Assessing and Communicating Uncertainty Effectively in a Rapidly Changing World

David R. Mandel, Ph.D.
Defence Research and Development Canada
1133 Sheppard Avenue West, Toronto, Ontario
CANADA

David.Mandel@drdc-rddc.gc.ca

ABSTRACT

Command and policy decision-making alike depend on timely and accurate intelligence. As Sherman Kent had noted in the 1950s and 60s, intelligence assessments are seldom cold, hard facts but rather judgments made by experts under conditions of uncertainty. Since Kent, history has taught us that misjudging and miscommunicating uncertainties threaten prospects for operational and strategic successes. Nevertheless, NATO and its members' intelligence communities have persisted in using inadequate methods for assessing and communicating uncertainty, which rely on vague linguistic probabilities (e.g., "likely" or "unlikely"). Here, drawing on recent scientific evidence including from NATO SAS-114 and other research, I describe the principal reasons why the intelligence community should change course and consider using numeric probabilities for estimates that support important decisions. This is arguably more important now than ever since changes in the global security environment, which augment the importance of non-munitions targeting, call for characterization of deep uncertainties related to second- and higher-order effects.

1.0 INTRODUCTION

The need to assess and communicate uncertainty is as old as the second oldest profession. In spite of fundamental transformations of human civilization from agrarian societies, through the industrial era, and into the current post-industrial information age, intelligence assessments offered to key decision makers has remained vague, ambiguous and unnecessarily imprecise. This is so even though the collections side of intelligence has undergone leaps and bounds in technology and accompanying capability.

Even within the relatively short history of modern intelligence, the problem of effectively communicating uncertainty casts a long shadow. Following a 1951 National Intelligence Estimate on the former Yugoslavia, Sherman Kent was asked by a senior State Department official what was meant by the expression "serious possibility" in one of the key judgments in the report [1]. When Kent said that he took it to mean about a 65% chance in favour of Soviet aggression and 35% against, the official was deeply surprised. He had thought it meant a probability considerably lower. This experience prompted Kent to ask his analysts what "serious possibility" meant to them—after all, they were the one issuing judgments with such language. Kent was surprised by the results: estimates from analysts that range all the way from 20% (1:4 odds) to 80% (4:1 odds). This prompted Kent and his colleague, Max Foster, to develop a standard for communicating "words of estimative probability". The Kent-Foster approach is important not only because it was the first attempt within the US intelligence community to formalize a language of uncertainty for intelligence, but because the fundamental approach has hardly changed in 70 years, despite several intermittent lulls [2].

Current national and international standards, such as NATO's joint intelligence doctrine (i.e., AJP-2.1 [3]), rely on a curated set of verbal probability terms that are presented as a scale ordered in terms of likelihood. Some of these ordered lexical sets were given without any accompanying numeric probability equivalencies. However, nowadays, most standards assign numeric probability ranges in some form so that the stipulated meaning of each verbal probability term is bounded. For example, NATO joint intelligence doctrine uses the

terms *highly unlikely* (less the 10% [chance]), *unlikely* (10%-40%), *even chance* (40%-60%), *likely* (60%-90%), and *very likely* (greater than 90%). In the US, the Office of the Director of National Intelligence through Intelligence Community Directive 203 advises analysts to use the following terms: *almost no chance* or *remote* (1%-5%), *very unlikely* or *highly improbable* (5%-20%), *unlikely* or *improbable* (20%-45%), *roughly even chance* or *roughly even odds* (45%-55%), *likely* or *probable* (55%-80%), *very likely* or *highly probable* (80%-95%), and *almost certain* or *nearly certain* (95%-99%) [4]. Similarly, in the UK, the Cabinet Office's Professional Head of Intelligence Assessment advises analysts to use the following terms: *remote chance* (less than 5% or less than 1/20), *highly unlikely* (10%-20% or 1/10-1/5), *unlikely* (25%-35% or 1/4 -1/3), *realistic possibility* (40% to less than 50% or 4/10 to less than 1/2), *likely* or *probably* (55%-75% or 4/7-3/4), highly likely (80%-90% or 4/5-9/10), or *almost certain* (95% or more or 19/20 or more) [5].

The curated-list approach, as I discuss below, has always been woefully inadequate. However, its ineffectiveness and costliness will be greatly amplified as the information- and influence-centricity of the modern world continues to skyrocket. Modern warfare relies increasingly on a joint munitions and non-munitions targeting approach that requires the estimation of second- and higher-order effects, all of which are characterized by deep uncertainties. Continued reliance on vague verbiage ostensibly tamed by rigid impositions of meaning on terms that are familiar from vernacular usage reflects a dangerous pipe dream—a form of denial or wishful thinking that the status quo is good enough. It isn't.

2.0 THE PROBLEM

Kent's intention to transform the vagueness of natural language into something clearer and perhaps more precise was noble, but the curated-list approach has proven to be flawed in practice and in principle. Some of the flaws in practice are evident in the three examples given earlier. Each standard uses the same approach, yet each differs sufficiently to undermine interoperability among key collaborative partners. For instance, an *even chance* issued by NATO could mean *unlikely, roughly even chance,* or *likely* in the US system. NATO and the US have terms that overlap each other on the numeric probability scale. For instance, in the NATO standard, a 40% chance could be described as *unlikely* or an *even chance*. In contrast, the UK standard leaves chunks of the probability scale off the map; there are no words to express probabilities between 5% and 10%, between 20% and 25%, between 35% and 40%, and so on. A similar example of arbitrary decision making is excising the analyst's ability to communicate probabilities less than 1% or greater than 99%. So much for distinguishing "one in a hundred" from "one in a million"—in the US approach, "one in a hundred" is the smallest probability there is, while in the NATO and UK standards, "one in a million" would in any case be indistinguishable from "one in ten." *Orders of magnitude*, what of it?

The curated-list approach has many other examples of arbitrariness and institutionalized coarseness that preclude analysts from providing the level of precision and clarity they may be capable of providing to decision makers. Several chapters in the forthcoming SAS-114 final report detail these and other limitations (e.g., [6]-[8]). Moreover, there is now ample evidence that curated-lists impede judgment quality. When geopolitical forecasters' numeric probability forecasts are forced into the US curated-list standard, forecasting accuracy takes a substantial hit, and the loss in accuracy tends to be greatest for the most competent forecasters [9]. This important finding should signal to all that the curated-list approach is flawed not only in practice, but also in principle. Since providing decision makers with a clear indication of future events is a vital function of intelligence, the curated-list approach institutionalizes a reduction in forecasting accuracy. These methods force analysts to be less informative to decision makers (and other analysts) than they could be.

There is also ample evidence that the curated-list approach fails terribly in its central function to standardize the meaning of the items of the curated list. Several studies now show that when people are presented with such standards and then receive estimates using the probabilistic terms, a staggering proportion (often the majority) do not interpret the terms in ways that fall within the numeric ranges that are supposed to bound their meaning, even when those ranges are large [10]-[12]. Moreover, allowing people to consult the standard does

little to improve consistency rates. Putting the numeric equivalencies in the text helps more, but that runs the risk of confusing decision-makers who might take the intervals to represent confidence intervals on the substantive assessment—after all, the substantive issue is what the decision-maker cares about. It turns out language is not so easily tamed.

2.1 Inconvenient Truths

One might ask why intelligence communities would adopt standards that, in principle, detract from analytic assessment quality. The extent to which this state of affairs is due to motivated reasoning or sheer ignorance is debatable. It does not appear to be an either-or scenario and there is perhaps limited value in analysing the exact contributions of these alternative explanatory sources of the current state of methodology for communicating uncertainty. What is clear, however, is that intelligence communities persist in believing several convenient fictions that preclude the kind of disruptive change required to exit the ineffective local minima in which they are currently trapped. For each of these there is a corresponding inconvenient truth, which intelligence communities will need to understand and accept before real progress can be made.

First, intelligence communities must stop telling themselves that analysts are incapable of providing granular probability assessments about threats and opportunities in geopolitical areas of interest to their leaders. This has been thoroughly debunked in recent years by research funded by the US intelligence community through the Intelligence Advanced Research Projects Activity (IARPA) Aggregative Contingent Estimation (ACE) program. The scientific findings from this well-funded program not only show how intelligence communities can foster better analytic judgments through selection, training, elicitation, recalibration and aggregation methods (e.g., [13]-[14]), they clearly reveal (based on an enormous quantity of scientific data) the costs to accuracy caused by forcing geopolitical forecasters' numeric probability estimates into the coarse curated lists used by virtually all modern intelligence organizations [9]. Intelligence communities should desist in believing that their analysts cannot do better than discriminating between about 5-7 degrees of probability. Not only can they do so, practicing doing so will make them better at it. Barnes [15] provides a good example of how these beliefs can melt away with direct experience using numeric probabilities. When he instituted the practice of assigning numeric probabilities to forecasts in intelligence products, many of his analysts (in a strategic intelligence unit of the Canadian government) balked at the idea, but Barnes observed that such opposition soon turned into fierce debates among analysts regarding whether an assessment should be, say, a 70% chance versus an 80% chance (even in Barnes' approach, though, analysts only had up to 11 degrees of probability).

Second, intelligence communities must stop insisting that they know that their consumers don't want precision in assessments. This is contradicted not only by scientific evidence showing that in several domains, receivers of communications about uncertain estimates prefer to receive numeric estimates rather than verbal estimates of probability (e.g., [16]-[19]), but also by research on US presidential decision making [20]. Moreover, intelligence communities should disabuse themselves of the false notion that precision will trigger unwarranted risk taking in decision makers. Recent evidence shows the opposite—decision makers are more cautious when they receive clearer information presented in numeric form than when they receive it via verbal probabilities [21]. Intelligence professionals might be concerned about such outcomes not only for virtuous reasons such as wanting to avoid false pretences for war, as in the 2003 Iraq weapons of mass destruction fiasco, but also for political cover from blame. Yet here too, recent evidence suggests that misestimates issued with verbal probabilities incur greater reputational harm than the same misestimates issued with corresponding numeric probabilities [22-23].

Third, and related to the previous point, intelligence communities must stop assuming that precision isn't valuable for national security decision-making—it is whether or not decision-makers realize it. One of the intelligence community's great clichés is that the world is "complex" and that the intelligence community must come to terms with "increasing complexity" [24]. Yet the irony of clinging to vague verbiage seems lost on the same community. Intelligence professionals not only swear that decision-makers don't want precision, they also claim that such precision is rarely, if ever, beneficial. But how then can analysts and decision-makers

properly grapple with the deep uncertainties that pervade the complex world they inhabit? Complex realities require models of complexity and these invariably require interactions among probabilistic causal pathways to be estimated. Reliance on vague verbiage all but precludes the analyst from doing so. After all, how can the analyst estimate the conjunctive probability of even independent factors that represent necessary conditions for a threat scenario to manifest if each factor is merely coded with words like "remote" or "likely"? The curated-list approach simply precludes all necessary computation required for estimating the effects of complex (and even most comparatively simple) events.

Finally, intelligence communities must educate themselves about the nature of probability and brush away their many misconceptions that have thwarted a clear understanding of the issue. For instance, they should understand that the use of numeric probabilities does not imply that it is a scientific estimate. The foundations of modern decision theory are based on the idea that degrees of belief in a proposition can be quantified in terms of numeric probabilities, either directly or through the elicitation of bets [25]. Subjective probabilities most certainly can be quantified, and Bayesianism is a theoretical framework for doing so coherently [26]. The intelligence community should be teaching this to analysts, especially since there is evidence that such training even in a single brief session pays off [27].

3. THE SOLUTION

The solution to the assessment and communication challenges faced by intelligence communities is to use numeric probabilities wherever feasible. Numeric probabilities are not only convey probability information more clearly than verbal probabilities, they also are less prone to conveying implicit recommendations to decision-makers [28]. Therefore, they are less likely to conflate informing decision-makers with biasing their policy preferences. Based on over a decade of experience working with intelligence professionals, I am sure that this recommendation will not sit well with most of them. Just as research shows that receivers prefer numeric expressions of uncertainty, the same studies show that most senders prefer to use verbal probabilities to communicate their uncertainty (e.g., [16]-[19]). Most analysts and their managers appear to be no different. However, in the context of national-security intelligence, the aim should not be to make the intelligence professional's life as comfortable as possible but rather to make intelligence assessments as relevant, accurate and informative as possible for decision-makers whose choices ultimately affect national and international security.

Intelligence professionals who brush off the idea of using numeric probabilities on a granular level as merely the whims of scientific geeks out of touch with "the real world" are doing their organizations and the citizens they represent a great disservice. It is their responsibility to do better, to be better informed about the relevant science, and to sponsor research needed to fill critical gaps in knowledge. At present, the intelligence community relies much too heavily on the "good ideas" of "tradecraft" insiders who possess little scientific training [29]. These so-called experts are almost invariably dilettantes when it comes to the task of improving expert judgment. There is a genuine expert population that deals with such prescriptive challenges and they typically go by the name of decision scientists. Remarkably, the realization that most good ideas nevertheless fail has not permeated senior levels of leadership, as most tradecraft methods (a fancy term for the dilettantes' heuristics) remain scientifically untested [30]-[32] and recent tests indicate that tradecraft methods can impede rather than boost analysts' judgment quality [33-34]. The reliance on dilettantes for so many decades has also fostered the current level of incompatibility among curated-list approaches. Of course, each so-called expert or expert committee believes it has the best available method, even though it has been shown that the existing methods are easily improved upon using low-cost behavioral research methods [35].

The world is rapidly changing but the intelligence community's approach to analytic methods, including the assessment and communication of uncertainty, has been glacial in its progress. It is time for intelligence tradecraft to enter the 21st century or at least catch up with the intellectual developments of the 20th century.

REFERENCES

[1] Kent, S. (1964). Words of estimative probability. *Studies in Intelligence*, *8*(4), 49-65.

[2] Marchio, J. (2014). Analytic tradecraft and the intelligence community: Enduring value, intermittent emphasis. *Intelligence and National Security*, *29*(2), 159-183.

[3] North Atlantic Treaty Organization. (2016). *Allied Joint Doctrine for Intelligence Procedures AJP-2.1*. Brussels, Belgium.

[4] Office of the Director of National Intelligence. (2015). *Intelligence Community Directive ICD 203: Analytic Standards*. Retrieved from: https://fas.org/irp/dni/icd/icd-203.pdf

[5] UK Defence Intelligence. (n.d.). *PHIA probability yardstick.* London: author.

[6] Dhami, M. K., & Mandel, D. R. (in press). UK and US policies for communicating probability in intelligence analysis: A review. In D. R. Mandel (Ed.), *Assessment and Communication of Uncertainty in Intelligence to Support Decision Making: Final Report of Research Task Group SAS-114*. Brussels, Belgium: NATO Science and Technology Organization. doi: 10.14339/STO-TR-SAS-114

[7] Irwin, D. & Mandel, D. R. (in press). Variants of vague verbiage: Intelligence community methods for communicating probability. In D. R. Mandel (Ed.), *Assessment and Communication of Uncertainty in Intelligence to Support Decision Making: Final Report of Research Task Group SAS-114*. Brussels, Belgium: NATO Science and Technology Organization. doi: 10.14339/STO-TR-SAS-114

[8] Rein, K. (in press). Issues of uncertainty in natural language communications. In D. R. Mandel (Ed.), *Assessment and Communication of Uncertainty in Intelligence to Support Decision Making: Final Report of Research Task Group SAS-114*. Brussels, Belgium: NATO Science and Technology Organization. doi: 10.14339/STO-TR-SAS-114

[9] Friedman, J. A. (2019). *War and Chance: Assessing Uncertainty in International Politics.* New York, NY: Oxford University Press.

[10] Budescu, D. V., Por, H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change*, *113*(2), 181-200.

[11] Budescu, D. V., Por, H., Broomell, S. B., & Smithson, M. (2014). The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, *4*(6), 508-512.

[12] 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-e0213522.

[13] Karvetski, C. W., Olson, K. C., Mandel, D. R., & Twardy, C. R. (2013). Probabilistic coherence weighting for optimizing expert forecasts. *Decision Analysis, 10*(4), 305-326.

[14] Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. New York, NY: Crown Publishing Group.

[15] Barnes, A. (2016). Making intelligence analysis more intelligent: Using numeric probabilities. *Intelligence and National Security*, *31*, 327–344.

[16] Brun, W., & Teigen, K. (1988). Verbal probabilities: Ambiguous, context dependent, or both? *Organizational Behavior & Human Decision Processes*, *41*, 390-404.

[17] Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior & Human Decision Processes, 44,* 1-18.

[18] 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,* 135-138.

[19] Olson, M. J., & Budescu, D. V. (1997). Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making, 10,* 117-131.

[20] Friedman, J. A., & Zeckhauser, R. (2015). Handling and mishandling estimative probability: likelihood, confidence, and the search for bin Laden", *Intelligence and National Security, 30*(10), 77-99.

[21] 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.

[22] Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2017). Maintaining credibility when communicating uncertainty: The role of communication format. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 582-587). Austin, TX: Cognitive Science Society.

[23] Jenkins, S. C., Harris, A. J. L., & Lark, R. M. (2018). When unlikely outcomes occur: the role of communication format in maintaining communicator credibility. Journal of Risk Research, 22, 537-554.

[24] Kerbel, J. (2015, August), The complexity challenge: The U.S. government's struggle to keep up with the times. *The National Interest*. Retrieved from: https://nationalinterest.org/feature/the-complexity-challenge-the-us-governments-struggle-keep-13698

[25] Savage, L. J. (1954). *The Foundations of Statistics.* New York, NY: Wiley.

[26] Navarrete, G., Mandel, D. R., eds. (2016). *Improving Bayesian Reasoning: What Works and Why?* Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-745-3.

[27] Mandel, D. R. (2015). Instruction in information structuring improves Bayesian judgment in intelligence analysts. *Frontiers in Psychology, 6:*387, 1-12. doi: 10.3389/fpsyg.2015.00387

[28] Collins, R. N., & Mandel, D. R. (in press). Cultivating credibility with probability words and numbers. *Judgment and Decision Making.*

[29] Mandel, D. R. (2019). Can decision science improve intelligence analysis? In S. Coulthart, M. Landon-Murray, & D. Van Puyvelde (Eds.), *Researching National Security Intelligence: Multidisciplinary Approaches* (pp. 117-140). Washington, DC: Georgetown University Press.

[30] 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: http://dx.doi.org/10.1080/02684527.2017.1400230

[31] Dhami, M. K., Mandel, D. R., Mellers, B. A., & Tetlock, P. E. (2015). Improving intelligence analysis with decision science. *Perspectives on Psychological Science, 106*(6), 753-757.

[32] 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

[33] Dhami, M. K., Belton, I., & Mandel, D. R. (2019). The 'Analysis of Competing Hypotheses' in intelligence analysis. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3550

[34] Mandel, D. R., Karvetski, C. & Dhami, M. K. (2018). Boosting intelligence analysts' judgment accuracy: what works, what fails? *Judgment and Decision Making, 13*(6), 607-621.

[35] Ho, E. Budescu, D. V., Dhami, M. K., & Mandel, D. R. (2015). Improving the communication of uncertainty in climate science and intelligence analysis. *Behavioral Science & Policy, 1*(2), 43-55.