Climate scientists' wide prediction intervals may be more likely but are perceived to be less certain

Article

Accepted Version

Climate scientists' wide prediction intervals are more likely but perceived to be less certain

Erik Løhre[1,2], Marie Juanchich[3], Miroslav Sirota[3], Karl Halvor Teigen[1,4], and Theodore G. Shepherd[5]

[1]Simula Research Laboratory, Oslo, Norway

[2]Inland Norway University of Applied Sciences, Lillehammer, Norway

[3]Department of Psychology, University of Essex, UK

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

[5]Department of Meteorology, University of Reading, UK

Short title: Wide intervals are perceived as uncertain

Corresponding author:
Erik Løhre
Inland Norway University of Applied Sciences
Postboks 400, 2418 Elverum, Norway
Telephone: +47 41 42 01 35
E-mail: erik.lohre@gmail.com

Abstract (149 words)

The use of interval forecasts allows climate scientists to issue predictions with high levels of certainty even for areas fraught with uncertainty, since wide intervals are objectively more likely to capture the truth than narrow intervals. However, wide intervals are also less informative about what the outcome will be than narrow intervals, implying a lack of knowledge or subjective uncertainty in the forecaster. In six experiments, we investigate how lay people perceive the (un)certainty associated with wide and narrow interval forecasts, and find that the preference for accuracy (seeing wide intervals as "objectively" certain) vs. informativeness (seeing wide intervals as indicating "subjective" uncertainty) is influenced by contextual cues (e.g., question formulation). Most importantly, we find that people more commonly and intuitively associate wide intervals with uncertainty than with certainty. Our research thus challenges the wisdom of using wide intervals to construct statements of high certainty in climate change reports.

Keywords: uncertainty, intervals, IPCC, climate change, communication

1. Introduction

The knowledge of general principles governing the climate system is sufficient to make strong qualitative predictions about climate change. For instance, the Intergovernmental Panel on Climate Change (IPCC) leaves little room for doubt when concluding that "continued emissions of greenhouse gases will cause further warming and changes in all components of the climate system" (IPCC 2013). In contrast, it is not possible to make precise quantitative predictions of exactly how the climate will change, even under a given forcing scenario (such conditional predictions are typically called projections). Thus, climate scientists generally issue predictions in the form of interval (range) forecasts (e.g., 0.3 to 1.7°C temperature rise[1], 0.26 to 0.55 m sea level rise) rather than point forecasts (e.g., 1.0°C temperature rise). Interval estimates allow a tradeoff between forecast precision and forecast certainty, or what Yaniv and Foster (1995) has described as a tradeoff between informativeness and accuracy. If a high degree of certainty (accuracy) is desired, one can forecast a wide interval (the rate of sea level rise [during the 21st century] will *very likely* exceed that observed during 1971 to 2010 [meaning more than a 20 cm rise]). This is commonly done in the IPCC reports when summary statements of high certainty are sought. Alternatively, if a high level of precision (informativeness) is desired, one can forecast a narrower interval with a lower degree of certainty (it is *likely* the sea level will rise between 26 and 55 cm).

While a large body of research shows that people often misunderstand the verbal probability expressions (e.g., "very likely", "unlikely") used by the IPCC (Budescu et al. 2009; Budescu et al. 2012; Budescu et al. 2014; Harris and Corner 2011; Harris et al. 2017; Harris et al. 2013; Ho et al. 2015; Juanchich and Sirota 2017), few studies have examined

[1] All examples are taken from IPCC, 2013: Summary for policymakers. *Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, and Coauthors, Eds., Cambridge University Press.

how lay people respond to the use of intervals to communicate degrees of (un)certainty in the climate change domain (Dieckmann et al. 2015; Dieckmann et al. 2017; Joslyn and LeClerc 2016; Løhre and Teigen 2017). We argue and demonstrate in this paper that the relationship between interval width (i.e., forecast precision) and certainty is ambiguous: a wide interval (an imprecise forecast) is "accurate" in the sense that it has a high probability of capturing the actual outcome, but its width also signals greater uncertainty about what the outcome will be, in comparison to a narrow interval (a more precise and hence more informative forecast). This ambiguity makes it important for forecasters to know whether lay people see wide intervals as more (or less) certain than narrow ones, and which of these two perspectives on intervals is more frequent and more intuitively appealing.

The two perspectives on the relationship between interval width and certainty may rely on two forms of certainty (Fox and Ülkümen 2011; Hacking 1975; Kahneman and Tversky 1982). On the one hand, certainty refers to our state of knowledge or belief. Such internal or subjective certainty is often expressed by statements where the subject is a sentient being ("I am 90% certain"), and using subjective terms like being confident, or sure (Fox and Ülkümen 2017; Ülkümen et al. 2016). But certainty can also be used in an external, more objective sense, reflecting variability, predictability and randomness in the outside world. Degrees of certainty are in these contexts often embedded in statements with an impersonal subject ("it is 90% certain"), and are used synonymously with degrees of probability, likelihood, or chance (Juanchich et al. 2017; Løhre and Teigen 2016).

With interval predictions, a wider interval allows for a greater degree of objective certainty (more hits and fewer misses). Even if the exact number of hits vs. misses can be assessed only retrospectively, after the outcomes are known, this general relationship can be claimed prospectively on purely logical grounds. Subjective certainty, however, might not increase with interval width. In fact, people may see wide intervals as cueing *un*certainty and lack of knowledge, for two reasons. First, more knowledge about a topic enables one to be more precise in one's statements about it (Yaniv and Foster 1997). Second, conversational norms suggest that people seek to maximize informativeness in communication (Grice 1975). The prediction "The temperature in Oslo will be between -35 and +35°C tomorrow" is true, with close to 100% certainty, but is also far too vague to be useful for someone preparing for a visit. A forecaster with higher subjective confidence may make a more precise, informative prediction ("The temperature at noon will be between 15 and 18°C"), which can be seen as conveying more certain expectations about tomorrow's weather.

Thus, different concepts of certainty might lead to different views on the implications of wide vs. narrow interval predictions. Those who find a wide interval to be more certain, by being more likely to include the true (actual) values, will in this paper be referred to as showing a *preference for accuracy*. In contrast, those who consider a wide interval to be less certain, by being less informative and expressing lower confidence about expected outcomes, display a *preference for informativeness*.

Previous research has found support for both types of preference (or "mindsets"). In line with the informativeness mindset, lay people expect experts to give narrower interval estimates than novices (McKenzie et al. 2008). Recipients of information prefer precise statements (Du et al. 2011; Jørgensen 2016), with narrow intervals occasionally preferred over wide intervals even when the wide interval includes the correct answer while the narrow interval does not (McKenzie and Amin 2002; Yaniv and Foster 1995). Teigen (1990) found that people placed more confidence in precise statements than in vague statements, but also that people chose the more precise statement when asked which statement they would be more skeptical about. Participants in a recent study received high and low probability forecasts made by climate change experts, and completed the forecasts by filling in corresponding intervals (Løhre and Teigen 2017). Some associated high probabilities with wide intervals, but many did the opposite and assigned *narrow* intervals to high probabilities. Similar results were obtained when people were given wide and narrow interval forecasts, and asked to fill in missing probability values. Some participants assumed wide intervals were more probable, whereas others felt they were *less* probable than narrow intervals.

These studies leave open several important questions that we address in the present paper. (1) Is one "mindset" more prevalent than the other? (2) Can contextual and linguistic cues, which are known to change the way people think about probabilities (Løhre and Teigen 2016; Nisbett et al. 1983; Reeves and Lockhart 1993; Ülkümen et al. 2016), also influence people's views on the relationship between interval width and certainty? These two questions were investigated in Experiments 1-5, where we manipulated the focus of a question about certainty. We predicted that a question about which of two intervals is "more certain to be correct", would promote reflections about objective certainty, accuracy and the probability of hits and misses, and should accordingly be answered in favour of the wide interval. On the other hand, a question about which interval "conveys more certainty", would make thoughts about informational value and subjective certainty more salient, and induce people to find wide intervals to imply *less* certainty than narrow ones. (3) A third issue is which mindset people find more intuitive. Experiment 6 investigated people's lay theories about interval width and probability, and asked people to rate how intuitively appealing two statements compatible with the two mindsets were.

<Insert Table 1 about here>

2. Experiments 1-5: Effects of question type on the perception of wide vs. narrow intervals

a. Participants

The participants in these experiments (total $N = 923$, see Table 1) were university students from the UK and Norway who volunteered to participate or who received course credits for participation, and Amazon MTurk workers from the US who were paid to complete the questionnaires. Both of these types of convenience samples are typical in psychology experiments, and are often reasonably similar to community samples (Goodman et al. 2013; Paolacci et al. 2010). For the purpose of the current studies, namely to investigate subjective perceptions of interval forecasts of climate change, we would expect that participants from these samples should be at least as well-equipped (if not better) to interpret the information as more representative samples.

b. Materials and procedure

In all experiments, the participants received interval forecasts of sea level rise and temperature rise by the end of the century from two different teams of climate scientists. One team issued a forecast with a wide interval (e.g., “The temperature will increase between 1.1° Celsius and 6.4° Celsius”), while the other team gave a forecast with a narrower interval (e.g., “The temperature will increase between 2.2° Celsius and 5.4° Celsius”). The participants were asked, in three to four different conditions in the different experiments, to choose which prediction “conveys more uncertainty [certainty]” or which prediction “is more likely [certain, uncertain] to be correct”. These questions were formulated to focus on informativeness or on accuracy, respectively. An overview of the questions used in the different experiments is provided in Table 2, and more detailed descriptions of the procedure for each experiment is provided below. The full description of the scenarios, as well as separate statistical analysis of each experiment, can be found in the Supplementary materials (in the Results section, only the overall results are described). Several of the experiments also investigated secondary hypotheses, which are briefly described below, while more detailed descriptions and analyses are provided in the Supplementary materials.

<Insert Table 2 about here>

1) MATERIALS AND PROCEDURE VARIATIONS IN EXPERIMENTS 1-5

In Experiment 1, we manipulated question type and reasons for variability in a 2 x 2 within-subject design. Participants completed a daily survey for 14 days. On the third day, the participants received questions about which interval "is most likely to be correct" and on day 6 which interval "conveys most uncertainty". The same questions were repeated on days 9 and 11, but here, participants also received an explanation for the variability in the expert forecasts. The variability was explained by referring to temperature rise "in different countries" and sea level rise "in different parts of the world". On day 14 participants rated their belief in climate change by answering four questions taken from Heath and Gifford (2006). For each scenario (temperature and sea level rise), participants could choose one of the two predictions or rate them as equal.

Participants in Experiment 2 received the same questions as in Experiment 1, but this was a 2 x 2 design with question type and reason for variability varied between subjects. Hence, participants in different groups received questions either about which interval "conveys most uncertainty" or which interval "is most likely to be correct", and either received an explanation for the variability in estimates or did not receive such an explanation.

In Experiment 3, we attempted to control for some potential confounding factors in Experiments 1 and 2. Beside their focus on informativeness or accuracy, the questions used in the first two experiments differed in several respects. First, the term "uncertainty" was used in the informativeness-focus condition and the term "likely" was used in the accuracy-focus condition. These terms were assumed to be associated with different sources of uncertainty, with "uncertainty" being an internal/epistemic term, and "likely" an external/aleatory term

(Ülkümen et al. 2016). Second, the two terms differ in their *directionality* (Teigen and Brun 1995, 1999). While "uncertain" has a negative directionality (i.e., it points towards the possibility that an outcome might not occur), "likely" has a positive directionality (i.e., it points towards the possibility that an outcome might occur). To better control for the source of uncertainty and directionality of the verbal probabilities used in the question, we used the two terms "uncertain(ty)" and "certain(ty)", which are usually considered as reflecting epistemic uncertainty (Fox and Ülkümen 2011; Teigen and Løhre 2017; Ülkümen et al. 2016). The word stem was hence kept constant, while directionality and question type varied between-subjects, with different groups of participants receiving the question about which prediction "conveys more [un]certainty" and which prediction is "more [un]certain to be correct".

In Experiment 4, we removed the (arguably incorrect) "equal" option, so the participants chose between the wide and the narrow interval in each condition. Participants read the same temperature rise and sea level rise vignettes as in previous experiments in one of three conditions: uncertainty conveyed, certainty conveyed, and certain to be correct.

In Experiment 5, we added a third prediction that featured a narrower interval to each vignette, for two reasons: first, to highlight even more strongly that the teams differ in width of prediction intervals; and second, since the intervals in previous experiments were both quite wide, to include a very narrow interval that suggests high precision, but might be "too good to be true". Participants read the sea level and temperature rise scenarios and for each selected one of the three forecasts as the one that conveyed more certainty, conveyed more uncertainty or was more certain to be correct, in three between-subjects conditions.

2) SECONDARY HYPOTHESES

In addition to investigating the prevalence of the informativeness and accuracy mindsets and their associations with different kinds of questions, Experiments 1-5 also addressed some additional hypotheses. In Experiments 1 and 2, we investigated whether the accuracy mindset would be seen as more appropriate (i.e., wide intervals associated with certainty) in contexts where interval width could be related to variability. Predictions concerning a class of multiple outcomes might induce more distributional ("outside view") thinking, with wide intervals reflecting external variability, in contrast to predictions of a singular outcome, where wide intervals are more easily taken to reflect the forecaster's ignorance (Kahneman and Tversky 1982; Kahneman and Lovallo 1993; Nisbett et al. 1983; Reeves and Lockhart 1993). Hence, participants in different conditions in Experiments 1 (within-subjects) and 2 (between-subjects) were told that the intervals described temperature rise "in different countries" and sea level rise "in different parts of the world", while no explanation for the variability in the estimate was given in the other conditions.

In Experiment 3, we investigated whether perceptions of expertise could be influenced by question type, with the hypothesis that questions highlighting informativeness would lead to a stronger preference for experts giving narrow interval forecasts, as compared to questions highlighting accuracy. Therefore, after selecting the prediction that conveys more (un)certainty/is more (un)certain to be correct, participants in Experiment 3 rated which team seemed more trustworthy, seemed to have most knowledge (about temperature rise or sea level rise), seemed to have the best models (for predicting temperature rise or sea level rise), and which team seemed to be most competent. These ratings were done on scales from 1 (definitely the team with the wide interval) to 5 (definitely the team with the narrow interval).

Experiment 4 investigated factors that might explain people's preference for narrow intervals: their fluency and the perceived expertise of the speaker. Previous research has found that statements that are more fluent (i.e., easier to process), for example due to repetition or to heightened visibility, are judged as more truthful than less fluent statements (Arkes et al. 1989; Reber and Schwarz 1999). We expected that predictions with narrower intervals might be easier to process than predictions with wider intervals, and that this heightened fluency could be a reason why people prefer narrow intervals. Narrow intervals might also be preferred due to the association between precision and expertise. Hence, participants in Experiment 4 rated the fluency of the predictions featuring a narrow and a wide interval, as well as the perceived expertise of the teams (see Supplementary materials for more details about the rating scales).

For exploratory purposes, we included in Experiment 5 three measures of individual differences that might be related to the degree of perception of wide intervals as more uncertain and narrow intervals as more certain. Specifically, strong climate change beliefs could explain a preference for wide intervals as certain, since wide intervals can incorporate more extreme climate change values. In addition, people who are more numerate, and people who are able to understand the probability of occurrence of more than one event (i.e., people who correctly assess that the probability of one of two events is greater than the probability of occurrence of each of those events), might be better able to appraise that a wider interval means a greater likelihood to be correct. Hence, we included a climate change belief scale (Heath and Gifford 2006), a numeracy scale (Lipkus et al. 2001), and a disjunction task (adapted from Costello 2009).

c. Results

1) EFFECTS OF QUESTION FOCUS

Participants in Experiments 1-5 received wide and narrow interval forecasts of sea level rise and temperature rise from two different (fictional) teams of climate scientists, and indicated which interval *conveyed more (un)certainty* (question focused on informativeness) or was *more likely [(un)certain] to be correct* (question focused on accuracy).

<Insert Figures 1, 2, and 3 about here>

Question focus strongly influenced certainty judgments (Figures 1 and 2). Participants largely chose the wide interval as the one that conveyed more uncertainty, and indicated that the narrow interval conveyed more certainty. Responses to questions about which interval was more likely or more certain to be correct were mixed: some experiments showed a small preference for the wide interval, while narrow and wide intervals were seen as equally certain in other experiments.

Figure 3 summarizes the overall results (for all experiments with three response options, i.e., all experiments except Experiment 4), with responses coded according to whether wide intervals are seen as more certain (consistent with the accuracy mindset), narrow intervals are seen as more certain (consistent with the informativeness mindset), or both intervals are seen as equally likely. Analysis of Experiments 2, 3 and 5, where question focus was varied between-subjects and three response alternatives (wide more certain, narrow more certain, equal/"medium" interval more certain) were provided, showed a clear effect of question focus, $\chi^2(2, N=1080) = 213.373$, $p < .001$. While wide intervals were clearly associated with uncertainty after informativeness-focused questions, more participants associated wide intervals with certainty after accuracy-focused questions. However, even for questions about correctness, where wide intervals should logically be chosen as more certain, only about 40% of the participants did so.

2) RESULTS FOR SECONDARY HYPOTHESES

In Experiments 1 and 2, we investigated whether giving people an explanation for variability, for instance by telling them that the forecasts concerned sea level rise "in different parts of the world", would facilitate the accuracy mindset (i.e., would make more people associate wide intervals with certainty). However, this hint about variability did not affect participants' interval choice in either Experiment 1 ($p = .150$) or Experiment 2 ($p = .303$).

We further examined whether the accuracy and informativeness mindsets led to different inferences about the forecaster. Participants in Experiment 3 rated whether they found teams giving wide or narrow interval forecasts to have more expertise, on scales from 1 (definitely the team with the wide interval) to 5 (definitely the team with the narrow interval). The average of the ratings of the experts across scenarios (i.e., an average of the four questions per scenario) were slightly higher in the "conveys more"-conditions ($M = 3.50$, $SD = .73$) than in the "to be correct"-conditions ($M = 3.29$, $SD = .87$), and this difference was significant, $F(1,234) = 3.991$, $p = .047$, $\eta^2_p = .017$. In other words, the team with narrow intervals was rated more positively after informativeness-focused questions, indicating that making one or the other mindset salient can influence how well both the prediction and the communicator is received.

Experiment 4 investigated whether people find narrow intervals easier to process (i.e., more fluent) and more related to expertise than wide intervals. As predicted, participants judged the narrow interval as being easier to process and as reflecting more expertise than the wide interval (see Supplementary materials for more details about these findings).

Finally, in Experiment 5 we set out to investigate individual differences that might be related to the preference for informativeness vs. accuracy. Specifically, we asked participants about their climate change beliefs, and gave them a test measuring numeracy, and a test measuring their understanding of disjunctive probabilities. However, there were no clear correlation patterns between interval choice and any of these three measures across groups, and the experiment did not have enough power to detect differences within each condition.

3. Experiment 6: Is it more intuitive to associate wide intervals with uncertainty than with certainty?

Experiments 1-5 demonstrated that different question focus promotes different views about the relationship between certainty and interval width. However, the fact that only about 40% endorsed wide intervals as “more certain to be correct”, indicates that it is more common to associate wide intervals with (subjective) uncertainty than with (objective) certainty. This raises the possibility that the lay view about the relationship between interval width and certainty is more in line with the informativeness mindset than with the accuracy mindset.

In support of this idea, research on confidence intervals has repeatedly shown that people produce intervals that are too narrow for the assigned degree of certainty (Moore et al. 2016). This consistent overprecision (Moore and Healy 2008) is very hard to eliminate and suggests that the preference for informativeness may be a dominant intuitive response. Studies showing that recipients of information in general prefer narrow intervals illustrate a similar point (Du et al. 2011; Jørgensen 2016; McKenzie and Amin 2002; Yaniv and Foster 1995), as does the preliminary finding that people with higher numeracy can (sometimes) better appreciate the trade-off between precision and certainty than those with lower numeracy (Løhre and Teigen 2017). Hence, we ran Experiment 6 to test the hypothesis of an intuitive preference for informativeness among lay people.

a. Materials and procedure

The opening paragraph of the survey in Experiment 6 explained that climate scientists sometimes use intervals when giving their predictions of future outcomes, and presented two predictions concerning the expected sea level rise in the Oslo fjord. One of the predictions contained a wide interval (minimum 20 and maximum 60 cm sea level rise) and the other prediction contained a narrow interval (minimum 30 and maximum 50 cm sea level rise). Participants (students at the University of Oslo, N = 105, see Table 1) were randomly assigned to either the wide condition, where it was pointed out that one prediction is wider than the other, or to the narrow condition, where it was pointed out that one prediction is narrower than the other.

The text then explained that there are two different ways that one can think about the relationship between interval width and uncertainty, using the following formulation in the wide condition:

> " – On the one hand, WIDE intervals indicate that it is MORE UNCERTAIN what the outcome will be (the sea level could rise by anything from 20 to 60 cm, compared to 30 to 50 cm for the narrow interval)
>
> - On the other hand, it is MORE CERTAIN that projections using WIDE intervals will be correct (the forecast is correct if the sea level rises by anything from 20 to 60 cm, compared to 30 to 50 cm for the narrow interval)"

In other words, the accuracy mindset (seeing the wide interval as more certain to be correct) and the informativeness mindset (seeing the wide interval as indicating that it is more uncertain what the outcome will be) were explained to the participants. In the narrow condition, the text explained that narrow intervals could be seen as indicating that it is more certain what the outcome will be, or that it is more uncertain that predictions using narrow intervals will be correct. The order of the statements was counterbalanced in both conditions.

After reading the description of the different ways of thinking about intervals and uncertainty, participants were asked to rate how intuitive, natural, appealing, logical, and complicated they found the two ways of thinking, on scales from 1 (not intuitive/natural etc. at all) to 7 (very intuitive/natural etc.). Next, the participants were given tests of numeracy (Cokely et al. 2012; Schwartz et al. 1997) and cognitive reflection (Frederick 2005) to see whether individual differences in these abilities were related to a preference for informativeness or accuracy. Finally, participants were asked if they had already seen or responded to the cognitive reflection test online or in other experiments.

<Insert Figures 4 and 5 about here>

b. Results

Figures 4 and 5 display the ratings of the different mindsets for both wide and narrow intervals, and show that the view that wide intervals convey uncertainty was judged as more intuitive, natural, appealing, logical, and less complicated than the view that wide intervals are more certain to be correct. For simplicity we refer to this combination of attributes as more “intuitively appealing”. We also computed an average difference score to measure the degree to which one “mindset” was judged as more intuitively appealing than the other, by taking the “wide = uncertain” and “narrow = certain” ratings, which are in line with the informativeness mindset, and subtracting the corresponding “wide = certain” and “narrow = uncertain” ratings, which are in line with the accuracy mindset.[2] Thus, positive difference scores indicate that the informativeness mindset is seen as more intuitively appealing than the accuracy mindset. The average difference score for the five items (Cronbach’s $\alpha = .74$) did not differ between conditions, $F(1,103) = .144$, $p = .706$, $\eta^2_p = .001$. More interestingly, the average difference score across conditions was positive, $M = .42$, $SD = 1.32$, and differed significantly from 0, $t(104) = 3.290$, $p = .001$, 95% CI [.17, .68]. Hence, participants overall judged the informativeness mindset as more intuitively appealing than the accuracy mindset.

There was no significant correlation with the average difference score for either the cognitive reflection test ($r = .01$, $p = .958$) or numeracy ($r = .09$, $p = .355$). However, people with higher cognitive reflection and numeracy perceived *both* mindsets as more intuitive, as shown by positive correlations between CRT and the informativeness ($r = .20$, $p = .040$) and accuracy mindsets ($r = .21$, $p = .037$), and between numeracy and the informativeness ($r =$

[2] The only exception was for the ratings of how complicated the participants found the two ways of thinking. Here the “wide = uncertain” ratings were subtracted from the “wide = certain” ratings, and the “narrow = certain” ratings were subtracted from the “narrow = uncertain” ratings.

.24, $p = .014$) and accuracy mindsets ($r = .14$, $p = .161$). Hence, higher scores on these measures indicate a tendency to find it intuitive to use intervals to express both certainty and uncertainty.

4. General Discussion

The experiments reported in this paper fill a gap in the literature about climate change communication (Moser 2010; Pidgeon and Fischhoff 2011) by investigating lay perceptions of the relationship between interval width (forecast precision) and certainty. We found evidence of two alternative ways of thinking. Overall, independent of question focus, 45% of our participants[3] perceived narrow intervals as giving more certain knowledge about what the outcome will be, in line with what we have called a *preference for informativeness*; while 26% of the participants perceived that wide intervals have a higher certainty of capturing the true value, displaying a *preference for accuracy*. These two opposite "mindsets" can be made more or less salient by drawing attention to different types of uncertainty. Questions about which interval conveys more (un)certainty (i.e., focusing more on subjective uncertainty) led to a consistent preference for informativeness, while questions about which interval is more certain/likely to be correct (i.e., focusing more on objective certainty) led to a response pattern more in line with the accuracy mindset.

Questions focused on informativeness led to a clearer response pattern (wide intervals seen as uncertain and narrow ones as certain) than did questions focused on accuracy. It is somewhat puzzling that people were so divided in their answers to the question about which interval is more likely/certain to be correct. Logically, wider intervals are objectively more likely to capture the outcome value that will occur, as they cover both central (likely) and more peripheral (unlikely) values. Our results indicate that (perhaps for good reasons) people

[3] These percentages are based on all experiments with three response alternatives (wide more certain, narrow more certain, equal), i.e., Experiments 1, 2, 3, and 5.

would like to know more precisely what the expected values are, and hence find it more intuitive to adopt the informativeness than the accuracy mindset, as shown in Experiment 6. Although the generalizability of the results should be investigated in non-western samples, we find it noteworthy that they are replicated in two different languages (Norwegian vs. English), in three different countries (Norway, UK, USA), and with both student and MTurk samples. Note also that our participants should be more educated and arguably more knowledgeable about these topics than more representative samples. Hence, one might expect an even stronger preference for informativeness in a more representative sample.

These results have important theoretical implications, particularly for the literature on overprecision (Moore et al. 2016). The intuitive preference for informativeness means that wide intervals are usually associated with uncertainty, and as a result, people may not understand or agree that they should widen their intervals to increase their certainty. This can be said to strengthen the conversational norms/informativeness account of overprecision (Kaesler et al. 2016; Yaniv and Foster 1995, 1997).

Climate scientists may choose to give wide intervals in order to present predictions with high certainty. Yet, our results show that wide intervals are a stronger signal of (subjective) uncertainty than of (objective) certainty, and the use of wide intervals may therefore undermine trust in climate scientists and their predictions. Although language that accentuates the accuracy mindset may make wide intervals more acceptable to the public (see Experiment 3), our results suggest that many recipients will still prefer narrow intervals, as suggested by 25% of the participants given accuracy-focused questions in our experiments (see Figure 3). Note, however, that in the current experiments, the participants only received intervals, and were asked about their perceptions of (un)certainty. In statements from the IPCC, intervals are often accompanied by verbal or numerical probability statements (e.g., "During the last interglacial period, the Greenland ice sheet *very likely* contributed between

1.4 and 4.3 m to higher global mean sea level”) (IPCC 2013). A recent study showed that explicitly mentioning the high certainty of wide intervals can counteract the tendency of lay people to see such intervals as uncertain, with most people stating that a wide interval with 90% probability was more certain than a narrow interval with 50% probability (Teigen et al. 2018).

Nevertheless, the current evidence gives reason to be skeptical about the use of wide intervals to achieve high certainty in statements about climate change. However, presenting a precise interval along with a statement about the low certainty of such an interval is arguably not a much better option. One compromise solution would be to provide two intervals rather than one: a narrow (informative) interval paired with a wide (confident) interval, to satisfy both camps of readers. The drawback is that presenting two intervals simultaneously adds complexity to the communication of an already complex topic. Using graphical representations could be useful to simultaneously communicate informativeness and accuracy in a relatively simple way (Spiegelhalter et al. 2011). In any case, communicators should be aware that the current practice of claiming to be very certain about a very wide interval will to many readers sound like a contradiction in terms, which might damage rather than strengthen the public’s belief in climate science.

Acknowledgements

This research was supported by grant no.235585/E10 from The Research Council of Norway. The authors report no conflicts of interest. All datafiles and materials are available on OSF (https://osf.io/h95cj/?view_only=3ddfde9204f3497eaebf42aa0c091ade).

References

Arkes, H. R., C. Hackett, and L. Boehm, 1989: The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, **2,** 81-94.

Budescu, D. V., S. Broomell, and H. H. Por, 2009: Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, **20,** 299-308.

Budescu, D. V., H. H. Por, and S. B. Broomell, 2012: Effective communication of uncertainty in the IPCC reports. *Climatic Change*, **113,** 181-200.

Budescu, D. V., H.-H. Por, S. B. Broomell, and M. Smithson, 2014: The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, **4,** 508-512.

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero, 2012: Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, **7,** 25-47.

Costello, F. J., 2009: Fallacies in probability judgments for conjunctions and disjunctions of everyday events. *Journal of Behavioral Decision Making*, **22,** 235-251.

Dieckmann, N. F., E. Peters, and R. Gregory, 2015: At home on the range? Lay interpretations of numerical uncertainty ranges. *Risk Anal*, **35,** 1281-1295.

Dieckmann, N. F., R. Gregory, E. Peters, and R. Hartman, 2017: Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning. *Risk Anal*, **37,** 471-486.

Du, N., D. V. Budescu, M. K. Shelly, and T. C. Omer, 2011: The appeal of vague financial forecasts. *Organizational Behavior and Human Decision Processes*, **114,** 179-189.

Fox, C. R., and G. Ülkümen, 2011: Distinguishing two dimensions of uncertainty. *Perspectives on Thinking, Judging, and Decision Making*, W. Brun, G. Keren, G. Kirkebøen, and H. Montgomery, Eds., Universitetsforlaget, 21-35.

——, 2017: Comment on Løhre & Teigen. *Thinking & Reasoning*, **23,** 483-491.

Frederick, S., 2005: Cognitive reflection and decision making. *Journal of Economic Perspectives*, **19,** 25-42.

Goodman, J. K., C. E. Cryder, and A. Cheema, 2013: Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, **26,** 213-224.

Grice, H. P., 1975: Logic and conversation. *Syntax and semantics, Vol. 3: Speech acts*, P. Cole, and J. L. Morgan, Eds., Academic Press, 41-58.

Hacking, I., 1975: *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference.* Cambridge University Press.

Harris, A. J. L., and A. Corner, 2011: Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology-Learning Memory and Cognition*, **37,** 1571-1578.

Harris, A. J. L., H. H. Por, and S. B. Broomell, 2017: Anchoring climate change communications. *Climatic Change*, **140,** 387-398.

Harris, A. J. L., A. Corner, J. M. Xu, and X. F. Du, 2013: Lost in translation? Interpretations of the probability phrases used by the Intergovernmental Panel on Climate Change in China and the UK. *Climatic Change*, **121,** 415-425.

Heath, Y., and R. Gifford, 2006: Free-market ideology and environmental degradation - The case of belief in global climate change. *Environment and Behavior*, **38,** 48-71.

Ho, E. H., D. V. Budescu, M. K. Dhami, and D. R. Mandel, 2015: Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy*, **1,** 43-55.

IPCC, 2013: Summary for policymakers. *Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, and Coauthors, Eds., Cambridge University Press.

Joslyn, S. L., and J. E. LeClerc, 2016: Climate projections and uncertainty communication. *Topics in Cognitive Science*, **8,** 222-241.

Juanchich, M., and M. Sirota, 2017: How much will the sea level rise? Outcome selection and subjective probability in climate change predictions. *Journal of Experimental Psychology: Applied*, **23,** 386-402.

Juanchich, M., A. Gourdon-Kanhukamwe, and M. Sirota, 2017: «I am uncertain» vs «it is uncertain». How linguistic markers of the uncertainty source affect uncertainty communication. *Judgment and Decision Making*, **12,** 445-465.

Jørgensen, M., 2016: The use of precision of software development effort estimates to communicate uncertainty. *Software Quality. The Future of Systems- and Software Development*, D. Winkler, S. Biffl, and J. Bergsmann, Eds., Springer International Publishing, 156-168.

Kaesler, M., M. B. Welsh, and C. Semmler, 2016: Predicting overprecision in range estimation. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, Austin, TX, Cognitive Science Society, 502-507.

Kahneman, D., and A. Tversky, 1982: Variants of uncertainty. *Cognition*, **11,** 143-157.

Kahneman, D., and D. Lovallo, 1993: Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, **39,** 17-31.

Lipkus, I. M., G. Samsa, and B. K. Rimer, 2001: General performance on a numeracy scale among highly educated samples. *Med Decis Making*, **21,** 37-44.

Løhre, E., and K. H. Teigen, 2016: There is a 60% probability, but I am 70% certain: Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, **22,** 369-396.

——, 2017: Probabilities associated with precise and vague forecasts. *Journal of Behavioral Decision Making*, **30,** 1014-1026.

McKenzie, C. R. M., and M. B. Amin, 2002: When wrong predictions provide more support than right ones. *Psychonomic Bulletin & Review*, **9,** 821-828.

McKenzie, C. R. M., M. J. Liersch, and I. Yaniv, 2008: Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, **107,** 179-191.

Moore, D. A., and P. J. Healy, 2008: The trouble with overconfidence. *Psychological Review*, **115,** 502-517.

Moore, D. A., E. R. Tenney, and U. Haran, 2016: Overprecision in judgment. *Handbook of Judgment and Decision Making*, G. Wu, and G. Keren, Eds., Wiley.

Moser, S. C., 2010: Communicating climate change: history, challenges, process and future directions. *Wiley Interdisciplinary Reviews-Climate Change*, **1,** 31-53.

Nisbett, R. E., D. H. Krantz, C. Jepson, and Z. Kunda, 1983: The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, **90,** 339-363.

Paolacci, G., J. Chandler, and P. G. Ipeirotis, 2010: Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, **5,** 411-419.

Pidgeon, N., and B. Fischhoff, 2011: The role of social and decision sciences in communicating uncertain climate risks. *Nature Climate Change*, **1,** 35-41.

Reber, R., and N. Schwarz, 1999: Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, **8,** 338-342.

Reeves, T., and R. S. Lockhart, 1993: Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology-General*, **122,** 207-226.

Schwartz, L. M., S. Woloshin, W. C. Black, and H. G. Welch, 1997: The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, **127,** 966-972.

Spiegelhalter, D., M. Pearson, and I. Short, 2011: Visualizing uncertainty about the future. *Science*, **333,** 1393-1400.

Teigen, K. H., 1990: To be convincing or to be right: A question of preciseness. *Lines of Thinking*, K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos, Eds., Wiley.

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

——, 1999: The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, **80,** 155-190.

Teigen, K. H., and E. Løhre, 2017: Expressing (un)certainty in no uncertain terms: Reply to Fox and Ülkümen. *Thinking & Reasoning*, **23,** 492-496.

Teigen, K. H., E. Løhre, and S. M. Hohle, 2018: The boundary effect: Perceived post hoc accuracy of prediction intervals. *Judgment and Decision Making*, **13,** 309-321.

Yaniv, I., and D. P. Foster, 1995: Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, **124,** 424-432.

——, 1997: Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making*, **10,** 21-32.

Ülkümen, G., C. R. Fox, and B. F. Malle, 2016: Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, **145,** 1280-1297.

Tables

Table 1. Demographics for the samples used in the different experiments.

Experiment no.	n	Sample	Mean age (SD)	Female	Male
1	81	University of Essex students	24.0 (6.5)	80.2%	19.8%
2	201	Amazon Mechanical Turk	37.9 (12.0)	51.7%	48.3%
3	238	Amazon Mechanical Turk	37.7 (11.2)	47.9%	52.1%
4	302	Amazon Mechanical Turk	34.6 (10.4)	44.4%	55.6%
5	101	University of Essex, snowball sampling	28.0 (13.1)	36.6%	62.4%
6	105	University of Oslo students	23.1 (4.9)	76.2%	23.8%

Table 2. Overview of questions, response options and design used in the different experiments regarding interval predictions of climate change outcomes.

Experiment no.	Question(s)/ statements focused on informativeness: "Which interval conveys…"	Question(s)/statements focused on accuracy: "Which interval is…"	Response options	Design
1	"…most uncertainty"	"… most likely to be correct"	Wide, narrow, equal	Within-subjects
2	"… most uncertainty"	"… most likely to be correct"	Wide, narrow, equal	Between-subjects
3	"… more uncertainty" "… more certainty"	"… more certain to be correct" "… more uncertain to be correct"	Wide, narrow, equal	Between-subjects
4	"… more uncertainty" "… more certainty"	"… more certain to be correct"	Wide, narrow	Between-subjects
5	"… more uncertainty" "… more certainty"	"… more certain to be correct"	Wide, "medium", narrow	Between-subjects
6	"Wide intervals indicate that it is more uncertain what the outcome will be" "Narrow intervals indicate that it is more certain what the outcome will be"	"It is more certain that projections using wide intervals will be correct" "It is more uncertain that projections using narrow intervals will be correct"	Ratings of the intuitive appeal of both statements	Within-subjects

Figures

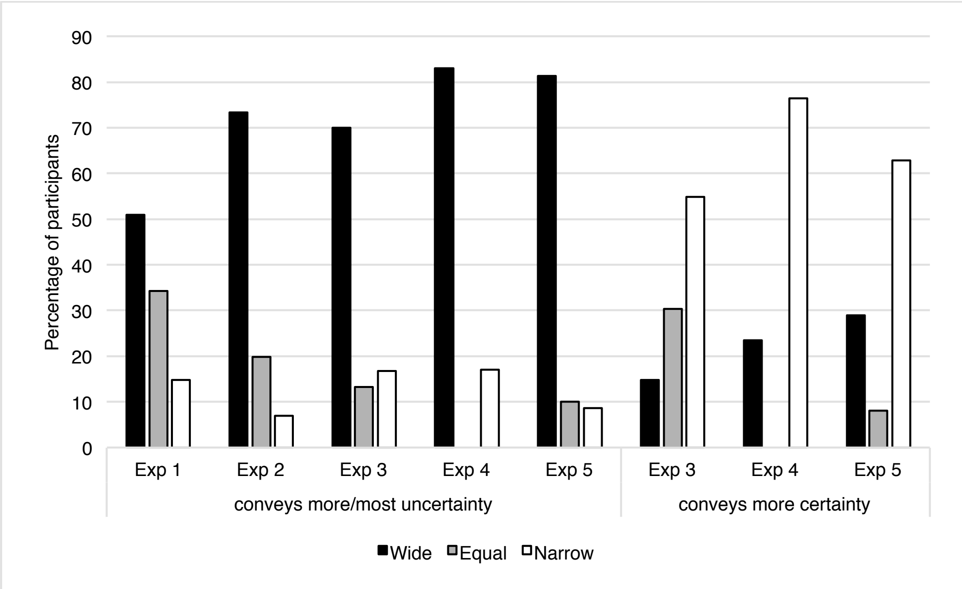

Figure 1. Choices of which interval conveys more certainty and uncertainty.

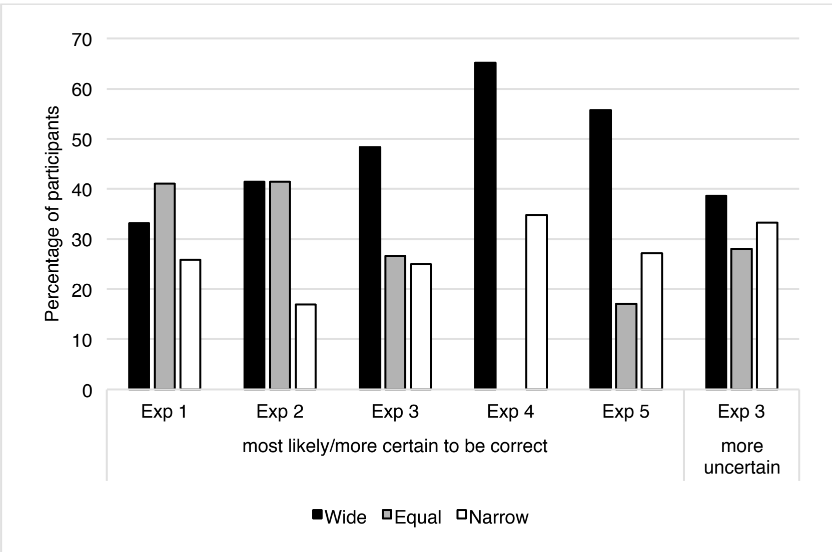

Figure 2. Choices of which interval is more certain/likely and more uncertain to be correct.

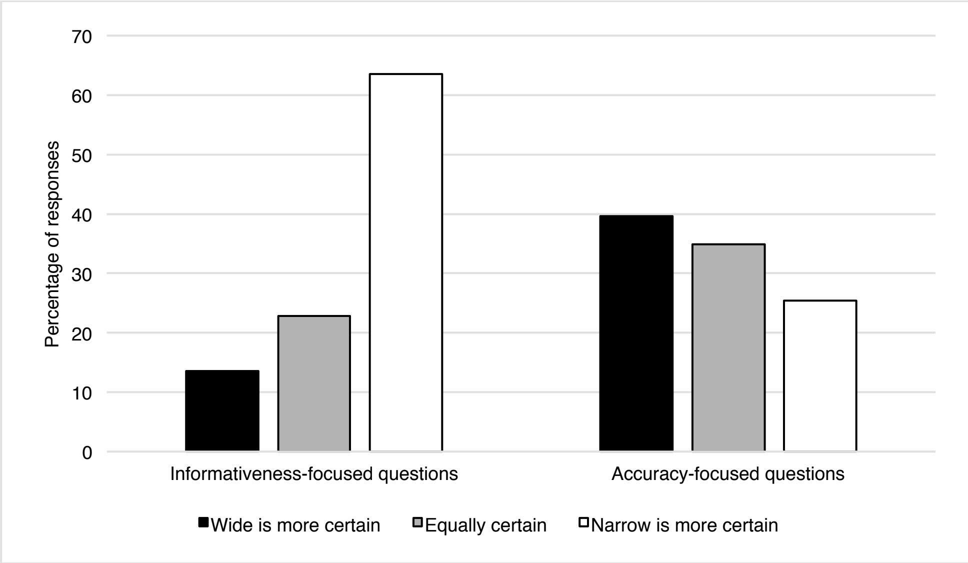

Figure 3. Overall preference for wide vs. narrow intervals as “more certain” for all experiments with three response options (Experiments 1, 2, 3, and 5).

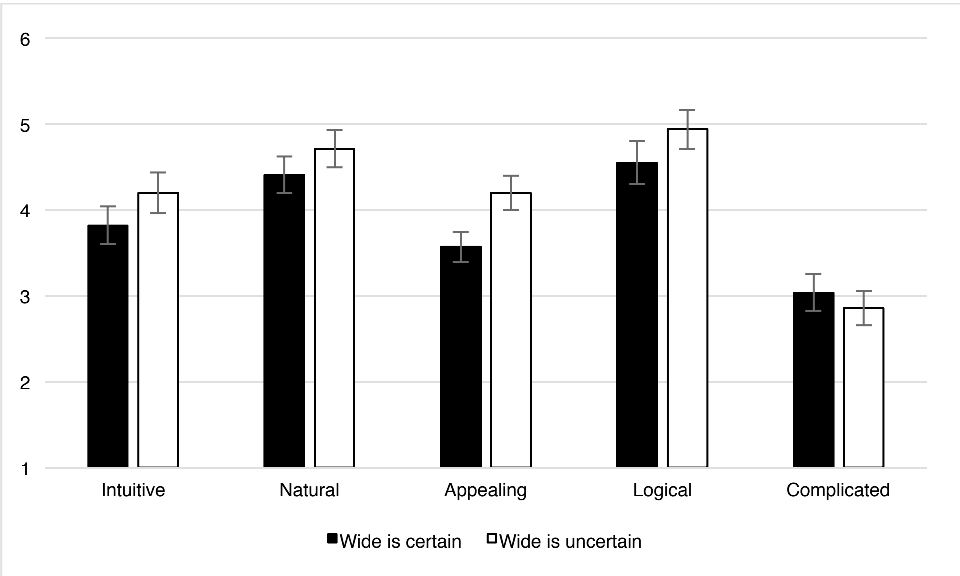

Figure 4. Mean perceptions of two ways of thinking about wide intervals (wide is certain vs. wide is uncertain) in Experiment 6, error bars +/- 1 SEM.

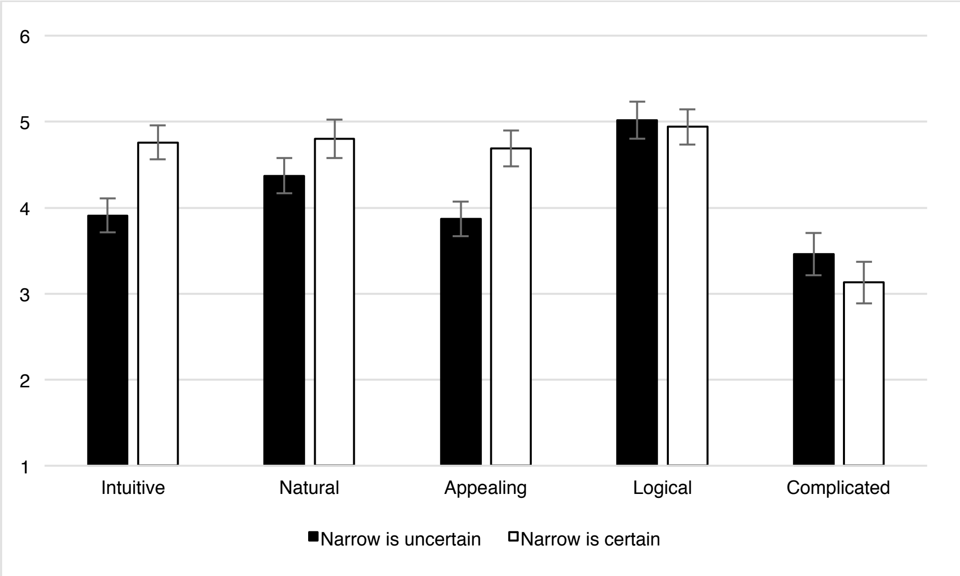

Figure 5. Mean perceptions of two ways of thinking about narrow intervals (narrow is uncertain vs. narrow is certain) in Experiment 6, error bars +/- 1 SEM.