



Assessing Uncertainty in Intelligence

Citation

Friedman, Jeffrey A., and Richard Zeckhauser. 2012. Assessing Uncertainty in Intelligence. HKS Faculty Research Working Paper Series RWP12-027, John F. Kennedy School of Government, Harvard University.

Published Version

<http://web.hks.harvard.edu/publications/citation.aspx?PubId=8427>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:9359827>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Assessing Uncertainty in Intelligence

Faculty Research Working Paper Series

Jeffrey A. Friedman
Harvard Kennedy School

Richard Zeckhauser
Harvard Kennedy School

June 2012
RWP12-027

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

ASSESSING UNCERTAINTY IN INTELLIGENCE

Jeffrey A. Friedman

Ph.D. Candidate in Public Policy, Harvard Kennedy School

Jeffrey_Friedman@harvard.edu

Richard Zeckhauser

Frank Ramsey Professor of Political Economy, Harvard Kennedy School

Richard_Zeckhauser@harvard.edu

Revised May 21, 2012

ABSTRACT

This article addresses the challenge of managing uncertainty when producing estimative intelligence. Much of the theory and practice of estimative intelligence aims to eliminate or reduce uncertainty, but this is often impossible or infeasible. This article instead argues that the goal of estimative intelligence should be to assess uncertainty. By drawing on a body of nearly 400 declassified National Intelligence Estimates as well as prominent texts on analytic tradecraft, this article argues that current tradecraft methods attempt to eliminate uncertainty in ways that can impede the accuracy, clarity, and utility of estimative intelligence. By contrast, a focus on assessing uncertainty suggests solutions to these problems and provides a promising analytic framework for thinking about estimative intelligence in general.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the input of Daniel Altman, Fulton Armstrong, Richard Betts, Richard Cooper, Jeffrey Edmonds, Burton Gerber, Andrew Harter, William Hogan, Robert Jervis, Loch Johnson, Jason Matheny, Peter Mattis, Sean Nolan, Joseph Nye, Randolph Pherson, Paul Pillar, Stephen Rosen, Daniel Simons, Mark Thompson, Gregory Treverton, two anonymous reviewers, and intelligence officials who asked to remain unnamed.

ASSESSING UNCERTAINTY IN INTELLIGENCE

Managing uncertainty is essential to making foreign policy. To understand and confront the challenges facing the United States and its allies, policymakers need to make assumptions about the likelihood and potential consequences of various events and scenarios. A critical function of the intelligence community is to help policymakers form these assumptions, especially when they depend on factors that are uncertain and complex.

When intelligence products make uncertain judgments, they fall under the category of estimative intelligence. The production of estimative intelligence is a large and important subset of the intelligence process. It involves predicting the behavior of terrorist groups, determining the security of Pakistan's nuclear arsenal, judging China's capabilities and intentions, assessing the risk of Iran building nuclear weapons, and addressing many other issues of great importance for international security.

Any analyst studying these issues would wish to make judgments that are as certain as possible. This impulse, however, can impair the accuracy, clarity, and utility of intelligence estimates. These problems frequently fall under one of two complementary categories. *Consequence neglect* occurs when collectors, analysts, and consumers of intelligence focus too much on the probability of each possible scenario and too little on the magnitude of those scenarios' potential consequences. *Probability neglect* is the reverse problem, arising when intelligence focuses predominantly on the potential consequences of various possibilities while giving less attention to their respective likelihoods. When likelihoods and consequences are not identified separately and then considered together, estimative intelligence will be incomplete, unclear, and subject to misinterpretation.

The main argument in this article is that trying to *eliminate* uncertainty fosters these problems, while attempting to *assess* uncertainty helps to avoid them. In particular, the following sections argue against the view that analysts should view multiple possibilities as different and competing hypotheses; discuss the way that the concepts of likelihood and confidence are used and sometimes conflated; and describe potential shortcomings in the way that information is evaluated and filtered along the analytic chain. These subjects engage debates about the basic purposes and principles of intelligence estimating. They illuminate the challenges of dealing with uncertainty in intelligence, and provide an opportunity to discuss how those challenges can be addressed.

Eliminating uncertainty versus assessing uncertainty

Estimative intelligence often focuses on issues that are impossible to address with certainty. Matters such as the eventual outcome of the war in Afghanistan, or China's stance toward the United States in fifteen years, or the ultimate result of political turmoil in Syria involve such a

vast range of factors that no one can foresee exactly what will happen in each instance. Estimative intelligence also deals with questions where certain answers are possible in principle, but infeasible in practice. A prominent example would be assessing the status of Iran's nuclear program. It is theoretically possible to know this conclusively, but there are practical limits to obtaining this knowledge, as the intelligence community lacks direct access to high-level officials and relevant facilities. Any intelligence analyst studying Iran's nuclear program will have to work with incomplete information, and that information will not be sufficient to draw definitive conclusions.¹

Of course, intelligence analysts will rarely possess determinative evidence about every aspect of a given situation – almost all intelligence products entail uncertainty of some kind. Uncertainty is most important when analysts are wrestling with multiple possibilities that have meaningfully different implications for policy. This is the kind of challenge where estimative intelligence can play an important role in helping policymakers to form and revise their assumptions, and it is the subject of this article.

Roughly speaking, there are two distinct, ideal-type views about what the goals of estimative intelligence should be in dealing with this kind of uncertainty. One ideal-type view of estimative intelligence is that its goal should be to *eliminate uncertainty*, or at least to reduce it as much as possible. In this view, estimative questions have correct answers. Those answers might be difficult to ascertain, but the analyst's goal should be to make the strongest possible determination about what those answers are. As analysts gain information and improve their conceptual frameworks, they should be able to reduce the amount of uncertainty they face. The more uncertainty that remains, the more this indicates that analysts need better information or better concepts.

This view of intelligence is widespread. A recent ethnography of the intelligence community found that many analysts believe intelligence errors are 'factual inaccuracies resulting from poor or missing data' or from 'incorrect, missing, discarded, or inadequate hypotheses'.² Another scholar argues that Americans have an 'unbounded faith' in the ability of the intelligence

¹ A related example was the intelligence community's ten-year effort to locate Osama bin Laden. There was a correct answer to the question of where bin Laden was hiding, but uncertainty about bin Laden's location remained even up to the point where President Obama authorized a raid against his compound.

² Rob Johnston, *Analytic Culture in the U.S. Intelligence Community* (Washington, DC: Center for the Study of Intelligence 2005), p.xviii. Cf. Michael Heazle, 'Policy Lessons from Iraq on Managing Uncertainty in Intelligence Assessment: Why the Strategic/Tactical Distinction Matters', *Intelligence and National Security* 25/3 (2010), p.297, which refers to a similar aspect of analytic culture that the author calls 'positivist perceptions of knowledge' expressed as 'the truth is out there and it can be known'. This aspect of analytic culture is as old as the estimative process itself. Roger Hilsman, Jr., 'Intelligence and Policy-Making in Foreign Affairs', *World Politics* 5/1 (1952), pp.11 and 13, criticizes 'the implied assumption... that truth is obvious once all the facts are known', a point of view 'accepted with so little question'.

community to ‘hold accurate images of the outside world’.³ This perspective has understandable appeal. Certainty simplifies decision making. It is not surprising that consumers of intelligence wish to have it, and that producers of intelligence seek to provide it. As the following sections will show, a push to eliminate or reduce uncertainty characterizes many standard methods of intelligence analysis.

A second ideal-type view of estimative intelligence is that its goal should be to *assess uncertainty*. In this view, it makes little sense to seek a single ‘right answer’ to many estimative questions. In fact, good intelligence often reveals new uncertainties: as analysts gain information and improve their conceptual frameworks, they may identify additional possibilities that they had not previously considered. That should not be seen as a problem, since the goal of intelligence is to describe the uncertainty that surrounds a particular question, and not to eliminate or to reduce this uncertainty per se.

Since the very *definition* of estimative intelligence is that it involves making uncertain judgments, it would seem as though the subject inherently belongs in this second category. Yet while the theory and practice of estimative intelligence often appear to favor assessing uncertainty in an accurate manner, many standard practices actually push in a different direction, albeit in ways that are often subtle and possibly unintended.

For example, scholars and practitioners often speak about evaluating intelligence in terms of an analyst’s ‘batting average’. Many authors argue that the batting average is an appropriate metaphor because it accepts the idea that intelligence estimates, like baseball players, will fall short much of the time. Flawless performance is an unreasonable goal, but analysts should at least try to ‘raise the batting average’ of their intelligence estimates.⁴ This metaphor is explicitly aimed to avoid pressuring analysts to state their estimates with certainty.

This proposal, though seemingly intuitive, is misguided – for if analysts truly wished to maximize their batting averages, then they would end up offering judgments that are as certain as possible. To see this, consider a situation in which an analyst might have been asked in spring 2012 to predict whether or not unrest in Syria would unseat Bashar al-Assad within the year. Say she believed it was 60 percent likely that al-Assad would stay and 40 percent likely that al-Assad would go. Now imagine that the analyst would be scored based on whether she got her prediction

³ Paul R. Pillar levels this critique in his book *Intelligence and U.S. Foreign Policy: Iraq, 9/11, and Misguided Reform* (NY: Columbia 2011), p.4.

⁴ For examples, see Allen Dulles, *The Craft of Intelligence* (NY: Harper and Row 1963), p.155; Harold P. Ford, *Estimative Intelligence: The Purposes and Problems of National Intelligence Estimating* (Lanham, MD: Defense Intelligence College 1993), p.86; Mark M. Lowenthal, *Intelligence: From Secrets to Policy*, 4th ed. (Washington, DC: CQ Press 2009), p.148; Kristan J. Wheaton, ‘Evaluating Intelligence: Answering Questions Asked and Not’, *International Journal of Intelligence and CounterIntelligence*, 22/4 (2009), pp.614-631; John A. Gentry, ‘Assessing Intelligence Performance’ in Loch K. Johnson (ed.), *Oxford Handbook of National Security Intelligence* (NY: Oxford 2010); Thomas J. Fingar, *Reducing Uncertainty: Intelligence Analysis and National Security* (Stanford: 2011), p.3.

right. One way to do this would be to ask her to make a single prediction. If this prediction turned out to be correct, the analyst would earn a ‘hit’. Of course, the analyst would then predict al-Assad will stay, since she believes this to be the most likely outcome, resulting in an expected score of $0.6 \times 1.0 = 0.6$. This is a certain, single-outcome prediction. In order to encourage this analyst to take multiple outcomes into account, one might instead ask her to assess the likelihoods *both* that al-Assad will stay *and* that al-Assad will go. To give her an incentive to make an accurate assessment, one could say that if she predicted these outcomes with respective probabilities of 60 percent and 40 percent, then she would earn 0.6 ‘points’ if al-Assad stayed and 0.4 ‘points’ if al-Assad left. But then her expected score would fall to $0.6 \times 0.6 + 0.4 \times 0.4 = 0.52$. This is less than the expected score of 0.6 that the analyst would get by simply stating that al-Assad will stay in power. A self-interested analyst aiming to maximize her expected payoff would place all the probability on the outcome she believes to be most likely.⁵

Thus the batting-average metaphor, despite appearing to promote the careful study of uncertainty, actually provides incentives to make all-or-nothing predictions. This dissuades analysts from studying possibilities that are unlikely but important, and it is thus an example of consequence neglect. Few people would have predicted in the summer of 2010 that there was a substantial risk of regime-threatening uprisings throughout the Middle East. An analyst concerned with her batting average would have had little reason to study these possibilities, and this might have caused her to miss important warning signs of instability.⁶

This article demonstrates that the batting-average metaphor is not alone in superficially embracing the assessment of uncertainty while following a logic that is actually very different.⁷

⁵ This is not to say that the analyst stakes her entire reputation on each estimate, but rather to point out that the strategy which maximizes the analyst’s payoff across many predictions is to make single-outcome judgments on each one.

⁶ For a related example regarding the fall of the Shah of Iran, see Robert Jervis, *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War* (Ithaca, NY: Cornell 2010), ch.2. On consequence neglect more generally, see Alan Berger et al., ‘The Five Neglects: Risks Gone Amiss’ in Howard Kunreuther and Michael Useem (eds.), *Learning from Catastrophes* (Philadelphia, PA: Wharton 2010), pp.83-99.

⁷ Similarly, Mark Lowenthal’s book, *Intelligence*, contains thoughtful analyses of many issues raised in this article. But he also writes (p.148) that the ‘accuracy’ of a prediction should be judged according to whether its likelihood is ‘something more than 50 percent and something less than 100 percent’ – even though it can, of course, be perfectly accurate to say that something has a probability lower than 50 percent. Ford’s *Estimative Intelligence* reinforces many of this article’s themes, while at the same time stating that ‘one of the main purposes of national intelligence estimating is to lessen policymakers’ uncertainties about the world’(p.179). Richard K. Betts’s article, ‘Analysis, War, and Decision: Why Intelligence Failures Are Inevitable,’ *World Politics* 31/1 (1978), pp.61-89 is one of the most well-cited treatments of the dangers of making estimates with too much certainty, yet he does imply that certainty should be the goal: ‘It is the role of intelligence to extract certainty from uncertainty’(p.69). Kristan Wheaton’s article, ‘The Revolution Begins on Page Five: The Changing Nature of NIEs’, *International Journal of Intelligence and CounterIntelligence* 25/2 (2012), pp.330-349 discusses the challenges of conveying uncertainty in an accurate fashion, but states (p.331) that the purpose of National Intelligence Estimates is ultimately ‘to reduce national security policymakers’ level of uncertainty’.

The following sections highlight similar tensions in other aspects of intelligence theory and tradecraft.

To provide empirical support for its discussion, this article examines a broad range of National Intelligence Estimates (NIEs). Though NIEs comprise only a small fraction of overall estimative intelligence, their production is so highly scrutinized that it is reasonable to assume that their flaws would characterize lower-profile estimates as well.⁸ In addition to examining more than a dozen specific estimates, the following sections describe general patterns across a database of 379 declassified NIEs that were written between 1964 and 1994 and that were released through the Central Intelligence Agency's Historical Review Program.⁹ Throughout the following sections, this combination of deductive and inductive analysis helps to draw out the tensions between eliminating uncertainty and assessing uncertainty in estimative intelligence.

Analyzing Alternatives

Intelligence analysts often wrestle with alternatives. To return to the example of an analyst studying Syria in 2012, this analyst would have had to consider a wide range of possibilities. If al-Assad were to survive the year, his domestic and international standing could presumably change in any number of ways. If al-Assad left power, the transition might be stable but it might also descend into widespread violence, while al-Assad's eventual successors could be relatively friendly or hostile to the United States. There are many relevant scenarios here. An intelligence analyst who seeks to eliminate uncertainty would argue that one of them constitutes the 'correct answer', and that her ideal goal should be to identify what that answer is. Intelligence tradecraft often encourages this kind of thinking when it comes to analyzing alternative possibilities.

When estimates focus on a single possibility, it is called 'single-outcome forecasting'. This practice has been criticized for some time, because analysts often have insufficient information for making definitive predictions, and because policymakers should be aware of a range of potential contingencies.¹⁰ Yet even when analysts do consider multiple possibilities, several elements of standard practice reveal a tendency to assume that one of those scenarios should be considered the most important or the most correct.

⁸ As one intelligence scholar recently stated, 'The NIE is arguably the highest form of the intelligence art'. Wheaton, 'The Revolution Begins on Page Five', p.340. A recent discussion of NIEs can be found in Loch K. Johnson, *The Threat on the Horizon: An Inside Account of America's Search for National Security after the Cold War* (NY: Oxford 2011), pp.164-85.

⁹ There are actually 426 entries in the database, but 47 were dropped for various reasons: some entries are not estimates but update memoranda, some remain heavily classified, and some appear in the database twice. The database was accessed through <foia.cia.gov> between October 2010 and May 2011.

¹⁰ See, for example, Willis C. Armstrong et al., 'The Hazards of Single-Outcome Forecasting', *Studies in Intelligence* 28/3 (1984), pp.57-70.

For instance, most NIEs highlight a ‘Best Estimate’ or ‘Key Judgments’. In principle, calling some estimates ‘Best’ does not exclude the idea that there are other possibilities. Meanwhile, the Key Judgments section of an NIE is generally intended to serve the function of an executive summary, so it does not inherently privilege one alternative over another. In practice, however, these sections often highlight a subset of relevant possibilities and encourage consumers to give these possibilities special attention. NIEs typically present their judgments in sequence, often with one or two possibilities receiving the bulk of explanation and support. Many NIEs contain a distinct section enumerating ‘Alternatives’ that often receive relatively limited discussion.

This treatment can suggest that certain alternative views are relatively insignificant, and the 2002 NIE on *Iraq’s Continuing Programs for Weapons of Mass Destruction* serves as a prominent example. The Key Judgments section of this NIE begins with 42 paragraphs supporting the assessment that Iraq ‘has continued its weapons of mass destruction (WMD) programs’ and that ‘if left unchecked, it probably will have a nuclear weapon during this decade’. That conclusion, as is now widely known, was based on controversial information. Doubts about the NIE’s main judgment were raised in a two-paragraph text box at the end of the opening section, arguing that the evidence does not ‘add up to a compelling case that Iraq is currently pursuing... an integrated and comprehensive approach to acquire nuclear weapons’. It is almost impossible to miss this objection – but it is equally difficult to miss the disparity in emphasis between these very different points of view about Iraq’s nuclear program.

The tendency to privilege particular judgments goes beyond the structure of intelligence estimates. Ironically, it permeates many of the conceptual frameworks that are intended to encourage analysts to consider multiple possibilities in the first place. For instance, one prominent text on analytic tradecraft recommends that analysts approach complex questions by generating multiple hypotheses, evaluating the ‘credibility’ of each hypothesis, sorting hypotheses ‘from most credible to least credible’, and then ‘select[ing] from the top of the list those hypotheses most deserving of attention’.¹¹ Though its authors intend for this method of ‘multiple hypothesis generation’ to ensure that important alternatives do not get overlooked, the instruction to focus on the ‘most credible’ predictions indicates an assumption that unlikely possibilities are less ‘deserving of attention’. Yet that is often untrue. The most consequential events (such as major terrorist attacks, the outbreak of conventional wars, and the collapse of state governments) are often perceived as unlikely before they occur, yet they can have such enormous impact that they deserve serious consideration. The overall significance of any event is a product of its probability *and* its consequences, and so both of these factors must be considered when comparing alternative scenarios.

To give another example, many intelligence analysts are trained in a practice called Analysis of Competing Hypotheses (ACH). ACH seems to embrace the goal of assessing uncertainty. It

¹¹ Richards J. Heuer and Randolph H. Pherson, *Structured Analytic Techniques for Intelligence Analysis* (Washington, DC: CQ Press 2011), ch.7.1.

instructs analysts to form a matrix of potential hypotheses and available information that helps to show how much evidence supports (or contradicts) each possibility. This practice combats first impressions and promotes alternative thinking.¹² Yet the word ‘Competing’ is important here. Competing for what? The original description of ACH explains its goal as being to determine ‘Which of several possible explanations is the correct one? Which of several possible outcomes is the most likely one?’¹³ A recent manual introduces ACH as a tool for ‘selecting the hypothesis that best fits the evidence’.¹⁴ To this end, ACH instructs analysts to ‘focus on disproving hypotheses’.¹⁵ This does not mean that ACH always generates single-outcome estimates, and the method is designed to indicate places where the evidence sustains multiple interpretations. When this occurs, ACH tells analysts to rank remaining possibilities from ‘weakest’ to ‘strongest’.

An analyst seeking to assess uncertainty would approach the issue differently. She would not see the relevant possibilities as rival or competing. She would say that no possibility merits attention for being ‘correct’ and that focusing on disproving hypotheses places unnecessary emphasis on eliminating relevant scenarios from consideration. Moreover, she would argue that it makes little sense to say that any possibility is ‘weak’ or ‘strong’ so long as analysts accurately assess its likelihood.¹⁶ This is not to claim that all scenarios have equal significance. But to repeat, the significance of any possibility is the product of its likelihood *and* its potential consequences. For that reason, ACH’s method of ranking hypotheses based on probability exposes analysts to consequence neglect.

Other prominent analytic methods introduce similar tensions. In some cases, analysts are instructed to choose and flesh out their best estimate, with critiques and alternatives raised later as the estimate receives formal review. In other cases, the process of considering alternatives is deliberately adversarial or contrarian from the start, relying on ‘devil’s advocates’ or dividing analysts into ‘Team A/Team B’ groupings, so as to foster a clash of viewpoints.¹⁷

These methods help to ensure that diverse perspectives are considered during the analytic process, and this is a valuable goal. Nevertheless, these methods still imply that hypotheses are rivals, and that they must compete with each other for pride of place. And the standards for

¹² U.S. Government, *A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis* (Washington, DC: U.S. Government Printing Office 2009), p.14.

¹³ Heuer, *The Psychology of Intelligence Analysis* (Washington, DC: Center for the Study of Intelligence 1999), p.95.

¹⁴ Heuer and Pherson, *Structured Analytic Techniques*, p.160.

¹⁵ U.S. Government, *Tradecraft Primer*, p.15.

¹⁶ By way of analogy, if the best economic models predict a ten percent chance of a recession in a given year, then this does not constitute a ‘weak possibility’ or a scenario that should be disproved. The important thing for economic forecasters (both in the private and the public sectors) is to assess the chances of recession accurately. The same is true for doctors assessing potential complications during surgery, meteorologists predicting the chances of inclement weather, and experts in many other fields.

¹⁷ More broadly, an increasing amount of recent intelligence work has been devoted to techniques of ‘competitive analysis,’ which are designed to pit different viewpoints against one another. See Richard L. Russell, ‘Competitive Analysis’ in Johnson (ed.), *Oxford Handbook of National Security Intelligence*.

judging these competitions – for determining which judgments are truly the ‘best’ – are often subjective and contentious. For instance, the NIE, *Likelihood of an Attempted Shoot-Down of a U-2* (1964) predicts a ‘significant and, over time, growing threat’ of a spy plane coming under fire. The NIE, *South Africa: Weathering the Storm* (1992) argues that ‘The recent surge in factional violence and the African National Congress (ANC) suspension of talks have dealt the negotiation process a serious – but we believe not fatal – blow’. But what constitutes a ‘significant threat’ or a ‘fatal blow’? Reasonable people can disagree on these issues, and a wide range of intelligence literature addresses similar questions of how analysts should know when to ‘sound the alarm’ about potential threats, and when they should refrain from ‘crying wolf’. As one scholar writes, these debates are ‘in effect theological disputes’ and they can lead to a great deal of friction among analysts and the agencies they represent.¹⁸

Yet these disputes are only relevant to the extent that analysts believe that they need to judge whether or not some threat is ‘significant’ enough to warrant policymakers’ attention. Consider instead a situation where analysts see no reason to highlight any particular prediction relative to the alternatives. If the analyst seeks simply to describe several relevant possibilities, along with their respective likelihoods and potential consequences, then that obviates the need to make value judgments about whether some threat is ‘significant’.

A good example of this kind of analysis is the NIE, *The Deepening Crisis in the USSR* (1990). A single figure from this document – which is reproduced at the end of this article as an appendix – lays out four different ‘scenarios for the next year’. Each scenario receives two or three explanatory bullet points along with a ‘Rough Probability’. The most likely scenario is presented first, but none is given more attention than the others. This NIE avoids both probability neglect and consequence neglect; it conveys no notion that one possibility deserves pride of place as being the best or most correct; it does not require analysts or readers to debate the meaning of concepts like ‘significant’, ‘serious’, or ‘important’; and it allows readers to decide for themselves which possibilities to prioritize. Moreover, even though the figure contains a wide range of information, it is still clear and concise.

Yet this kind of multi-faceted estimate is rare. Of the 379 declassified NIEs surveyed for this article, 200 (53 percent) examine a single possibility without explaining potential alternatives. Only 112 (30 percent) explicitly consider three or more possible judgments. The next section also demonstrates how the *Deepening Crisis* NIE is especially rare in the way it conveys the estimated likelihood of each potential scenario in a reasonably precise fashion.

A delineation of multiple possibilities with attached likelihoods is called a *probability distribution*. This is one of the cornerstone concepts of decision theory, a field that was

¹⁸ Betts, *Enemies of Intelligence: Knowledge and Power in American National Security* (NY: Columbia 2007), p.101.

developed to help individuals make effective choices under uncertainty.¹⁹ The goal of using probability distributions (such as the example in this article's appendix) is to state clearly the likelihoods and potential consequences that are associated with relevant scenarios. This provides information that is important to managing uncertainty in many professions.²⁰

Using probability distributions removes the inclination to believe that alternative views are necessarily at odds with each other. A probability distribution is not a collection of competing hypotheses: it is a *single* hypothesis about the way uncertainty is spread across multiple possibilities. The presence of multiple possibilities does not indicate disagreement, dissent, or confusion. None of these possibilities is either right or wrong. The true best estimate is the one that accurately describes the distribution of possibilities and their likelihoods, the equivalent of well-calibrated odds in a horse race.

This perspective is the bedrock of attempting to assess uncertainty, and it avoids many pitfalls in existing tradecraft. For instance, when analysts attempt to 'make the call' about which of several possibilities is the best or most important, they often subconsciously view subsequent evidence in ways that support this preconceptions.²¹ But confirmation bias should only be an issue when analysts are under pressure to identify and support a subset of possibilities. If the analyst were to view the entire probability distribution as a single, coherent hypothesis, then there would be no reason to confirm one part of it to the exclusion of the rest. Similarly, the goal of assessing uncertainty reduces the need for analysts to serve as advocates. When alternative possibilities need not compete, the risk that certain points of view will be marginalized is reduced. So long as analysts properly assess the likelihood of each possibility, there is no reason to think that one is any more valid than others. In short, the *likelihood* of a hypothesis is something different from the *validity* of a hypothesis, but intelligence tradecraft often conflates these attributes.

¹⁹ A fundamental introductory text on this subject is Howard Raiffa's *Decision Analysis: Introductory Lectures on Choices under Uncertainty* (Reading, MA: Addison-Wesley 1968). A more recent text is John W. Pratt, Raiffa, and Robert Schlaifer, *Introduction to Statistical Decision Theory* (Cambridge, MA: MIT 1995). See also Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases* (NY: Cambridge University Press 1982) and Kahneman, *Thinking Fast and Slow* (NY: Farrar, Straus, and Giroux 2011) which examines behavioral aspects of decision making under uncertainty and discuss ways to mitigate relevant problems.

²⁰ The decision theory literature often deals with subjects such as doctors prescribing medical treatments, business executives making economic forecasts, and gamblers assessing their prospects. In these cases and many others, decision makers face uncertainties about the current and future states of the world that will influence their choice of action. Sherman Kent explains how intelligence analysis and foreign policy relate to other kinds of decision making under uncertainty in *Strategic Intelligence for American World Policy* (Hamden, CT: Archon 1965), pp.58-61. See also Walter Laqueur, 'The Question of Judgment: Intelligence and Medicine', *Journal of Contemporary History* 18/4 (1983), pp.533-548; Charles Weiss, 'Communicating Uncertainty in Intelligence and Other Professions', *International Journal of Intelligence and CounterIntelligence* 2/11 (2008), pp.57-85; and David T. Moore, *Sensemaking: A Structure for an Intelligence Revolution* (Washington, DC: National Defense Intelligence College 2011), pp.95-99.

²¹ Heuer, *Psychology of Intelligence Analysis*, ch.10.

Critics of the probability distribution approach might argue that it induces too much complexity by having analysts identify a larger range of possibilities, but this is not true. The probability distribution simply imposes structure on what many analysts do anyway. For instance, the NIE, *Cuba's Changing International Role* (1975) predicts that 'there is a better-than-even chance that a partial reduction in the scope of US sanctions would be enough to lead Castro to engage in substantive negotiations'. This indicates that there are alternative ways Castro might respond to reduced sanctions. The combined likelihood of these alternatives might even approach 50 percent. These possibilities comprise what decision theorists sometimes call a *catch-all hypothesis*, covering any contingency not otherwise mentioned. Does it make matters more complex to explain what the catch-all hypothesis entails? On the contrary, if analysts explain these alternatives – even in a concise format like the figure in this article's appendix – it would help to clarify what these analysts are saying already.

Moving from judgments to distributions brings an additional benefit: it reduces pressures to 'water down' estimates in order to achieve consensus. Intelligence analysts often complain about this pressure, which is amplified by the notion that a subset of the relevant possibilities should be emphasized or given pride of place. This creates a tendency to broaden estimative judgments in order to accommodate different views.²² Because techniques such as ACH revolve around settling on some answers to the relative exclusion of others, analysts are explicitly instructed to ask questions like 'Does it make sense to combine two hypotheses into one?'²³ Framing estimates in terms of probability distributions dampens the incentive to do this. If different possibilities need not compete, and if no possibility receives a special imprimatur, then there is no reason to hedge, merge, or exclude particular views.

This is another way in which the probability distribution not only *provides* more information than standard approaches, but it also lessens the difficulty of *producing* and *presenting* that information, by removing prominent grounds for disagreement. And while analysts will no doubt disagree about the probability that should be attached to each of several possibilities, this challenge is already a prominent element of intelligence analysis, and that is the subject of the next section.

²² Ford, in *Estimative Intelligence* (p.21 cf. pp.78, 101) writes that NIEs are often criticized for being 'wishy-washy', because the coordination process tends to produce 'coordinated mush'. His rough dichotomy between mush and split decisions, however, is a dichotomy that is not necessary when thinking in terms of probability distributions, rather than of discrete, single-outcome judgments. Cf. Gregory Treverton, *Reshaping National Intelligence for an Age of Information* (Santa Monica, CA: RAND 2001), p.204; Roger Z. George, 'Beyond Analytic Tradecraft', *International Journal of Intelligence and CounterIntelligence*, 23/2 (2010), pp.300-301; Betts, *Enemies of Intelligence*, pp.32, 101.

²³ Heuer and Pherson, *Structured Analytic Techniques*, p.163.

Likelihood and confidence

Intelligence estimates often use the terms likelihood and confidence. These concepts differ, and it is important to keep them separate. As the 2007 NIE, *Iran: Nuclear Intentions and Capabilities* explains, ‘estimates of likelihood’ constitute ‘probabilistic language’ framing an analyst’s judgments. But since analysts typically estimate likelihood based on evidence that is both incomplete and ambiguous, it is useful to assess how reliable those estimates may be. For that purpose, the *Iran* NIE defines a ‘high confidence’ assessment as one that is based on ‘high quality’ information; a ‘moderate confidence’ assessment ‘generally means that the information is credibly sourced and plausible but not of sufficient quality or corroborated sufficiently to warrant a higher level of confidence’; and a ‘low confidence’ assessment relies on evidence that is ‘too fragmented or poorly corroborated to make solid analytic inferences’. In short, likelihood describes the probability that analysts assign to some judgment. Confidence is then a way of qualifying that statement by describing the ‘scope, quality, and sourcing’ that supports it.

This is a reasonable classification, but it is not the way that the concepts of likelihood and confidence are actually used in the *Iran* NIE. For example, here is the opening paragraph of the NIE’s Key Judgments:

We judge with high confidence that in fall 2003, Tehran halted its nuclear weapons program; we also assess with moderate-to-high confidence that Tehran at a minimum is keeping open the option to develop nuclear weapons. We judge with high confidence that the halt, and Tehran’s announcement of its decision to suspend its declared uranium enrichment program and sign an Additional Protocol to its Nuclear Non-Proliferation Treaty Safeguards Agreement, was directed primarily in response to increasing international scrutiny and pressure resulting from exposure of Iran’s previously undeclared nuclear work.

Throughout this paragraph, confidence is not being used to *qualify* expressions of likelihood. Rather, it is being used to *make* expressions of likelihood. In all, the Key Judgments section of the *Iran* NIE uses the term ‘confidence’ 19 times in order to convey the probability that some statement is true.²⁴ The Key Judgments express likelihoods in several places, too.²⁵ But there is no instance where the NIE conveys both the probability *and* the confidence that the authors

²⁴ For example, ‘We assess with high confidence that until fall 2003, Iranian military entities were working under government direction to develop nuclear weapons’; ‘We judge with high confidence that the halt lasted at least several years’; ‘We continue to assess with moderate-to-high confidence that Iran does not currently have a nuclear weapon’.

²⁵ It seems as though the real operative distinction between words of likelihood and words of confidence in the *Iran* NIE is that the latter are used to make predictions about events that had already happened (such as whether or not Iran had stopped and/or restarted its nuclear research) while the former is used to make predictions about the future (e.g., ‘Iran probably would use covert facilities’ for a given purpose; ‘Iran probably would be technically capable’ of producing a weapon in a given time frame).

assign to some prediction, even though the NIE's front matter explains why those concepts convey different, important ideas.

This tendency to conflate likelihood and confidence follows logically from the perspective that estimative questions have right answers, and that an analyst should seek to define those answers with minimal uncertainty. With this goal in mind, it is reasonable to think that likelihood and confidence converge. To the extent that an analyst believes a particular estimate is likely to be correct, this means she has a relatively large amount of relatively sound evidence to support that assessment. To the extent that the analyst attaches a *low* likelihood to an assessment, this indicates that there is relatively little or relatively unreliable information to support it. If an analyst is striving to achieve certainty, then it makes sense to use the language of confidence when talking about the concept of likelihood.

An analyst who aims to assess uncertainty would have a very different perspective. She would argue that estimative questions rarely have single, right answers, and even if they do (the status of Iran's current nuclear program is in principle a knowable fact), the available evidence is often ambiguous enough to sustain multiple interpretations.²⁶ Either way, likelihood and confidence are different concepts. If an analyst aims to assess uncertainty, then it is perfectly logical to be highly confident that some outcome should be viewed in probabilistic terms.

Consider some hypothetical cases. An analyst should be highly confident when stating that the odds of a coin flip coming up heads are one in two, and that the odds of drawing a face card from a shuffled deck are three in thirteen.²⁷ The chances of a U.S. bomber striking its target from high altitude are much less precise, but they can also be estimated in rigorous ways that analysts can discuss with relatively high levels of confidence. And while any predictions about political turmoil will be less reliable still, there are some aspects of the situation that analysts can judge more confidently than others. For example, analysts in the spring of 2012 could presumably state with confidence that the chances were low that Bashar al-Assad would finish the year in power *and* restore the international standing he lost during the crisis. Analysts would presumably have had less confidence in any predictions that they made about the likelihood that al-Assad will be ousted. They would probably have been even less confident in their predictions about the odds that any given candidate would be al-Assad's successor. That does not mean that analysts should shy away from making these kinds of assessments. It only means that analysts need to convey *both* the probability *and* the confidence associates with their estimates.

Melding confidence and likelihood not only leaves intelligence analysis incomplete, but it also clouds what that analysis means. When the *Iran* NIE assessed 'with moderate confidence' that 'Tehran had not restarted its nuclear weapons program as of mid-2007', did this reflect that the

²⁶ None of this discussion implies that intelligence *collectors* should not seek to eliminate uncertainty. But when *analysts* confront uncertainty, they must assess their information in light of its limitations.

²⁷ High confidence implies the estimate will change little in response to new information, such as close inspection of the coin. This is an important point to which the article returns below.

odds that Iran was currently pursuing nuclear weapons more like 20 percent or 60 percent? And was the term ‘confidence’ being used here to represent inferences about likelihood, or to qualify the reliability of those inferences?

These kinds of concepts and terms have received much debate within the intelligence community in recent years, and the NIEs surveyed for this article provide some historical context.²⁸ Of the 379 declassified NIEs surveyed for this article, only 16 (four percent) discuss probability using quantitative indicators of any kind. This category was coded broadly to include not only percentages, but also bettors’ odds such as ‘1 in 5’, as well as statements like ‘close to even’.²⁹ 70 NIEs (18 percent) discuss a range of potential outcomes but do not give even a *qualitative* sense of their probabilities: saying, for instance, that some outcome is the ‘most likely’ but not conveying what its likelihood actually is.³⁰ The *Deepening Crisis* NIE discussed in the previous section is the only one of the 379 NIEs surveyed for this article that discussed more than two potential outcomes *and* that described their probabilities numerically.³¹

Expressing a quantitative phenomenon in qualitative terms – or not expressing it at all – is unavoidably vague. And in many respects, such vagueness is unnecessary. For example, the *Iran* NIE defines seven different terms for expressing likelihood, and arranges them on a spectrum. The word ‘remote’ is at the extreme left of this spectrum, such that it is reasonable to think that the term should only apply to events whose likelihood of occurring is less than 10 percent. The word ‘very likely’ is placed on the right end of the spectrum, at a position indicating a likelihood of perhaps 75 percent. In order to know which term to use, analysts need to estimate probabilities, so that they go to the right place on the spectrum and pick the word that appears most appropriate. Why not simply report those probability estimates in the first place?³²

²⁸ See Wheaton, ‘The Revolution Begins on Page Five’ for a recent and thorough discussion of this debate, which also covers the 2007 Iran NIE.

²⁹ For example, *Prospects for the South African Transition* (1994) estimates a 70 percent chance that an election will occur on schedule; *Russia Over the Next Four Years* (1992) traces several contingencies and identifies the chances of two of them as ‘slightly better than even’ and ‘one in three’; and *Soviet Ballistic Missile Defense* (1982) assesses the odds that the Soviets will abrogate the Anti-Ballistic Missile Treaty at 10-20 percent.

³⁰ For instance, several estimates discuss different ‘illustrative force models’ for how Soviet strategic forces might evolve without giving a sense of the likelihood that each model is correct. Examples of NIEs that lay out more than three possible scenarios while providing the reader with almost no sense of their likelihoods include *Soviet Policy toward the West* (1989); *The Changing Sino-Soviet Relationship* (1985); and *Soviet Military Options in the Middle East* (1975).

³¹ This is not to say that other estimates do not share some of its characteristics. For instance the NIE *Implications of Alternative Soviet Futures* (1991) presents a range of potential scenarios in a concise table, though without explicit probabilities. The NIE *Russia over the Next Four Years* (1992) gives numeric probabilities over three or more possible outcomes but does not present the information in a concise table.

³² Sherman Kent wrote one of the earliest and best-known articles on this subject: ‘Words of Estimative Probability’, *Studies in Intelligence* 8/4 (1964); more recent examples include Weiss, ‘Communicating

The most common explanation for why analysts smudge probabilities in this way is that they wish to avoid giving their estimates an undue appearance of precision. But this is exactly why it is important to convey both likelihood *and confidence* when making a prediction. For instance, the *Iran NIE* discusses whether Tehran might enrich enough uranium to make a nuclear weapon by 2013. The estimate states that ‘Iran is unlikely to achieve this capability... because of foreseeable technical and programmatic problems’. Here are four assessments that are consistent with that statement:

There is roughly a 10 percent chance that Iran will achieve this capability, though we have low confidence in this estimate since it is based on speculation about Iran’s technical and programmatic potential.

There is roughly a 40 percent chance that Iran will achieve this capability, though we have low confidence in this estimate since it is based on speculation about Iran’s technical and programmatic potential.

There is roughly a 10 percent chance that Iran will achieve this capability, and we have high confidence in this estimate because it is based on reliable information about Iran’s technical and programmatic potential.

There is roughly a 40 percent chance that Iran will achieve this capability, and we have high confidence in this estimate because it is based on reliable information about Iran’s technical and programmatic potential.

None of these statements is particularly precise.³³ But each is unambiguous when it comes to predicting likelihoods and saying how reliable those predictions may be. These statements might have very different implications for policy, and the relevant distinctions are essentially lost in conventional estimative language.

Moreover, there is an important class of situations in which little is lost – and much is gained – by expressing likelihood more concretely. When analysts deal with small probabilities, it becomes especially important to give a clear sense of what they are. For instance, the risk that terrorists will capture nuclear weapons in Pakistan within the next year is presumably fairly ‘low’. But because that event could have such enormous consequences, policymakers need to know just how ‘low’ that probability is. Stating that this probability is ‘remote’ – or even stating that it is ‘less than five percent’ – allows for interpretations that range over multiple orders of magnitude. The odds could be one in a hundred, one in a thousand, or one in a million. Standard

Uncertainty’ and Joab Rosenberg, ‘The Interpretation of Probability in Intelligence Estimation and Strategic Assessment’, *Intelligence and National Security* 23/2 (2008), pp.139-152.

³³ A reasonable compromise between clearly expressing likelihood and not giving an undue sense of scientific precision might be for analysts to assess most probabilities in intervals of five percentage points (e.g., 20 percent or 85 percent), while expressing small probabilities in tighter intervals (e.g., 1 percent or 2 percent).

intelligence terminology cannot distinguish among these possibilities, even though the differences among them can be critical.

To illustrate, NIEs throughout the Cold War regularly discuss the risk of the Soviet Union launching a conventional or nuclear attack on the United States or its allies. In all cases, the likelihood of major conflict is predicted to be low. To give some examples, the NIE *Soviet Forces and Capabilities for Strategic Nuclear Conflict* (1987) states that ‘the Soviets have strong incentives to avoid risking nuclear war’. *Implications of Recent Soviet Military-Political Activities* (1984) tells readers that despite rhetoric from the Kremlin to the contrary, ‘we are confident that, as of now, the Soviets see not an imminent military clash’. *Soviet Capabilities for Strategic Nuclear Conflict* (1983) states ‘The Soviets, in our judgment, are unlikely to initiate nuclear conflict on a limited scale’. *Warsaw Pact Concepts and Capabilities for Going to War in Europe* (1978) judges it to be ‘highly unlikely that the Warsaw Pact nations, or the Soviets alone, would deliberately decide to attack member countries of the North Atlantic Treaty Organization’.

Just as policymakers today understand that the probability of a terrorist group capturing a Pakistani nuclear weapon is low, most people in the policy community throughout the Cold War understood that the risk of a Soviet attack on the United States or its allies was small. But then as today, the operative question was *how* small those chances really were. Assuming that the risk of Soviet attack in a given year was one percent would have made for very different policy than assuming that it was one hundredth of one percent, yet Cold War NIEs provided little guidance on this matter.

When an estimate simply says some event is ‘unlikely’, it is difficult for readers to weight the prediction properly. Some threats are not worth worrying about if they are too unlikely, yet policymakers may overreact if their consequences are especially dangerous or vivid. Conversely, policymakers often have trouble thinking about what small probabilities mean, and sometimes effectively treat them as if they were zero. These are both important examples of probability neglect. There is a large scholarly literature on how to mitigate these problems by defining and presenting probabilities in rigorous, practical, and easily interpretable ways.³⁴ But mathematics is not necessary to understand that it is difficult to make wise judgments when policymakers lack critical information. Presenting likelihoods vaguely – or not presenting them at all – creates this very problem.³⁵

³⁴ For literature reviews on this subject, see Paul H. Garthwaite, Joseph B. Kadane, and Anthony O’Hagan, ‘Statistical Methods for Eliciting Probability Distributions’, *Journal of the American Statistical Association* 100/470 (2005), pp.680-700, and Frederick Mosteller and Cleo Youtz, ‘Quantifying Probabilistic Expressions’, *Statistical Science* 5/1 (1990), pp.2-34.

³⁵ Another set of cases that is relevant to this argument is a series of NIEs written about potential security concerns accompanying presidential visits to various countries: e.g., *Security Conditions in China* (1972), *Security Conditions in Mexico* (1970); *Security Conditions in Mexico City* (1968); and *The President’s Trip to Central America* (1968). Each of these estimates the threat to the president’s safety to be low – but

The 2002 NIE, *Iraq's Continuing Programs for Weapons of Mass Destruction* provides a prominent illustration. As previously mentioned, the NIE stated, 'We judge that Iraq has continued its weapons of mass destruction (WMD) programs'. There was substantial debate in the intelligence community about the extent to which the available evidence supported this claim. Yet this assessment contains no information about likelihood. By leaving likelihood vague, the authors made it easier for readers to focus on potential consequences rather than expected consequences, and clarifying this distinction was the most important function this estimate could have served. Policymakers at the time were already concerned with and well aware of the potential consequences of Iraq's pursuit of WMD. Far less clear was the *likelihood* that Iraq was then pursuing nuclear weapons, and the odds that Iraq would actually obtain them. Knowing analysts' assessments of the matter would reveal the extent to which the potential and expected consequences of this outcome might have diverged.

Making this kind of determination is bound to be difficult and contentious. But the intelligence community is better positioned than policymakers to make an informed and objective determination on this score. By effectively declining to distinguish between potential and expected consequences, the 2002 NIE failed to steer its consumers away from probability neglect, or to prevent them from interpreting the report in ways that were consistent with their initial preconceptions. This type of misinterpretation can be mitigated by explicitly assessing uncertainty.

A final reason why it is important to separate likelihood and confidence – and to be explicit about each – is that this provides information about how those predictions might change if new information emerged. In making predictions such as whether Bashar al-Assad would be ousted from the presidency of Syria in 2012, it is possible that analysts might be presented with substantial evidence suggesting al-Assad would stay, alongside substantial evidence that he would go, with the evidence on each side being reasonably compelling. If analysts are dealing with a large amount of high-quality information, then they might state that the evidence is ambiguous but extensive. They might predict with relatively high confidence that the odds of al-Assad being ousted are roughly even, and these estimates would shift little as a result of gathering new, small pieces of information. But if analysts only have a few slivers of information to work with, then while they might still say that al-Assad has an even chance of being ousted, they should also report that these predictions entail relatively low confidence. When analysts report low confidence, this implies that gathering more evidence may dramatically alter their perceptions. Policymakers might then be inclined to refrain from acting on that assessment until more intelligence could be collected, and predictions could be refined.

This kind of scenario is common, as policymakers often make important decisions about whether to act or to wait for more information. Clearly expressing both likelihood and

this is exactly the kind of issue for which consumers of intelligence rely on having a more fine-grained sense of what the phrase 'low probability' actually means.

confidence may not eliminate the difficulty of making these kinds of decisions, but it can improve the process by giving policymakers a better understanding of what estimates mean and how reliable they are. The next section goes into more detail about the reliability of intelligence, and how this affects the way that information is filtered through the intelligence cycle.

Filtering information

Estimative intelligence depends on information that is filtered through myriad layers. NIEs, for instance, draw on reports from different intelligence agencies, and only a tiny fraction of information that these agencies collect is actually considered in drafting an estimative product. The way information is evaluated and filtered thus provides the foundation for producing quality estimates.

Consider the flow of human intelligence, from the point where it is collected by a case officer to the point where that information gets included in a published NIE. This process has several stages: a case officer typically receives the information and decides whether to report it, agency reviewers receive the information and decide what to pass along to analysts, analysts decide whether to use the information in their reports, reports are vetted by colleagues and superiors, and some of those reports are then considered in drafting an NIE. At each of these decision nodes throughout the analytic chain, some information gets passed along but much information gets pared out. Some useful information will inevitably be discarded and some misleading or irrelevant information will presumably get through. The question is not whether this filtering process is perfect, but whether it systematically favors some kinds of information over others in a problematic fashion. Such favoritism could be labeled *biased attrition*.

One of the most frequently-criticized forms of biased attrition is the way the intelligence cycle tends to prioritize tactical information that can be identified with precision, and to exclude political or strategic information that is harder to know with certainty, but which is often more important for making major policy decisions.³⁶ This ‘fetish for precision over relevance’ is a common concern in the intelligence literature.³⁷ It is also a pattern to be expected when the intelligence community prioritizes eliminating uncertainty. If you believe that *estimates* should ideally be as close to certain as possible, then it makes sense to use *information* that is as close to certain as possible. Yet an analyst aiming to assess uncertainty would approach the issue far

³⁶ This has been a long-standing critique: see, for example, Senate Committee on Armed Services, Preparedness Investigating Subcommittee [Stennis Report], *Investigation of the Preparedness Program*, 88th Cong., 1st Sess. (1963), pp.5, 10. For a more recent example, see MG Michael Flynn, *Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan* (Washington, DC: Center for a New American Security 2010), p.7.

³⁷ Jennifer E. Sims, ‘A Theory of Intelligence and International Politics,’ in Treverton and Wilhelm Agrell (eds.), *National Intelligence Systems: Current Research and Future Prospects* (NY: Cambridge 2009), p.81.

differently. If the analyst wishes to make her estimates as *accurate* as possible, then perfect information is no longer necessary.

To amplify this point, imagine that a case officer receives two reports, from sources perceived to be equally unreliable. The first report is that the government of Iran is roughly three years from building a nuclear bomb. The second report is that the government of *Saudi Arabia* is roughly three years away from building a nuclear bomb. An intuitive reaction may be to discard the second report on the grounds that policymakers and intelligence analysts generally believe that Saudi Arabia does not currently have (or wish to have) a nuclear weapons program. But all else being equal, the surprising nature of the second report makes it *more* valuable, not less, because it could shift analysts' and policymakers' views more significantly. Even if the information comes from a questionable source, it still might be important to follow up on it because of its potential impact on the way analysts and policymakers think about security and stability in the Middle East. By contrast, the U.S. intelligence community is fully aware of Iran's nuclear program. Having watched the issue closely for years, a large body of information is already available for estimating Tehran's progress toward a functioning nuclear weapon, and properly incorporating an unreliable judgment about Iran's nuclear capabilities is unlikely to shift expectations significantly.

This example is hypothetical, but it helps to frame a problem with the way the intelligence community filters information on a daily basis. Most intelligence – and especially human intelligence – passes through the intelligence cycle accompanied by an explicit 'source assessment'. For example, the U.S. Army Field Manual on *Human Intelligence Collector Operations* instructs practitioners to assign an 'alphanumeric designator' to each piece of information. The letters (A through F) represent 'source reliability', and the numbers (1 through 6) represent 'information content'.³⁸ Since the importance of any piece of information depends on both its reliability and its content, this seems to provide an adequate foundation for judging whether a piece of intelligence is significant.

In practice, however, 'source reliability' and 'information content' are really just two different ways of assessing the 'probable accuracy of the information reported'.³⁹ Source reliability is defined by an informant's personal characteristics, such as whether they are 'trustworthy', 'authentic', and 'competent'. Information content is then defined by whether the source's report has been 'confirmed by other independent sources', whether that information is 'logical in itself', and whether it is 'consistent with other information on the subject'.⁴⁰ These are each important

³⁸ U.S. Army Field Manual 2-22.3, *Human Intelligence Collector Operations* (Washington, DC: Department of the Army 2006), par. 12-12 and appendix B.

³⁹ *Ibid.*, par. 12-13.

⁴⁰ *Ibid.*, par. B-2. As a further indication that 'information content' is a way of judging a report's accuracy, it is worth noting that the categories along this scale are defined as 'confirmed', 'probably true', 'possibly true', 'doubtfully true', 'improbable', and 'cannot be judged'.

characteristics of an intelligence report, but they neglect to provide information about how consequential the report may be.

In fact, this source assessment framework may actually *reduce* the chances that consequential information gets passed along. Note how the ‘information content’ of a report is assumed to decrease if it has not been confirmed by other sources, or if it is illogical, or if it is not ‘consistent with other information on the subject’. Even though intelligence that is inconsistent with prior beliefs can be the most important information to consider because it can have the greatest impact on overall assessments, analysts are explicitly instructed to downgrade surprising reports. The source assessment framework says that ‘the degree of confidence’ an analyst places on a given piece of information ‘decreases if the information is not confirmed’; it suggests that if an intelligence report is ‘contradicted by other information on the subject’ then it is ‘improbable’; and the Field Manual says that analysts should ‘[treat] that information with skepticism’.⁴¹ Yet it is a mistake to interpret inconsistencies between new information and prior beliefs solely (or even predominantly) through the lens of reliability.

The problem here is that the reliability of a given piece of evidence is just one component of its value. The overall importance of a given piece of information also depends (among other factors) on what it says in relation to estimates derived from the body of evidence that is already available. The more limited is that body of evidence, or the more that new information appears to contradict it, the greater the potential for that information to be important. Just as it is obviously problematic to focus on the message and neglect its reliability, it is problematic to focus on the reliability and give the message short shrift. This is an important form of consequence neglect that existing source assessment frameworks encourage.⁴²

The most straightforward way to modify this system would be to change the definition of ‘information content’ so that it captures the extent to which each piece of intelligence provides a new or original perspective on the collection requirement. The current category of ‘source reliability’ could be broadened to include information about the trustworthiness of the source along with whether that source’s report is logical and independently confirmed. This would not add undue complexity to the source assessment system, as collectors and analysts are already accustomed to using a two-part designation. But the recommendation here would help to ensure that this designation captures both the reliability of the information *and* the potential consequences of that information should it be true. Juggling these attributes requires making judgment calls when deciding what to pass along the analytic chain. But appropriate tradecraft

⁴¹ Ibid, Appendix B and par. 12-13.

⁴² This problem is compounded by the way case officers and analysts are trained to determine what information to pass along. As intelligence professionals make this decision, they are instructed to think in terms of ‘thresholds’ for what makes intelligence significant. It certainly makes sense to say that NIEs should represent the most *significant* information, but the threshold concept blurs the line between significance and reliability.

can at least ensure that these judgments are based on the relevant inputs, and that none of these inputs is systematically neglected.

To conclude this discussion, it is worth returning to the Japanese attack at Pearl Harbor. U.S. intelligence had encountered an explicit warning that Japan planned to attack Hawaii in 1941. The report came from Joseph Grew, the American ambassador in Tokyo. Grew had received the information from his Peruvian counterpart. He cabled the message to the State Department, which then forwarded it to Army and Navy intelligence. Upon further investigation, the Navy determined that the Peruvian ambassador had originally received the information from his chef. The source was deemed unreliable, and the report was ‘discarded and forgotten’.⁴³

There is no doubt, even in hindsight, that the U.S. intelligence community should not have grounded strategic warning on the basis of a report from a Peruvian chef. However, it is hardly clear that the right move was to ‘discard and forget’ the information. The report had a low degree of reliability, but its potential consequences were enormous. Had there been several independent reports bearing similar information then there would have been substantial grounds for taking the threat seriously. The point here is that it would have been essentially impossible to know that such reports existed if they were systematically filtered out of the process before they could be considered together. This problem is hardly unique to the events at Pearl Harbor. For instance, a U.S. Senate report on the Cuban Missile Crisis concluded that the intelligence community was slow to react to the placement of Soviet missiles in Cuba because it disregarded reports by Cuban exiles that were deemed to be insufficiently reliable. The Israeli military disregarded warnings of an Egyptian attack in 1973 because they were perceived to have a low probability of being true. The CIA and FBI did not follow up on a range of questionable leads surrounding the plotters of the 9/11 terrorist attacks. These examples are controversial, but they indicate how privileging certainty can hinder the flow of information that supports the estimative process.⁴⁴

Conclusion

This article has argued that the goal of estimative intelligence should be to assess uncertainty, and not to eliminate uncertainty. Certainty has intuitive appeal: consumers of intelligence naturally demand it, and producers of intelligence naturally wish to provide it. But this is precisely why it is important to recognize how striving for certainty can expose the intelligence process to numerous flaws, and why it is important to deal with those flaws in structured ways. Those flaws affect the way analysts compare alternative hypotheses, the way they express

⁴³ Ariel Levite, *Intelligence and Strategic Surprises* (NY: Columbia 1987), p.72, cf. Roberta Wohlstetter, *Pearl Harbor: Warning and Decision* (Stanford: 1964), pp.368, 386.

⁴⁴ See Stennis Report, pp.5, 10; Klaus Knorr, ‘Failures in National Intelligence Estimates: The Case of the Cuban Missiles’, *World Politics* 16/ 3 (1964), pp.455-467; Shlomo Nakdimon, *Low Probability* (Tel Aviv: Revivim 1982); Amy B. Zegart, *Spying Blind: The CIA, the FBI, and the Origins of 9/11* (Stanford: 2007), ch.5; and Betts, *Enemies of Intelligence*, p.22.

likelihood and confidence, and the way they filter information. In each case, a push for certainty can systematically bias the intelligence process. This article suggests ways of adapting existing methods in order to *assess* uncertainty, explains how this represents a useful conceptual framework for thinking about estimative intelligence in general, and shows how this framework contrasts with many standard aspects of intelligence theory and practice.

It is important to make clear that the arguments in this article are not solely relevant to analysts and other members of the intelligence community. The goal of estimative intelligence is to help policymakers deal with situations that are uncertain and complex. If estimative intelligence does not address these situations appropriately, then this can adversely affect major decisions. Consumers of intelligence have an interest in making sure that this does not happen, and they have an important role to play in this process by encouraging (or at least not resisting) improvements in tradecraft. Ultimately, this article aims to stimulate discussion about the basic analytic foundations of estimative intelligence, and that subject is relevant to a wide range of stakeholders.

In some ways, assessing uncertainty in intelligence entails accepting complexity in the service of realistic analysis. For instance, many aspects of estimative intelligence today focus on either likelihoods or consequences, while objective analysis requires judging both. Few intelligence estimates present both the probability and the confidence of their predictions, and this article recommends assessing these factors independently and explicitly.

Yet it is important to note that the recommendations in this article should also help to avoid many existing difficulties. For instance, probability distributions avoid the difficulty of reconciling opposing viewpoints; they reduce the challenge of judging which hypotheses are ‘better’ than others; and they help to obviate debates about what constitutes a ‘significant’ threat. Clearly expressing likelihoods helps to prevent confusion about what estimates say and what they imply for policymaking. Assessing both likelihood and confidence helps to resolve disagreements about whether estimates make reliable predictions, whether they are actionable, and how much policymakers would benefit by waiting for more information. Improving source assessments helps to ensure that analysts and policymakers base their judgments on the most important information that is available. In each of these instances, the framework of assessing uncertainty helps to mitigate salient problems by improving the conceptual structure surrounding challenging issues that intelligence analysts deal with already.

Estimative intelligence will always be as much art as science, and the result will always be imperfect. This makes it all the more important to keep the process focused on the correct objectives, and to avoid unnecessary obstacles or biases. All else being equal, certainty should be welcomed when it comes to informing foreign policy. Yet uncertainty is bound to persist on critical issues, and when it does, the ultimate goal of estimative intelligence should be to assess this uncertainty in a clear and accurate manner.

Appendix. NIE 11-18-90, *The Deepening Crisis in the USSR*, p. iv

Figure 1
Scenarios for the Next Year

Scenario ^a	Factors That Could Lead to Scenario	Rough Probability
Deterioration Short of Anarchy	<p>Failure to agree upon and implement effectively a far-reaching marketization plan; or the broad resistance of the population to such a course.</p> <p>Failure of the center and the republics to move to new, mutually acceptable political and economic relations.</p> <p>Inability of political institutions to adapt to changing political realities; and ineffectiveness of new democratically elected leaders in governing.</p> <p>Continued, though diminished, viability of the central government.</p>	Close to even
Anarchy	<p>A precipitous decline of the economy.</p> <p>Massive social protests or labor strikes that proved to be beyond the security services' ability to control.</p> <p>The assassination of Gorbachev or Yel'tsin.</p> <p>The complete breakdown of relations between the center and the republics--especially the Russian Republic.</p>	1 in 5 or less
Military Intervention (ranging from a coup to civilian-directed martial law)	<p>Breakdown of key elements of the national economy, such as the transportation system.</p> <p>Violence against central government institutions.</p> <p>A situation approaching collapse of central authority.</p> <p>Anarchy.</p>	1 in 5 or less overall; much lower for a coup
"Light at the End of the Tunnel"	<p>Substantial progress toward:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Developing a new set of relationships allowing the republics to deal constructively with each other and the center. <input type="checkbox"/> The filling of the political power vacuum by new political institutions and parties. <input type="checkbox"/> Establishing new economic relations based on the market. 	1 in 5 or less

^a These scenarios are analytical constructs describing overall directions the USSR could take over the next year and are not mutually exclusive.

~~Secret NOFORN~~

328484 11-90

~~Secret~~