Middlesex University London

Middlesex University Research Repository

An open access repository of

Middlesex University research

http://eprints.mdx.ac.uk

Dhami, Mandeep K. ORCID: https://orcid.org/0000-0001-6157-3142 and Mandel, David R. (2020) Words or numbers? Communicating probability in intelligence analysis. American Psychologist . ISSN 0003-066X (Published online first) (doi:10.1037/amp0000637)

Final accepted version (with author's formatting)

This version is available at: http://eprints.mdx.ac.uk/29655/

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: http://eprints.mdx.ac.uk/policies.html#copy

Middlesex University Research Repository

An open access repository of

Middlesex University research

http://eprints.mdx.ac.uk

Dhami, Mandeep K. ORCID: https://orcid.org/0000-0001-6157-3142 and Mandel, David R. (2020) Words or Numbers? Communicating Probability in Intelligence Analysis. American Psychologist . ISSN 0003-066X (Accepted/In press)

Final accepted version (with author's formatting)

This version is available at: http://eprints.mdx.ac.uk/29658/

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: http://eprints.mdx.ac.uk/policies.html#copy

Accepted: *American Psychologist*

Words or Numbers? Communicating Probability in Intelligence Analysis

Mandeep K. Dhami[1] and David R. Mandel[2]

[1]Department of Psychology, Middlesex University

[2]Intelligence, Influence and Collaboration Section, Defence Research Development Canada

Author Note

Funding for this work was provided to the first author by HM Government and to the second author by Canadian Department of National Defence projects 05da and o5fa and Canadian Safety and Security Program project CSSP-2018-TI-2394. This work contributed to NATO System Analysis and Studies Panel Research Technical Group 114 on Assessment and Communication of Uncertainty in Intelligence to Support Decision Making. We thank Daniel Irwin and Serena Tran for their research assistance.

We have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Mandeep K. Dhami, Department of Psychology, School of Science and Technology, Middlesex University, The Burroughs, Hendon, London, NW4 4BT.

Abstract

Intelligence analysis is fundamentally an exercise in expert judgment made under conditions of uncertainty. These judgments are used to inform consequential decisions. Following the major intelligence failure that led to the 2003 war in Iraq, intelligence organizations implemented policies for communicating probability in their assessments. Virtually all chose to convey probability using standardized linguistic lexicons in which an ordered set of select probability terms (e.g., *highly likely*) is associated with numeric ranges (e.g., 80-90%). We review the benefits and drawbacks of this approach, drawing on psychological research on probability communication and studies that have examined the effectiveness of standardized lexicons. We further discuss how numeric probabilities can overcome many of the shortcomings of linguistic probabilities. Numeric probabilities are not without drawbacks (e.g., they are more difficult to elicit and may be misunderstood by receivers with poor numeracy). However, these drawbacks can be ameliorated with training and practice, whereas the pitfalls of linguistic probabilities are endemic to the approach. We propose that, on balance, the benefits of using numeric probabilities outweigh their drawbacks. Given the enormous costs associated with intelligence failure, the intelligence community should reconsider its reliance on using linguistic probabilities to convey probability in intelligence assessments. Our discussion also has implications for probability communication in other domains such as climate science.

Keywords: Subjective Probability, Uncertainty, Verbal Probabilities, Policy-Making, Intelligence Analysis

Significance Statement

Psychological research on probability communication suggests that using numbers such as 70% (or 65%-75%) rather than words such as *likely* provides a more clear and unambiguous way of communicating uncertainty in judgments. Therefore, we recommend that the intelligence community changes its current policy for probability communication in its assessments from the verbal to numeric mode, in order to mitigate intelligence failures such as that which led to the 2003 war in Iraq.

Intelligence assessments are vital to decision-making in several consequential domains including law enforcement, defense and national security. Analysts must answer questions of strategic, tactical and operational importance for a variety of consumers including commanders, government officials and other analysts. For example, “How will North Korea’s ballistic missile capability develop over the next three years?” And, “Where are Islamic State’s financiers located?” Such assessments are typically made under conditions of uncertainty (Fingar, 2011).[1] This is because relevant information may be missing or even unknowable (such as a foreign leaders’ intentions), information collection may be biased, and information may be unreliable as well as purposefully misleading. Consequently, most assessments are subjective probability judgments (Kent, 1964). Not only are analysts expected to judge probabilities accurately, they are also expected to communicate them clearly to consumers. Even the best judgments, if poorly communicated, can fail in their primary objective of supporting sound decision-making.

The deleterious consequences of poor probability communication were acutely evident in the 2003 invasion of Iraq that expected to find weapons of mass destruction (WMD), which in fact did not exist. Post-mortem analyses of this major intelligence failure criticized intelligence organizations for understating the probability in their assessments and suggesting greater certainty than was warranted by the available information (see UK House of Commons Foreign Affairs Committee, 2003; UK Intelligence and Security Committee, 2003; US Congressional Select Committee on Intelligence, 2005). The 2004 Butler review further questioned whether intelligence products were “drafted and presented in a way which best helps readers to pick up

---

[1] In intelligence analysis, uncertainty addresses both the probability of events and the confidence an analyst has in his/her assessment. Probability refers to the degree of certainty assigned to the belief that an event will or will not occur, and this is bounded from 0 (impossibility) to 1 (certainty), whereas confidence is treated as a multi-dimensional construct referring to variables such as source reliability and information credibility (Friedman & Zeckhauser, 2018). In this article, we examine how the intelligence community assesses and communicates probability. The topic of confidence is beyond the scope of this article.

the range of uncertainty attaching to intelligence assessments" (p. 144). The issue of knock-on errors in reporting was highlighted in the 2005 US Congressional report, which concluded that building on prior intelligence "without carrying forward the uncertainty from the first layer, … gave the impression of greater certainty about its judgments than was warranted" (p. 173). The damaging effects on decision-making of this intelligence and the misinterpretation of the probability attached to it, was emphasized in the Chilcot (2016) inquiry which noted that former UK Prime Minister Tony Blair mistakenly believed that intelligence organizations were "sure" about Iraq possessing WMD.

Similar conclusions were reached decades earlier in a now declassified 1983 Central Intelligence Agency (CIA) report into major intelligence failures. This concluded that analysts "lacked a doctrine or a model for coping with improbable outcomes. Their difficulty was compounded in each case by reluctance to quantify their theories of probability or their margins of uncertainty. Findings such as "likely," "probable," "highly probable," "almost certainly," were subjective, idiosyncratic, ambiguous between intelligence producer and consumer, uncertain in interpretation from one reader to another, and unchallenged by a requirement to analyze or clarify subordinate and lesser probabilities" (p. 5). The report went on to recommend the quantification of probability in future intelligence assessments.

The intelligence community has long debated the pros and cons of using words (e.g., terms such as *likely* and *unlikely*) and numbers to communicate probability (see Marchio, 2014). Rarely have these debates been informed by pertinent evidence. In the present article, we examine current policies for communicating probability in intelligence assessments—policies that continue to rely on linguistic rather than numeric probabilities. Probability communication has received considerable attention in psychological research, and the extant evidence can be used to inform policy-making. Our discussion also has implications for the assessment and communication of threats and risk in the intelligence analysis domain (see Friedman, 2019), and for probability communication in other domains such as medicine and climate science (e.g., see Budescu, Por, & Broomell, 2012; Budescu, Por, Broomell, & Smithson, 2014; Mazur & Merz, 1994; Timmermans, 1994). The use of linguistic probabilities to communicate probability in these domains is akin to the approach adopted by the intelligence community.

The paper is divided into three sections. We first discuss the intelligence community's reliance on a circumscribed set of probability terms, and we consider the benefits and drawbacks of this approach to probability communication. In the second section, we highlight the limitations of standardization—a common solution to the pitfalls of using words to communicate probability. Finally, we propose numeric probabilities as an alternative to linguistic probabilities, we consider the pros and cons of this approach, and we review research comparing the effects of verbal and numeric modes of probability communication on decision-making.

Communicating Probability Using Words

Traditionally, the intelligence community has taken an unstructured approach to probability communication. As a result, often probabilities were simply not communicated. Friedman and Zeckhauser (2012) revealed that a notable proportion (i.e., 18%) of the 379 declassified National Intelligence Estimates (NIEs), which are the most authoritative intelligence reports produced in the US, written between 1964 and 1994, did not provide any assessment of the probability associated with the outcome being forecast. On the other hand, as we saw with the earlier WMD example, suggesting greater certainty than warranted is also not uncommon. Analyses of the contents of 120 NIEs written between the 1950s and 2000s (i.e., 20 per decade, Kesselman, 2008) and of 2,013 Canadian intelligence assessments from 2006 to 2011 (Mandel &

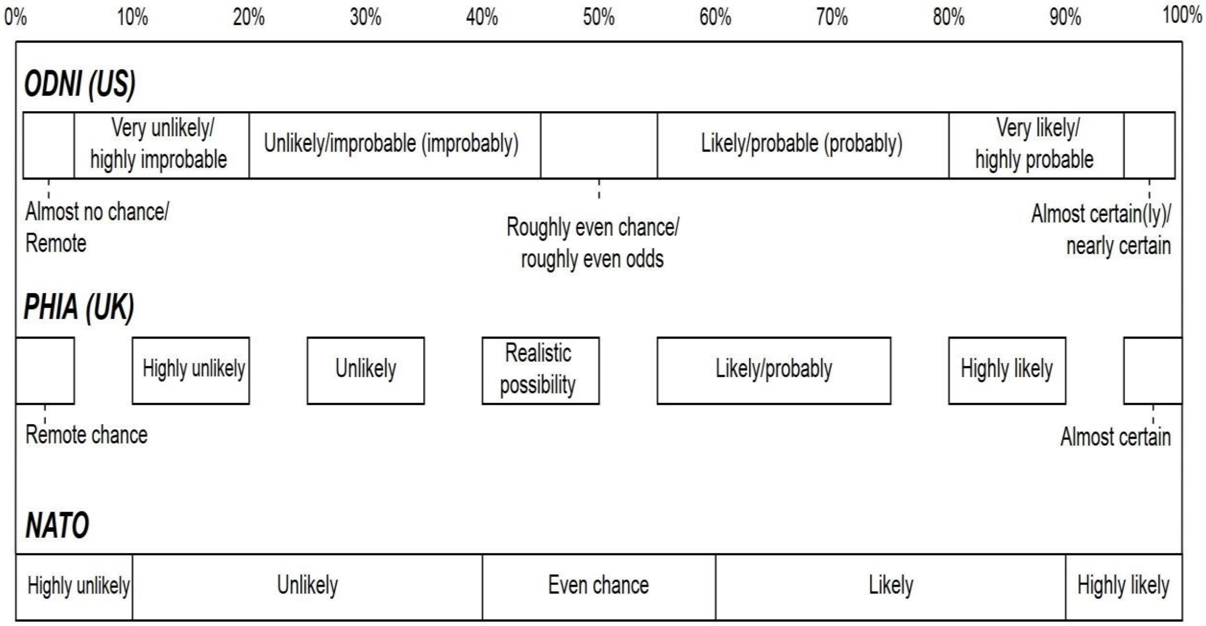
Barnes, 2018) show that the term *will*, which represents certainty, was most commonly used by analysts. However, probability statements should not be avoided. The 2005 US Congressional report states, "As much as they hate to do it, analysts must be comfortable facing up to uncertainty and being explicit about it in their assessments" (p. 408).

Intelligence Community Policies for Probability Communication

The US and the UK introduced formal policies for communicating probability in intelligence assessments in the mid-2000s following recommendations made by inquiries into the Iraq WMD intelligence failure. In the UK, the 2004 Butler review stated, "While not arguing for a particular approach to [expressing]…uncertainty, …we recommend that the intelligence community review their conventions again to see if there would be advantage in refreshing them" (p. 146). In the US, the 2005 Congressional report stated, "Whatever device is used to signal the degree of certainty—mathematical percentages, graphic representations, or key phrases—all analysts in the Community should have a common understanding of what the indicators mean and how to use them" (p. 419).

Policies for communicating probability in intelligence assessments have been developed at both national and international levels. National policies include those developed in the US by the Office of the Director of National Intelligence (ODNI), and in the UK by the Professional Head of Intelligence Analysis (PHIA); both organizations responsible for overseeing the intelligence community in their respective countries. Internationally, NATO has a policy for probability communication to be used by member states. Despite the various options available, policy-makers chose to convey probabilities in assessments using a standardized lexicon of linguistic probabilities. These lexicons have undergone revisions over the years, and Figure 1 presents the most recent versions of the US, UK and NATO lexicons (NATO Standardization Office, 2016; ODNI, 2015; PHIA, 2018 as published in College of Policing, n.d.).

Figure 1. ***Lexicons for communicating probability in intelligence assessments***



As Figure 1 shows, analysts are required to use a list of select words or probability terms that are ordered from the lowest to the highest degree of probability. These are combined with numeric ranges. Sherman Kent (1964), who played a key role in founding the CIA's Office of National Estimates (the predecessor of today's US National Intelligence Council), was the first to advocate this type of standardization, although it was not implemented at the time. He was prompted by the observation that both analysts and consumers varied considerably in their understanding of the term *serious possibility* which was used to communicate the probability of a

Soviet invasion of Yugoslavia in 1951. Over a half-century later, current solutions to the problem of probability communication in intelligence analysis are largely the same.

Benefits of Linguistic Probabilities

It appears that intelligence organizations strongly prefer to communicate probability linguistically rather than numerically. In their analysis of NIEs, Friedman and Zeckhauser (2012) revealed that only four percent contained numeric expressions of probability (e.g., percentages, odds). Intelligence organizations are not alone in their inclination for using linguistic probabilities. Several studies show that people may prefer to communicate probability linguistically rather than numerically (e.g., Erev & Cohen, 1990; Wallsten, Budescu, Zwick, & Kemp, 1993; but see Brun & Teigen, 1988 and Vahabi, 2010).

Some scholars have pointed to the potential benefits of using linguistic probabilities. Zimmer (1983) proposed that people know the rules of language better than the rules of probability, and they find it easier and more natural to use words when dealing with probability. Wallsten and Budescu (1995) suggest that linguistic probabilities should be preferred when the underlying uncertainty in a task is epistemic or internal (i.e., based on one’s knowledge). There is some evidence to support these viewpoints. In one study, Wallsten et al. (1993) observed that some people said it was easier and more natural to use language rather than numbers, and for some, this preference was strongest when the issue was deemed to be unimportant and/or the information was unreliable (see also Olson & Budescu, 1997).

The evidence for other claims, however, is either lacking or conflicting. For instance, Zimmer (1984) suggests that people process information verbally through argumentation and so asking them to respond in the verbal mode (as opposed to the numeric mode) requires less cognitive effort and makes them less susceptible to bias and unreliability. Budescu, Weinberg, and Wallsten (1988) did not find evidence to support this claim. Wallsten (1990) notes that linguistic probabilities allow the decision-maker to retain control of the decision by choosing how much risk to take. To our knowledge, this assertion has not been directly examined.

Drawbacks of Linguistic Probabilities

Notwithstanding the potential benefits that linguistic probabilities may offer, a substantial body of evidence points to the drawbacks of using this approach to probability communication. Research has repeatedly demonstrated considerable variability in how people understand linguistic probabilities (e.g., Beyth-Marom, 1982; Brun & Teigen, 1988; Clarke, Ruffin, Hill, & Beaman, 1992; Dhami & Wallsten, 2005; Lichtenstein & Newman, 1967). This variability is evident both across- and within-individuals. Individuals have broad or fuzzy interpretations of particular probability terms. In addition, different people may use different terms to refer to the same probability value(s) and/or may use the same term to refer to different values.

Variability has also been demonstrated in the intelligence analysis context (Dhami, 2018; Ho, Budescu, Dhami, & Mandel, 2015; Johnson, 1975; Wallsten, Shlomi, & Ting, 2008; Wark, 1964). For instance, Dhami (2018) reported that 145 unique terms were used to represent the 0 to 1 probability interval by a sample of 26 UK analysts. Eighteen terms were used to represent 10%. *Unlikely* was one of the most common terms used (i.e., by 58% of the sample), but it was used to represent a wide interval (i.e., 10%-40%). Thus, the communication of probability using words can mislead because they can be interpreted in different ways. This imprecision can also mask disagreement in judgments and create the illusion of consensus.

Research also shows that the interpretation of linguistic probabilities may be affected by the context in which they are used. Contexts can be externally provided such as the topic being considered (Brun & Teigen, 1988: Mellers, Baker, Chen, Mandel, & Tetlock, 2017), the order in which terms are presented (Bergenstrom & Sherr, 2003; Hamm, 1991), the base-rate of the event being judged (Wallsten, Fillenbaum, & Cox, 1986; Weber & Hilton, 1990), the severity of an outcome (Harris & Corner, 2011; Mazur & Merz, 1994; Merz, Druzdzel, & Mazur, 1991) and the outcome's valence (Cohen & Wallsten, 1992). Contexts may also be internal to the person such as one's attitude to the subject matter (Budescu et al., 2012; see also Piercey, 2009) and even one's locus of control (Hartsough, 1977).

Although more research is needed, context effects have been documented in the intelligence analysis domain. In one study, Mandel (2015a) reported that a sample of 17 analysts (and a student sample of 40 who did not differ in responses from the analyst sample) provided significantly less discriminating numeric interpretations of probability terms in a Canadian intelligence organization's lexicon when the event in question was described as a failure than when it was described as a success. For example, *will not* was interpreted as close to a 0% chance when it referred to a success, but it was interpreted as roughly a 10% chance when it referred to a failure. Conversely, *will* was interpreted as closer to a 100% chance in the success than the failure condition. In another study, using a mixed sample of 596 "superforecasters" (i.e., forecasters at or above the 98th percentile in accuracy in a geopolitical forecasting tournament), regular forecasters and undergraduates, Mellers et al. (2017) found that numeric interpretations of probability terms were affected by the geopolitical context in which they appeared. For instance, across five topics, the average interpretation of *almost certain* ranged from a 20% point spread (for superforecasters in Study 1) to a 39% point spread (for undergraduates in Study 2). Such contextual effects suggest that the interpretation of probability terms cannot simply be anchored through standards that stipulate a fixed (context independent) meaning for each term.

Not only do linguistic probabilities vary in their interpretation, as we noted earlier, they are also imprecise, and this coarsens the probability scale, creating fewer distinguishable levels of probability that may be communicated to receivers. Such coarsening can artificially inflate (or decrease) the expressed likelihood of low (or high) probability events occurring. A particular concern is with rare events, tail risk, or so-called “black swans”—namely, events that have extremely low probabilities but extremely severe consequences (Makridakis & Taleb, 2009). While terms such as *remote chance* represent the lowest probability categories in the lexicons, these do not convey the exceedingly small chances required to accurately characterize tail risks, and they may be orders of magnitude off. Indeed, a study of 34 Canadian and 27 UK analysts found that the average best interpretation of *remote chance* in the UK was about 23% and it was about 17% in Canada (Ho et al., 2015). Clearly, an analyst could not use this term to effectively communicate a 1% chance, let alone a one in a million chance.

Beyond making it difficult to communicate tail risks, coarsening the probability scale has also been shown to adversely affect forecasting accuracy, especially amongst the most competent forecasters (Friedman, Baker, Mellers, Tetlock, & Zeckhauser, 2018). That is, when precise forecasts were binned into varying numbers of categories, accuracy declined with decreasing bin size. Efforts to constrain the meanings of terms by attaching numeric ranges as in the current UK, US and NATO lexicons (see Figure 1) does not overcome the imprecision as it simply divides the 0 to 1 probability interval into a small number of categories (usually less than ten).

It is also doubtful that linguistic probabilities can be effectively combined because words do not easily lend themselves to arithmetic operations (see Budescu, Zwick, Wallsten, & Erev, 1990; Wallsten, Budescu, & Tsao, 1997; Zwick, Budescu, & Wallsten, 1988). This is of particular concern in intelligence assessment where analysts are often required to express how the probability of a number of (independent as well as interacting) events may combine to lead to an outcome or set of outcomes or where a decision-maker receives multiple assessments and wants to know their average value. Imagine a chain of four independent events that must occur in order for a particular threat scenario to manifest. In one case, the probabilities of the events are communicated numerically as being .75, .10, .70, and .01. In another case, these probabilities are given using the US standard equivalents, *likely*, *very unlikely*, *probable*, and *almost no chance*. In two recent experiments comparing these two modes of communication, participants were significantly more accurate in computing averages and products from values provided in numeric rather than comparable verbal form (Mandel, Dhami, Tran, & Irwin, 2020).

Finally, linguistic probabilities convey more than probability. In particular, probability terms convey "directionality" to recipients (Teigen & Brun, 1995, 1999, 2003; see also Budescu, Karelitz, & Wallsten, 2003; Honda & Yamagishi, 2006). Directionality is a pragmatic feature of probability statements that subtly conveys the speakers' attitude towards the focal event. For instance, two speakers who agree that the probability of a given event is low may nevertheless differ by communicating an optimistic attitude using a directionally positive term such as *some chance* or by communicating a pessimistic attitude using a directionally negative term such as *doubtful*. Directionality can, in turn, shape receivers' beliefs about the sender's implicit recommendations (Teigen & Brun, 1999, 2003). For example, Collins and Mandel (2019) found that whereas probability information was rated as significantly clearer when it was conveyed numerically rather than verbally, the clarity of implicit recommendations (which were not explicitly communicated) was significantly greater when it was conveyed verbally rather than numerically. These findings suggest that it will not only be harder to infer probability levels from linguistic probabilities, but also that assessments communicated using probability terms are more likely to suggest recommendations to decision-makers. This runs contrary to, and threatens, the intelligence community's longstanding mandate to remain policy neutral.

Is Standardization the Solution?

The approach (i.e., standardized lexicons) adopted by intelligence organizations to communicating probability in intelligence assessments cannot mitigate the pitfalls of using linguistic probabilities. In addition to the concerns raised above, research shows that people cannot easily suppress their normal, context-dependent meanings of probability terms and adopt new (mandated) meanings that deny context dependence (e.g., Wallsten et al. 1986; see also Budescu & Wallsten, 1990). For instance, studies on the lexicons used by the Inter-Governmental Panel on Climate Change (IPCC) have demonstrated that recipients of terms in lexicons often do not interpret the words as intended when reading statements that contain them, and default to their personal interpretation of the terms, thus defeating the purpose of the lexicon (e.g., Budescu et al., 2012, 2014; Wintle, Fraser, Wills, Nicholson, & Fidler, 2019). Budescu et al. (2014) found that the proportion of respondents whose numeric interpretations of IPCC terms that fell in the stipulated ranges varied from 21%-35% across 25 countries, with roughly 200 respondents per country. In addition, in an attempt to increase compliance rates, researchers have examined the effect of presenting the stipulated numeric ranges alongside the probability terms used in the statements (e.g., "likely [55%-80%]"; Budescu et al., 2012, 2014; Wintle et al., 2019). Although this hybrid approach yields higher compliance rates, they still fall short of acceptable levels. For example, Budescu et al. (2014) found that with the hybrid approach, the proportion of compliant respondents ranged from 28%-54% in the same set of 25 countries.

However, even if compliance rates were much higher, using numeric ranges alongside probability terms in assessments can be confusing to receivers who may interpret them as credible intervals that pertain directly to the assessment. Imagine applying this approach to the US lexicon, a decision-maker who reads that “The Xs are very likely [80%-95%] to attack country Y in the next month” might think that the analyst believes there is an 80%-95% chance that the attack will occur. Yet, the analyst may, if pressed, give lower and upper bounds that cut across stipulated ranges, such as 70%-85%. Another analyst, who agrees perfectly with this latter range, may nevertheless assign the term likely, in which case the decision-maker would instead receive the estimate, “The Xs are likely [55%-80%] to attack country Y in the next month.”

In the intelligence community, the problem with standardization is no doubt further compounded by having multiple lexicons in operation at one time. Figure 1 illustrates the lexical differences across nations that share intelligence. In addition, national lexicons may be implemented with slight variations across organizations within a country. For instance, the National Crime Agency in the UK uses a visual representation of PHIA’s lexicon (National Crime Agency, 2018). Probabilities below 50% are depicted in shades of blue, whereas probabilities above 50% are depicted in shades of purple. Due to organizational differences in lexicons, analysts (and other intelligence consumers) must rapidly shift their use and understanding of probability terms and mentally juggle different lexicons. Furthermore, revisions of lexicons impose additional mental juggling due to intra-organizational changes over time. All of these undermine interoperability and can exacerbate communication errors.

Problems with the intelligence community’s use of standardized lexicons have also been highlighted in recent years by researchers who have examined the effectiveness of these lexicons. Studies have employed quantitative methods inspired by Zadeh’s (1965) fuzzy sets theory in mathematics, to elicit analysts' lexicons for probability communication and to measure how people numerically interpret linguistic probabilities. The emerging evidence points to discrepancies between policy (what lexicons stipulate) and practice (what analysts do). Specifically, interpretations of terms in the current US and UK lexicons do not map directly onto the prescribed numeric ranges (Dhami, 2018; Ho et al., 2015, Study 2; Mandel, 2015a; Wallsten et al., 2008). In some cases, interpretations were either below or above the category ranges. For instance, in the UK lexicon, whereas *highly unlikely* represents 10%-20%, the corresponding 95% confidence interval on the median best estimate in Mandel's (2015a) study of a combined sample of Canadian analysts and students, was 8%-10%. Similarly, whereas *likely* is intended to represent 55%-75% in the US lexicon, the 95% confidence interval on the median estimate of this term was 75%-80%. Furthermore, terms that the lexicons intend to be substitutable such as *likely* and *probably/probable* were not so in participants' minds (see also Dhami, 2018). Finally, although the rank order of terms in analysts' lexicons may generally correspond to that in the US and UK lexicons, analysts whose lexicons contained terms that appeared in these lexicons did not use them as mandated. For example, Dhami (2018) found that while *likely* was ranked before *highly unlikely* in UK analysts' lexicons, they used the terms to represent ranges with higher maximum probabilities than that mandated (see also Wallsten et al., 2008).

In an early survey examining the effect of the verbal mode of probability communication on recipients, Wark (1964) compared 240 US analysts' and 63 consumers' interpretations of specific probability terms. He reported that consumers had lower numeric interpretations of terms than did analysts. In other contexts, studies also show differences in communicators' and receivers' interpretations of probability terms. For instance, in the climate change context, Budescu et al. (2012) revealed that the public consistently misinterprets the probabilistic statements in IPCC reports in a regressive fashion (i.e., they underestimate high probabilities and overestimate low probabilities; see also Budescu et al., 2014). The public thus does not interpret the reports' conclusions as intended by the authors of the reports.

Some have suggested that standardization could be effective if the lexicons were empirically derived. Ho et al. (2015, Study 2) used data from 34 Canadian analysts to derive evidence-based lexicons that utilized the terms common to the US and UK lexicons. The lexicons were developed using statistical methods that optimized the fit between the stipulated numeric ranges and analysts' numeric interpretations of the terms. The resulting lexicons were then validated using data from a sample of 27 UK analysts by examining the proportion of their numeric interpretations that fell in the ranges stipulated by the various lexicons. Ho et al. (2015) found that they could improve agreement between analysts' interpretations and the standards by using an evidence-based lexicon. Specifically, they showed that both the UK lexicon and their best performing evidence-based lexicon outperformed the US lexicon at the extremes (i.e., for the lowest and highest probabilities), whereas the US and evidence-based lexicons outperformed the UK lexicon for the less extreme probabilities. Wintle et al. (2019) replicated the advantage of using an evidence-based lexicon over the existing US standard, using a larger, non-expert sample.

To the best of our knowledge (which draws, in part, from extensive consultation with intelligence analysts and managers in several intelligence organizations), the current US, UK and NATO lexicons are not empirically grounded and were not rigorously tested prior to adoption. Policy development in the intelligence community has historically occurred without consideration of relevant evidential bases beyond the anecdotal "lessons learned" following intelligence failures (Chang, Berdini, Mandel, & Tetlock, 2018; Dhami, Mandel, Mellers, & Tetlock, 2015; Mandel & Tetlock, 2018). Regardless, for the reasons mentioned, we do not view evidence-based lexicons as an effective solution to the intelligence community's requirement to communicate probabilities clearly to consumers. While such lexicons can improve compliance rates, they cannot circumvent the fact that linguistic probabilities are context dependent, coarse, difficult to combine, and imply unstated policy recommendations.

Communicating Probability Using Numbers

An obvious alternative to the intelligence community's current approach to probability communication is to use numeric probabilities. This can be accomplished by using precise numeric point values (e.g., "the probability is .75" or "there is a 75% chance") or imprecise values (e.g., 65% chance plus or minus 10% or 20%-40% chance). Given that probability terms in the current US, UK and NATO lexicons have numeric ranges assigned to them suggests that numbers are deemed necessary to help analysts convey probability in their assessments.

In fact, research shows that while most communicators prefer to send linguistic probabilities, most recipients prefer to receive numeric probabilities (Brun & Teigen, 1988; Erev & Cohen, 1990; Murphy, Lichtenstein, Fischhoff, & Winkler, 1980; Olson & Budescu, 1997; Vahabi, 2010; Wallsten et al., 1993). The preference for receiving numeric probabilities appears to be especially pronounced when those expecting to receive information (e.g., decision-makers) consider the issue to be important and when precision was considered possible. Although more research is required with intelligence consumers, this preference has been documented by early studies in the intelligence community (see Marchio, 2014; see also Barnes, 2016).

Benefits of Numeric Probabilities

There are several benefits of using numeric probabilities beyond reducing the resource burdens and logistical challenges associated with standardization efforts. Whereas linguistic probabilities have unreliable ordinal scale properties, numeric probabilities have reliable ratio scale properties. Consumers are less likely to misinterpret probabilities expressed numerically because numbers are explicitly scaled. This reduces the intra- and inter-individual variability in the values they represent. All agree that 25% is greater than 10%, and that the difference between these two values is the same as that between 75% and 90%.

In addition to reducing the misinterpretation of probability, numeric probabilities confer other important advantages. In many cases, it may be difficult to provide a precise probability estimate, but considerably easier to estimate a range. In effect, using linguistic probabilities involves assigning fuzzy ranges, whereas estimating lower and upper numeric bounds involves assigning crisp ranges. A crisp range can be easily converted to a point estimate with an uncertainty interval using interval analysis (see Moore, Kearfott, & Cloud, 2009). For instance, a 55%-80% chance could easily be reframed as a 67.5% chance ± 12.5%. Using the same approach, numeric probabilities can be easily combined (i.e., added, subtracted, multiplied and divided) even if they are expressed as imprecise ranges. By contrast, this would be difficult if estimates were given using the existing US, UK or NATO lexicons.

Unlike linguistic probabilities, as mentioned, numeric probabilities can be used to describe rare events. An analyst could easily distinguish one in ten from one in a million—something that is impossible to do with the current standards. Collapsing all orders of magnitude less than 1 in 10 will not serve policy-makers tasked with planning for and responding to low-probability, high-consequence-severity threats. A single black swan that could have been better prepared for through more granular analyses is likely sufficient to merit adopting communication methods that enable quantification and discrimination between orders of magnitude.

The numeric approach is also verifiable, allowing the intelligence community to track the accuracy of intelligence assessments and measure components of analytic skill (Mandel, 2015a;

National Research Council, 2011). Studies on geopolitical forecasting using numeric probabilities have illustrated the feasibility of measuring the accuracy of such forecasts using Brier scores and other quantitative measures of forecasting skill such as calibration and discrimination (e.g., Chang, Chen, Mellers, & Tetlock, 2016; Mandel & Barnes, 2014, 2018; Mellers et al. 2015). Similarly, quantifying forecasts has enabled comparisons to be made between the accuracy of traditional analytic methods and alternative methods, such as classified prediction markets for analysts (Mandel, 2019; Stastny & Lehner, 2018). Thus, adopting a numeric approach can make the intelligence community both more accountable and informed (e.g., Dhami et al., 2015; Friedman et al., 2018), which may render it less susceptible to the "blame game" that typically ensues after intelligence failures (Tetlock & Mellers, 2011).

Drawbacks of Numeric Probabilities

Although numeric probabilities confer many advantages over linguistic probabilities as a basis for probability communication, they are not without drawbacks. Numbers or numeric quantifiers are not immune from context effects (Bilgin & Brenner, 2013; Mandel, 2014; Verplanken, 1997), although their resistance to them is greater than for linguistic probabilities, which derive their meaning from the context in which they are used. Numeric probabilities can also convey directionality (Teigen & Brun, 2000), although compared to linguistic probabilities they are directionally more ambiguous and therefore less prone to conveying implicit recommendations for action by decision-makers (Bilgin & Brenner, 2013; Collins & Mandel, 2019). Accordingly, while numeric probabilities are not immune from exploitation, they are less susceptible than linguistic probabilities to being misused. For instance, in research directly comparing numeric and linguistic probabilities, Piercey (2009) found that under conditions of motivated reasoning, judgments made using numeric probabilities were less biased than those made using linguistic probabilities. Budescu et al. (2012) similarly observed that motivated reasoning had less impact on interpretations of statements about climate change containing probability terms when these were presented along with numeric ranges than alone.

Another potential limitation of numeric probability use is that individuals may find it difficult to quantify their uncertainty. For instance, Lanir and Kahneman (2006) described the difficulties that 19 intelligence analysts, academic experts and foreign affairs personnel had in judging conditional probabilities in a 1975 case study where Israeli experts numerically forecasted the consequences of alternative outcomes of US brokered negotiations between Israel and Egypt. These researchers also noted that one of the two consumers of the report misunderstood the numeric estimate. However, over time, people may adapt to using numeric probabilities. In recounting efforts to require Canadian analysts who were under his direction to communicate probability numerically, Barnes (2016, p. 7) states that, “With experience, analysts became more comfortable using numeric probabilities.” Nevertheless, the use of numbers will likely require investment in upskilling analysts, their managers and consumers. Fortunately, statistical and probabilistic reasoning skills can be improved through training (Chang et al., 2016; Sedlemeier & Gigerenzer, 2001). In fact, in the late 1970s a course called “Statistical Concepts for Analysts and Managers” was introduced in the US (Marchio, 2014), although this was short-lived. More recently, Mandel (2015b) showed that even brief training in Bayesian reasoning using natural-sampling trees can improve the coherence and accuracy of analysts’ probability judgments. However, such studies have yet to demonstrate how long these improvements endure beyond the immediate post-training period.

Decision-Making Using Numeric Versus Linguistic Probabilities

In further considering the pros and cons of numeric versus linguistic probabilities, the intelligence community may be enlightened by research comparing the effects of these two modes of probability communication on decision-making. Evidence suggests that moving from probability words to numbers is unlikely to adversely affect decision-making, and it may have a positive effect. Studies comparing the effects of the verbal and numeric modes find little difference in outcomes between the two modes (Budescu & Wallsten, 1990; Erev & Cohen, 1990; Gonzalez-Vallejo & Wallsten, 1992; Wallsten, Budescu, & Zwick, 1993). In some cases, researchers report that the best performing mode depends on the structure of the task (Gonzalez-Vallejo, Erev, & Wallsten, 1994; Olson & Budescu, 1997), whereas in other cases the numeric mode leads to better outcomes such as accuracy relative to a Bayesian calculation when judging the likelihood of alternative hypotheses and expected monetary gain in a gambling task (Budescu et al., 1988; Rapoport, Wallsten, Erev, & Cohen, 1990; see also Budescu & Wallsten, 1990). This is partly because the numeric mode results in less variability than the verbal mode.

Although more research is needed on the effects of numeric and verbal modes of probability communication, research applied to professional domains has also demonstrated the potential positive effects of the numeric mode compared to the verbal mode. For instance, Timmermans (1994) found that medical professionals were more likely to agree on treatment decisions and were better at Bayesian reasoning when probabilistic information was presented numerically rather than verbally. In a study comparing 407 US national security officials' responses to verbal and numeric probability estimates, Friedman, Lerner, and Zeckhauser (2017) found that quantifying probabilities did not prompt officials to take riskier actions, even for optimistic scenarios. Nor did quantification result in overconfidence. Rather, quantification led to greater willingness to gather additional information.

Barriers to Using Numeric Probabilities in Intelligence Assessments

Past calls for the use of numbers to communicate probability in intelligence assessments (e.g., CIA, 1983; Johnson, 1973) have resulted in “false starts”. Barnes (2016) concluded that “Overall, the division’s experience in using numeric probabilities was positive…. It enhanced the discussion between director and analyst during the review process: both now had a common understanding of the degree of certainty attached to a judgment, which allowed for a more effective discussion of the key factors and chain of logic that underpinned the analyst’s conclusion.” (p. 7). Despite this, the use of numeric probabilities only lasted from 2004 to 2011, and ended when Barnes retired. Marchio (2014) describes a trial conducted within the US Defence Intelligence Agency in the late 1970s where analysts, among other things, used precise percentage values to represent the uncertainty in their conclusions. The majority of the 128 consumers who were later questioned about these assessments favored quantification of probability. These helped to increase their confidence in the analytic conclusions and gave greater credibility to briefings. Nevertheless, the brief foray to quantification lost momentum and soon ended. Even the statistical training given to analysts and managers at that time was not supported when graduates were back on the job (Marchio, 2014).

Much of the intelligence community’s reluctance to quantify probability reflects widespread misconceptions about probability and its quantification. Common arguments include that it would be wrong to use numeric probabilities because they convey a false sense of precision and suggest a scientific basis to the estimate. Neither of these are reasonable inferences. As already noted, numeric probabilities can be stated as precisely or imprecisely as a sender intends them to be. Precision may actually be useful as it can reveal disagreement that can be informative (e.g., it may suggest the need for more intelligence collection, or even the need for lowering one’s confidence in an assessment). Barnes (2016) recalled that his analysts debated the appropriateness of assigning specific values to an event in question and focused on the judgment process. Regardless of the degree of precision, quantification does not imply anything about the use of the scientific method. The belief that quantification necessitates science suggests that many in the intelligence community do not understand the concept of subjective probability. Barnes (2016) also noted that his analysts needed to gain a better understanding of subjective probability before being comfortable with using numeric probabilities.

Part of the problem is that the intelligence community has traditionally considered analysis as an “art” and analysts as “poets” rather than “mathematicians” (Kent, 1964). Consequently, analysts and intelligence organizations may be resistant to quantifying probability. However, these barriers are not insurmountable. Barnes (2016), for example, observed that the initial apprehension his analysts had in using numeric probabilities soon dissipated. Analysts outside his unit who criticized the use of numbers to communicate probability nevertheless used them, and those consumers who were provided with numeric estimates also found them useful. In fact, the barriers to communicating probability numerically may be lowering in the age of ‘big data’ as intelligence organizations find themselves increasingly needing to recruit data scientists. The mathematician-to-poet ratio may thus be increasing. Regardless, it is clear that moving from probability words to numbers would constitute a major shift from the intelligence community’s current policy and organizational culture. We believe that by exploiting the extant psychological evidence on probability communication, the community can improve its current policies for communicating probability in its assessments. This can better mitigate the risk of future intelligence failures.

References

Barnes, A. (2016). Making intelligence analysis more intelligent: Using numeric probabilities. *Intelligence and National Security, 31*(3), 327-344. doi: 10.1080/02684527.2014.994955

Bergenstrom, A., & Sherr, L. (2003). The effect of order of presentation of verbal probability expressions on numerical estimates in a medical context. *Psychology, Health & Medicine, 8*(4), 391-398. doi: 10.1080/1354850310001604522

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting, 1*(3), 257-269. doi: 10.1002/for.3980010305

Bilgin, B., & Brenner, L. (2013). Context affects the interpretation of low but not high numerical probabilities: A hypothesis testing account of subjective probability. *Organizational Behavior and Human Decision Processes, 121*(1), 118-128. doi: 10.1016/j.obhdp.2013.01.004

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

Budescu, D. V., Karelitz, T. M., & Wallsten T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making, 16*(3), 159-180. doi: 10.1002/bdm.440

Budescu, D. V., Por, H. H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change, 113*(2), 181-200. doi: 10.1007/s10584-011-0330-3

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*(6), 508-512. doi: 10.1038/nclimate2194

Budescu, D. V., & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior and Human Decision Processes, 46*(2), 240-263. doi: 10.1016/0749-5978(90)90031-4

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 281-294. doi:10.1037/0096-1523.14.2.281

Budescu, D. V., Zwick, R., Wallsten, T. S., & Erev, I. (1990). Integration of linguistic probabilities. *International Journal of Man-Machine Studies, 33*(6), 657-676. doi: 10.1016/S0020-7373(05)80068-9

Butler, Chilcot, J., Marshall, Mates, M., & Taylor, A. (2004). *Review of Intelligence on Weapons of Mass Destruction: Implementation of its Conclusions*. https://fas.org/irp/world/uk/butler071404.pdf

Chang, W., Berdini, E., Mandel, D. R., & Tetlock, P. E. (2018). Restructuring structured analytic techniques in intelligence. *Intelligence and National Security, 33*(3), 337-356. doi: 10.1080/02684527.2017.1400230

Chang, W., Chen, E., Mellers, B., & Tetlock, P. E. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making, 11*(5), 509-526.

Chilcot, J. (2016). *The Report of the Iraq Inquiry. Executive Summary.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535407/The_Report_of_the_Iraq_Inquiry_-_Executive_Summary.pdf

Clarke, V. A., Ruffin, C. L., Hill, D. J., & Beaman, A. L. (1992). Ratings of orally presented verbal expressions of probability by a heterogeneous sample. *Journal of Applied Social Psychology, 22*(8), 638-656. doi: 10.1111/j.1559-1816.1992.tb00995.x

Cohen, B. L., & Wallsten, T. S. (1992). The effect of constant outcome value on judgments and decision making given linguistic probabilities. *Journal of Behavioral Decision Making, 5*(1), 53-72. doi: 10.1002/bdm.3960050107

College of Policing (n.d.). *Delivering effective analysis*. https://www.app.college.police.uk/app-content/intelligence-management/analysis/delivering-effective-analysis/

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. (2018). Towards an evidence-based approach to communicating uncertainty in intelligence analysis. *Intelligence and National Security*, *33*(2), 257-272. doi: 10.1080/02684527.2017.1394252

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. doi: 10.1177/1745691615598511

Dhami, M. K., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities *Memory & Cognition, 33*(6), 1057-1068. doi: 10.3758/BF03193213

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. doi:10.1016/0749-5978(90)90002-Q

Fingar, T. (2011). *Reducing uncertainty: Intelligence analysis and national security*. Stanford, CA: Stanford Security Studies.

Friedman, J. A. (2019). *War and chance: Assessing uncertainty in international politics*. New York, NY: Oxford University Press. doi:10.1093/oso/9780190938024.001.0001

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. doi: 10.1093/isq/sqx078

Friedman, J. A., Lerner, J. S., & Zeckhauser, R. (2017). Behavioral consequences of probabilistic precision: Experimental evidence from national security professionals. *International Organization, 71*(4), 803-826. doi:10.1017/S0020818317000352

Friedman, J. A., & Zeckhauser, R. (2012). Assessing uncertainty in intelligence. *Intelligence and National Security, 27*(6), 824-847. doi: 10.1080/02684527.2012.708275

Friedman, J. A., & Zeckhauser, R. (2018). Analytic confidence and political decision-making: Theoretical principles and experimental evidence from national security professionals. *Political Psychology, 39*(5), 1069-1087. doi: 10.1111/pops.12465

González-Vallejo, C., Erev, I., & Wallsten, T. S. (1994). Do decision quality and preference order depend on whether probabilities are verbal or numerical? *The American Journal of Psychology, 107*(2), 157-172. doi: 10.2307/1423035

González-Vallejo, C., & Wallsten, T. S. (1992). Effects of probability mode on preference reversal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 855- 864. doi: 10.1037/0278-7393.18.4.855

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes, 48,* 193-223. doi: 10.1016/0749-5978(91)90012-I

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. doi: 10.1037/a0024195

Hartsough, W. R. (1977). Assignment of subjective probabilities to verbal probability phrases as a function of locus of control and set conditions. *The Journal of Psychology, 95*(1), 87-97. doi: 10.1080/00223980.1977.9915864

Ho, E. H., Budescu, D. V., Dhami, M. K., & Mandel, D. R. (2015). Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science and Policy, 1*(2), 43-55. doi:10.1353/bsp.2015.0015

Honda, H., & Yamagisih, K. (2006). Directional verbal probabilities: Inconsistencies between preferential judgments and numerical meanings. *Experimental Psychology, 53*(3), 161-170. doi: 10.1027/1618-3169.53.3.161

Johnson, E. M. (1973). *Numerical encoding of qualitative expressions of uncertainty*. Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Kent, S. (1964). *Words of Estimative Probability*. 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

Kesselman, R. F. (2008). *Verbal probability expressions in national intelligence estimates: A comprehensive analysis of trends from the fifties through post 9/11*. (Unpublished Masters dissertation). Mercyhurst College, Erie, PA.

Lanir, Z., & Kahneman, D. (2006). An experiment in decision analysis in Israel in 1975. *Studies in Intelligence, 50*(4), 11-19.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science, 9*(10), 563-564. doi:10.3758/BF03327890

Makridakis, S., & Taleb, N. (2009). Living in a world of low levels of predictability. *International Journal of Forecasting, 25*(4), 840-844. doi: 10.1016/j.ijforecast.2009.05.008

Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of Experimental Psychology. General, 143*(3), 1185-1198. doi: 10.1037/a0034207

Mandel, D. R. (2015a). Accuracy of intelligence forecasts from the intelligence consumer's perspective. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 111-120. doi: 10.1177/2372732215602907

Mandel, D. R. (2015b). Instruction in information structuring improves Bayesian judgment in intelligence analysts. *Frontiers in Psychology, 6,* 387. doi: 10.3389/fpsyg.2015.00387

Mandel, D. R. (2019). Too soon to tell if the US intelligence community prediction market is more accurate than intelligence reports: Commentary on Stastny and Lehner (2018). *Judgment and Decision Making, 14*(3), 288-292.

Mandel, D. R., & Barnes, A. (2014). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 111*(30), 10984-10989. doi: 10.1073/pnas.1406138111

Mandel, D. R., & Barnes, A. (2018). Geopolitical forecasting skill in strategic intelligence. *Journal of Behavioral Decision Making, 31*(1), 127-137. doi: 10.1002/bdm.2055

Mandel, D. R., Dhami, M. K., Tran, S., & Irwin, D. (2020). *Arithmetic computation with probability words and numbers.* Manuscript submitted for publication.

Mandel, D. R., & Tetlock, P. E. (2018). Correcting judgment correctives in national security intelligence. *Frontiers in Psychology*, 9, 2640. doi: 10.3389/fpsyg.2018.02640

Marchio, J. (2014). "If the weatherman can...": The intelligence community's struggle to express analytic uncertainty in the 1970s. *Studies in Intelligence, 58*(4), 31-42.

Mazur, D. J., & Merz, J. F. (1994). How age, outcome severity, and scale influence general medicine clinic patients' interpretations of verbal probability terms. *Journal of General Internal Medicine, 9*(5), 268-271.

Mellers, B. A., Baker, J. D., Chen, E., Mandel, D. R., & Tetlock, P. E. (2017). How generalizable is good judgment? A multi-task, multi-benchmark study. *Judgment and Decision Making, 12*(4), 369-381.

Mellers, B., Stone, E., Atanasov, P., Roghbaugh, N., Metz, S. E., Ungar, L.,… & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*(1), 1-14. doi: 10.1037/xap0000040

Merz, J. F., Druzdzel, M. J., & Mazur, D. J. (1991). Verbal expressions of probability in informed consent litigation. *Medical Decision Making*, *11*(4), 273-281. doi:10.1177/0272989X9101100405

Moore, R. E., Kearfott, R. B., & Cloud, M. J. (2009). *Interval analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Murphy, A. H., Lichtenstein, S., Fischhoff, B., & Winkler, R. L. (1980). Misinterpretation of precipitation probability forecasts. *Bulletin of the American Meteorological Society, 61*(7), 695-701. doi: 10.1175/1520-0477(1980)061<0695:MOPPF>2.0.CO;2

National Crime Agency (2018). *National strategic assessment of serious and organized crime 2018*. http://www.nationalcrimeagency.gov.uk/publications/905-national-strategic-assessment-for-soc-2018/file

National Research Council (2011). *Intelligence analysis for tomorrow: Advances from the behavioral and social sciences*. Washington, DC: National Academies Press.

NATO Standardization Office (2016). *AJP-2.1, Edition B, Version 1: Allied Joint Doctrine for Intelligence Procedures*. Brussels, Belgium.

Olson, M. J., & Budescu, D. V. (1997). Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making*, *10*(2), 117-131. http://doi.org/dc7kfh

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. doi: 10.1016/j.obhdp.2008.05.004

Rapoport, A., Wallsten, T. S., Erev, I., & Cohen, B. L. (1990). Revision of opinion with verbally and numerically expressed uncertainties. *Acta Psychologica, 74*(1), 61-79. doi: 10.1016/0001-6918(90)90035-E

Sedlemier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*(3), 380-400. doi: 10.1037//0096-3445.130.3.380

Stastny, B. J., & Lehner, P. E. (2018). Comparative evaluation of the forecast accuracy of analysis reports and a prediction market. *Judgment and Decision Making, 13*(2), 202-211.

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. doi: 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. doi: 10.1006/obhd.1999.2857

Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: when does p = 0.3 reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making, 13*(3), 345-362. http://doi.org/fprqrs

Teigen, K. H., & Brun, W. (2003). Verbal probabilities: A question of frame? *Journal of Behavioral Decision Making, 16*(1), 53-72. doi: 10.1002/bdm.432

Tetlock, P. E., & Mellers, B. A. (2011). Intelligent management of intelligence agencies: Beyond accountability ping-pong. *American Psychologist, 66*(6), 542-554. doi: 10.1037/a0024285

Timmermans, D. (1994). The roles of experience and domain expertise in using numerical and verbal probability terms in medical decisions. *Medical Decision Making, 14*(2), 146-156.

UK House of Commons Foreign Affairs Committee (2003). *The Decision to Go to War in Iraq.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/272087/6062.pdf

UK Intelligence and Security Committee (2003). *Iraqi Weapons of Mass Destruction – Intelligence and Assessments.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/272079/5972.pdf

US Central Intelligence Agency. (1983). *Report on a study of intelligence judgments preceding significant historical failures: The hazards of single-outcome forecasting.* https://www.cia.gov/library/readingroom/docs/CIA-RDP86B00269R001100100010-7.pdf

US Congressional Select Committee on Intelligence (2005). *Report on the U.S. Intelligence Community's Prewar Intelligence Assessments on Iraq.* https://fas.org/irp/congress/2004_rpt/ssci_iraq.pdf

US Office of the Director of National Intelligence (ODNI, 2007). *Prospects for Iraq's Stability: A Challenging Road Ahead*. National Intelligence Estimate. https://www.dni.gov/files/documents/Newsroom/Press%20Releases/2007%20Press%20Releases/20070202_release.pdf

US Office of the Director of National Intelligence (ODNI, 2015). *Intelligence Community Directive 203, Analytic Standards*. https://fas.org/irp/dni/icd/icd-203.pdf

Vahabi, M. (2010). Verbal versus numerical probabilities: Does format presentation of probabilistic information regarding breast cancer screening affect women's comprehension? *Health Education Journal, 69*(2), 150-163. doi: 10.1177/0017896909349262

Verplanken, B. (1997). The effect of catastrophe potential on the interpretation of numerical probabilities of the occurrence of Hazards 1. *Journal of Applied Social Psychology, 27*(16),1453-1467. doi: 10.1111/j.1559-1816.1997.tb01608.x

Wallsten, T. S. (1990). Measuring vague uncertainties and understanding their use in decision making. *In acting under uncertainty: Multidisciplinary conceptions* (pp. 377-398). Springer, Dordrecht.

Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *The Knowledge Engineering Review, 10*(1), 43-62. doi: 10.1017/S0269888900000725

Wallsten, T. S., Budescu, D. V., & Tsao, C. (1997). Combining linguistic probabilities. *Psychologische Beitrage, 39*, 27-55.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence numerical and verbal probability judgments. *Management Science, 39*(2), 176-190. doi: 10.1287/mnsc.39.2.176

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. doi: 10.3758/BF03334162

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

Wallsten, T. S., Shlomi, Y., & Ting, H. (2008). Exploring intelligence analysts' selection and interpretation of probability terms. Final report for research contact 'expressing probability in intelligence analysis'. Sponsored by the CIA.

Wark, D. L. (1964). The definition of some estimative expressions. *Studies in Intelligence 8*(4), 67-80.

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. doi: 10.1037/0096-1523.

Wintle, 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. doi: 10.1371/journal.pone.0213522

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338-353. doi: 10.1016/S0019-9958(65)90241-X

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probabilities. *Advances in Psychology* (Vol. 16, pp. 159-182). North-Holland.

Zimmer, A. C. (1984). A model for the interpretation of verbal predictions. *International Journal of Man-Machine Studies*, *20*(1), 121-134. doi: 0.1016/S0020-7373(84)80009-7

Zwick, R., Budescu, D. V., & Wallsten, T. S. (1988). An empirical study of the integration of linguistic probabilities. In T. Zétényi (Ed.), *Advances in Psychology* (Vol 56, pp. 91-125). Amsterdam: North-Holland.