Numerically Bounded Linguistic Probability Schemes Are Unlikely to Communicate Uncertainty Effectively

D. R. Mandel1, T. S. Wallsten2, and D. V. Budescu3

1Defence Research and Development Canada, Toronto, Ontario, Canada, 2Department of Psychology, University of Maryland, College Park, MD, USA, 3Department of Psychology, Fordham University, New York, NY, USA

Abstract In a recent issue of Earth’s Future (vol. 7, pp. 1020–1026), S. C. Lewis et al. (2019, https://doi.org/10.1029/2019EF001273) recommended a numerically bounded linguistic probability (NBLP) scheme for communicating probabilistic information in extreme event attribution studies. We provide a critique of NBLP schemes in general and of Lewis et al.’s in particular, noting two key points. First, evidence from voluminous behavioral science research on the interpretation of linguistic probabilities indicates that NBLP schemes are an ineffective means of communicating uncertainty to others. Second, where the motivation to implement such schemes nevertheless persists, the schemes should be developed through an evidence-based approach that seeks to optimize interpretational agreement between the scheme and users.

1. Introduction

In recent years, there has been considerable interest in, and much importance attached to, accurately communicating information about climate change and its causes to the public. In that spirit, Lewis et al. (2019) suggest that publications aimed at lay readers employ numerically bounded linguistic probability (NBLP) schemes to convey scientists’ uncertain and probabilistic forecasts and causal statements. (Lewis et al., 2019, use the term calibrated language to refer to such schemes. However, because the term calibrated has disparate technical meanings in related fields, we prefer to avoid the ambiguity, and we denote them with the term NBLP.) Specifically, Lewis et al. (2019) recommend use of a seven-point, ordered scale with probabilistic statements (e.g., “Virtually certain the event would not have happened without climate change”) that have associated risk ratios (RRs) and fractions of attributable risk (FARs), both expressed as intervals. Their recommendation is consistent with similar NBLP schemes used for several decades in a variety of organizational contexts. These include the Intergovernmental Panel on Climate Change (IPCC; Mastrandrea et al., 2011), the U.S. Environmental Protection Agency (Morgan, 1998), and most intelligence organizations, including in all five-eye countries and in NATO (Friedman, 2019; for an early example, see Kent, 1964). The key feature of NBLP schemes is an ordered list of linguistic probability terms or phrases accompanied by numeric ranges intended to define their imprecise meanings.

We believe that in the absence of appropriate empirical work, such schemes are more likely to obfuscate than to illuminate the uncertainties associated with the events in question. The primary aim of communicating scientific thinking about climate and such events to the public should be to maximize the degree to which the end users of communications interpret the message as intended by the writers. Below, we review key research findings that undermine the use of NBLP schemes in general. Then we focus on the Lewis et al. (2019) scheme in particular. Our critiques lead to two recommendations. The first is that communicators of probabilistic information should consider means other than NBLP schemes to improve the clarity and usability of their communications to the relevant audiences. The second point is that when NBLP schemes are used, they should be developed with empirical testing aimed at optimizing the correspondence between end users’ understanding of the focal terms and the stipulated ranges adopted in the schemes. They should not, as is often the case, be constructed by fiat.

2. Problems With NBL Schemes

One problem with using NBLP schemes to describe scientific uncertainties to the public is the well-documented finding that people generally prefer to receive such communications in numerical form, even while they generally prefer to communicate them verbally (Erev & Cohen, 1990; Juanchich & Sirota, 2019;
Mandel (2019) found that individuals perceive linguistic probabilities as communicating probability levels but frequently also imply a directionality that in turn suggests expected actions. For example, terms that convey comparable probability levels can signal optimism (e.g., likely) regarding desirable future events (Teigen & Brun, 1995, 1999). Collins and Mandel (2019) found that individuals perceive linguistic probabilities as communicating probability levels less clearly than do numerical probabilities. They also found that, for low-probability terms, individuals may misinterpret them as assessment terms and not on how to interpret the probability terms. Therefore, the proportion of end users having an interpretation consistent with the IPCC standard after viewing it ranged from 21% to 35% across 25 countries (based on approximately 200 respondents per country). Even when the stipulated numeric ranges were printed beside the linguistic probability terms within the text, the proportion of end users who gave interpretations consistent with the terms ranged from 28% to 54% (again, based on separate samples of approximately 200 respondents per country). (See Figure 3 in Budescu et al., 2014.) These findings were well replicated using a NBLP scheme adopted by the U.S. intelligence community (Wintle et al., 2019), and similar interpretational discrepancies have been documented in frequency based language schemes (Berry et al., 2003). Note, however, that most NBLP schemes do not require numeric ranges to be printed in assessments. Therefore, the proportion of end users that interpret probability terms as stipulated is likely to be small. Even when the ranges are printed alongside the terms, there is a risk that they will be interpreted as assessment-specific credible intervals rather than as assessment-general stipulated ranges. After all, end users are likely to be focused on the substance of the text rather than on how to interpret the probability terms.

Yet another problem with all NBLP schemes, and a likely explanation for the results summarized just above, is that linguistic probabilities are context dependent in a variety of respects, which is to be expected given their function as relative adjectives (Clark, 1990). For instance, their meaning is influenced by perceived base rates of the forecasted events (Wallsten et al., 1986; Weber & Hilton, 1990), event severity (Harris & Corner, 2011), outcome valence (Mandel, 2015; Mullet & Rivet, 1991), and content domain (Brun & Teigen, 1988; Mellers et al., 2017). Such effects imply that the meaning of linguistic probabilities will vary not only across individuals but also within individuals as they encounter identical expressions in different contexts. The Wallsten et al. (1986) study is especially compelling as its respondents were professionals working for the National Weather Service who were using probability terms with assigned probability intervals in their daily work, but when the same expressions were embedded in different contexts, they ignored this translation scheme. To put it bluntly, it is practically impossible to mandate the use of natural language.

Finally, NBLP schemes can implicitly communicate action recommendations, either inadvertently or intentionally but, in any case, inappropriately. This is because linguistic probabilities not only convey probability levels but frequently also imply a directionality that in turn suggests expected actions. For example, terms that convey comparable probability levels can signal optimism (e.g., some chance) or pessimism (e.g., doubtful) regarding desirable future events (Teigen & Brun, 1995, 1999). Another problem with NBLP approaches is the evidence that they often are ineffective. Several studies have shown that after being presented with a NBLP scheme, end users fail to interpret the identified terms within their stipulated ranges. Budescu et al. (2009, 2012, 2014) studied this issue extensively in the context of the IPCC standard. The proportion of end users having an interpretation consistent with the IPCC standard after viewing it ranged from 21% to 35% across 25 countries (based on approximately 200 respondents per country). When the stipulated numeric ranges were printed beside the linguistic probability terms within the text, the proportion of end users who gave interpretations consistent with the terms ranged from 28% to 54% (again, based on separate samples of approximately 200 respondents per country). This is because linguistic probabilities not only convey probability levels but frequently also imply a directionality that in turn suggests expected actions. For example, terms that convey comparable probability levels can signal optimism (e.g., some chance) or pessimism (e.g., doubtful) regarding desirable future events (Teigen & Brun, 1995, 1999). Furthermore, in the climate change risk assessment context, there is a risk that they will be interpreted as assessment-specific credible intervals rather than as assessment-general stipulated ranges. After all, end users are likely to be focused on the substance of the text rather than on how to interpret the probability terms.

Perhaps of greatest interest, however, is our respondents' flexibility in communicating about uncertainty as a function of the nature of the issue and the strength of the data. Generally, people indicated a preference for numerical communication when the situation was important or when numerical estimates were supported by the information base. In contrast, they preferred verbal communication when the situation was unimportant or the information base weak. (p. 138)

This suggests that the schematic translation scheme is not the only route to interpreting linguistic terminologies. Rather, it is the context in which these terminologies are used that is most important. This suggests that communicators need to be cautious in their choice of translation schemes and to ensure that they are consistent with the context in which they are used.

Yet another problem with all NBLP schemes, and a likely explanation for the results summarized just above, is that linguistic probabilities are context dependent in a variety of respects, which is to be expected given their function as relative adjectives (Clark, 1990). For instance, their meaning is influenced by perceived base rates of the forecasted events (Wallsten et al., 1986; Weber & Hilton, 1990), event severity (Harris & Corner, 2011), outcome valence (Mandel, 2015; Mullet & Rivet, 1991), and content domain (Brun & Teigen, 1988; Mellers et al., 2017). Such effects imply that the meaning of linguistic probabilities will vary not only across individuals but also within individuals as they encounter identical expressions in different contexts. The Wallsten et al. (1986) study is especially compelling as its respondents were professionals working for the National Weather Service who were using probability terms with assigned probability intervals in their daily work, but when the same expressions were embedded in different contexts, they ignored this translation scheme. To put it bluntly, it is practically impossible to mandate the use of natural language.

Finally, NBLP schemes can implicitly communicate action recommendations, either inadvertently or intentionally but, in any case, inappropriately. This is because linguistic probabilities not only convey probability levels but frequently also imply a directionality that in turn suggests expected actions. For example, terms that convey comparable probability levels can signal optimism (e.g., some chance) or pessimism (e.g., doubtful) regarding desirable future events (Teigen & Brun, 1995, 1999). Collins and Mandel (2019) found that individuals perceive linguistic probabilities as communicating probability levels less clearly than do numerical probabilities. They also found that, for low-probability terms, individuals...
perceived implicit recommendations more clearly from the verbal probability term than from the numeric probability, even though the communicators provided no explicit recommendations. This tendency can add an influence function, such as nudging policy-makers or the public toward a particular viewpoint or policy, when that is outside the communicator’s mandate (Piercey, 2009).

3. Specific Comments on Lewis et al. (2019)

We turn now to Lewis et al.’s (2019) article, which suggests language for communicating RR or FAR estimated from extreme event attribution (EEA) studies. Two features of this paper merit discussion as illustrations of the problems we outlined above. What makes their paper particularly notable is that the EEA approach provides a model-based means for estimating probabilities of well-defined events having specific effects (National Academies of Sciences, Engineering, and Medicine, 2016). Lewis et al. suggest that lay readers cannot fathom the RR or FAR results per se but will understand the linguistic interpretations they suggest.

The first problem that we address is Lewis et al.’s (2019) specific language choices (see their Table 1 or Figure 1). Five of the seven alternatives seem difficult to interpret. For brevity we illustrate the point with only two of the five: The event was very much more likely due to climate change. The event was exceptionally less likely due to climate change. (We assume the word unlikely in line 6 of Lewis et al.’s Table 1 was a typographical error and they intended the word likely). These are comparative descriptors and beg the question, the event is much more likely or exceptionally less likely due to climate change than due to what? They are not interpretable in an absolute sense.

The bigger point is that the choice of terms is not backed by research. Do people understand them as Lewis et al. (2019) intend? Is their interpretation insensitive to context differences? One cannot just compile a set of probability terms and assume they convey the intended values even if they are explicitly defined (Ho et al., 2015).

Our second concern with Lewis et al.’s (2019) approach is that it seems counterproductive within the EEA framework. Linguistic probabilities are vague, they have context-dependent meanings, and they often convey unintended recommendations to end users, as we pointed out in discussing NBLP schemes in general. But EEA results are typically drawn from scientific research in which the probabilities and uncertainties have already been quantified (Lewis et al., 2019; National Academies of Sciences, Engineering, and Medicine, 2016). Thus, NBLP schemes in this context seem to make the information both more vague and coarse. This effect contradicts sound advice from the European Food Safety Authority (EFSA), which recommends against converting quantitative scientific estimates into linguistic probabilities for public consumption of risk information (European Food Safety Authority et al., 2018).

It is possible that our criticism is misplaced and the lay public does understand a suitably selected set of terms more clearly than they do quantitative RR and FAR values, but this can only be determined with appropriate empirical work, which does not appear to have yet been done.

4. Recommendations

Given the diversity of aims and capabilities among targeted audiences, such as the media, policy-makers, and the public, we caution against the use of communication strategies that presume how the information communicated will be understood. In some important cases, quantitative information can better support decision-making than can linguistic information. For instance, when end users receive multiple estimates from different sources (e.g., different advisors), they may wish to know the aggregate estimate, such as the mean or median value. That easily can be done with numbers, but not with probability words. Recent research shows that accuracy is greater when nonexpert end users estimate averages and products from numeric than from linguistic probabilities (Mandel et al., 2019). Visual methods, such as icon arrays, have been shown to improve comprehension of risk information by less numerate end users (Galesic et al., 2009) as well as by those with relatively low graph literacy (Okan et al., 2015).

Although we caution against the use of NBLP schemes, we realize that some organizations will nevertheless adopt them. Accordingly, we recommend that such schemes draw on relevant empirical research on linguistic probability interpretation and be further empirically tested before being adopted. For instance,
Ho et al. (2015) demonstrated that by eliciting numeric translations of probability terms in the IPCC and in various intelligence community NBLP schemes, they could improve the degree of agreement between end users’ interpretations of the probability terms and the stipulated ranges in these schemes (for a conceptual replication, see Wintle et al., 2019). Usually, however, as in national security intelligence (Dhami et al., 2015), NBL schemes are devised by fiat by a small group of insiders, without a clear methodological justification and without empirical testing.

Data Availability Statement

Data were not used nor created for this research.

References

BERRY, D., RAYNER, D. K., KAPP, P., & BERSELLINI, E. (2003). Patients’ understanding of risk associated with medication use: Impact of European Commission guidelines and other risk scales. Drug Safety, 26(1), 1–11. https://doi.org/10.2165/00002018-200302601-00001

BRUN, W., & TEIGEN, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? Organizational Behavior and Human Decision Processes, 41(3), 390–404. https://doi.org/10.1016/0749-5978(88)90036-2

BRUDER, D. V., BROWNELL, S., & POR, H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. Psychological Science, 20(3), 299–308. https://doi.org/10.1111/j.1467-9280.2009.02284.x

BRUDER, D. V., POR, H., & BROWNELL, S. B. (2012). Effective communication of uncertainty in the IPCC reports. Climatic Change, 113(2), 181–200. https://doi.org/10.1007/s10584-011-0530-3

BRUDER, D. V., POR, H., BROWNELL, S. B., & SMITHSON, M. (2014). The interpretation of IPCC probabilistic statements around the world. Nature Climate Change, 4(6), 508–512. https://doi.org/10.1038/nclimate2394

CLARK, H. H. (1990). Quantifying probabilistic expressions: Comment. Statistical Science, 5(1), 12–16. https://doi.org/10.1214/ss/1177012243

COLLINS, R. N., & MANDEL, D. R. (2019). Cultivating credibility with probability words and numbers. Judgment and Decision making, 14(6), 683–695.

DHAMI, M. K., MANDEL, D. R., MELLERS, B. A., & TETLOCK, P. E. (2015). Improving intelligence analysis with decision science. Perspectives on Psychological Science, 10(6), 753–757. https://doi.org/10.1177/1745691615598511

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. Organizational Behavior and Human Decision Processes, 45(1), 1–18. https://doi.org/10.1016/0749-5978(90)90002-Q

European Food Safety Authority, Hart, A., Maxim, L., Siegrist, M., von Goetz, N., da Cruz, C., et al. (2018). Guidance on communication of uncertainties: A common approach across the working groups. Decision Processes, 5(1), 67–171. https://doi.org/10.1037/cop0000127

FRIEDMAN, J. A., BAKER, J. D., MELLERS, B. A., TETLOCK, P. E., & ZECKHAUSER, R. (2018). The value of precision in probability assessment: Evidence from a large-scale geopolitical forecasting tournament. International Studies Quarterly, 62(2), 410–422. https://doi.org/10.1093/isq/sqx078

GALESIC, M., GARCIA-RETMERO, R., & GIGERENZER, G. (2009). Use of icon arrays to communicate medical risks: Overcoming low numeracy. Health Psychology, 28(2), 210–216. https://doi.org/10.1037/a0014474

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(6), 1571–1578. https://doi.org/10.1037/a0024195

HO, E., BRUDER, D. V., DHAMI, M. K., & MANDEL, D. R. (2015). Improving the communication of uncertainty in climate science and intelligence analysis. Behavioral Science & Policy, 2(2), 43–55. https://doi.org/10.1033/j.bsp.2015.0015

JUANICH, M., & SIORTA, M. (2019). Do people really prefer verbal probabilities? Psychological Research Psychologische Forschung, 84(8), 2325–2338. https://doi.org/10.1007/s00426-019-02107-0

KEST, S. (1964). Words of estimative probability. Studies in Intelligence, 8(4), 49–65.

LEWIS, S. C., KING, A. D., GREENE, W. S., & RICE, M. (2019). Toward calibrated language for effectively communicating the results of extreme event attribution studies. Earth’s Future, 7, 1020–1026. https://doi.org/10.1002/2019EF001273

MANDEL, D. R. (2015). Accuracy of intelligence forecasts from the intelligence consumer’s perspective. Policy Insights From the Behavioral and Brain Sciences, 2(1), 111–120. https://doi.org/10.1093/pibs/2015.1.0127

MANDEL, D. R., DHAMI, M. K., TRAN, S., & IRWIN, D. (2019). Arithmetic computation with probability words and numbers. PsyArXiv. https://doi.org/10.31234/osf.io/hduf4

MASTRANDREAS, M. D., MACH, K. J., PLOTTNER, G., & MELLERS, B. A. (2015). Improving intelligence analysis with decision science. Journal of Experimental Psychology: Applied, 21(2), 178–194. https://doi.org/10.1037/xap0000045

MORGAN, M. G. (1998). Commentary: Uncertainty analysis in risk assessment. Human and Ecological Risk Assessment, 4(1), 25–39. https://doi.org/10.1080/10807039.1998.11906880

MUHLE, T., & RIVET, J. L. (1991). Comprehension of verbal probability expressions in children and adolescents. Language & Communication, 11(3), 217–225. https://doi.org/10.1016/0163-6383(91)90007-1

MURPHY, A. H., LIGHTENSTEIN, S., FISCHOFF, B., & WINKLER, R. L. (1980). Misinterpretation of precipitation probability forecasts. Bulletin of the American Meteorological Society, 61(7), 695–701. https://doi.org/10.1175/1520-0477(1980)061<0695:MOPPPF>2.0.CO;2

National Academies of Sciences, Engineering, and Medicine (2016). Attribution of extreme weather events in the context of climate change. https://doi.org/10.17226/21852

OKUZ, Y., GARCIA-RETMERO, R., COLEY, E. T., & MALDONADO, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. Journal of Experimental Psychology: Applied, 21(2), 178–194. https://doi.org/10.1037/xap0000045

Funding for this research was provided by Canadian Safety and Security Program Project CSSP-2018-TI-2394 to the first author. We thank Daniel Irwin for his research assistance. We also thank Karl Teigen and Marie Juanchich for their constructive comments on an earlier draft of this article.

Acknowledgments

Data were not used nor created for this research.
Olson, M. J., & Budescu, D. V. (1997). Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making, 10*(2), 117–131. https://doi.org/10.1002/(SICI)1099-0771(199706)10:2<117::AID-BDM251>3.0.CO;2-7

Piercey, M. D. (2009). Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organizational Behavior and Human Decision Processes, 108*(2), 330–341. https://doi.org/10.1016/j.obhdp.2008.05.004

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica, 88*(3), 233–258. https://doi.org/10.1016/0001-6918(93)E0071-9

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes, 80*(2), 155–190. https://doi.org/10.1006/obhd.1999.2857

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31*(2), 135–138. https://doi.org/10.3758/BF03334162

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language, 25*(5), 571–587. https://doi.org/10.1016/0749-596X(86)90012-4

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 781–789. https://doi.org/10.1037//0096-1523.16.4.781

Winkle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E., & Fidler, F. (2019). Verbal probabilities: Very likely to be somewhat more confusing than numbers. *PLoS ONE, 14*(4), e0213522. https://doi.org/10.1371/journal.pone.0213522