Anchoring Climate Change Communications

Adam J. L. Harris

University College London

Han-Hui Por

Fordham University

Stephen B. Broomell

Carnegie Mellon University

Author Note

Adam J. L. Harris, Department of Experimental Psychology, University College London; Han-Hui Por, Department of Psychology, Fordham University; Stephen B. Broomell, Department of Social and Decision Sciences, Carnegie Mellon University.

Han-Hui Por is now at Educational Testing Service, Princeton, New Jersey.

We thank David Budescu for discussions and comments on a previous draft.

Correspondence concerning this article should be addressed to Adam J. L. Harris, Department of Experimental Psychology, University College London, London, WC1E 6BT, UK. E-mail: adam.harris@ucl.ac.uk

Abstract

Verbal probability expressions (VPEs) are frequently used to communicate risk and uncertainty. The Intergovernmental Panel on Climate Change attempts to standardise the use and interpretation of these expressions through a translation scale of numerical ranges to VPEs. A common issue in interpreting VPEs is the tendency for individuals to interpret VPEs around the mid-point of the scale (i.e., around 50%). Previous research has shown that compliance with the IPCC's standards can be improved if the numerical translation is presented simultaneously with the VPE, reducing the regressiveness of interpretations. We show that an explicit statement of the lower or upper bound implied by the expression (e.g., 0-33%; 66-100%) leads to better differentiated estimates of the probability implied by 'likely' and 'unlikely' than when the bound is not explicitly identified (e.g., less than 33%; greater than 66%).

Keywords: risk communication; verbal probability expressions; pragmatics; Intergovernmental Panel on Climate Change; International Accounting Standards;

Anchoring Climate Change Communications

Tackling climate change is a global challenge that requires a unified understanding of potential risks and losses attributable to human activities. The Intergovernmental Panel on Climate Change (IPCC) is the body charged with the dissemination of information about climate change to both policy makers and the general public. As with any scientific evidence, there exists some degree of uncertainty in any particular observation or prediction. In some instances, the amount of agreement or evidence will be insufficient to quantify this uncertainty. In these instances, standardised qualitative reports of confidence are prescribed (see Figure 1 in Mastrandrea et al., 2010). Where such quantification is, however, possible, the IPCC prescribes the use of words, also known as verbal probability expressions (VPEs), rather than numbers to communicate likelihood (e.g., “It is *very likely* that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent” (IPCC, 2007, p. 15).

VPEs effectively convey the understanding that probability estimates are often fuzzy concepts (e.g., Wallsten, 1990). It has long been known, however, that there is considerable interpersonal variation in people’s interpretation of VPEs (e.g., Budescu & Wallsten, 1985, 1995; Beyth-Marom, 1982; Dhami & Wallsten, 2005; Karelitz & Budescu, 2004), suggesting that VPEs can give rise to an “illusion of communication” (Budescu & Wallsten, 1995, p. 299). Additionally, the usage of VPEs can change depending on context, adding another layer of complexity to standardizing the use of VPEs (e.g., Beyth-Marom, 1982).

In an effort to reduce the variability in the interpretation of its VPEs, the IPCC provides guidelines for the numerical ranges that should be communicated with each VPE (Table 1). Recent research on the interpretations of VPEs in the IPCC reports has demonstrated large amounts of between person variability in these interpretations (Budescu, Broomell, & Por, 2009; Budescu, Por, & Broomell, 2012; Budescu, Por, Broomell, & Smithson, 2014; Harris, Corner, Xu, & Du, 2013). Moreover, overall, interpretations are typically highly regressive (i.e., interpretations tend to be closer to 50% than the prescribed meaning of the phrase). The regressiveness of interpretations results in less differentiation between phrases such as 'likely' and 'unlikely' than is intended by the IPCC (since estimates of both are 'pulled' towards 50%). For example, in Budescu et al. (2009), 64% of 'best estimates' of the terms 'very unlikely', 'unlikely', 'likely' and 'very likely' were regressive and outside the prescribed range for those terms.

Efforts to standardize the meaning of VPEs by providing a translation table (Table 1) somewhat reduce the variability in interpretations and increase correspondence with the IPCC guidelines (54% were inconsistent with the prescribed range - Budescu et al., 2009). Budescu and colleagues (Budescu et al., 2009; Budescu et al., 2012; Budescu et al., 2014) have additionally shown that the correspondence between interpretations and the IPCC's guidelines can be further increased with the use of a joint (verbal-numerical) presentation format. This format reduces the variability of interpretations across participants as well as the regressiveness in interpretations of VPEs. The joint presentation format provides the numerical definition directly alongside each usage of a VPE (e.g., "It is very likely (greater than 90%) that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent"). Despite the greater differentiation between VPEs, Budescu and colleagues found interpretations to remain highly regressive, even with the joint verbal-numerical format (47% of responses were still inconsistent with the prescribed range). We build upon this past work, testing whether another presentation difference can further reduce the regressiveness of interpretations.

The IPCC (2007) guidelines for the fourth assessment report (AR4; see Table 1) were somewhat ambiguous as to whether the numerical ranges for different VPEs were intended to overlap. Indeed, a pragmatic interpretation of the IPCC's meaning of 'likely' might lead one to the assumption that (for example) the range for 'likely' is really 67-90% (i.e., suggesting a lack of overlap with the range prescribed for 'very likely'). For if the communicator knows the probability is greater than 90%, they should maximise the informativeness of their communication by choosing the more precise term (e.g., Grice, 1975/2001). We term this a 'curtailed range' assumption. The guidelines for AR5 (Mastrandrea et al., 2010; see Table 1) were amended to make clear, for example, that the range of acceptable values for 'likely' extended as far as 100%, and did not stop at 90%. In the present paper, we test the effectiveness of this strategy by comparing interpretations of verbal-numerical presentation formats with numerical labels presented as in AR4 (single-anchor) versus AR5 (two-anchor).

There are two reasons to predict that interpretations should be less regressive in the two-anchor condition than the single-anchor condition:

Firstly, in line with the intentions of Mastrandrea et al. (2010), making explicit the fact that the range of (e.g.) 'likely' extends to 100%, rather than being curtailed at 90%, effectively increases the upper limits of the estimate, allowing estimates to be spread over a larger range. We term this the 'extended range' account where the midpoint of the perceived range is higher in the explicit extended range than in the ambiguous curtailed range.

Secondly, the effect might be seen as an instance of anchoring (e.g., Tversky & Kahneman, 1974), where the bounds pull judgments towards them. By not explicitly stating the implied lower bound (0%) for 'very unlikely' (or upper bound of 100% for 'very likely') the single-anchor presentation draws attention to the upper bound (10%) for 'very unlikely' (and the lower bound of 90% for 'very likely'). Such an effect would be countered by the value of 0 or 100 presented in a two-anchor condition. Anchoring effects have been demonstrated in the laboratory using a variety of methodologies (for a review see Furnham & Boo, 2011). Most commonly, participants first determine whether a target value is greater or less than an anchor value. For example, demonstrating anchoring in probability judgments, Plous (1989) asked participants ‘Is the chance of nuclear war between the United States and the Soviet Union greater or less than 1%.’ Participants who first answered this question later judged the likelihood of nuclear war as 9%, compared with an estimate of 19% for those who didn’t first answer this question. Other studies have, however, observed anchoring effects in consequential applied domains *without* an initial comparison question. Stewart (2009; see also, Navarro-Martinez, Salisbury, Lemon, Stewart, Matthews, & Harris, 2012), for example, observed that participants paid off less of a hypothetical credit card statement when a minimum payment was specified than when it was not. Stewart proposed that the minimum payment amount acted as an anchor, which reduced people’s estimates of how much they should repay.[1]

On the basis of the mechanisms outlined above, we predict that best estimates of the numerical probability will be less regressive with a two-anchor presentation than with a single-anchor presentation. ‘Less regressive’ means that estimates of low probability expressions (below 50%) should be lower, whilst those of high probability expressions (above 50%) should be higher. We therefore predict an interaction between verbal probability expression and presentation format, such that numerical estimates for ‘likely’ and ‘very likely’ are predicted to be higher and estimates for ‘unlikely’ and ‘very unlikely’ are predicted to be lower with a two-anchor presentation (such that both move further from 50%).

Although the current study is not intended to tease apart the extended range and anchoring explanations, there are certain patterns of results predicted to be generated by each mechanism. Consider a hypothetical participant who believed that ‘unlikely’ and ‘very

[1] Strictly speaking, ‘anchoring’ is an effect rather than a mechanistic explanation. We use the term here, however, to refer to a general assimilative effect of a provided numerical value on an estimate, rather than being concerned with the precise underlying mechanism (for discussions of the major theories of anchoring see e.g., Furnham & Boo, 2011; Mochon & Frederick, 2013; Newell & Shanks, 2014).

unlikely' were not intended to overlap and who picked the central value of the range as their best estimate. A possible response is one whereby the minimum, best and maximum estimates of 'unlikely' in the single anchor condition are 10%, 21% and 33% respectively. Upon understanding that the lower end of the range extended all the way to zero (in the two-anchor condition for example), a participant with this response strategy would update their estimates to 0%, 16% and 33%. Although consistent with an anchoring account, the most parsimonious explanation for such an effect (whereby the maximum estimate is unchanged for 'unlikely' and the minimum estimate is unchanged for 'likely') would seem to be the extended range account. In contrast, if both minimum and maximum estimates are similarly affected by the manipulation, this result would seem to be more consistent with a general anchoring account.

Our conceptualisation of the AR4 guidelines as a single-anchor format and AR5 as a two-anchor format can be thought of as synonymous with Teigen, Halberg and Fostervold's (2007a, 2007b) terminology of single bound and range, respectively. Teigen et al. (2007a, Study 2) reported that best estimates of the price of skis described as costing less than 1500 Norwegian Krone (NOK 1500) were higher than estimates of skis described as costing between NOK 500 and 1500 NOK. Similarly, estimates for shoes described as costing more than NOK 500 were lower than for shoes costing between NOK 500 and NOK 1500. The direction of effects is therefore as predicted in the current study. The situation is, however, rather different. This difference arises from our focus on a probability scale, which is *bounded*. With unbounded scales (at least at the upper end) such as price, there is no indication as to what a plausible range is. Consequently, a Gricean interpretation would be that the price should be quite close to the given value, otherwise a range would have been specified. The range presentation thus provides *additional* information in such situations. In Table 1, and the forthcoming experiments, the bounded probability scale ensures that an upper and lower bound is present in both presentation formats. Notably, in the information provided to participants, this bound is formally equivalent in the single- and two-anchor conditions. As a result of this equivalency, there is no guarantee that the results observed in Teigen et al. (2007a) will generalise to the present scenarios.

Judgments about climate change are highly politicised (e.g., Leiserowitz, Maibach, Roser-Renouf, & Hmielowski, 2011), and may provide a difficult and unique context for communicating uncertainty. VPEs can be (and have been) used in a number of contexts to present uncertainty information. To enhance the generality of the present research, we additionally test our manipulation of the single and two anchor formats in sentences taken from the International Accounting Standards (IAS; Deloitte, 2008).

Method

Participants

Two hundred and eighty two US-based Mechanical Turk workers completed the experiment. Sixty one of these failed the attention check (or did not complete it as they did not finish the survey). Of the remaining 221 participants, 69 were female, and the age range was 18-71 (median = 30 years; IQR = 11 years).

Design and Materials

A 2 (anchor) x 4 (VPE) mixed design was employed, with anchor condition manipulated between-participants and VPE manipulated within-participants. The anchor condition corresponded to whether the IPCC translations for the VPEs were presented with a single anchor (e.g., “less than 10%” or “more than 90%”) or with two anchors (e.g., “0-10%” or “90-100%”). The 4 VPEs used were ‘very unlikely’, ‘unlikely’, ‘likely’, and ‘very likely.’

Each VPE was embedded in two separate statements from the IPCC (2007, see Table 2). The VPEs and their numerical translations were highlighted in yellow in the provided text (see Table 2). The order of presentation of the sentences was randomised across participants. Four additional sentences containing the terms 'likely' and 'unlikely' from the IAS (Deloitte, 2008) were also used, and these items were presented in the same anchor format as the IPCC items. The IPCC items were always presented before the IAS items, as the IPCC items were the main focus of the study.

All VPEs were presented with their numerical translations next to them (see Table 1), and so the presentation format in the single-anchor condition was identical to the verbal-numerical condition of Budescu et al. (2009). The IAS items were presented with the same numerical translations as the IPCC items.

Participants were asked to indicate the minimum, best and maximum probabilities that they thought "*the authors intended to communicate*" [emphasis added] in each sentence. Responses were constrained such that the best estimate was equal or more than the minimum estimate and less than or equal to the maximum estimate. Responses were made by moving sliders to provide estimates between 0 and 100% (see Figure 1).

At the end of the experiment, participants completed the same 5-item numeracy test (Online Resource 1)[2] as in Budescu et al. (2012). Participants also completed a short demographic questionnaire, which included asking for participants' year of birth, gender and political affiliation: Strong Republican; Lean Republican; Independent; Lean Democrat; Strong Democrat; Others. In analyses including this covariate, the first five options were coded 1-5, whilst respondents reporting 'other' were excluded.

[2] In this pre-print, all online resources are included at the end of this document.

Procedure

After participants consented to participate in the study, they were asked to indicate their age and gender. At the start of both the IPCC items and the IAS items, participants were introduced to these organisations and their guidelines for the interpretation of their probability terms (in a table format, corresponding to the appropriate anchor condition – see Table 1, although the inequality sign was presented verbally, i.e., “greater than / less than”). Before proceeding to the main experimental task, participants were provided with a practice example using the phrase “about as likely as not (33-66%)”, to ensure they were comfortable using the response sliders. At the end of the IPCC and IAS tasks, participants completed the numeracy test and the demographic questionnaire. Consistency between responses to the age question at the start of the experiment, and the year of birth question in the final demographic questionnaire served as an attention check.

Results

We first report analyses of the ‘best estimates’, before considering the range endorsed by participants. We focus our analyses on the items taken from the IPCC report, and subsequently report the analysis including the IAS context for ‘likely’ and ‘unlikely’ (as these were the only two expressions included in the IAS context). The latter analysis reveals no differences between the two contexts. All analyses used the average of participants’ interpretations for each VPE, across the items within each individual context.

IPCC

Mean ‘best estimates’ for the four VPEs across both anchor conditions are plotted in Figure 2. A visual inspection of Figure 2 shows that, directionally, estimates are further from

50% (less regressive) in the two-anchor condition than the single-anchor condition for all four VPEs, as predicted. A 2 (anchor condition) x 4 (VPE) mixed ANOVA revealed a main effect of VPE, $F(1.3, 283.2) = 3736$, $p < .001$, $\text{eta}_p^2 = .95$ (Greenhouse-Geisser correction applied in cases when sphericity is violated.). The main effect of anchor condition was not significant, $F(1, 219) = 1.43$, $p = .233$, but the predicted VPE x anchor condition interaction was, $F(1.3, 283.2) = 6.71$, $p = .006$, $\text{eta}_p^2 = .03$. Simple effects tests (following Howell, 1997) showed that estimates were significantly different (and further from 50%) in the two-anchor condition for both 'unlikely', $F(1, 873.3) = 4.67$, $p = .03$, $\text{eta}_p^2 = .02$, and 'likely', $F(1, 873.3) = 16.32$, $p < .001$, $\text{eta}_p^2 = .08$. There was no anchor effect for either 'very likely' or 'very unlikely' (*F*s < 1).

IAS and IPCC

In an analysis including the IAS context, interpretations of 'likely' and 'unlikely' did not differ between the contexts: main effect of context, $F < 1$, interaction between context and VPE, $F(1, 219) = 2.18$, $p = .142$. Figure 3 therefore plots the mean estimates for 'likely and 'unlikely' in both anchor conditions, collapsed across context. Directionally, estimates are further from 50% in the two-anchor condition than the single-anchor condition. This result was borne out with a significant VPE x anchor condition interaction, $F(1, 219) = 21.55$, $p < .001$, $\text{eta}_p^2 = .09$, but this was not qualified by a 3-way interaction with context, $F(1, 219) = 1.32$, $p = .251$, suggesting that the effect is comparable across both the IPCC and IAS contexts. Separate ANOVAs performed on 'likely' and 'unlikely' suggested that the effect of anchor condition was significant for both: 'likely', $F(1, 219) = 24.41$, $p < .001$, $eta_p^2 = .10$; 'unlikely', $F(1, 219) = 9.14$, $p = .003$, $eta_p^2 = .040$, with no effects of, or interactions involving, context[3]. Finally, an ANCOVA confirmed that the overall pattern of results was consistent when controlling for numeracy, political affiliation, age and gender (see Online Resource 2 for distributions of political affiliations and numeracy scores).

To better understand the nature of the effect, we considered the range endorsed by participants for the VPEs. To determine this range, participants' minimum estimates were subtracted from their maximum estimates. Considering only 'likely' and 'unlikely', in a 2 (context) x 2 (VPE) x 2 (anchor condition) ANOVA, there was a significant effect of anchor condition, $F(1, 219) = 44.92$, $p < .001$, $eta_p^2 = .17$, and a context x anchor condition interaction, $F(1, 219) = 5.03$, $p = .026$, $eta_p^2 = .022$. We therefore analysed the endorsed range for the IPCC and IAS contexts separately.

Figure 4 plots the 'minimum', 'best' and 'maximum' estimates across the anchor conditions for the four expressions used in the IPCC context. A visual inspection suggests that the results are more consistent with the predictions of the extended range account. For 'very likely' and 'likely', the maximum estimate appears to increase more than the minimum estimate. For 'unlikely' and 'very unlikely', the minimum estimate appears to decrease more than the maximum estimate. A 4x2 (VPE x anchor condition) ANOVA revealed a significant effect of VPE, $F(2.2, 486.2) = 501.73$, $p < .001$, $eta_p^2 = .70$. Of more interest, there was also a main effect of anchor condition, $F(1, 219) = 13.71$, $p < .001$, $eta_p^2 = .06$, as well as a significant VPE x anchor condition interaction, $F(2.2, 486.2) = 8.60$, $p < .001$, $eta_p^2 = .04$ (there was no main effect of anchor condition, $F < 1$). Simple effects revealed that there was a significant effect of anchor condition for 'likely', $F(1, 614.9) = 26.7$, $p < .001$, $eta_p^2 = .14$, and 'unlikely', $F(1, 614.9) = 17.5$, $p < .001$, $eta_p^2 = .09$, with a larger range endorsed in the two-anchor condition than the single-anchor condition. There was no effect of anchor

[3] For 'likely', the main effect of context approached significance, $F(1, 219) = 3.17$, $p = .076$, with slightly higher estimates in the IAS context (mean = 78.7) than the IPCC context (mean = 78.0).

condition on the endorsed range for either 'very likely' or 'very unlikely' (*F*s < 1). As suggested in Figure 4, the increased range for 'likely' in the two-anchor condition stems from participants providing higher maximum estimates, $t(164.3) = 5.70$, $p < .001$, $d = 0.77$, with no corresponding change in their minimum estimates, $t(219) = 1.21$, $p = .230$. Likewise, the increased range for 'unlikely' in the two-anchor condition stems from participants providing lower minimum estimates, $t(193.1) = 4.27$, $p < .001$, $d = 0.57$, with no change in maximum estimates ($t < 1$). The results of the overall ANOVA held when numeracy, political affiliation, age and gender were included as covariates in an ANCOVA.

Despite the interaction with context, the pattern of results from the IAS context mirrored those from the IPCC context. A significant effect of VPE was observed, $F(1, 219) = 6.56$, $p = .011$, $\text{eta}_p^2 = .03$. More importantly, however, a significant effect of anchor condition was also observed, $F(1, 219) = 20.40$, $p < .001$, $\text{eta}_p^2 = .19$, with a larger range endorsed in the two-anchor condition than the single-anchor condition. As with the IPCC, the increased range for 'likely' in the two-anchor condition stemmed from participants providing higher maximum estimates (97.9% vs. 90.0%, $t[163.7] = 6.601$, $p < .001$, $d = 0.90$) with no difference in minimum estimates across anchor conditions (64.6% vs. 64.0%, $t < 1$). Likewise, the increased range for 'unlikely' in the two-anchor condition stemmed from participants providing lower minimum estimates (1.9% vs. 8.6%, $t[151.8] = 5.63$, $p < .001$, $d = 0.77$) with no difference in maximum estimates (33.7% vs. 34.3%, $t < 1$; see Online Resource 3 for figure displaying full descriptive statistics). The overall effect of anchor condition was also significant in an ANCOVA controlling for numeracy, political affiliation, age and gender as covariates.[4] Thus, the results from both the IPCC and IAS contexts are in line with the predictions of the extended range account.

[4] In contrast to the ANOVA results, there was no main effect of VPE on endorsed range in the ANCOVA.

Consistency with prescribed ranges

The reduced regressiveness of interpretations might lead to a greater number of best estimates and ranges consistent with the prescribed ranges of the IPCC. Across all items, 94% of best estimates were consistent in the two-anchor condition, versus 90% in the single-anchor condition, $\chi^2(1) = 6.2$, $p = .013$, with more estimates consistent in the two-anchor condition across all VPEs (Online Resource 4, Table A). Following Budescu et al. (2009), we defined an endorsed range as consistent if both upper and lower bounds were within the prescribed range, as inconsistent if both were outside the prescribed range, and as partially consistent otherwise. Eighty one percent of endorsed ranges were consistent in the two-anchor condition, compared with 77% in the single-anchor condition, $\chi^2(2) = 6.23$, $p = .044$, with more estimates consistent in the two-anchor condition across all VPEs (Online Resource 4, Table B).

General Discussion

The overall pattern of results is clear, and consistent with the results of an additional experiment, which recruited university students (Online Resource 5). For ‘likely’ and ‘unlikely,’ best estimates were less regressive with the two-anchor presentation than with the single-anchor presentation. These results are aligned with the findings using absolute values (e.g., cost, number of tables) from Teigen et al. (2007a, 2007b). Considering the analysis of the possible range endorsed by participants, the pattern of results is as predicted by the extended range account, with an increased endorsed range for both ‘likely’ and ‘unlikely.’

The lack of any effects for the extreme expressions, ‘very likely’ and ‘very unlikely’ is also in line with the extended range account. In AR4, an individual assuming that the range only extended to 99% (in the case of ‘very likely’) would only have this range extended by a single percentage point when the extended range is made explicit (as in AR5 – although it should be noted that simple ceiling and floor effects might also explain these results).[5]

Although the aim of the present paper was not to choose between two plausible explanations for the effects we observe, the pattern of results is more consistent with the extended range account, although we cannot rule out the additional potential influence of a more general anchoring contribution, which, if present, would appear to exert a smaller effect than the perceived extended range. Nonetheless, the effects themselves seem robust, holding when controlling for the influence of potential covariates across two experiments from two different populations.

The pattern of results observed was consistent across the IPCC and IAS contexts. The similar pattern across contexts makes us confident that providing two anchors rather than one in a VPE translation will reduce the regressiveness of interpretations across a variety of contexts, not solely the ones considered here. Because interpretations of the IPCC’s VPEs have typically been shown to be too regressive, the reduced regressiveness observed in the two-anchor condition is an improvement, and indeed results in greater consistency with the prescribed interpretations. The evaluation might, however, be considered more complex than this. Although we observed greater differentiation between interpretations of ‘unlikely’ and ‘likely’, because no effect was observed for ‘very unlikely’ and ‘very likely’ this effect simultaneously reduces the differentiations between ‘(un)likely’ and ‘very (un)likely.’ Ultimately, it is for policy makers to decide which terms are more important to differentiate.

[5] Despite obtaining broadly consistent results, the experiment reported in Online Resource 5 found that a lower best estimate for ‘unlikely’ was not associated with an increase in range. This experiment also did not report a significant difference in consistency rates between conditions, although numerically the trend was in the predicted direction in 11 out of 12 instances.

Currently, however, with ‘very (un)likely’ nested within ‘(un)likely’ (e.g., 5% risk can be represented by either ‘unlikely’ or ‘very unlikely’), people’s interpretations in the two-anchor condition would appear to be more in line with the intentions of the IPCC. Ultimately, it would be beneficial for future research to identify a means for reducing the regressiveness of interpretations of the extreme terms ‘very (un)likely,’ which are typically those for which the most regressive responses have been observed in the past (e.g., Budescu et al., 2009).

The current research has been concerned with numerical interpretations of VPEs, in the tradition of much work in this area (e.g., Budescu & Wallsten, 1985; Budescu et al., 2009, 2012, 2014; Harris & Corner, 2011; Harris et al., 2013; Ho, Budescu, Dhami, & Mandel, in press; Mandel, 2015; Smithson, Budescu, Broomell, & Por, 2012). The effect of the current format manipulation (or indeed those in Budescu et al., 2009, 2012, 2014) on *decision making* is, however, less clear, and is an important topic for future research. Previous findings that are likely to be relevant in this context include those showing that the use of only upper bounds in describing a range (e.g., ‘less than 33%’ in the case of probability estimates) encourages downwards comparisons, in the case of probabilities presumably directing attention to an event’s non-occurrence, whilst lower bounds (e.g., ‘greater than 66%’) encourage upwards comparisons (Teigen 2008; Teigen et al., 2007a). Such pragmatic influences suggest further potential advantages of the use of the two-anchor format, since reducing these influences can be expected to enhance standardisation in interpretation of the IPCC’s probability phrases.

Conclusion

A number of researchers have criticised the verbal probability scale used by the IPCC (Table 1). In light of such criticism, it is important that researchers not only highlight improvements that *could* be made (e.g., Ho et al., in press, who argued that organisations should use VPEs to represent the probabilities that empirical research shows people best associate them with), but also acknowledge where changes made by the IPCC *are* improvements. Following work demonstrating a benefit of using a verbal-numerical joint presentation format (e.g., Budescu et al., 2014), we provide evidence that the explicit upper and lower boundaries prescribed in IPCC AR5 further reduce the regressiveness of people's interpretations. The high profile of the reports produced by the IPCC, combined with their global readership, ensures the importance of attention to any factor that can enhance communication effectiveness. The inclusion of items from the IAS suggests that the benefit conferred by a two-anchor format is not unique to climate related contexts. We therefore recommend that such verbal-numerical presentations explicitly state both the upper and lower bounds wherever a standardised treatment of VPEs is intended.

References

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, *1*, 257-269.

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

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

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

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes, 36,* 391-405.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. R. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 32: Decision Making from a Cognitive Perspective)* (pp. 275-318). San Diego, CA: Academic Press.

Deloitte (2008). *Assets held for sale and discontinued operations: A guide to IFRS 5.* Retrieved October 29, 2014 from http://www.iasplus.com/en-us/publications/global/guides/pub1923

Dhami, M. K. & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities. *Memory & Cognition, 33,* 1057-1068.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40,* 35-42.

Grice, H. P. (1975/2001). Logic and conversation. In A. P. Martinich (Ed.), *The Philosophy of Language (4th Edition)* (pp. 165-175). Oxford, UK: Oxford University Press.

Harris, A. J. L., & Corner, A. (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., Corner, A., Xu, J., & Du, X. (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.

Ho, E. H., Budescu, D. V., Dhami, M. K., & Mandel, D. R. (in press). Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy.*

Intergovernmental Panel on Climate Change. (2007). *A Report of Working Group I of the Intergovernmental Panel on Climate Change: Summary for policymakers.* Retrieved April 2010 from http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-spm.pdf

Intergovernmental Panel on Climate Change (2013). *Climate Change 2013: The Physical Science Basis. Summary for Policymakers.* Retrieved June 2015 from http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_SPM_FINAL.pdf

Karelitz, T. M., & Budescu, D. V. (2004). You say “probable” and I say “likely”: Improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied, 10,* 25-41.

Leiserowitz A, Maibach E, Roser-Renouf C, Hmielowski JD (2011) Politics & global warming: democrats, republicans, independents, and the tea party. Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication. Retrieved from http://environment.yale.edu/climate/files/PoliticsGlobalWarming2011.pdf

Mandel, D. R. (2015). Accuracy of intelligence forecasts from the intelligence consumer's perspective. *Policy Insights from the Behavioral and Brain Sciences*, *2,* 111-120.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Held, H., et al. (2010). *Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties.* IPCC cross-working group meeting on consistent treatment of uncertainties, Jasper Ridge, CA. [retrieved from http://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf August 14th, 2013]

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, *122,* 69-79.

Navarro-Martinez, D., Salisbury, L. C., Lemon, K. N., Stewart, N., Matthews, W. J., & Harris, A. J. L. (2011). Minimum required payment and supplemental information disclosure effects on consumer debt repayment decisions. *Journal of Marketing Research, 48,* S60-S77.

Newell, B. R., & Shanks, D. R. (2014). Prime numbers: Anchoring and its implications for theories of behavior priming. *Social Cognition, 32,* 88-108.

Plous, S. (1989). Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology, 19,* 67-91.

Smithson, M., Budescu, D. V., Broomell, S. B., & Por, H-H. (2012). Never say “not”: Impact of negative wording in probability phrases on imprecise probability judgments. *International Journal of Approximate Reasoning, 53,* 1262-1270.

Stewart, N. (2009). The cost of anchoring on credit-card minimum repayments. *Psychological Science, 20,* 39-41.

Teigen, K. H. (2008). More than *X* is a lot: Pragmatic implicatures of one-sided uncertainty intervals. *Social Cognition, 26,* 379-400.

Teigen, K. H., Halberg, A-M., & Fostervold, K. I. (2007a). Single-limit interval estimates as reference points. *Applied Cognitive Psychology, 21,* 383-406.

Teigen, K. H., Halberg, A-M., & Fostervold, K. I. (2007b). More than, less than, or minimum, maximum: how upper and lower bounds determine subjective interval estimates. *Journal of Behavioral Decision Making, 20,* 179 - 201.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124 –1131.

Wallsten, T. S. (1990). The costs and benefits of vague information. In R. M. Hogarth (Ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn* (pp. 28-43). Chicago, IL: The University of Chicago Press.

Witteman, C., & Renooij, S. (2003). Evaluation of a verbal-numerical probability scale. *International Journal of Approximate Reasoning*, *33*, 117-131.

Table 1. *Likelihood scale of the IPCC.*

Term	Likelihood of the Outcome	
	AR4 (single-anchor)	AR5 (2 anchor)
Virtually certain	> 99%	99-100%
Very likely	> 90%	90-100%
Likely	> 66%	66-100%
About as likely as not	33 to 66%	33 to 66%
Unlikely	< 33%	0-33%
Very unlikely	< 10%	0-10%
Exceptionally unlikely	< 1%	0-1%

Table 2. *The statements used in the experiment (examples shown are as the text would appear in the single anchor condition).*

No.	Type	Statement
1	IPCC	Continued greenhouse gas emissions at or above current rates would cause further warming and induce many changes in the global climate system during the 21st century that would *very likely (greater than 90%)* be larger than those observed during the 20th century.
2	IPCC	It is *very likely (greater than 90%)* that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent.
3	IPCC	The Greenland ice sheet and other Arctic ice fields *likely (greater than 66%)* contributed no more than 4 m of the observed sea level rise.
4	IPCC	Temperatures of the most extreme hot nights, cold nights and cold days are *likely (greater than 66%)* to have increased due to anthropogenic forcing. [Note: Anthropogenic forcing refers to the influences on the environment by human, rather than natural, factors.]
5	IPCC	Over the past 3,000 to 5,000 years, oscillations in global sea level on time-scales of 100 to 1,000 years are *unlikely (less than 33%)* to have exceeded 0.3 to 0.5 m.
6	IPCC	Reconstructions of climate data for the past 1,000 years also indicate that this warming was unusual and is *unlikely (less than 33%)* to be entirely natural in origin.
7	IPCC	It is *very unlikely (less than 10%)* that the MOC will undergo a large abrupt transition during the 21st century. [Note: MOC stands for Meridional Overturning Circulation, and refers to the global ocean currents.]
8	IPCC	It is *very unlikely (less than 10%)* that climate changes of at least the seven centuries prior to 1950 were due to variability generated within the climate system alone.
9	IAS	IAS 36(21) notes that the fair value less costs to sell of an asset to be disposed of will often approximate its value in use, as the value in use calculation will consist mainly of the net disposal proceeds. This is because the future cash flows from continuing use of the asset until its disposal are *likely (greater than 66%)* to be negligible. [IFRS5]
10	IAS	Investment property shall be recognized as an asset when, and only when it is *likely (greater than 66%)* that the future economic benefits or service potential that are associated with the investment property will flow to the entity. [IPSAA 9]
11	IAS	Conversely, where an asset is still in the course of construction, and significant activities will need to be performed before it can be transferred, it is *unlikely (less than 33%)* that it could be regarded as available for immediate sale. [IFRS5]
12	IAS	During the initial one-year period, circumstances arise that were previously considered unlikely *(less than 33%)* and, as a result, a non-current asset (or disposal group) previously classified as held for sale is not sold by the end of that period. [IFRS5]

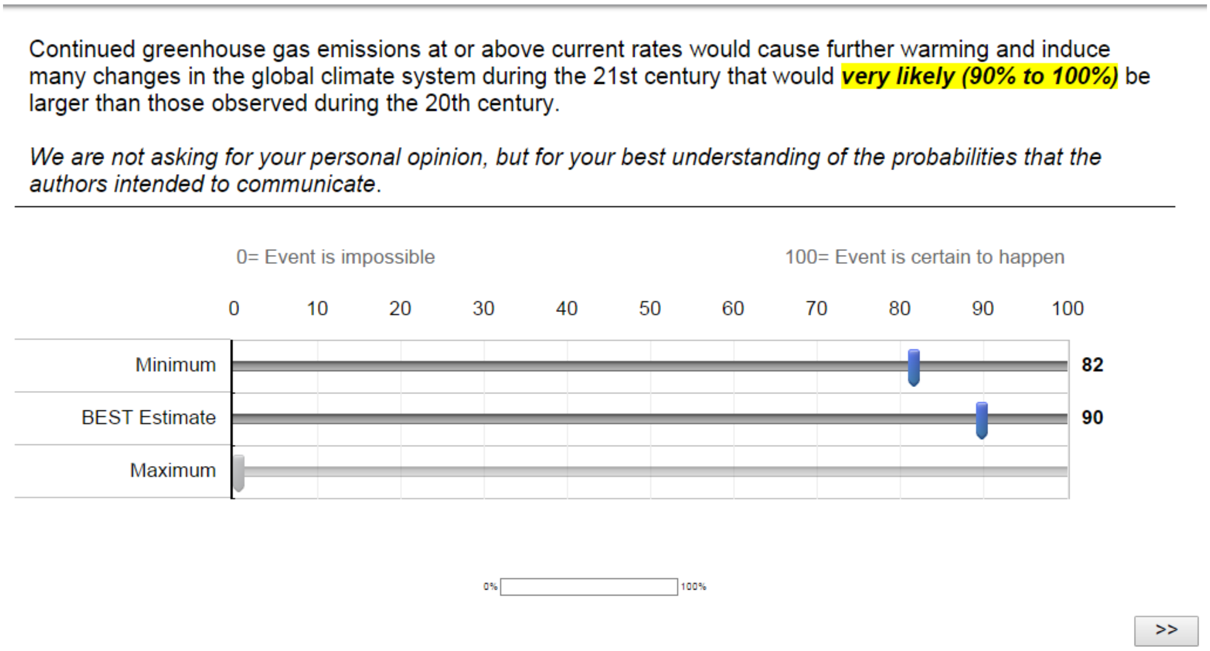

Figure 1. Example of an IPCC item with two anchors. The order of minimum, maximum and best estimates was always as shown here. The sliders turned blue after they were moved and their numerical value was shown on the right.

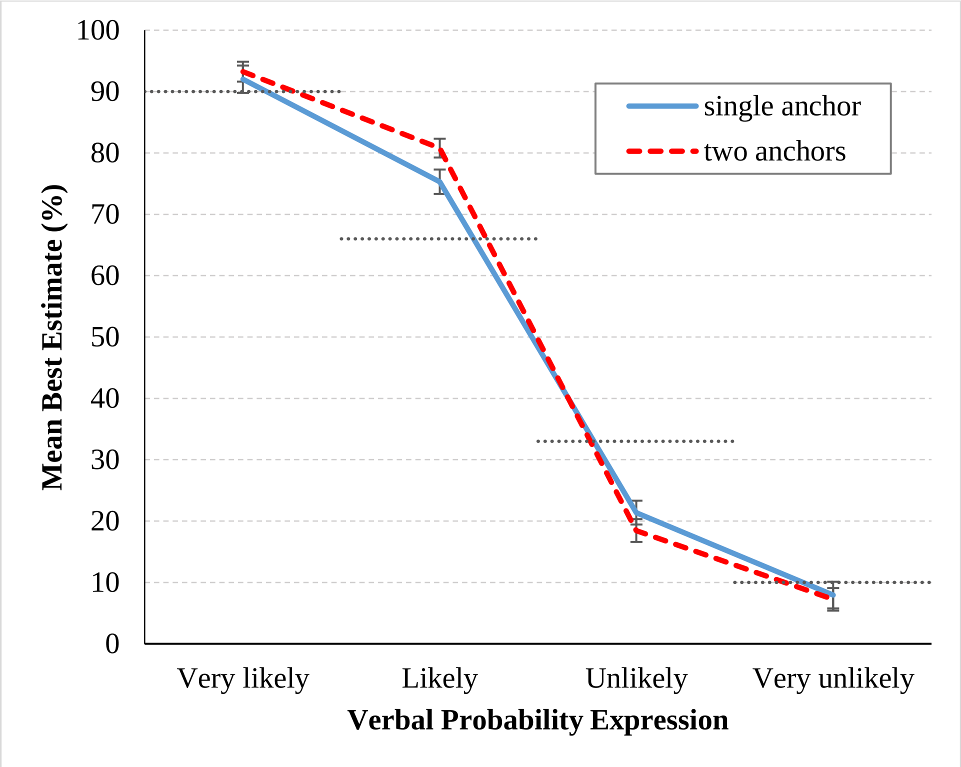

Figure 2. Mean 'best estimates' provided in response to the IPCC sentences. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bounds of '(very) likely' and the upper bounds of '(very) unlikely' (Table 1).

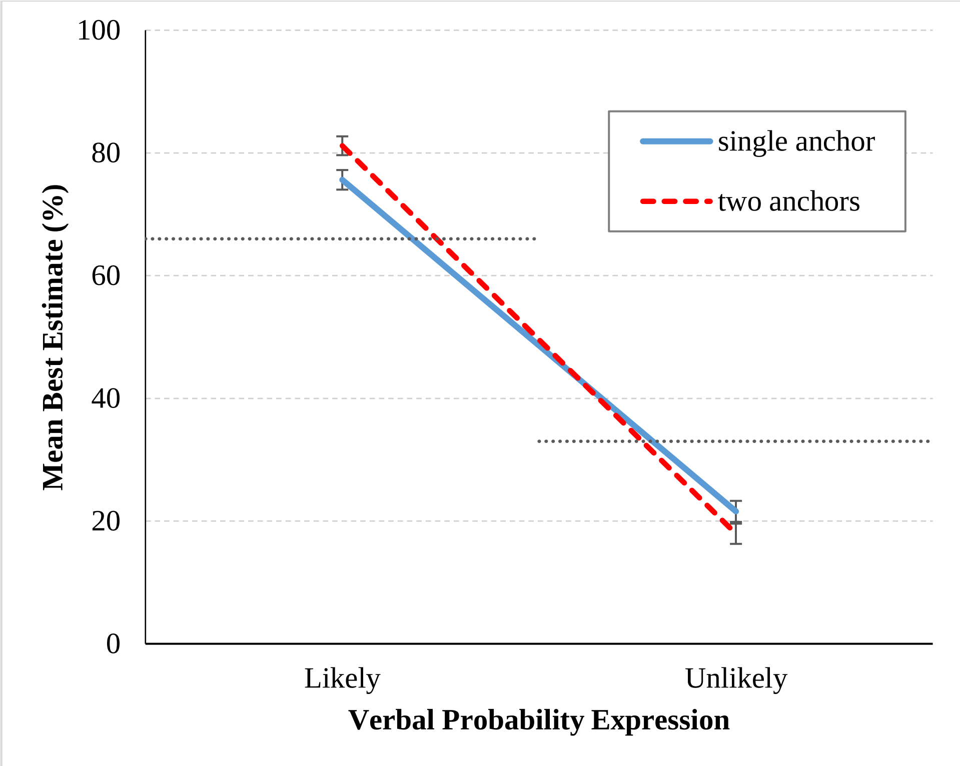

Figure 3. Mean ‘best estimates’ for interpretations of Likely and Unlikely, collapsed across the IPCC and IAS contexts. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bound of ‘likely’ and the upper bound for ‘unlikely’ (Table 1).

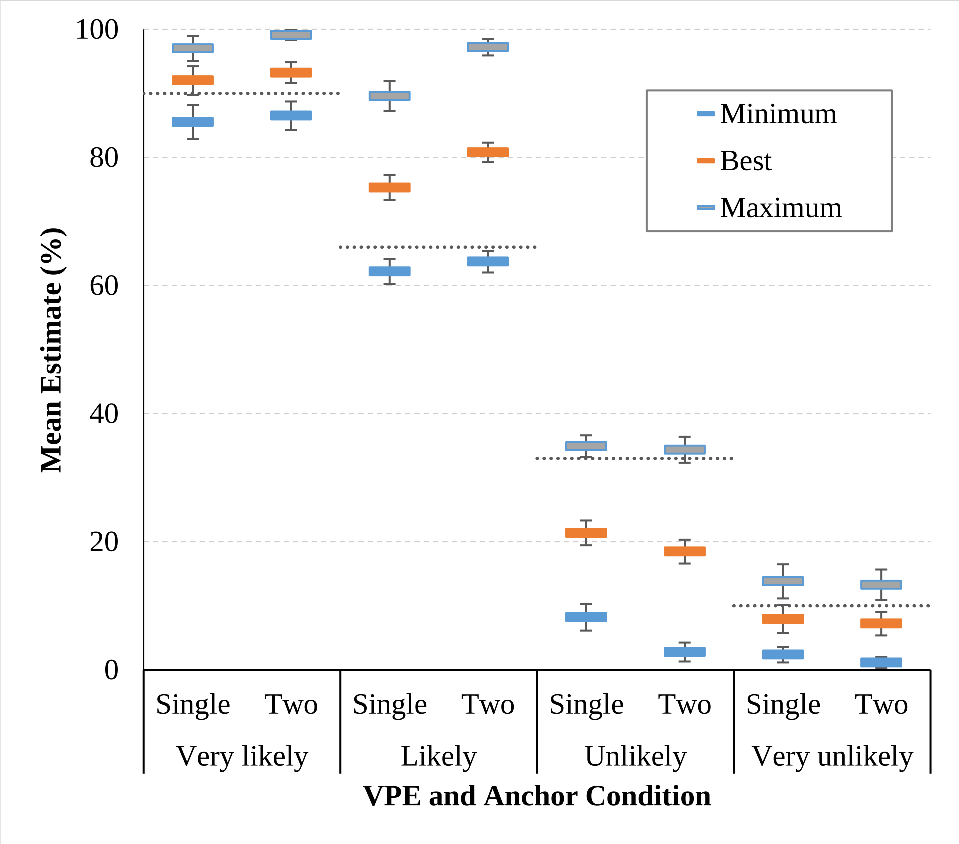

Figure 4. Mean estimates in the IPCC context. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bounds of '(very) likely' and the upper bounds of '(very) unlikely' (Table 1).

Online Resource 1

Numeracy Items used in the questionnaire (questions taken from Frederick, 2005; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006).

No.	Question	Answer
N1.	Imagine that we roll a fair, six-sided die 1,000 times. (That would mean that we roll one die from a pair of dice.) Out of 1,000 rolls, how many times do you think the die would come up as an even number?	Number of times: 500
N2	In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?	Number of people: 10
N3	A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	Cost of ball: $0.05
N4	In a lake, there is a patch of lilypads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how many days would it take for the patch to cover half of the lake?	Number of days: 47
N5	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	Number of minutes: 5

References

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19,* 24-42.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17,* 407-413.

Online Resource 2

Distribution of political affiliations.

Political affiliation	N	Cumulative percent
Strong Right Wing	14	6%
Right to Center	33	21%
Center	56	47%
Center to Left	68	77%
Strong Left Wing	45	98%
Other	5	100%
Didn’t answer question	0	--

Distribution of numeracy scores.

Number of questions correct (/5)	N	Cumulative percent
Zero	6	2.7
One	19	11.3
Two	32	25.8
Three	30	39.4
Four	50	62.0
Five	84	100.0

Online Resource 3

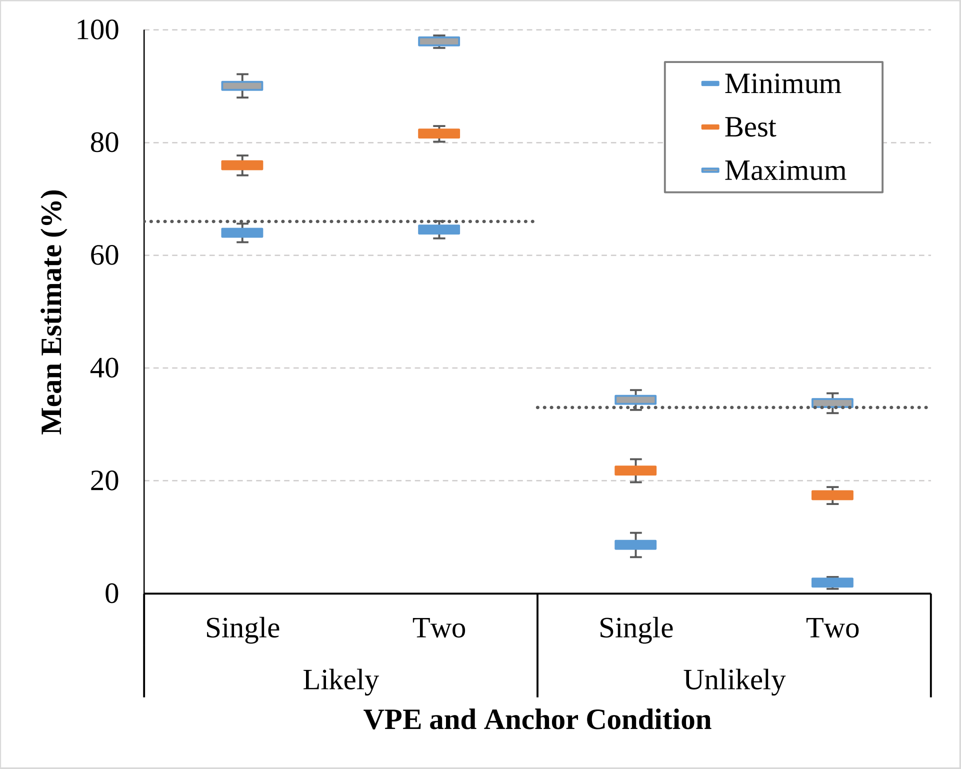

Figure. Mean estimates in the IAS context. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bound of 'likely' and the upper bound for 'unlikely' (Table 1).

Online Resource 4

Online Resource 4: Consistency of estimates with the prescriptions in Table 1.

Best estimates were labelled as consistent if they were within the range prescribed by the IPCC (Table 1). For example, best estimates for ‘unlikely’ were considered to be consistent if they were less than or equal to 33% (see Table A for consistency rates of best estimates)

Ranges were labelled as consistent if both upper and lower bounds were within the prescribed range, as inconsistent if both were outside the prescribed range, and as partially consistent otherwise. For example, endorsed ranges for ‘unlikely’ were considered to be consistent if the higher bound was 33% or lower, inconsistent if the lower bound was higher than 33%, and partially consistent if the lower bound was 33% or lower but the higher bound was greater than 33% (see Table B for consistency rates of endorsed ranges).

Note that there were no significant differences between conditions for any of the individual probability phrases, although all were numerically in the predicted direction.

Table A. *Consistency of best estimates with the prescriptions of Table 1.*

	Single-anchor format		Two-anchor format	
Probability phrase	Consistent	Inconsistent	Consistent	Inconsistent
IPCC very likely	88%	12%	91%	9%
IPCC likely	89%	11%	93%	7%
IPCC unlikely	91%	9%	96%	4%
IPCC very unlikely	91%	9%	92%	8%
IAS likely	92%	8%	97%	3%
IAS unlikely	92%	8%	96%	4%
All terms	90%	10%	94%	6%

Table B. *Consistency of endorsed ranges with the prescriptions of Table 1.*

	Single-anchor format			Two-anchor format		
Probability phrase	Consistent	Partially consistent	Inconsistent	Consistent	Partially consistent	Inconsistent
IPCC very likely	73%	21%	6%	81%	17%	2%
IPCC likely	71%	25%	4%	75%	24%	1%
IPCC unlikely	81%	16%	3%	84%	12%	4%
IPCC very unlikely	78%	16%	6%	82%	15%	3%
IAS likely	75%	23%	2%	79%	20%	1%
IAS unlikely	81%	16%	3%	86%	13%	1%
All terms	77%	19%	4%	81%	17%	2%

Online Resource 5

Additional experiment with university students

Method

Participants

Nineteen male and 98 female (aged 18-22; median = 19 years; IQR = 1.0) first year psychology undergraduates at University College London (UCL) participated in the experiment as part of a course requirement. A further 30 male and 32 female (aged 18-33; median = 20.5 years; IQR = 2.0) social science undergraduates at Carnegie Mellon University (CMU) participated for course credit.

Design and Materials

The design was the same as in the main experiment.

The only difference in the materials was in the introduction to the IPCC and the IAS that participants were presented with prior to a practice trial using “about as likely as not (33-66%).” The translation table presented on this page in the single-anchor condition presented the range with an inequality sign (e.g., “>99%”) instead of in words (i.e., “greater than 99%”). Words were used elsewhere throughout the experiment. Replacing the inequality signs with words was a ‘fix’ employed in the main experiment, to ensure that there could be no confusion with participants misunderstanding the directions of the inequalities.

Procedure

Participants at UCL completed the experiment in two large (approximately equal-sized) groups. Participants at CMU accessed the experiment from the online undergraduate participant pool in their own time. The experiment used the same qualtrics computer program and was completed on individual desktop computers. Participants were provided with the experimental link and all instructions were presented on the computer.

Other than this, all aspects of the procedure were identical to the main experiment except for the fact that no consistency check was conducted, with the age and gender questions being asked at the end of the experiment, with no additional year of birth question.

Results

We follow the same procedure and structure for reporting our results as in the main experiment.

IPCC

Mean 'best estimates' for the four VPEs across both anchor conditions are plotted in Figure A. A visual inspection of Figure A shows that the predicted result of estimates being further from 50% in the two-anchor condition than the single-anchor held, directionally, for the expressions 'likely', 'unlikely' and 'very unlikely.' A 2 (anchor condition) x 4 (VPE) mixed ANOVA revealed a main effect of VPE, $F(1.4, 251.5) = 3579$, $p < .001$, $\text{eta}_p^2 = .95$ (Greenhouse-Geisser correction applied), as well as a main effect of anchor condition, $F(1, 177) = 4.89$, $p = .028$, $\text{eta}_p^2 = .03$. The main effect was qualified by the predicted interaction between VPE and anchor condition, $F(1.4, 251.5) = 4.37$, $p = .024$, $\text{eta}_p^2 = .02$. Simple effects tests showed that the effect was significant only for 'unlikely', $F(1, 704.8) = 12.36$, $p < .001$.

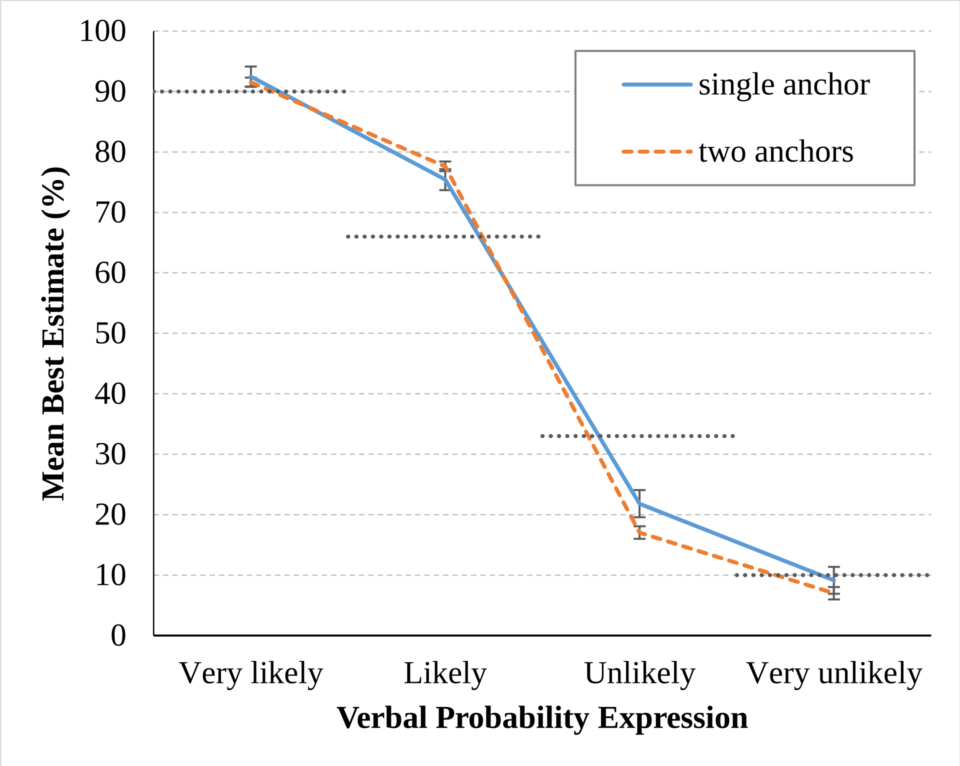

Figure A. Mean ‘best estimates’ provided in response to the IPCC sentences. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bounds of ‘(very) likely’ and the upper bounds of ‘(very) unlikely’ (Table 1).

IAS and IPCC

In the analysis of ‘likely’ and ‘unlikely’ including the IAS context, interpretations of ‘likely’ and ‘unlikely’ did not significantly differ between the contexts: main effect of context, $F < 1$, interaction between context and VPE, $F(1, 177) = 2.77$, $p = .098$, $\text{eta}_p^2 = .015$. Figure B therefore plots the mean estimates for ‘likely and ‘unlikely’ across both anchor conditions, collapsed across context. Directionally, estimates are further from 50% in the two-anchor condition than the single-anchor condition. This result was borne out with a significant VPE x anchor condition interaction, $F(1, 177) = 13.20$, $p < .001$, $\text{eta}_p^2 = .069$. Separate ANOVAs performed on ‘likely’ and ‘unlikely’ suggested that the anchor condition was significant for both: ‘likely’, $F(1, 177) = 7.44$, $p = .007$, $eta_p^2 = .040$; ‘unlikely’, F(1,

177) = 12.14, p = .001, eta_p^2 = .064, with no effects of, or interactions involving, context. Finally, an ANCOVA confirmed that the overall pattern of results was consistent when controlling for numeracy, political affiliation[6], age and gender. Thus, overall these results are consistent with those of the main experiment.

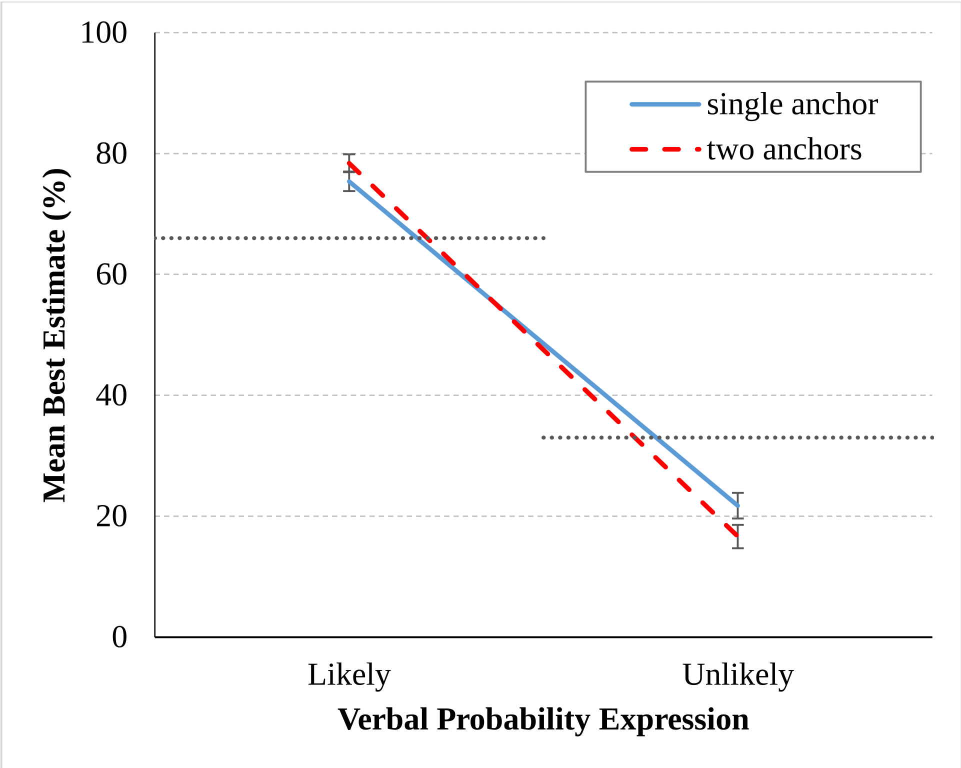

Figure B. Mean 'best estimates' for interpretations of 'likely' and 'unlikely', collapsed across the IPCC and IAS contexts. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bound of 'likely' and the upper bound for 'unlikely' (Table 1).

To better understand the nature of the effect, participants' minimum and maximum estimates of 'likely' and 'unlikely' were analysed. Figure C plots the 'minimum', 'best' and

[6] The distribution of political affiliations and numeracy scores are included in an appendix at the end of this document.

'maximum' estimates across the anchor conditions for these expressions. For 'likely', the maximum estimate increased to a greater degree than the minimum estimate in the two-anchor condition, resulting in a greater endorsed range in the two-anchor condition, as in the main experiment and predicted by the Extended Range hypothesis. This pattern appears less clear for 'unlikely', however, as both the minimum and maximum estimates appear to decrease in the two-anchor condition – a result more consistent with an anchoring account.

To determine the range endorsed by participants for each VPE, participants' minimum estimates were subtracted from their maximum estimates. In a 2 (context) x 2 (VPE) x 2 (anchor condition) ANOVA, a main effect of VPE was observed, $F(1, 177) = 11.89$, $p = .001$, $\text{eta}_p^2 = .063$. The interaction between VPE and anchor condition was again significant, $F(1, 177) = 12.88$, $p < .001$, $\text{eta}_p^2 = .001$. There was no main effect of context, $F(1, 177) = 1.64$, $p = .202$, nor were there any interactions involving context (all Fs < 1, except VPE x context, $F(1, 177) = 1.05$, $p = .307$), suggesting that the documented effects do not systematically differ between the IPCC and IAS contexts. Separate ANOVAs for 'likely' and 'unlikely', revealed that the effect of anchor condition was significant for 'likely', $F(1, 177) = 7.49$, $p = .007$, $\text{eta}_p^2 = .041$, but not for 'unlikely', $F < 1$. As suggested in Figure C, the increased range for 'likely' in the two-anchor condition stems from participants providing higher maximum estimates, $F(1, 177) = 19.62$, $p < .001$, $eta_p^2 = .100$, with no corresponding change in their minimum estimates, $F(1, 177) = 1.36$, $p = .246$. By contrast, the consistent range endorsed for 'unlikely' in the two conditions, coupled with the effect of anchor condition for 'best estimates' shows that both the minimum and maximum estimates were also less regressive in the two-anchor condition (minimum: $F(1, 177) = 15.07$, $p < .001$, $\text{eta}_p^2 = .078$; maximum: $F(1, 177) = 4.41$, $p = .037$, $eta_p^2 = .024$). In an ANCOVA including numeracy, political affiliation, age and gender as covariates, the VPE x anchor condition interaction remained significant, with no interactions or effects of context.[7]

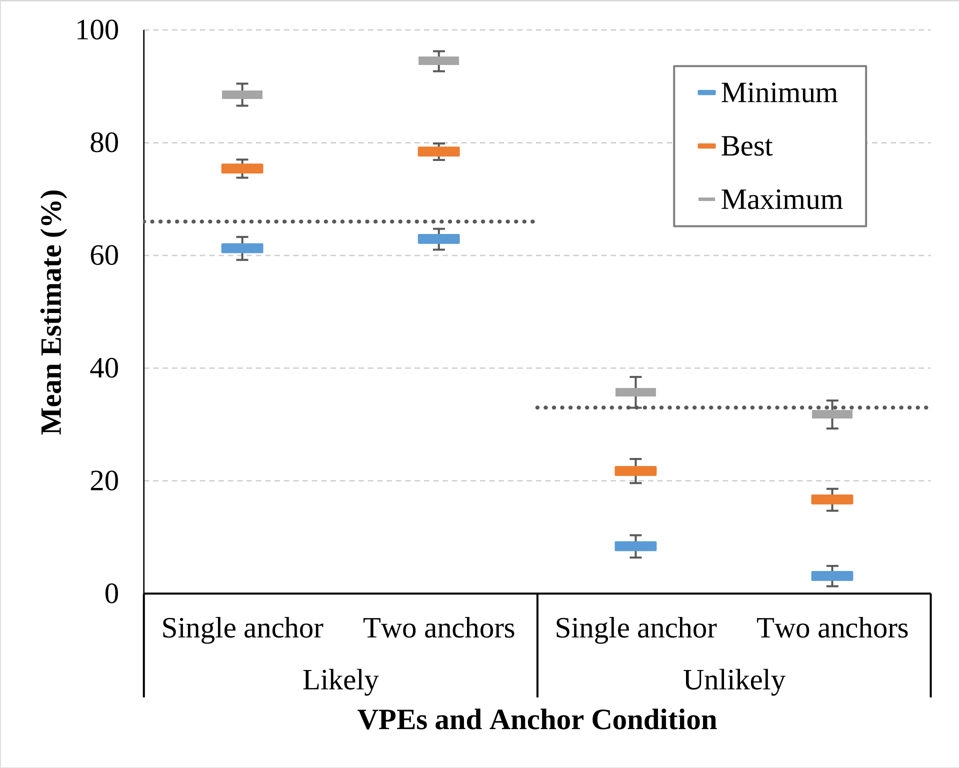

Figure C. Mean estimates for 'likely' and 'unlikely', collapsed across the IPCC and IAS contexts. Error bars are 95% confidence intervals. Dashed horizontal lines represent the prescriptions of the IPCC for the lower bound of 'likely' and the upper bound for 'unlikely' (Table 1).

Consistency with prescribed ranges

Best estimates were labelled as consistent if they were within the range prescribed by the IPCC (Table 1). For example, best estimates for 'unlikely' were considered to be consistent if they were less than or equal to 33% (see Table A for consistency rates of best estimates). Unlike in the experiment reported in the main text, the effect of anchor condition on consistency rates was not significant, $\chi^2(1) = 0.61$, $p = .44$, although the numerical trend

[7] In contrast to the ANOVA, there was no main effect of VPE on endorsed range in the ANCOVA.

was in the predicted direction for all instances apart from 'very likely' in the IPCC context (Table A).

Table A. *Consistency of best estimates with the prescriptions of Table 1.*

	Single-anchor format		Two-anchor format	
Probability phrase	Consistent	Inconsistent	Consistent	Inconsistent
IPCC very likely	83%	17%	82%	18%
IPCC likely	91%	9%	93%	7%
IPCC unlikely	94%	6%	96%	4%
IPCC very unlikely	86%	14%	89%	11%
IAS likely	91%	9%	94%	6%
IAS unlikely	95%	5%	96%	4%
All terms	90%	10%	91%	9%

Ranges were labelled as consistent if both upper and lower bounds were within the prescribed range, as inconsistent if both were outside the prescribed range, and as partially consistent otherwise. For example, endorsed ranges for 'unlikely' were considered to be consistent if the higher bound was 33% or lower, inconsistent if the lower bound was higher than 33%, and partially consistent if the lower bound was 33% or lower but the higher bound was greater than 33% (see Table B for consistency rates of endorsed ranges). Unlike in the experiment reported in the main text, the effect of anchor condition on consistency rates was not significant, $\chi^2(2) = 5.65$, $p = .059$, although the numerical trend was in the predicted direction for all instances (Table B).

Table B. *Consistency of endorsed ranges with the prescriptions of Table 1.*

	Single-anchor format			Two-anchor format		
Probability phrase	Consistent	Partially consistent	Inconsistent	Consistent	Partially consistent	Inconsistent
IPCC very likely	55%	41%	4%	59%	36%	5%
IPCC likely	54%	45%	1%	60%	38%	2%
IPCC unlikely	75%	22%	3%	86%	12%	2%
IPCC very unlikely	67%	28%	5%	77%	20%	3%
IAS likely	58%	37%	5%	64%	36%	0%
IAS unlikely	79%	17%	4%	83%	13%	2%
All terms	65%	32%	3%	72%	26%	2%

Discussion

Generally, the pattern of results was the same as in the main experiment. There were three differences observed in the patterns of results, which we draw attention to here.

Firstly, the regressiveness of participants' best estimates of 'likely' in the IPCC context was not significantly attenuated in this experiment, whilst it was in the main experiment. In the analysis combining the IPCC and IAS data, however, the same result was observed as in the main experiment, and there was no interaction with context. We are therefore confident in our overall conclusion that the two-anchor presentation reduces the regressiveness of ‘best estimate’ interpretations of ‘likely’ and ‘unlikely.’

In the main experiment, we posited that the extended range account seemed to provide a better account of the present findings on the basis of the range of plausible estimates increasing in the two-anchor condition. The present experiment yielded one result that did not follow this pattern. This was for ‘unlikely.’ In this instance, the lower best estimate is not associated with an increase in range, and is also associated with a lower maximum estimate (as well as lower minimum estimate). In the Introduction to the paper, we stated that such a pattern of results was difficult to predict with the extended range account, suggesting a role for anchoring. As we state in the General Discussion of the main manuscript, although we believe that the data are generally more consistent with the extended range hypothesis, we could not have ruled out an influence of anchoring in any case, and this result potentially underscores that point. Nonetheless, the effects themselves seem robust, holding when controlling for the influence of potential covariates across two experiments from two different populations.

Finally, we do not observe a significant effect in the consistency analyses. We have no clear explanation for the difference between the two experiments in this regard, but do note that numerically, across both experiments, the trends were in the predicted direction for 23 out of 24 comparisons. Consequently, we do not perceive there to be a strong discrepancy in the overall conclusions that can be drawn from the two experiments.

Appendix

Distribution of political affiliations in the additional experiment.

Political affiliation	N	Cumulative percent
Strong Right Wing	4	2%
Right to Center	25	16%
Center	58	49%
Center to Left	47	76%
Strong Left Wing	17	85%
Other	26	100%
Didn't answer question	2	--

Distribution of numeracy scores in the additional experiment.

Number of questions correct (/5)	N	Cumulative percent
Zero	2	1.1
One	13	8.4
Two	47	34.6
Three	41	57.5
Four	40	79.9
Five	36	100.0