Improving Intelligence Analysis with Decision Science

Mandeep K. Dhami, Middlesex University, London, UK
David R. Mandel, Defence Research and Development Canada, Toronto Research Centre and York University, Canada
Barbara A. Mellers, University of Pennsylvania, USA
Philip E. Tetlock, University of Pennsylvania, USA

Abstract

Intelligence analysis plays a vital role in policy decision-making. Key functions of intelligence analysis include accurately forecasting significant events, appropriately characterizing the uncertainties inherent in such forecasts, and effectively communicating those probabilistic forecasts to stakeholders. We review decision research on probabilistic forecasting and uncertainty communication, drawing attention to findings that could be used to reform intelligence processes and contribute to more effective intelligence oversight. We recommend that the intelligence community regularly and quantitatively monitor its forecasting accuracy to better understand how well it is achieving its functions. We also recommend that the intelligence community use decision science to improve these functions (namely, forecasting and communication of intelligence estimates made under conditions of uncertainty). In the case of forecasting, decision research offers suggestions for improvement that involve interventions on data (e.g., transforming forecasts to debias them) and behavior (e.g., via selection, training, and effective team structuring). In the case of uncertainty communication, the literature suggests that current intelligence procedures, which emphasize the use of verbal probabilities, are ineffective. The intelligence community should, therefore, leverage research that points to ways in which verbal probability use may be improved as well as exploring the use of numerical probabilities wherever feasible.

Introduction

Intelligence analysis involves searching for, selecting, processing, and interpreting data in order to gain an awareness of current situations and forecast potentially important future developments in areas of interest to decision-making stakeholders. Although intelligence analysis is not a branch of science, it bears some important similarities. As in science, it involves generating and testing hypotheses, and accurately characterizing the degrees of uncertainty in both the evidence and conclusions reached. Although there are standard processes by which the intelligence community (IC) directs analysis, and collects, processes, and disseminates intelligence (e.g., Department of Defense, 2013), these tend not to be based on scientific methods, theories, or past research findings (see National Research Council, 2011; Pool, 2010). In this paper, we identify two relevant bodies of literature in the field of decision science that can be used to inform the IC's policies and practices for intelligence analysis and the dissemination of analytic products. The first research area examines methods for assessing and improving forecasting accuracy, whereas the second examines communication of uncertainty using verbal and numerical probabilities. In practice, these two areas are closely related because intelligence forecasts must be accurately qualified by degrees of uncertainty, and those probabilistic forecasts must in turn be communicated with high fidelity to decision-makers.

Decision Research on Forecasting Accuracy

The IC instructs analysts to be accurate (e.g., Office of the Director of National Intelligence, 2015). However, the IC does not routinely and quantitatively track predictive accuracy to verify that its forecasts are accurate. There are at least three good reasons the IC should track accuracy in a proactive manner: First, forecast accuracy is an empirical issue, and without proper quantitative tracking, one cannot know how good accuracy is or whether improvement is possible. Moreover, various methods for scoring aspects of forecasting skill are readily available (e.g., Swets, 1986; Yaniv, Yates, & Smith, 1991) and can also be applied to rank-ordered data, such as where a verbal probability scale is used to characterize uncertainty (Liberman & Tversky, 1993). Second, political experts have been shown to be overconfident in their forecasts of geopolitical events, and they are easily outperformed by all but the most basic statistical models (Tetlock, 2005). In fact, overconfidence in judgment has been documented in other areas of expert judgment, such as medical diagnosis (e.g., Dawson et al., 1993) and legal judgments (Goodman-Delahunty, Granhag, Hartwig, & Loftus, 2010). Third, the few studies that have examined the forecast accuracy of actual intelligence forecasts made by analysts have yielded mixed results. The most comprehensive study to date found very good performance among strategic intelligence analysts that used numerical probabilities (Mandel & Barnes, 2014) as well as among those who used verbal probabilities to communicate uncertainty (Mandel, Barnes, & Richards, 2014). However, substantially weaker performance in forecasts qualified by verbal probabilities was found in another study with fewer forecasts (Lehner, Michelson, Adelman, & Goodman, 2012).

In addition to assessing forecast accuracy, decision science could help the IC improve its forecast accuracy. Recent psychological research on forecasting has examined statistical interventions that substantially improve the accuracy of probabilistic predictions. These include aggregation algorithms that show an improvement over unweighted linear opinion pools by giving more weight to forecasters who exhibit greater coherence in their judgments of related topics (Karvestski, Olson, Mandel, & Twardy, 2013) or have better track records (Satopaa et al, 2014). Likewise, transformation rules that improve calibration, an important facet of forecasting skill, have been recently documented (Baron et al., 2014), including in use with intelligence forecasts (Mandel & Barnes, 2014).

Other forecasting research has examined behavioral interventions that improve accuracy. For instance, probabilistic judgments are more accurate with training in probabilistic reasoning, cautionary tales about errors and biases (e.g., overconfidence and confirmation biases), and practical advice for making predictions, such as considering multiple reference classes (Mellers et al. 2014). Likewise, training using visual representations of nested sets in the form of natural frequencies is effective for improving analysts' probability judgments (Mandel, 2015). Other research has shown that individual-difference testing can be used to identify attributes of better forecasters, such as actively open-minded thinking, knowledge of task-specific skills, and numeracy (Karvetski et al., 2013; Mellers et al, 2015a). Forecasting can also be improved by using well-coordinated teamwork that allows forecasters to share information, debate rationales, and motivate each other to perform well (Mellers et al., 2014). Finally, assignment of the best forecasters into so-called "superteams", allowing elite performers to interact online with each other, yields additional improvements to forecasting accuracy (Mellers et al., 2015b; Tetlock & Gardner, 2015).

Decision Research on Communicating Uncertainty

As noted earlier, analysts work under conditions of uncertainty and are expected to accurately characterize uncertainties regarding their conclusions. Because their judgments are intended for decision-makers, they have to both characterize the uncertainties and express them clearly to end-users. Uncertainty in intelligence assessments is typically communicated using verbal probabilities. For instance, the National Intelligence Estimate on weapons of mass destruction in Iraq stated, "if left unchecked, it [Iraq] *probably will* have a nuclear weapon during this decade" (Friedman & Zeckhauser, 2012, p. 829, italics added). An important question is whether decision-makers interpret the meaning of such statements in the same way as it was intended by the intelligence organization producing the assessment.

Decision research clarifies several points about the use of verbal probabilities pertinent to that question, which policymakers should consider. First, people receiving communications about uncertain estimates prefer them to be expressed numerically even though communicators prefer to use words to convey uncertainty (Brun & Teigen, 1988; Murphy, Lichtenstein, Fischhoff, & Winkler, 1980; Wallsten, Budescu, Zwick, & Kemp, 1993). Second, because of the vagueness of verbal probabilities, most terms have a wide range of permissible meanings when scaled on the 0-1 interval (Dhami & Wallsten, 2005; Karelitz & Budescu, 2004). Third, and more problematic, individuals vary greatly in the ranges and best estimates they assign to probability terms (Budescu, Weinberg, & Wallsten, 1988; Dhami & Wallsten, 2005). Finally, people's interpretations of verbal probabilities are affected by several contextual factors, such as whether the event whose uncertainty is being characterized has a low or high base-rate (for review, see Wallsten & Budescu, 1995).

Decision research also sheds light on how uncertainty communication might be improved. For instance, given the vagueness inherent in verbal probabilities, methods have been developed for translating verbal expressions in a communicator's lexicon to equivalently ranked expressions in a listener's lexicon (Dhami & Wallsten, 2005; Karelitz & Budescu, 2004). Another approach is to establish a standardized lexicon of verbal probabilities, and indeed many intelligence organizations have done so (e.g., Barnes, 2015; Dhami, 2013). Behavioral research methods have been used to develop standardized lexicons that people are more likely adhere to (Ho, Budescu, Dhami & Mandel, in press). Communication might also be improved through presentational methods. For instance, when people are given a lexicon equating verbal probabilities with numerical ranges, they often lose track of the equivalents (Budescu, Por, Broomell, & Smithson, 2014). However, consistency with the lexicon can be substantially improved by providing the numerical range equivalents each time a probability term is used in a statement.

Policy Implications

Forecast Accuracy

Our recommendations for improving forecast accuracy within the IC are two-pronged (see Table 1). First, the IC should take immediate steps to monitor its forecast accuracy. This would require intelligence organizations to collect probabilistic forecasts, outcomes (i.e., did the forecasted event occur or not?), and putative moderators of forecasting skill (e.g., analyst experience, forecast time-frame, forecast difficulty). These data would enable the IC to quantitatively score analytic forecasts so that key aspects of forecasting skill, such as calibration and discrimination, could be tracked over time, agencies, and other key variables, some of which might prove to be important moderators of forecasting skill.

Secondly, we recommend that the IC leverage decision research, theory, and methods to improve its forecasting abilities. We outlined several recent developments that could be clustered into two routes of improvement. The first focuses on methods for improving forecast quality through interventions on data sources (including raw forecasts), whereas the second focuses on doing so through behavioral interventions (e.g., through selection, training, and team structuring). We also advise the IC to work with behavioral scientists to devise fair tests of the effectiveness of these methods as implemented. This is an important step that is sorely lacking in the IC's application of structured analytic techniques to promote analytic rigor (Pool, 2010).

Communication of Uncertainty

We recommend two courses of action for improving the communication of uncertainty (see Table 1). The first, as we noted earlier, focuses on exploiting means of improving communication using verbal probabilities, such as using translation methods or presentational methods that at least remind users of what the vague terms are supposed to mean. The IC should also monitor inconsistencies among standards promulgated in different organizations, and take steps to eliminate discrepancies that might proliferate rather than mitigate confusion.

The second course of action we recommend is to use numerical probabilities in place of verbal probabilities wherever feasible. To the extent that communications about uncertainty, risk, and probability are intended to be as accurate as possible, the scientific literature reviewed earlier makes a clear case for the use of numerical probabilities. Numerical probabilities can be precise (e.g., 87.5% chance) or imprecise (e.g., 75% chance of rain plus or minus 10%). However, even when such probabilities express imprecision, they are not vague.

Conclusion

The IC should adopt an evidence-based approach to monitoring and improving its performance. Such an approach would address the aforementioned key challenges and also strengthen the IC's accountability processes, enabling it to better protect itself from the adverse consequences of blame-games that ensue following significant errors (Tetlock & Mellers, 2011). Decision science is well positioned to assist the IC because of its quantitative methods for measurement and testing cognition and behavior, its theoretical models of human judgment and decision-making, and its history of dealing with applied problems. The IC should capitalize on it.

References

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

Baron J., Mellers, B. A., Tetlock, P. E., Stone, E., & Ungar, L. H. (2014). Two reasons to make aggregated probability forecasts more extreme. *Decision Analysis*, *11*, 133–145.

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

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

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

Dawson, N. V., Connors, A. F., Speroff, T., Kemka, A., Shaw, P., & Arkes, H. R. (1993). Hemodynamic assessment in managing the critically ill: Is physician confidence warranted? *Medical Decision Making, 13*, 258–266.

Department of Defense (2013). *Joint intelligence. Joint publication 2-0.* Retrieved from http://www.dtic.mil/doctrine/new_pubs/jp2_0.pdf

Dhami, M. K. (2013). *Understanding and communicating uncertainty in intelligence analysis.* Report prepared for HM Government, UK. (available from first author).

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

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

Goodman-Delahunty, J., Granhag, P.A., Hartwig, M., & Loftus, E.F. (2010). Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy & Law, 16*, 133–157.

Ho, E., Budescu, D., Dhami, M. K., & Mandel, D. R. (in press). On the effective communication of uncertainty: Lessons from the climate change and intelligence analysis domains. *Behavioral Science & Policy.*

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

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

Lehner, P., Michelson, A., Adelman, L., & Goodman, A. (2012). Using inferred probabilities to measure the accuracy of imprecise forecasts. *Judgment and Decision Making, 7,* 728-740.

Liberman, V., & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin*, *114*, 162–173.

Mandel, D. R., & Barnes, A. (2014). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences*, *111*, 10984-10989.

Mandel, D. R. (2015). Instruction in information structuring improves Bayesian judgment in intelligence analysts. *Frontiers in Psychology, 6*:387, 1-12.

Mandel, D. R., Barnes, A., & Richards, K. (2014). *A quantitative assessment of the quality of strategic intelligence forecasts*. Technical Report no. 2013-036. Toronto, Canada: Defence Research and Development Canada.

Mellers, B. A., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical tournament. *Psychological Science, 25*, 1106-1115.

Mellers, B. A., Stone, E., Atanasov, P., Roghbaugh, N., Metz, S.E., Ungar, L., Bishop, M., Horowitz, M., Merkle, E., & Tetlock, P. E. (2015a). *Journal of Experimental Psychology: Applied, 21*, 1-14.

Mellers, B. A., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L., & Tetlock, P. E. (2015b). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science, 10,* 267-281.

Murphy, A. H., Lichtenstein, S., Fischhoff, B., & Winkler, R. L. (1980). Misinterpretation of precipitation probability forecasts. *Bulletin of the American Meteorological Society, 6*, 695-701.

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

Office of the Director of National Intelligence (2015). *Intelligence community directive 203: Analytic standards*. Retrieved from http://www.dni.gov/files/documents/ICD/

Pool, R. (2010). *Field evaluation in the intelligence and counterintelligence context: Workshop summary*. Washington, DC: National Academies Press.

Satopaa, V. A., Jensen, S. T., Mellers, B. A., Tetlock, P. E., & Ungar, L. (2014). Aggregation in time-series: Dynamic hierarchical modeling of sparse expert beliefs. *The Annals of Applied Statistics, 8,* 1256-1280.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*, 181–198. 27.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York: Crown.

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

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

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.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110*, 611–617.