B2 (M = 61.2%). The average guesses differed significantly from 64% in all conditions (ps < .003). A one-way ANOVA showed an effect of condition on predicted future forecast, $F(2, 247) = 14.67$, $p < .001$, $\eta^2_p = .11$, and Bonferroni post hoc tests showed that the predicted forecast was lower for the B2 condition than for both condition B1 (p < .001) and A1 (p < .001), indicating less trend continuation in this condition. The difference between A1 and B1 was not significant (p = .843).

Thus, participants tended to believe that the next forecast would be more certain than the two previous ones, even when the first forecast had been produced by a different expert. Only when divergent forecasts were produced at the same time they suggested that the next forecast would be lower than the highest of the two.

In contrast to our hypothesis, participants' confidence in their predicted forecast did not differ by condition.

Expert evaluation. The perceived confidence of Scientist A differed between conditions, $F(2, 247) = 3.041$, $p = .050$, $\eta^2_p = .02$. Mean ratings suggested slightly higher confidence in B2 (M = 3.98) compared to A1 (M = 3.71) and B1 (M = 3.72), but none of the differences were significant with Bonferroni post hoc tests (p > .095). None of the other ratings of the scientist differed significantly across conditions (p > .14).

Climate change belief. The average score on the two climate change belief items indicated a strong belief in human-caused climate change (M = 4.00, SD = 1.10). Belief scores correlated weakly with predicted future forecast (r = .13, p = .042), and with the expert ratings: improvement of scientists, perceived confidence, trust, perceived competence, and importance of forecasts (r's = from .16 to .34, p's from < .001 to .004).

Discussion

When told about one expert's recent forecast, a recent forecast by another expert was considered more relevant to know than previous forecasts by the same or a different expert. This is arguably a good choice: both forecast recency and divergent source would be expected to give the best information overall. But even if the second expert's older forecast was judged to be least relevant it affected participants' beliefs about the target expert's future forecasts. That is, when expert A was more certain about a forecast than expert B was in the past, participants expected A to become still more certain in the future. Forecasts were only

averaged when both were equally recent and the difference between them could not be construed as a trend.

Study 3: Disrupting the trend

Linear trends are easy to detect and extrapolate when there are only two forecasts, as a straight line can always be drawn connecting two numbers. A single probability revision tells a compelling story: the event has become more likely or less likely than before. Two or more revisions might complicate the story by changes going up and down. Study 3 was designed to investigate receivers' responses to probabilistic forecasts made at three different points of time, which did not constitute a uniform trend. We tested whether participants would predict trend continuation based on the two most recent revisions, or whether a still earlier forecast might disrupt the trend.

Method and material

Participants were 125 unpaid volunteers recruited on campus at the University of Oslo, 71 women and 53 men (one did not indicate gender), median age 23 years. They were randomly assigned to three conditions A (n = 41), B (n = 54), and C (n = 30)[1].

All participants received questionnaires with the grain production scenario described in Study 1, except that the expert had made two prior forecasts. In Condition A, her first prediction was made in 2003, with 60% probability. In 2009, she gave a 70% probability for the same outcome, and in her most recent prediction, from 2015, she was back to 60%. In Condition B, the corresponding probabilities were 60%, 50%, and 60%. The reports were presented in chronological order. Participants in Condition C were also told that she had issued several reports, but received only her most recent probability estimate (60%).

Following the procedure of Study 1, and to avoid superficial responses, participants in Conditions A and B were then invited to write up to three reasons for why her probability estimates had changed. In Condition C, they suggested reasons for her most recent estimate only.

On the next page, all participants were asked to assess the degree to which they felt she had changed her opinion about future grain productivity and whether she appeared uncertain (on 1-7 rating scales). They were further asked to predict her next probabilistic

[1] The unequal numbers were due to a typo in a subset of the C questionnaires, which made one question unanswerable. These questionnaires were accordingly discarded.

forecast, which was scheduled to take place in 2021. This was followed by two Likert scales intended to measure the expert's perceived expertise: "I will feel confident in Rugstad's future forecasts", and "Rugstad appears competent", and two climate belief scales: "I am certain that climate changes occur", and "Claims about human activity causing climate changes are exaggerated" (reverse coded).

Participants in Condition C were in addition asked to indicate the perceived relevance of previous reports. If Rugstad's previous reports had contained different probability estimates than her current one, would this have played a role for their evaluations of her most recent report? Answers were given on a five-point scale from 0: not at all, to 4: very much.

Results

Next forecast. According to the two most recent forecasts, the expert in Condition A was becoming less certain and the expert in Condition B was becoming more certain than they were previously. From Study 1, one might expect these trends to continue into the future. However, these potential trends were cancelled by an even earlier, incompatible estimate. As a result, the projected future estimates were, on average, close to the expert's most recent estimate in all conditions (M = 59.2%, 56.1%, and 58.3%, in conditions A, B and C, respectively; $F(2, 122) = 0.973$, $p = .38$). However, a closer inspection of data reveals that estimates that equal 60% are modal responses only in Condition A. Frequency counts (bottom panel in Table 1) show that a majority of participants in all groups thought that the expert's next forecast would differ from 60%, but they disagreed about the direction of change. In Condition B, where the most recent forecast was higher than the previous one, most participants guessed that the next forecast would go back to 50%, in contrast to the trend effect. In Condition C, exact 60% responses were even less popular than in the other conditions, despite the fact that this was the only estimate these participants had received.

Expert evaluation. As shown in Table 2, the expert who gave only one estimate was believed to have changed her opinion more, and to be more uncertain than experts giving three forecasts; there was also a tendency for her to be less trusted.

Perceived relevance. Participants in Condition C did not think that another expert's previous estimates would play an important role for their evaluations. Answers ranged from "not at all" (0) to "very much" (4) with $M = 2.00$ ($SD = 1.08$).

Climate change belief. The combined climate change score indicated a strong belief in climate change ($M = 5.96$, $SD = 1.11$). Belief was related negatively to perceived uncertainty ($r = -.23$, $p = .003$) and positively to competence of the expert ($r = .22$, $p = .004$).

< Insert Table 2 about here>

Discussion

The present results show that when a third, inconsistent forecast predates the two more recent ones, the most recent revision is no longer construed as a trend, and predictions for future forecasts become more varied. One could therefore argue that this added piece of information made participants more rational, no longer extrapolating a trend from just two data points. If the three forecasts had formed a linear trajectory (e.g., 50%, 60%, 70%), it is more likely that the trend effect would have persisted. It is not the presence of an extra forecast that disrupts the effect, but its incompatibility with a coherent trend.

Somewhat surprisingly, the expert with one estimate was seen as more uncertain and slightly less trustworthy than those with several different previous estimates, perhaps because her single estimate of 60% must be regarded as an indeterminate value, quite far from certainty. In the other two conditions, certainty can also be inferred from the pattern of estimates over time, which did only vary with 10 percentage points from one occasion to the next. From this we can perhaps conclude that even uncertain estimates can give an impression of reliability and trustworthiness when they are repeated more than once.

Study 4: Agreement and post-hoc accuracy

The previous studies focused on the way people perceive probabilities revised by a single expert, or made by different forecasters at different points in time. But how will two forecasters be judged if both revise their forecasts, in the same or different directions? The trend effect suggests that changes in opposite directions induce larger differences over time. Two forecasters may both suggest that El Niño is 60% likely, but might still be perceived to differ if one of them has upgraded her estimate from a lower level, whereas the other was more certain before.

The trend effect could also influence perceived accuracy of a forecast. Previous research suggests that judged accuracy is not merely a function of probability level. One factor is congruency. In a study by Yeung (2014), participants received a prediction framed either as “there is a 70% chance that the proposition will pass” or “there is a 30% chance that the proposition will fail”. While these frames are logically equivalent, they differ by highlighting two different outcomes. When asked to indicate the accuracy of the prediction, participants found “70% chance of passing” more accurate if the proposition had passed,

while “30% chance of failing” more accurate if it failed. Thus, perceived accuracy was highest when the qualitative component of the statement matched the outcome.

In a related manner, Teigen (1988) found that a speaker saying an outcome was “possible” was judged as more correct if the event took place, than one who said “it is not quite certain” — although the second statement was deemed to indicate a higher probability than the first one. This has been explained as a *directionality* effect (Teigen & Brun, 1995, 1999). “It is possible” has a positive directionality; the speaker is emphasizing that the event may happen, and is thus on the right track despite a moderate probability. “Not quite certain”, is negative, and is accordingly pointing in the wrong direction. Similarly, increased probabilities suggest that forecasters are “on their way” towards 100% (X will occur), whereas decreased probabilities indicate a direction towards zero (X will not occur). Forecasts may therefore appear accurate to the extent that they have been revised in the “right” direction. Study 4 was designed as a test of this assumption.

Methods and material

US participants were recruited from Amazon Mechanical Turk to participate in an experiment with two parts. After excluding 12 participants who spent less than 1 minute on the survey or failed the attention check, and one participant who reported an age of 314, 328 participants remained for analysis, of which 153 identified themselves as women, 173 as men and 2 as other. A majority (81%) had at least some college education. The mean age was 35.5 years (*SD* = 11.9). All participants received two scenarios, and rated one item per scenario,

Part 1: Agreement. Participants were randomly assigned to three different conditions and told about two forecasters’ estimates about the probability that there will be oil drilling in a fictitious location. The experts, A and B, had both submitted two probability forecasts, one a year ago and one now. In the Same Direction condition, both had revised their probabilities upwards (A from 30% to 40%, and B from 45 to 55%). In the Opposite Direction condition, one had revised his probability upwards, the other downwards (A from 50 to 40% and B from 45 to 55%). Participants in a Control condition received only the most recent forecasts (A: 40%, B: 55%), while the previous forecasts were unknown. All participants rated agreement between experts, from 1: A and B disagree completely to 7: A and B agree completely.

Part 2: Post hoc accuracy. On the next page (separate screen), participants were randomly assigned to four new conditions and read a scenario describing a research institute’s estimate of El Niño developing in the fall. The likelihood was in all conditions 40% for the most recent forecast (April). In March, the month before, the chance had been

estimated to be 30% (Increase condition), 50% (Small Decrease condition), 60% (Large Decrease condition), or not specified (Control condition). All participants were told that El Niño actually developed in the fall, and rated "How accurate was the research institute?", from 1: completely mistaken to 7: completely accurate.

Belief in climate change. After responding an unrelated task, participants' beliefs in climate change were measured using an adapted version of Heath and Gifford's (2006) belief in global warming and perception of that global warming is caused by humans scale. The original phrasing "global warming" was changed to "climate change" because the partisan divide on climate change has been found to be larger under a "global warming" frame than under a "climate change" frame (Schuldt, Konrath, & Schwarz, 2011). Participants rated their agreement with eight statements on a 5-point Likert scale (example item: "Climate change is occurring now"). The scale had a satisfactory reliability ($\alpha = .94$). Thus, a mean scale was computed, with high values indicating high belief in manmade climate change.

Results

Part 1: Agreement. Perceived agreement was higher for same-direction than for opposite-direction revisions, as shown in Figure 1. A Welch one-way ANOVA[2] indicated that the effect of condition was significant, $F(2, 210.61) = 33.51$, $p < .001$, est. $\omega^2 = .17$. Planned contrasts showed higher agreement in the Same Direction than in the Opposite Direction conditions, $t(217.84) = 6.348$, $p < .001$, and in Control compared to the Opposite Direction conditions, $t(196.68) = 7.875$, $p < .001$, but no difference between the Same Direction and the Control conditions ($p = .64$). The bars in Figure 1 indicate that probabilities of 40% and 55% are in moderate agreement if both were upgraded from lower values, but in disagreement (ratings below 4) when the smaller of the two had been adjusted downwards from a previously higher value.

<Insert Figure 1 about here>

Part 2: Post hoc accuracy. When participants were told that El Niño had actually arrived, the perceived accuracy of the forecasts was affected by direction of change (see Figure 2), $F(2, 178.88) = 5.24$, $p = .002$, est. $\omega^2 = .037$[2]. Planned contrasts showed no difference between the Large and Small Decrease conditions, so these two conditions were compared jointly to the other conditions. Perceived accuracy was lower in the two decrease conditions than in Increase ($p = .013$), and lower in decrease

[2] Welch's adjusted F ratio is reported because the assumption of homogeneity of variance was not met.

conditions than in Control ($p < .001$), while Control and Increase did not differ ($p = .101$).

<Insert Figure 2 about here>

Climate change belief. Belief in climate change was high ($M = 4.01$, $SD = 1.00$), and not significantly correlated with perceived agreement or post hoc accuracy.

Discussion

This study indicated that perceived agreement between forecasters is a product not only of their most recent probability level (which was the same in all conditions), but also of their previous ones. Forecasters who revised their estimates in opposite directions produced sets of estimates that were, on average, more similar to each other, but were perceived to disagree more than those whose forecasts had been revised in the same direction or not at all. However, revising forecasts in the same direction did not lead to higher agreement scores than no revision. Perhaps the expected increase in agreement from parallel revisions was cancelled out by the fact that forecasters in this condition differed (in absolute terms) from each other not just once, but twice.

Previous forecasts also affected accuracy judgments. Downward revisions made the forecast appear less correct than the same forecast after upward revisions or as a single estimate. It did not matter whether the downward revision had been large or small. A similar finding was reported by Maglio and Polman (2016), who also found that large and small downward changes in probability estimates affected perceived uncertainty to about the same extent. Revisions going in the “wrong” direction may be considered more deviant and attract more attention than those that have become more “right”, in line with other negativity effects in judgment (for an overview, see Alves, Koch & Unkelbach, 2017). In this particular case, to miss a possibly disastrous El Niño (a false negative) may be worse than to incorrectly predict it (a false positive). Future research could explore if increased uncertainty looms larger than increased certainty also for topics other than for this risky climate-related event.

General discussion

The present studies support and expand the findings of a trend effect in lay people’s perceptions of changed probabilities. Replicating previous research (Hohle & Teigen, 2015; Maglio & Polman, 2016), forecast revisions were interpreted as trends that were expected to continue in the future. Trend continuation remained the dominant strategy even for participants who were prompted to think about reasons why the probabilities might have

changed (Study 1), and when the different forecasts stemmed from different experts (Study 2). However, the effect disappeared when an even more remote third forecast was introduced disrupting the trend (Study 3). The findings indicate that experts' past probability estimates also affect judgments of their present uncertainty, the accuracy of the forecast, and of agreement between different experts (Study 4).

The trend effect in forecasts implies that two consecutive probability estimates, like other numbers in a series (e.g., yearly population statistics), are viewed as expressions of an underlying pattern. They are not just independent summaries of the forecasters' current knowledge, but are conceived as a dynamic change with its own progression or momentum (Markman & Guenther, 2007), leading towards greater certainty, or increasing doubt, as the case might be. The difference between two forecasts, or risk estimates, given at different points in time, thus contain an implicit message in addition to the explicit numbers. As probabilities for single events are difficult to interpret and validate, people may pay more attention to the direction of change than to the numeric estimates, as conveying qualitative information or the *gist* (Reyna, 2004; Reyna & Brainerd, 1995) of the message. They then can base their guesses of what the future has in store upon this gist. Thus, revised probabilities appear to belong to a class of framed messages that comes with dynamic implications of what will happen next, in much the same way as directional verbal probabilities (Teigen & Brun, 1995), choice of lower vs. upper bound of an uncertainty interval (Hohle & Teigen, 2017), use of congruent vs. incongruent outcome terms (Yeung, 2014), and estimates of the likely vs. the unlikely alternative in binary prediction tasks (Bagchi & Ince, 2016).

For the trend effect to occur, an earlier forecast has to be available. In many real-life situations probabilistic forecasts become unavailable once they have served their purpose or have been replaced with more recent ones. However, even if lay "consumers" of forecasts cannot keep track of all such changes, they are often a central feature of risk communication in the media. Even small probabilities, for instance, for acts of terrorism, international conflicts, or natural disasters, are often reported in terms of whether the risks have increased or decreased, implying that the change itself contain an important message over and beyond the current probability level.

When experts change and disagree

Earning and maintaining public trust is pivotal for scientists and institutions communicating with the public (Myers et al., 2016; Poortinga & Pidgeon, 2003). Perceived

consensus is one dimension on which the public judge experts, sometimes with serious consequences. Information about expert disagreement on environmental problems weakens public belief in the problem and support for policy (Aklin & Urpelainen, 2014). Conversely, highlighting the strong consensus among climate scientists (Cook et al., 2016) increases acceptance of human-caused climate change and support for public action (Cook & Lewandowsky, 2016; Lewandowsky, Gignac, & Vaughan, 2013; van der Linden, Leiserowitz, Feinberg, & Maibach, 2015). While the effects as well as the causes of expert disagreement are widely studied (Hammond, 2000; Mumpower & Stewart, 1996; Shanteau, 2000), much less is known about which factors that shape perceived agreement. In studies of its effect on the public, consensus is often defined as a high proportion of experts agreeing versus disagreeing on a topic (Lewandowsky, 2011; van der Linden et al., 2015). But not all predictions can be expressed dichotomously as belief or disbelief, and disagreement may therefore take other forms than the proportion of experts on either side of a dispute. How does the public judge agreement between experts assigning different probabilities? While it may be difficult to judge whether two experts whose probability forecast diverge by, for example, 15 percentage points agree or disagree, their direction of change is easier to judge. Agreement between two forecasters might therefore be judged by the congruence between their forecast trends, in much the same way as the perceived accuracy of a forecast can be judged by the congruence between the forecast trend as “pointing” or not pointing in the direction of the outcome. Study 4 suggested that incongruence is more conspicuous than congruence, both for perceived agreement and for accuracy. Future studies need to test this observation.

Implications for communication

Results from the present studies have important implications in domains where the chances of future events are expressed probabilistically, such as in climate science, medical decision-making, stock market forecasts or election forecasts based on political polls. Forecasters and experts should be aware that for the public probability revisions may carry weight in addition to the actually probability level. A moderate likelihood will become more persuasive if the forecaster reported a lower probability before. On the other hand, the public may neglect a substantial risk if prior estimates were even more alarming.

Intriguingly, two estimates lend themselves more easily to trend projections than three. With two estimates, the change can unambiguously be described as either an increase or a decline. With three or more, one estimate can easily disrupt the trend, even

when located further back in back in time. As a result, one can predict that several estimates will make people less sure about what to expect and perhaps encourage more conservative or regressive predictions. We do not think that people are fully aware of how they utilize these previous assessments to draw inferences about the future. For instance, participants in Study 2 would rather know another expert's present opinion than forecasts issued in the past. Despite this, past forecasts made their predictions of the future more extreme and thus in a way bolder than otherwise. A better understanding of how past information moulds our thoughts about the future might improve the public's thinking about long-term risks.

References

Aklin, M., & Urpelainen, J. (2014). Perceptions of scientific dissent undermine public support for environmental policy. *Environmental Science & Policy, 38*, 173-177.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576.

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences, 21*(2), 69-79.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in cognitive science, 8*(1), 160-179.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R., Verheggen, B., Maibach, E. W., . . . Green, S. A. (2016). Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters, 11*(4), 048002.

Erlandsson, A., Hohle, S. M., Løhre, E., & Västfjäll, D. (2017). Rises and falls of scary numbers: The effect on future trend estimates, perceived severity and help-allocations. Manuscript in preparation.

Dessai, S., & Hulme, M. (2004). Does climate adaptation policy need probabilities? *Climate policy, 4*(2), 107-128.

Hammond, K. R. (2000). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*: Oxford University Press on Demand.

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 589-607.

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

Hohle & Teigen (2017). More than 50% or less than 70% chance: Pragmatic implications of single-bound probability estimates. *Journal of Behavioral Decision Making*. doi: 10.1002/bdm.2052

IUCN Red List of Threatenes Species (2015). *Panthera tigris ssp. altaica.* Retrieved from http://www.iucnredlist.org/details/15956/0

Ji, L.-J., Nisbett, R. E., & Su, Y. (2001). Culture, change, and prediction. *Psychological Science, 12*(6), 450-456.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Lerner, J. S. & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255-275.

Lewandowsky, S. (2011). Popular consensus: Climate change is set to continue. *Psychological Science, 22*(4), 460-463.

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change, 3*(4), 399-.

Løhre, E. (in press). Stronger, sooner, and more certain climate change: A link between certainty and outcome strength in revised forecasts. *The Quarterly Journal of Experimental Psychology*.

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

Markman, K. D., & Guenther, C. L. (2007). Psychological momentum: Intuitive physics and naive beliefs. *Personality and Social Psychology Bulletin, 33,* 800–812.

Mumpower, J. L., & Stewart, T. R. (1996). Expert judgement and expert disagreement. *Thinking & Reasoning, 2*(2-3), 191-212.

Myers, T. A., Kotcher, J., Stenhouse, N., Anderson, A. A., Maibach, E., Beall, L., & Leiserowitz, A. (2016). Predictors of trust in the general science and climate science research of US federal agencies. *Public Understanding of Science,* 0963662516636040.

Peters, R. G., Covello, V. T., & McCallum, D. B. (1997). The determinants of trust and credibility in environmental risk communication: An empirical study. *Risk Analysis, 17*(1), 43-54.

Poortinga, W., & Pidgeon, N. F. (2003). Exploring the dimensionality of trust in risk regulation. *Risk analysis, 23*(5), 961-972.

Reyna, V. F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science, 13*(2), 60-66.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences, 7*(1), 1-75.

Schuldt, J. P., Konrath, S. H., & Schwarz, N. (2011). “Global warming” or “climate change”? Whether the planet is warming depends on question wording. *Public Opinion Quarterly*, nfq073.

Shanteau, J. (2000). Why do experts disagree. In B. Green, R. Cressy, F. Delmar, T. Eisenberg, B. Howcroft, M. Lewis, D. Schoenmaker, J. Shanteau, & R. Vivian (Eds.), *Risk behaviour and risk management in business life* (pp. 86-196). Dordrecht, The Netherlands: Kluwer Academic Press.

Slovic, P. (1999). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk analysis, 19*(4), 689-701.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*(1), 3-13.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica, 68*(1), 27-38.

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica, 88*(3), 233-258.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes, 80*(2), 155-190.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The art and science of prediction*: Random House.

van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2015). The scientific consensus on climate change as a gateway belief: Experimental evidence. *PloS one, 10*(2), e0118489.

Weber, E. U. (2010). What shapes perceptions of climate change? *Wiley Interdisciplinary Reviews: Climate Change, 1*(3), 332-342.

Yeung, S. (2014). Framing effect in evaluation of others' predictions. *Judgment and Decision Making, 9*(5), 445-464.

Table 1. *Percentages of participants predicting a target forecaster's future probability estimates to be equal, lower, or higher than the most recent one, in four experiments. Forecasts indicating trend continuation are in bold.*

	Equal	Lower	Higher
Two forecasts without reason (Hohle & Teigen, 2015)			
Increasing condition (50-60%)	9%	18%	**73%**
Decreasing condition (70-60%)	4%	**88%**	8%
Two forecasts with reasons (Study 1)			
Increasing with reasons (50-60%)	9%	18%	**73%**
Decreasing with reasons (70-60%)	2%	**79%**	8%
Two forecasters, A and B (Study 2)			
Increasing same forecaster (A1: 52, A2: 64%)	1%	23%	**76%**
Increasing different forecasters (B1: 52, A2: 64%)	10%	31%	**59%**
Simultaneous forecasts (B2: 52%, A2: 64%)[a]	7%	61%	32%
Three forecasts (Study 3)			
Down before increasing (60-50-60%)	22%	53%	**25%**
Up before decreasing (60-70-60%)	41%	**31%**	28%
Single estimate (60%)[a]	14%	39%	46%

[a] In these conditions only the most recent estimates were reported, so no prediction can be construed as the continuation of a trend.

Table 2**.** Mean predictions of future forecast and ratings of experts who forecasts a 60% chance of future grain productivity, after a history of variable previous forecasts, Study 3.

	Condition A 60-70-60	Condition B 60-50-60	Condition C 60	*F*	df	*p*	η^2_p
Number of reasons	1.76	1.78	1.87	0.13	2, 122	.882	.002
Next probability	59.2%	56.1%	58.3%	0.97	2, 122	.381	.017
Changed opinion	3.56_a	3.81_a	4.96_b	6.56	2, 118	.002	.100
Perceived uncertainty	3.30_a	$3.89_{a,b}$	4.61_b	5.52	2, 119	.005	.085
Future trust	4.24	4.59	3.97	2.72	2, 121	.070	.043
Perceived competence	4.49	4.70	4.34	0.90	2, 121	.408	.015

Note. Different subscripts indicate that means differed significantly with Bonferroni post hoc tests, $p < .05$.

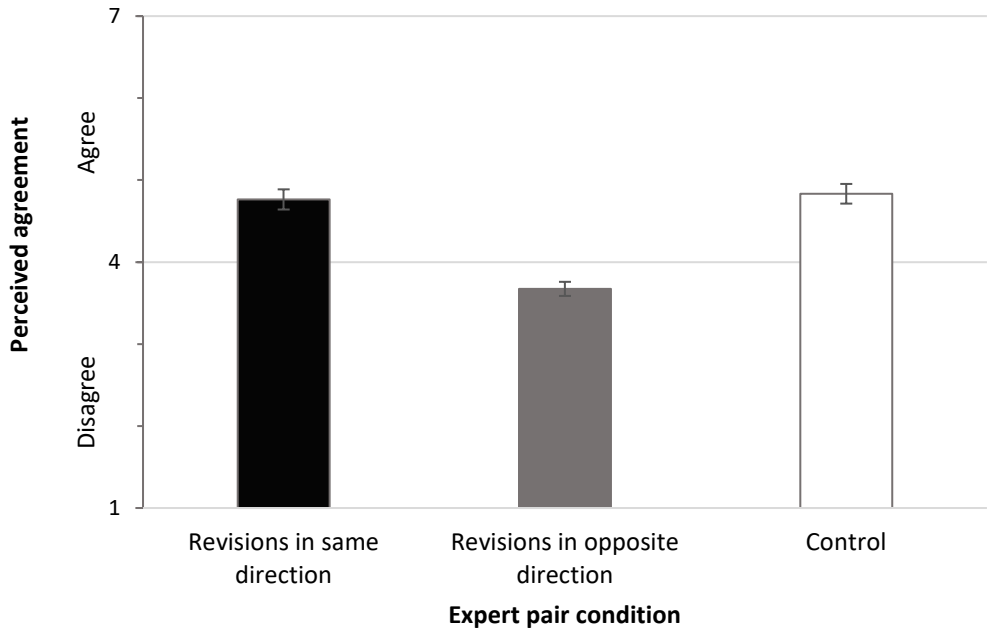

Figure 1. Mean ratings of agreement between experts who have revised their probabilistic forecasts in the same direction, in the opposite direction, or not at all, Study 4. Error bars show +/- 1 SE

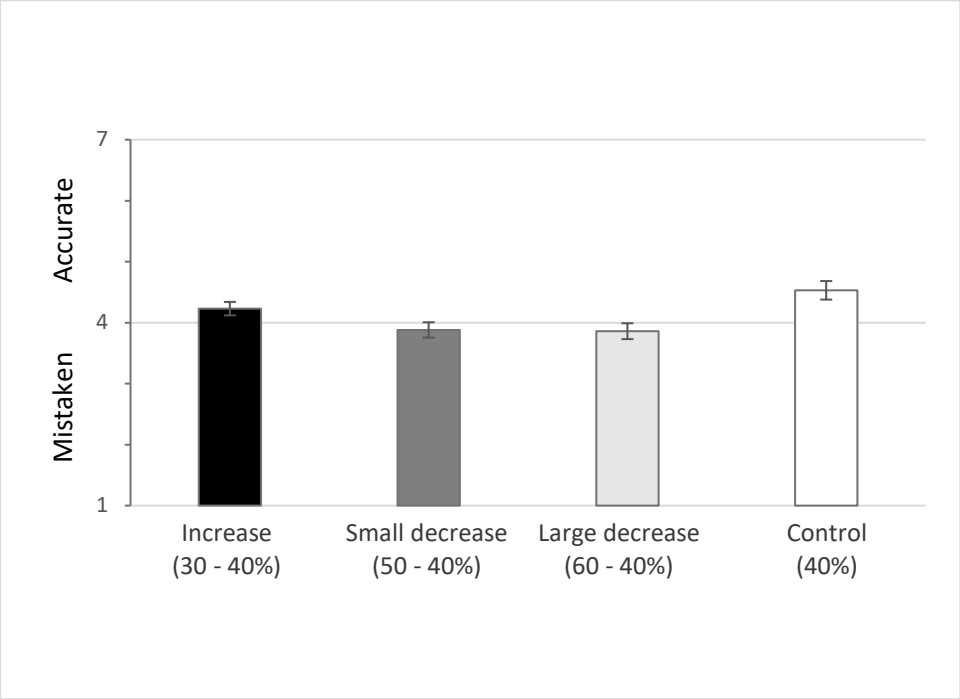

Figure 2. Mean ratings of forecast accuracy when the probability of El Niño is estimated to 40%, and the event occurred, Study 4. Error bars show +/- 1 SE.

Figure Captions

Figure 1. Mean ratings of agreement between experts who have revised their probabilistic forecasts in the same direction, in the opposite direction, or not at all, Study 4. Error bars show +/- 1 SE.

Figure 2. Mean ratings of forecast accuracy when the probability of El Niño is estimated to 40%, and the event occurred, Study 4. Error bars show +/- 1 SE.