improving the communication of uncertainty in climate science and intelligence analysis

Emily H. Ho, David V. Budescu, Mandeep K. Dhami, & David R. Mandel

abstract

Public policymakers routinely receive and communicate information characterized by uncertainty. Decisions based on such information can have important consequences, so it is imperative that uncertainties are communicated effectively. Many organizations have developed dictionaries, or *lexicons,* that contain specific language (e.g., *very likely*, *almost certain*) to express uncertainty. But these lexicons vary greatly and only a few have been empirically tested. We have developed evidence-based methods to standardize the language of uncertainty so that it has clear meaning understood by all parties in a given communication. We tested these methods in two policy-relevant domains: climate science and intelligence analysis. In both, evidence-based lexicons were better understood than those now used by the Intergovernmental Panel on Climate Change, the U.S. National Intelligence Council, and the U.K. Defence Intelligence. A well-established behavioral science method for eliciting the terms' full meaning was especially effective for deriving such lexicons.

Improving the communication of uncertainty in climate science and intelligence analysis

Emily H. Ho, David V. Budescu, Mandeep K. Dhami, & David R. Mandel

absract.   Public policymakers routinely receive and communicate information characterized by uncertainty. Decisions based on such information can have important consequences, so it is imperative that uncertainties are communicated effectively. Many organizations have developed dictionaries, or *lexicons,* that contain specific language (e.g., *very likely, almost certain*) to express uncertainty. But these lexicons vary greatly and only a few have been empirically tested. We have developed evidence-based methods to standardize the language of uncertainty so that it has clear meaning understood by all parties in a given communication. We tested these methods in two policy-relevant domains: climate science and intelligence analysis. In both, evidence-based lexicons were better understood than those now used by the Intergovernmental Panel on Climate Change, the U.S. National Intelligence Council, and the U.K. Defence Intelligence. A well-established behavioral science method for eliciting the terms' full meaning was especially effective for deriving such lexicons.

Decisions are often based on judgments made under conditions of uncertainty. That is true when people answer low-stakes questions such as, what is the chance it will rain tomorrow? It is also true with high-stakes national security queries such as, how likely is Russia's ground presence in Syria to trigger a military confrontation between the United States and Russia? And it applies to environmental queries with policy implications such as, if CO_2 emissions continue at current levels, what are the chances that rising sea levels will force a major population evacuation in Indochina in the next 50 years? Despite such high-stakes contexts, uncertainties are often communicated inappropriately, if at all.[1] In fact, the language of uncertainty may itself be a source of confusion.

Uncertainties can be communicated as precise values ("there is a 0.5 chance"), as ranges ("the probability is between 0.3 and 0.6"), as phrases ("it is not very likely"), or as a combination of phrases and ranges of numbers ("it is likely [between 0.60 and 0.85]").[2] But research has

shown that people overwhelmingly prefer to communicate uncertainty using vague verbal terms such as *almost certain* because these terms are perceived to be more intuitive and natural.[3,4] People may avoid the alternative of precise numerical values because they can imply a false sense of precision, particularly for scenarios in which uncertainty persists.[5] For example, in the legal domain, efforts to communicate the meaning of terms such as *reasonable doubt* focus on using other vague language (for example, phrases such as *firmly convinced*), partly because using numerical values (such as *90%*) may impose greater accountability and expose errors in judgment.[6,7]

People naively assume that others share their interpretation of the phrases they use to convey uncertainty. But research shows that interpretations of such phrases vary greatly across individuals.[8] This underappreciation of variability in people's intuitive understanding of phrases used to convey probability ultimately undermines communication.[9] Given the serious problems associated with communicating uncertainty using verbal terms, researchers have suggested that people either reduce and restrict the use of such terms or develop dictionaries, or *lexicons,* that tie the verbal terms to specific numerical values or ranges.[10,11] Indeed, some organizations have used such standardized lexicons with mixed results. In some cases, organizations develop multiple lexicons that assign different meanings to the same terms.[12]

For instance, Sherman Kent, a cofounder of the Central Intelligence Agency's Office of National Estimates, proposed the use of a standardized lexicon to reduce vagueness in the communication of uncertainty in intelligence estimates. The lexicon he proposed, however, had some limitations. For example, it contained numerical gaps and it was not based on systematic research on how analysts or intelligence users interpret these terms. More recently, the European Commission Pharmaceutical Committee released a guideline for communicating the risk of side effects of over-the-counter medications. However, research revealed that the language in the guideline did not match people's understanding of the terms.[13]

With a few exceptions,[14,15] these uncertainty lexicons are developed by fiat and reflect the perceptions, perspectives, and experiences of small committees of experts in a given field. Rarely do they adequately consider the wide diversity of backgrounds and perspectives of target audiences. It is no surprise that instead of enabling clear communication of uncertainty, such lexicons can be confusing and ultimately result in ill-informed decisions. We argue that when developing a new uncertainty lexicon or testing existing ones, research must focus on demonstrating the reliability and validity of evidence-based methods. However, few studies have done this (see reference 2).[16]

Our research strongly suggests that behavioral science can help people better communicate decision-critical uncertainties. In two studies, we established alternative approaches to developing lexicons and tested their effectiveness in communicating uncertainty in two domains: climate science and intelligence analysis. Linking phrases to numerical probabilities and then confirming that the phrases are understood accurately by target audiences is a promising approach to making murky communications more precise and reliable and therefore more meaningful. In our first study, we showed that our evidence-based lexicons are more effective than the lexicon used in the reports of the Intergovernmental Panel on Climate Change (IPCC) for communicating scientific results and conclusions to the public. Our second study applies a similar approach to communicating uncertainty in intelligence analysis among professional analysts. In both cases, evidence-based uncertainty lexicons improved the effectiveness of communication for experts and nonexperts alike.

Conveying Uncertainty in Climate Science

Climate change is one of the major challenges facing our society in the 21st century. The IPCC was assembled to collect and disseminate information about the causes and potential impacts of climate change, and strategies for mitigation and adaptation in response.[17] Climate science is complex, technical, and interdisciplinary. Projections about future temperatures, precipitation, sea levels, and storm surges are subject to uncertainties associated with a variety of variables, including physical (for example, climate sensitivity), social (for example, population growth rates), and economic (for example, the cost of reducing rates of greenhouse gas emissions). These uncertainties can influence important policy decisions. For example, organizations that seek to acquire land to protect the habitat of certain species

must decide which land to purchase to maximize the species' chances of survival. Such decisions rely on projections of future temperatures and precipitation in various locations, among many other factors.

Previous attempts to effectively communicate relevant climate-science results have been plagued by problems, such as increasing public confusion by not explicitly using uncertainty phrases, as was the case in the first three IPCC Assessment Reports.[18] The solution adopted by the IPCC in its reports was to use a set of verbal phrases to convey information about the relevant probabilities. For example, "It is *very likely* [emphasis added] that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent" or "It is *very unlikely* [emphasis added] that the Meridional Overturning Circulation [the system of global ocean currents] will undergo a large abrupt transition during the 21st century." To ensure that its verbal phrases were interpreted as intended, the IPCC published a conversion table that assigns numerical values to certain phrases (see Table 1). For example, in the IPCC lexicon, the term *very likely* denotes a likelihood of greater than 90% and the term *unlikely* denotes a likelihood of less than 33%.

Some researchers have argued that the IPCC's conversion table is ineffective, mainly because the lexicon is not grounded in people's intuitive and consensual understanding of what the phrases mean.[19] In Study 1, we developed two uncertainty lexicons that map phrases used by the IPCC to specific numerical ranges. The lexicons are evidence-based because we developed them using people's actual interpretations of probability phrases in the context of climate science.

Constructing Evidence-Based Lexicons

To construct evidence-based lexicons in the climate science domain, we reanalyzed survey data from participants in the United States, Australia, and the United Kingdom who were part of a large international study (see details in reference 3). Participants read eight sentences from the fourth IPCC Assessment Report. The sentences included the four probability phrases that were most frequently used in IPCC reports (*very unlikely, unlikely, likely, very likely*) and thus were the most relevant for policy efforts.[20] For example, one sentence read, "It is very unlikely that climate changes of at least the seven centuries prior to 1950 were due to variability generated within the climate system alone." Participants had access to the IPCC's conversion table when reading the sentences.

After reading each sentence, participants were asked to characterize each phrase's intended numerical meaning in the sentence by estimating its lower and upper bounds and offering their best estimate for its specific meaning. Later, participants were presented with the same four phrases again, outside the context of sentences, and were asked to indicate the same three numerical values. (Detailed descriptions of the methods and analysis used in this research are in our Supplemental Material published online.)

For both the contextualized and the stand-alone estimation tasks, we discarded cases where a participant's best estimate of a phrase's value fell outside of the participant's estimates of its upper and lower boundaries. Consequently, we used the stand-alone estimates provided by participants (n = 331 to 352,

Table 1. Abbreviated Intergovernmental Panel on Climate Change (IPCC) lexicon for translation of probability phrases and two evidence-based lexicons (Study 1)

Phrase	IPCC likelihood	Evidence-based methods	
		Peak Value (PV)	Member Function (MF)
Very likely	>90	65–100	75–100
Likely	>66	45–65	40–75
Unlikely	<33	15–45	15–40
Very unlikely	<10	0–15	0–15

All values represent percentages.

Figure 1. Illustration of determination of optimal cutoff points between two adjacent phrases using the PV and MF methods (Study 1)
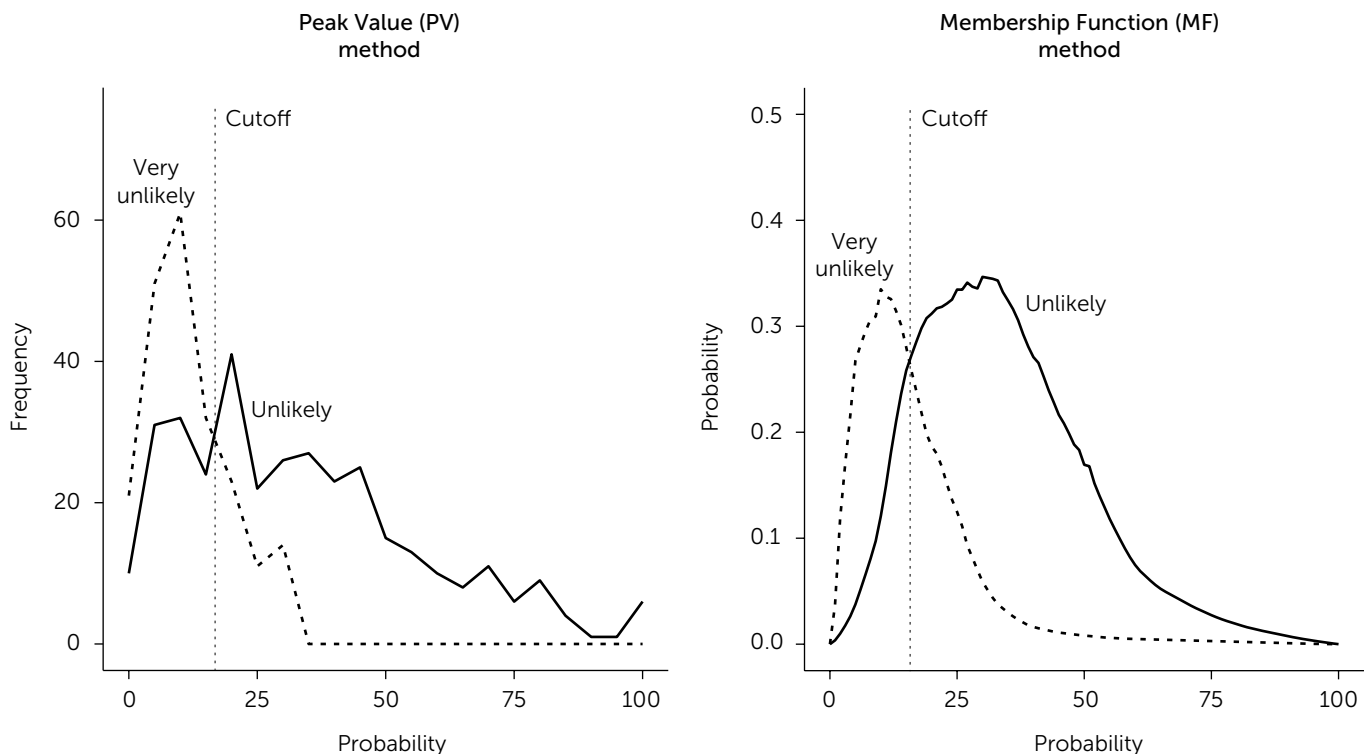
depending on the phrase) in the U.S. sample to construct two lexicons. Our primary goal was to identify regions of probability that maximize participants' consensus about the probability phrases' meaning. To do this, we applied two evidence-based approaches to associate each probability phrase with a range of numerical probabilities.

One approach, which we call the *peak value* (PV) method, reflects the distribution of participants' best estimates of the value of each phrase when it was presented alone, outside of the context of a sentence. The distributions of participants' best estimates of each phrase's numerical meanings were plotted, and we used the points where the distributions intersected to determine the cutoffs between adjacent phrases. The left panel of Figure 1 illustrates this process with the distributions of the phrases *very unlikely* and *unlikely*.

The second evidence-based approach we used is known as the *membership function* (MF) method, which uses a full description of the subjective meaning of a term. This is based on the work of Wallsten, Budescu, Rapoport, Zwick, and Forsyth, who demonstrated that

it is possible to quantify the meaning of probability terms by means of MFs that describe how well a certain numerical value defines or substitutes for a given phrase (for example, how well a probability of 0.20 defines the phrase *very unlikely*).[21] A membership of 0 denotes that the probability value does "not at all" substitute for (define) the phrase, whereas a membership of 1 indicates that a probability value "absolutely" substitutes for the phrase. A value with a membership of 1 is referred to as the *peak of the MF,* and an intermediate value reflects partial membership. The MF approach has been shown to be a reliable and valid method of measuring people's understanding and use of probability phrases (see references 10 and 12).[22,23]

Using each participant's three reported values of each probability phrase (lower bound, upper bound, and best estimate) when the phrases were presented alone, we computed MFs for each person and each term. The individual MFs of each term were averaged to obtain the sample's collective MF for each phrase. The optimal cutoff points between adjacent phrases were obtained by identifying the region of values for which the sample

Figure 2. Comparison of bounds in the IPCC guidelines with bounds from the evidence-based lexicons from Study 1
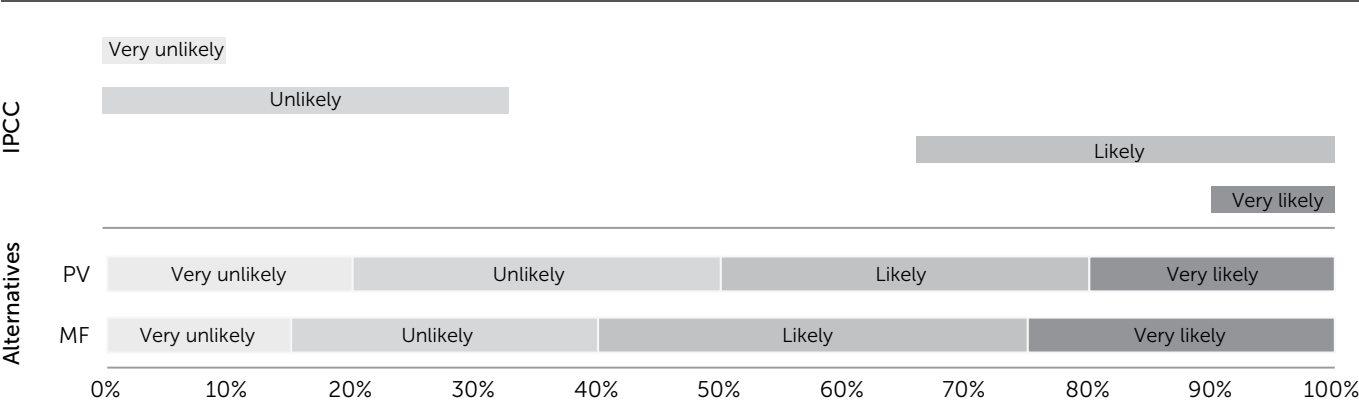
IPCC = Intergovernmental Panel on Climate Change; PV = peak value; MF = membership function. A color version of this figure is published in the Supplemental Material online.

membership of a given phrase was higher than all other phrases (see right panel of Figure 1).

Figure 2 displays the numerical cutoffs that the IPCC prescribes for each of the four probability phrases examined in Study 1 (*very unlikely, unlikely, likely, very likely*), as well as the values identified by the two evidence-based lexicons (based on context-free judgments of participants in the U.S. sample). The most prominent finding is that the IPCC's ranges for *very unlikely* and *very likely* are much narrower and more extreme (closer to the end points, 0 and 1) than are participants' intuitive and natural interpretations of these phrases.

Testing Evidence-Based Lexicons
in the Climate Science Domain

To compare how effectively the two evidence-based lexicons and the existing IPCC guidelines convey information about uncertainty, we analyzed evaluations of the phrases in the eight IPCC sentences, using the responses of the Australian and U.K. samples to the IPCC sentences. We were primarily interested in how consistently participants' best estimates of a phrase's numerical interpretation fell within the numerical range that defines that phrase in each of the three lexicons (IPCC, PV method, and MF method). These consistency rates can range from 0 to 1 (for example, a consistency rate of .3 indicates that 30% of the participants' estimates fell within the specified range).

For both the Australian and the U.K. samples, consistency rates were calculated separately for each phrase across all of the participants, which yielded an overall measure of lexicon consistency. The results of this analysis are shown in Table 2. The mean consistency rates (across the U.K. and Australian samples) were 40% for the PV method and 43% for the MF method. The evidence-based lexicons clearly outperformed the current IPCC lexicon in both samples (where the consistency rate was 27% for the U.K. sample and 25% for the Australian sample), even though participants had access to the IPCC lexicon.

Table 2. Consistency rates of Study 1

Sample	Method	Mean Consistency
Evidence-based lexicon		
United Kingdom	PV	44%
United Kingdom	MF	50%
Australia	PV	41%
Australia	MF	45%
Current IPCC lexicon		
United Kingdom		27%
Australia		25%

Sample sizes in the United Kingdom vary between 162 and 177 across the four terms. In Australia, they vary between 198 and 217 across the four terms. PV = peak value; MF = membership function; IPCC = Intergovernmental Panel on Climate Change.

Communicating Uncertainty in Intelligence Assessments

We tested the robustness of the evidence-based methods proposed in Study 1 by applying them to a second domain, intelligence analysis. Intelligence analysis plays a vital role in national security and defense decisionmaking. And, like climate science, intelligence assessment is characterized by uncertainty.[24] For example, Mandel and Barnes reported that only 29.5% of 1,514 strategic intelligence forecasts with quantified uncertainties implied certainty (that is, the analysts assigned probabilities of either 0 or 1 to event occurrence).[25] In a study of intelligence organizations, one manager of intelligence analysts stated, "There is a huge problem of language used to convey probability and importance/magnitude in terms of what the expressions mean to different people" (p. 23).[26,27]

After the 9/11 terrorist attacks, there were many calls for the intelligence community to include more explicit information about uncertainties surrounding forecasts and judgments in intelligence reports.[28] The National Intelligence Council (NIC), which is responsible for long-term strategic planning in the Office of the Director of National Intelligence,[29] developed a verbal probability lexicon that ranked eight terms used for communicating uncertainty but did not associate the terms with numerical values.[30] Recently, the Office of the Director of National Intelligence updated this lexicon to include numerical ranges, the efficacy of which remains untested. The NIC's counterpart in the United Kingdom, the Defence Intelligence (DI),[31] developed a different six-category lexicon in which phrases were translated into numerical ranges, although there are some numerical ranges for which there is no probability phrase (for example, 85%–90%).[32,33] Table 3 lists both institutional lexicons. Given that the United Kingdom and the United States are close and longtime NATO allies, it is startling that the lexicons of their respective intelligence organizations disagree for every phrase in the lexicon. It is equally puzzling that neither lexicon relies on systematic empirical research and that the communicative effectiveness of both lexicons is yet to be ascertained.

Constructing Evidence-Based Lexicons

To construct an evidence-based lexicon in the intelligence-assessment domain, we used the PV and MF methods described earlier. We recruited 34 Canadian intelligence analysts who were completing an intermediate intelligence course to serve as our initial calibration sample.

The analysts rated the degree to which specific numerical probabilities (from 0% to 100% in increments of 10%) can substitute for each of eight probability phrases that are used in both the NIC and the DI lexicons (*remote chance, very unlikely, unlikely, even chance, likely, probably, very likely,* and *almost certainly*). The ratings were on a scale from 0 (*not at all*) to 100 (*absolutely*), with each 10th point labeling an interval. Thus, each participant provided 11 ratings for each phrase, a refinement of the three ratings (upper and lower bounds and best estimate) used in the IPCC study. By simply linking the 11 ratings of each term—one for each probability—one can trace the term's MF.

Following a procedure similar to that used in the IPCC study, we used the Canadian analysts' responses to derive two lexicons, using the PV and MF methods to calculate value ranges for each of the eight phrases. We recruited 27 U.K. intelligence analysts to serve as a

Table 3. National Intelligence Council and Defence Intelligence lexicons (Study 2)

National Intelligence Council		Defence Intelligence	
Phrase	Numerical value (%)	Phrase	Numerical value (%)
Remote	1–5	Remote/highly unlikely	<10
Very unlikely	5–20	Improbable/unlikely	15–20
Unlikely	20–45	Realistic possibility	25–50
Even chance	45–55		
Probably/likely	55–80	Probable/likely	55–70
Very likely	80–95	Highly probable/very likely	75–85
Almost certainly	95–99	Almost certain	>90

Table 4. Mean peaks of group membership functions for all probability phrases (Study 2)

Sample	Remote chance	Very unlikely	Unlikely	Even chance	Likely	Probably	Very likely	Almost certainly
Canada								
mean	23.0	19.4	30.6	47.9	69.1	70.0	82.9	83.5
standard deviation	31.0	25.0	28.7	23.6	24.8	26.8	17.3	23.7
United Kingdom								
mean	16.8	18.2	28.3	50.0	74.4	78.8	82.9	88.6
standard deviation	31.5	26.9	27.8	18.2	19.5	18.7	11.3	15.1

In the validation sample, the phrase *highly unlikely* is a proxy for the phrase *very unlikely*. The sample size varied between 30 and 32 for various terms in the Canadian sample and between 24 and 25 for the United Kingdom sample. Because of an administrative error in our materials, *very unlikely* judgments were not collected, and we use the phrase *highly unlikely* as a substitute.

validation sample. They performed a similar task (see reference 32, Study 1).

Table 4 presents the mean peak MFs—the best numerical representations of the phrases' meanings— for the calibration and validation samples. It is reassuring that the order of both sets of mean estimates is consistent with the implied order of the phrases in the NIC and DI lexicons. An interesting feature of the evidence-based approach is that, in addition to determining optimal cutoff points between adjacent terms, it is possible to control the size of the lexicon by changing the number of phrases used in the lexicon. We illustrate this in two ways.

Synonyms. The NIC and DI lexicons deem certain phrases to be interchangeable (such as *probably* and *likely* in the NIC lexicon and *remote* and *highly unlikely*

in the DI lexicon), meaning they can be used to represent the same numerical range (see Table 3). We examined the validity of this assumption by comparing data provided by participants in the Canadian calibration sample. Specifically, we compared the average MFs for three phrase pairs in our evidence-based lexicon (*remote chance/very unlikely*, *probably/likely,* and *very likely/almost certain*). As shown in Figure 3, the items in each of these pairs are, for all practical purposes, indistinguishable and thus can be treated as synonyms. In light of this result, we accept the determination of implicit equivalence between terms in the NIC or DI in our evidence-based lexicons.

Abbreviated Lexicons. Can a simplified lexicon containing fewer but most frequently used terms provide better differentiation between adjacent phrases

Figure 3. Comparison of average membership function estimates for *remote chance* and *very unlikely* (left panel), *likely* and *probably* (middle panel), and *very likely* and *almost certainly* (right panel) in the Canadian sample (Study 2)
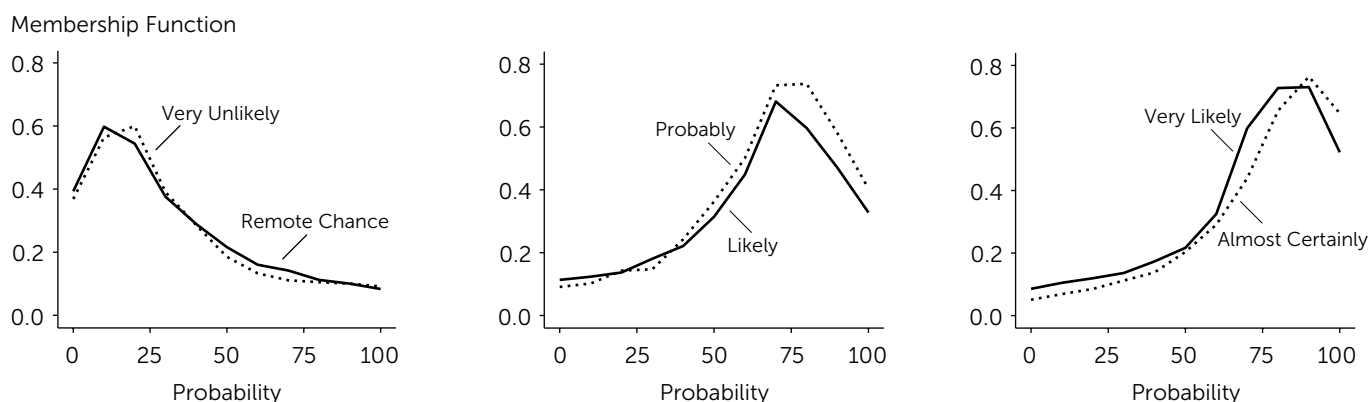
and more effective communication? To test this possibility, we analyzed the judgments of the Canadian analysts while excluding the terms *remote chance* and *almost certain*, which are rarely used, and (using the same procedures) derived a shorter, simplified lexicon.

Testing Evidence-Based Lexicons in the Intelligence Analysis Domain

As shown in Figure 4, probability phrases' numerical ranges in the DI lexicon are generally narrower than the ranges in the two evidence-based lexicons derived from responses provided by the Canadian analysts (our calibration sample), suggesting that the analysts attach broader meaning to the phrases in the middle of the lexicon than the creators of the DI lexicon assumed. In addition, the phrase ranges in the NIC are narrower at the extreme ends of the scale. On the whole, both institutional lexicons map onto different numerical values than do the same phrases in our two evidence-based lexicons.

The ranges in our two lexicons (MF and PV) are similar (see Figure 4). The most noticeable difference is that the MF method induces narrower and more extreme meanings to the end phrases (*remote chance*

and *almost certainly*), which is consistent with the evidence in the literature.[34] The abbreviated lexicon, which excludes these extreme phrases, eliminates this difference (see Figure 4).

To compare how frequently participants' subjective judgments of the numerical value of a phrase fell within the boundaries established by each lexicon, we computed consistency rates for each phrase within each lexicon. This analysis was based on the validation sample of U.K. analysts. Surprisingly, the U.K. analysts' judgments of the phrases were much more in line with the U.S. NIC lexicon than with the U.K. DI lexicon for most phrases, with the exception of the most extreme phrases (*remote chance* and *almost certainly*).

Figure 5 presents consistency rates for the PV and MF methods, both for the complete lexicon of seven phrases and the abbreviated version of five phrases. For most phrases, the intelligence analysts' judgments showed, on average, higher consistency with the evidence-based lexicons developed with the full MFs (79%) than with both the NIC's (53%) and the DI's (56%) existing lexicons. The PV-based empirical lexicon (58%) is only slightly better than its NIC and DI counterparts. More specifically, we see a much higher consistency rate for the extreme phrases in the evidence-based methods,

Figure 4. Abbreviated and optimal thresholds compared with Defence Intelligence (DI) and National Intelligence Council (NIC) lexicons (Study 2)
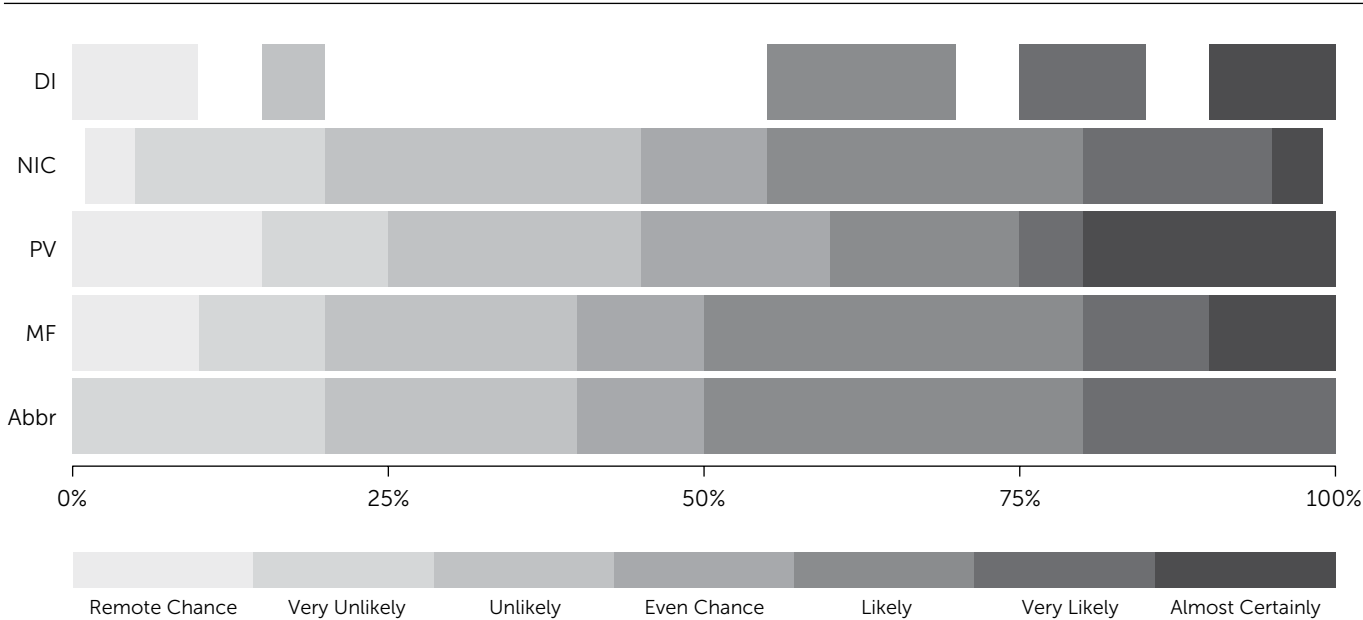
PV = peak value; MF = membership function; Abbr = abbreviated lexicon.

Figure 5. Consistency rates of National Intelligence Council (NIC), Defence Intelligence (DI), and evidence-based lexicons for the full and abbreviated evidence-based lexicons (Study 2)
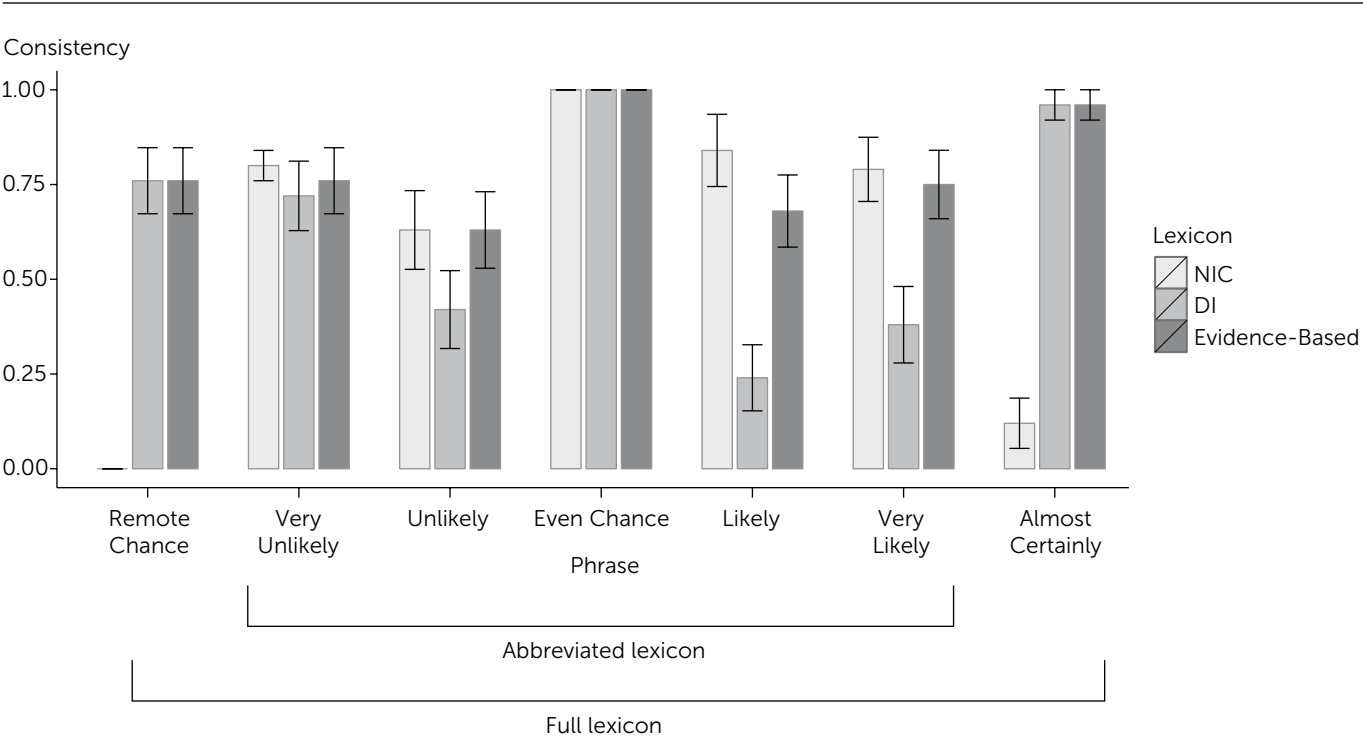
Error bars are included to indicate 95% confidence intervals. The bars display how much variation exists among data from each group. If two error bars overlap by less than a quarter of their total length (or do not overlap), the probability that the differences were observed by chance is less than 5% (i.e., statistical significance at p <.05).

a clear advantage over the existing NIC lexicon, and a higher consistency rate for the middle phrases, a clear advantage over the DI lexicon. The abbreviated evidence-based lexicons have high consistency rates, but the NIC lexicon is also highly consistent in this case, although this gain is greatly diminished in the extreme phrases, where the NIC lexicon is extremely narrow.

Using Evidence-Based Methods to Communicate Uncertainty

Our research shows that the current standard practices of communicating uncertainty in consequential policy domains as diverse as climate science and intelligence analysis can be ineffective. Those who receive the information may not understand it in the way it was intended by those who communicated it; as a consequence, decisions may be prone to errors and biases. Our research also shows that alternative evidence-based methods can be used to enhance understanding

and the effectiveness of communication. These findings have potential implications in many other domains, such as medicine and finance.

Recognizing that the effective communication of uncertainty can be challenging, some organizations have implemented uncertainty lexicons in an attempt to improve communication. However, most existing lexicons, such as those in used in the domains of climate science and intelligence analysis, are not based on empirical evidence, and their effectiveness has not been empirically tested. It is also striking that despite the close and long-standing political and military alliance between the United States and the United Kingdom, the official NIC and DI lexicons differ in their numerical prescriptions for every verbal term. Yet, the variance in approaches is not altogether surprising, given that such standards tend to be developed "in house," often based on whatever seems to make sense at the time.

We were able to quantify and document the problems associated with such arbitrary lexicons. We also

developed and tested evidence-based uncertainty lexicons, using two methods that require relatively little effort and involve low cost. Indeed, we showed that our lexicons were more effective than the existing institutional lexicons in communicating uncertainty. When people develop tools for communicating uncertainty to others—colleagues, peers, superiors, subordinates, or the general public—it is not ideal to rely on one's personal intuitions, local customs, or traditional norms. The quality of communication improves if one develops and adopts evidence-based communication tools focused on the target population (for example, patients in medical settings).

Our approach is intended to improve upon existing practices by minimizing the ambiguities inherent in communicating uncertainty in highly consequential public policy domains. We tested two methods: the PV method, which relies on a single estimate, and the MF method, which uses a full description of the subjective meaning of a term. MF performed better and its advantage was more pronounced in the intelligence application, which involves more phrases than does the IPCC. On the basis of these analyses, we consider MF to be a preferable method for developing evidence-based lexicons of uncertainty phrases, and we recommend its use. Our analysis also shows that the advantage of evidence-based methods is most pronounced for larger lexicons. In Study 2, the abbreviated NIC lexicon achieved a mean consistency rate similar to those found using our methods, but it failed to cover the entire range of interest. In particular, the NIC lexicon neglects very high and very low probabilities whose reporting to decisionmakers can be highly consequential in intelligence analysis.

Alternative evidence-based methods should be developed and tested. For example, Mandel (see reference 16) used an evidence-based approach to determine optimal numerical point probability values (as opposed to ranges) to assign to probability phrases used in a Canadian intelligence lexicon that assigns numerical point equivalents that are nevertheless meant to be interpreted as approximate (for example, a 90% probability is meant to be interpreted as "about 90%"; see also reference 13). Whereas lexicons with mutually exclusive and exhaustive range equivalents offer communicators the opportunity to convey any probability level, lexicons with point equivalents provide end users with more precise estimates of uncertainty.

It is important to recognize that there is no universal lexicon (see, for example, the ranges associated with the phrase *very likely* in Tables 1 and 3). The ranges depend on several factors, such as the number of terms used, the specific evidence-based method used (we see slight differences between the MF and PV methods), and the presence or absence of anchor terms at the two ends of the scale and at its center. The content of an effective lexicon may also be sensitive to its specific application. For example, a regulatory agency that seeks to develop a scale to communicate the likelihood of side effects associated with different drugs may find that *unlikely* is interpreted differently when the drug is a simple over-the-counter pain reliever compared to when it is an experimental treatment for a life-threatening condition (see reference 8).[35] It is best to derive application-specific lexicons that are tailored to the specific needs of any given domain and the expected level of precision.

Evidence-based methods do not completely eliminate the possibility of miscommunication. For example, recipients of probabilistic information tend to discount it somewhat and interpret it as less extreme than intended by communicators (see reference 8). Also, people typically process new information through the lens of their attitudes and beliefs. For example, Budescu, Por, and Broomell documented different interpretations of communications about climate science as a function of the recipients' ideology, attitudes toward the environment, and political affiliation (see reference 19). Indeed, research[36,37] has shown that probability phrases are subject to bias and distortion that fits one's expected and preferred conclusions (known as *motivated reasoning*).[38] However, the possibility of miscommunication is not fully eliminated by using numerical terms because the interpretation of numerical quantifiers is itself often imprecise and contextually dependent.[39] For instance, a point value, X, may be interpreted as lower ("at least X") or upper bound ("at most X"), as approximate ("about X"), or as precise ("exactly X").

Applying New Insights into Communicating Uncertainty

We recommend that behavioral science be used routinely in efforts to develop evidence-based tools for communicating uncertainty. In some domains, such as intelligence analysis, there has been longstanding opposition to quantifying uncertainty (see references

14, 24, and 28).[40] Empirical methods can delineate the limitations of existing approaches and offer sensible solutions to remedy them. The MF method, which elicited participants' full understanding of the key terms, was especially effective. In Study 1, we showed that the general public has a broader interpretation of probability phrases than the IPCC intended with its lexicon, and, in Study 2, we found that intelligence analysts' conceptions of probability terms simply do not match those of the organizational lexicons we examined. Moreover, we found that the U.S. and U.K. lexicons were inconsistent with each other, a factor that could undermine shared situational awareness and interoperability in allied military operations. Applying similar empirical verification processes to other domains, such as public health and finance, may reveal similar discrepancies in intended message and actual audience understanding, which opens the door to finding ways to increase communicative effectiveness.

When the nature and quality of the available evidence does not justify exclusive use of numerical communications, we recommend that probability phrases and numerical values be combined (see reference 2). This approach is sensitive to different people's preferences, and it also has the flexibility of adjusting the range of values in some cases to signal more precision. For example, in the context of an IPCC report, *very likely* may generally mean a probability greater than 85%, but in special cases, the communicator may feel sufficiently confident to say that the event is *very likely* but combine it with a more precise numerical meaning, say, between 90% and 95%.

Miscommunication is fertile ground for blame in public and private spheres. Organizations that inform policymakers about topics featuring uncertainty routinely get blamed when things go wrong. Ultimately, an evidence-based approach to communicating uncertainty would improve organizational accountability in such scenarios.[41] Let's face it: Uncertainty will never disappear. The job of expert assessors, such as intelligence analysts, is not to eliminate uncertainties but to assess them as accurately as possible.[42] Given that uncertainties are ubiquitous in the thorny issues that policymakers grapple with, it is incumbent on expert communities to use all effective means at their disposal to improve how they communicate such uncertainties. Evidence-based methods for doing so are their best bet.

author affiliation

Ho, and Budescu, Department of Psychology, Fordham University; Dhami, Department of Psychology, Middlesex University; and Mandel, Socio-Cognitive Systems Section, Toronto Research Centre, Defence Research and Development Canada. Corresponding author's e-mail: eho2@fordham.edu

supplemental material

- http://behavioralpolicy.org/vol-1-no-2/ho
- Methods & Analysis
- Additional References

References

1. Morgan, M. G., Henrion, M., & Small, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge, United Kingdom: Cambridge University Press.
2. 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. doi:10.1038/nclimate2194
3. Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context dependent, or both? *Organizational Behavior and Human Decision Processes, 41,* 390–404
4. Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in numeric or verbal terms. *Bulletin of the Psychonomic Society, 31,* 135–138. doi:10.3758/BF03334162

5. Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. In W. Brun, G. Keren, G. Kirkeboen, & H. Montgomery (Eds.), *Perspectives on thinking, judging and decision making* (pp. 21–35). Oslo, Norway: Univeristetforlaget.

6. Dhami, M. K. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied, 14,* 353–363. doi:10.1037/a0013344

7. Dhami, M. K., Lundrigan, S., & Mueller-Johnson, K. (2015). Instructions on reasonable doubt: Defining the standard of proof and the juror's task. *Psychology, Public Policy, and Law, 21,* 169–178.

8. Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *The Knowledge Engineering Review, 10,* 43–62. doi:10.1017/S0269888900007256

9. Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes, 36,* 391–405. doi:10.1016/0749-5978(85)90007-X

10. 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. doi:10.1037/1076-898X.10.1.25

11. Dhami, M., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities. *Memory & Cognition, 33,* 1057–1068. doi:10.3758/BF03193213

12. Mandel, D. R. (2007). *Toward a concept of risk for effective military decision making* (Technical Report 2007-124). Retrieved from Defence Research and Development Canada website: http://cradpdf.drdc-rddc.gc.ca/PDFS/unc68/p529014.pdf

13. Berry, D., Raynor, D. K., Knapp, P., & Bersellini, E. (2003). Patients' understanding of risk associated with medication use: Impact of European Commission guidelines and other risk scales. *Drug Safety, 26,* 1–11. doi:10.2165/00002018-200326010-00001

14. Barnes, A. (2015). Making intelligence analysis more intelligent: Using numeric probabilities. *Intelligence and National Security.* Advance online publication. doi:10.1080/02684527.2014.994955

15. Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science, 5,* 2–16.

16. 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.

17. Intergovernmental Panel on Climate Change. (n.d.). Organization. Retrieved December 13, 2015, from http://www.ipcc.ch/organization/organization.shtml

18. Swart, R., Bernstein, L., Ha-Duong, M., & Petersen, A. (2009). Agreeing to disagree: Uncertainty management in assessing climate change, impacts and responses by the IPCC. *Climatic Change, 92,* 1–29.

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

20. Ha-Duong, M., Swart, R., Bernstein, L., & Petersen, R. (2007). Uncertainty management in the IPCC: Agreeing to disagree. *Global Environmental Change, 17,* 8–11. doi:10.1007/s10584-008-9444-7

21. Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meaning of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365. doi:10.1037/0096-3445.115.4.348

22. Dhami, M. K. (2008). On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied, 14,* 353–363.

23. Lundrigan, S., Dhami, M. K., & Mueller-Johnson, K. (2013). Predicting verdicts using pre-trial attitudes and standard of proof. *Legal and Criminological Psychology.* Advance online publication. doi:10.1111/lcrp.12043

24. Kent, S. (1964). *Words of estimative probability.* Retrieved from Central Intelligence Agency website: https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html

25. Mandel, D. R., & Barnes, A. (2014). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences, USA, 111,* 10984–10989. doi:10.1073/pnas.1406138111

26. Derbentseva, N., McLellan, L., & Mandel, D. R. (2010). *Issues in intelligence production: Summary of interviews with Canadian managers of intelligence analysts* (Technical Report 2010-144). Retrieved from Defence Research and Development Canada website: http://pubs.drdc-rddc.gc.ca/PDFS/unc111/p534903_A1b.pdf

27. Adams, B. A., Thomson, M., Derbentseva, N., & Mandel, D. R. (2012). *Capability challenges in the human domain for intelligence analysis: Report on community-wide discussions with Canadian intelligence professionals* (Contract Report 2011-182). Retrieved from the Defence Research and Development Canada website: http://pubs.drdc-rddc.gc.ca/PDFS/unc118/p536570_A1b.pdf

28. Weiss, C. (2007). Communicating uncertainty in intelligence and other professions. *International Journal of Intelligence and CounterIntelligence, 21,* 57–85. doi:10.1080/08850600701649312

29. Office of the Director of National Intelligence. (n.d.). National Intelligence Council. Retrieved December 14, 2015, from http://www.dni.gov/index.php/about/organization/national-intelligence-council-who-we-are

30. National Intelligence Council. (2007). *Prospects for Iraq's stability: A challenging road ahead.* Retrieved from http://fas.org/irp/dni/iraq020207.pdf

31. Gov.UK. (n.d.). Defence Intelligence. Retrieved December 14, 2015, from https://www.gov.uk/government/groups/defence-intelligence

32. Dhami, M. K. (2013). *Understanding and communicating uncertainty in intelligence analysis.* Unclassified report prepared for Her Majesty's Government, United Kingdom. (Available from the author).

33. Defence Intelligence. (n.d). *Quick wins for busy analysts.* London, United Kingdom: United Kingdom Ministry of Defence.

34. Budescu, D. V., Karelitz, T. M., & Wallsten, T. S. (2003). Predicting the directionality of probability phrases from their membership functions. *Journal of Behavioral Decision Making, 16,* 159–180.

35. 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.

36. Fox, C. R., & Irwin, J. R. (1998). The role of context in the communication of uncertain beliefs. *Basic and Applied Social Psychology, 20,* 57–70.

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

38. Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498.

39. Mandel, D. R. (2015). Communicating numeric quantities in context: Implications for decision science and rationality claims. *Frontiers in Psychology, 6,* Article 537. doi:10.3389/fpsyg.2015.00537

40. Mandel, D. R., Barnes, A., & Richards, K. (2014). *A quantitative assessment of the quality of strategic intelligence forecasts* (Technical Report 2013-036). Retrieved from the Defence Research and Development Canada website: http://cradpdf.drdc-rddc.gc.ca/PDFS/unc142/p538628_A1b.pdf

41. Dhami, M. K., Mandel, D. R., Mellers, B., & Tetlock, P. (2015). Improving intelligence analysis with decision science. *Perspectives on Psychological Science, 106,* 753–757.

42. Friedman, J. A., & Zeckhauser, R. (2012). Assessing uncertainty in intelligence. *Intelligence and National Security, 27,* 824–847.