
ARTICLE

Handling and Mishandling Estimative Probability: Likelihood, Confidence, and the Search for Bin Laden

JEFFREY A. FRIEDMAN* AND RICHARD ZECKHAUSER

ABSTRACT In a series of reports and meetings in Spring 2011, intelligence analysts and officials debated the chances that Osama bin Laden was living in Abbottabad, Pakistan. Estimates ranged from a low of 30 or 40 per cent to a high of 95 per cent. President Obama stated that he found this discussion confusing, even misleading. Motivated by that experience, and by broader debates about intelligence analysis, this article examines the conceptual foundations of expressing and interpreting estimative probability. It explains why a range of probabilities can always be condensed into a single point estimate that is clearer (but logically no different) than standard intelligence reporting, and why assessments of confidence are most useful when they indicate the extent to which estimative probabilities might shift in response to newly gathered information.

Throughout Spring 2011, intelligence analysts and officials debated the chances that Osama bin Laden was living in Abbottabad, Pakistan. This question pervaded roughly 40 intelligence reviews and several meetings between President Obama and his top officials.¹ Opinions varied widely. In a key discussion that took place in March, for instance, the president convened several advisors to lay out the uncertainty about bin Laden's location as clearly as possible. Mark Bowden reports that the CIA team leader assigned to the pursuit of bin Laden assessed those chances to be as high as 95 per cent, that Deputy Director of Intelligence Michael Morell offered a figure of 60 per cent, and that most people 'seemed to place their confidence level at about 80 per cent' though 'some were as low as 40 or even 30 per cent'.² Other versions of the meeting agree that the president was given something like this

*Corresponding author. Email: jeffrey.a.friedman@dartmouth.edu

¹See Graham Allison, 'How It Went Down', *Time*, 7 May 2012.

²Mark Bowden, *The Finish: The Killing of Osama bin Laden* (NY: Atlantic Monthly 2012) pp.158–60.

range of probabilities, and that he struggled to interpret what this meant.³ President Obama reportedly complained that the discussion offered ‘not more certainty but more confusion’, and that his advisors were offering ‘probabilities that disguised uncertainty as opposed to actually providing you with more useful information’.⁴ Ultimately, the president concluded that the odds that bin Laden was living in Abbottabad were about 50/50.

This discussion played a crucial part in one of the most high-profile national security decisions in recent memory, and it ties into long-standing academic debates about handling and mishandling estimative probability. Ideas for approaching this challenge structure critical policy discussions as well as the way that the intelligence community resolves disagreements more generally, whether in preparing Presidential Daily Briefings and National Intelligence Estimates, or in combining the assessments of different analysts when composing routine reports. This process will never be perfect, but the president’s self-professed frustration about the Abbottabad debate suggests the continuing need for scholars, analysts, and policymakers to improve their conceptual understanding of how to express and interpret uncertainty. This article addresses three main questions about that subject.

First, how should decision makers interpret a range of probabilities such as what the president received in debating the intelligence on Abbottabad? In response to that question, this article presents an argument that some readers may find surprising: when it comes to deciding among courses of action, there is no logical difference between a range of probabilities and a single point estimate. In fact, this article explains why a range of probabilities always *implies* a point estimate, whether or not one is explicitly stated.

Second, what then is the value of presenting decision makers with a range of probabilities? Traditionally, analysts build ambiguity into their estimates to indicate how much ‘confidence’ they have in their conclusions. This article explains, however, why assessments of confidence should be separated from assessments of likelihood, and that expressions of confidence are most useful when they serve the specific function of describing how much predictions might shift in response to new information. In particular, this article introduces the idea of assessing the potential ‘responsiveness’ of an estimate, so as to assist decision makers confronting tough questions about whether it is better to act on existing intelligence, or delay until more intelligence can be gathered. By contrast, the traditional function of expressing confidence as a way to hedge estimates or describe their ‘information content’ should have little influence over the decision making process *per se*.

Third, what do these arguments imply for intelligence tradecraft? In brief, this article explains the logic of providing decision makers with both a point estimate and an assessment of potential responsiveness (but only one of each);

³David Sanger, *Confront and Conceal: Obama’s Secret Wars and Surprising Use of American Power* (NY: Crown 2012) p.93; Peter Bergen, *Manhunt: The Ten-Year Search for Bin Laden from 9/11 to Abbottabad* (NY: Crown 2012) pp.133, 194, 197, 204.

⁴Bowden, *The Finish*, pp.160–1; Bergen, *Manhunt*, p.198; and Sanger, *Confront and Conceal*, p.93.

it clarifies why it does *not* make sense to provide a range of predicted probabilities as was done in discussions about Abbottabad; and it demonstrates how current estimative language may exacerbate the problem of intelligence analysis providing ‘not more certainty but more confusion’, to use the president’s words.

These arguments are grounded in decision theory, a field that prescribes procedures for making choices under uncertainty. Decision theory has the advantage for our purposes that it is explicitly designed to help decision makers wrestle with the kinds of ambiguous and subjective factors that confronted President Obama and his advisors in debating the intelligence on Abbottabad.⁵ Decision theory cannot ensure that people make correct choices, nor that those choices lead to optimal outcomes, but it can at the very least help decision makers to combine available information and background assumptions in a consistent fashion.⁶ Logical consistency is a central goal of intelligence analysis as well and, as this article will show, there is currently a need for improving analysts’ and decision makers’ understanding of how to express and interpret estimative probability.⁷ The fact is that when the president and his advisors debated whether bin Laden was living in

⁵For descriptions of the basic motivations of decision theory, see Howard Raiffa, *Decision Analysis: Introductory Lectures on Choices under Uncertainty* (Reading, MA: Addison-Wesley 1968) p.x: ‘[Decision theory does not] present a *descriptive* theory of actual behavior. Neither [does it] present a *positive* theory of behavior for a superintelligent, fictitious being; nowhere in our analysis shall we refer to the behavior of an “idealized, rational, and economic man”, a man who always acts in a perfectly consistent manner as if somehow there were embedded in his nature a coherent set of evaluation patterns that cover any and all eventualities. Rather the approach we take *prescribes* how an individual who is faced with a problem of choice under uncertainty should go about choosing a course of action that is consistent with his personal basic judgments and preferences. He must consciously police the consistency of his subjective inputs and *calculate* their implications for action. Such an approach is designed to help us reason and act a bit more systematically – when we choose to do so!’

⁶Structured analytic techniques can also mitigate biases. Some relevant biases are intentional, such as the way that some analysts are said to hedge their estimates in order to avoid criticism for mistaken predictions. More generally, individuals encounter a wide range of cognitive constraints when assessing probability and risk. See, for instance, Paul Slovic, *The Perception of Risk* (London: Earthscan 2000); Reid Hastie and Robyn M. Dawes, *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making* (Thousand Oaks, CA: Sage 2001); and Daniel Kahneman, *Thinking, Fast and Slow* (NY: FSG 2011).

⁷Debates about estimative probability and intelligence analysis are long-standing; the seminal article is by Sherman Kent, ‘Words of Estimative Probability’, *Studies in Intelligence* 8/4 (1964) pp.49–65. Yet those debates remain unresolved, and they comprise a small slice of current literature. Miron Varouhakis, ‘What is Being Published in Intelligence? A Study of Two Scholarly Journals’, *International Journal of Intelligence and CounterIntelligence* 26/1 (2013) p.183, shows that fewer than 7 per cent of published studies in two prominent intelligence journals focus on analysis. In a recent survey of intelligence scholars, the theoretical foundations for analysis placed among the most under-researched topics in the field: Loch K. Johnson and Allison M. Shelton, ‘Thoughts on the State of Intelligence Studies: A Survey Report’, *Intelligence and National Security* 28/1 (2013) p.112.

Abbottabad, they were confronting estimative probabilities that they did not know how to manage. The resulting problems can be mitigated by steering clear of common misconceptions and by adopting basic principles for assessing uncertainty when similar issues arise in the future.

Assessing Uncertainty: Terms and Definitions

To begin, it is important to distinguish between risk and uncertainty. *Risk* involves situations in which decision makers know the probabilities of different outcomes. Roulette is a good example, because the odds of any bet's paying off can be calculated precisely. There are some games, such as blackjack or bridge, where expert players know most of the relevant probabilities as well. But true risk only characterizes scenarios that are simple and highly-controlled (like some forms of gambling), or circumstances for which there is a large amount of well-behaved data (such as what insurance companies use to set some premiums).

Intelligence analysts and policymakers rarely encounter true risk. Rather, they almost invariably deal with *uncertainty*, situations in which probabilities are ambiguous, and thus cannot be determined with precision.⁸ The debate about whether bin Laden was living in Abbottabad is an exemplar. Several pieces of information suggested that the complex housed Al Qaeda's leader, yet much of the evidence was also consistent with alternative hypotheses: the complex could have belonged to another high-profile terrorist, or to a drug dealer, or to some other reclusive, wealthy person. In situations such as these, there is no objective, 'right way' to craft probabilistic assessments from incomplete information. A wide range of intelligence matters, from strategic questions such as whether Iran will develop a nuclear weapon, to tactical questions such as where the Taliban might next attack US forces in Afghanistan, involve similar challenges in assessing uncertainty.

When decision makers and analysts discuss ambiguous odds, they are dealing with what decision theorists call *subjective probability*. (In the intelligence literature, the concept is usually called 'estimative probability'.) Subjective probability captures a person's degree of belief that a given statement is true. Subjective probabilities can rarely be calibrated with the precision of gambling odds or actuarial tables, but decision theorists stress that they can always be expressed quantitatively. The point of quantifying subjective probabilities is not to pretend that an analyst's beliefs are precise when they are not, but simply to be clear about what those beliefs entail, subjective foundations and all. Thus, when Deputy Director of Intelligence Michael Morell estimated the chances of bin Laden's residing in Abbottabad at roughly 60 per cent, it made little sense to think that this represented *the* probability of bin Laden's occupancy there, but it

⁸See Mark Phythian, 'Policing Uncertainty: Intelligence, Security and Risk', *Intelligence and National Security* 27/2 (2012) pp.187–205, for a similar distinction. *Ignorance* is another important concept, denoting situations where it is not even possible to define all possible answers to an estimative question. This situation often arises in intelligence analysis, but it is beyond the scope of this article.

was perfectly reasonable to think that this accurately represented *Morell's* personal beliefs about the issue.⁹

Decision theorists often use a thought experiment involving the comparison of lotteries to demonstrate that subjective probabilities can be quantified.¹⁰ Imagine that you are asked to make a prediction, such as whether the leader of a specified foreign country will be ousted by the end of the year. (As of this writing, for instance, Syrian rebels are threatening to unseat President Bashar al-Assad.) Now imagine that you are offered two options. The first option is that, if the leader is ousted you will receive a valuable prize; if the leader is not ousted, you will receive nothing. The second option is that an experimenter will reach into an urn containing 100 marbles, of which 30 are red and 70 are black. The experimenter will randomly withdraw one of these marbles. If it is red, you will receive the prize; if black, nothing. Would you bet on the leader's ouster or on a draw from the urn?¹¹

This question elicits subjective probabilities by asking you to evaluate your beliefs against an objective benchmark. A preference for having the experimenter draw from the urn indicates that you believe the odds of the leader's downfall by the end of the year are at most 30 per cent; otherwise you would see the ouster as the better bet. Conversely, betting on leadership change reveals your belief that the odds of this happening are at *least* 30 per cent. In principle, we could repeat this experiment a number of times, shifting the mix of colors in the urn until you became indifferent about which outcome to bet on, and this would reveal the probability you assign to the leader's ouster.

The point of discussing this material is not to advocate that betting procedures actually be used to assist with intelligence analysis,¹² but simply to make clear that even when individuals make estimates of likelihood based

⁹Sometimes subjective probability is called 'personal probability' to emphasize that it captures an individual's beliefs about the world rather than objective frequencies determined via controlled experiments. As Frank Lad describes, the subjectivist approach to probability 'represents your assessment of your own personal uncertain knowledge about *any* event that interests you. There is no condition that events be repeatable... In the proper syntax of the subjectivist formulation, you might well ask me and I might well ask you, "What is *your* probability for a specified event?" It is proposed that there is a distinct (and generally different) correct answer to this question for each person who responds to it. We are each sanctioned to look within ourselves to find our own answer. Your answer can be evaluated as correct or incorrect only in terms of whether or not you answer honestly'. Frank Lad, *Operational Subjective Statistical Methods* (NY: Wiley 1996) pp.8–9.

¹⁰On eliciting subjective probabilities, see Robert L. Winkler, *Introduction to Bayesian Inference and Decision*, 2nd ed. (Sugar Land, TX: Probabilistic Publishing 2010) pp.14–23.

¹¹To control for time preferences, one would ideally make it so that the potential resolutions of these gambles and their payoffs occurred at the same time. Thus, if the gamble involved the odds of regime change in a foreign country by the end of the year, the experimenter would draw from the urn at year's end or when the regime change occurred, at which point any payoff would be made.

¹²An issue of substantial controversy in recent years: see Adam Meirowitz and Joshua A. Tucker, 'Learning from Terrorism Prediction Markets', *Perspectives on Politics* 2/2 (2004) pp.331–6.

on ambiguous information, their beliefs can still be elicited coherently in the form of subjective probabilities. There is then a separate question of how *confident* people are in their estimates. A central theme of this article is that clearly conveying assessments of confidence is also critically important.¹³

Expressing and Interpreting Estimative Probability

This section offers an argument that some readers will find surprising, which is that when decision makers are weighing alternative courses of action, there is no logical difference between expressing a range of probabilities and articulating a single point estimate.¹⁴ Of course, analysts can avoid explicitly stating a point estimate; but this section will show that it is impossible to avoid implying one, either when analysts express their individual views or when a group presents its opinions.

This argument is important for two reasons. First, decision makers should know that, contrary to the confusion surrounding bin Laden's suspected location, there is a logical and objective way to interpret the kind of information that the president was offered. Second, if analysts know this to be the case, they have an incentive to express this information more precisely in order to avoid having decision makers draw misguided conclusions. As in the previous section, the argumentation here presents basic concepts using stylized examples, and then extends those concepts to the Abbottabad debate.

Combining Probabilities

Put yourself in the position of an analyst, knowing that the president will ask what you believe the chances are that bin Laden is living in Abbottabad.¹⁵ According to the accounts summarized earlier, some analysts assessed this probability as being 40 per cent, and some assessed it as being 80 per cent. Imagine you believe that each of these opinions is equally plausible. This simple case is useful for building insight, because the way to proceed here is straightforward: you should state the midpoint of the plausible range.

To see why this is so, imagine you are presented with a situation similar to the one described before, in which if you pull a red marble from an urn you win a prize and otherwise you receive nothing. Now, however, you are given

¹³On the distinction between likelihood and confidence in contemporary intelligence tradecraft, see Kristan J. Wheaton, 'The Revolution Begins on Page Five: The Changing Nature of NIEs', *International Journal of Intelligence and CounterIntelligence* 25/2 (2012) p.336.

¹⁴The separate question of how decision makers should determine when to act versus waiting to gather additional information is a central topic in the section that follows.

¹⁵The discussion below applies broadly to debating any question that has a yes-or-no answer. Matters get more complicated with broader questions that admit many possibilities, for example 'Where is bin Laden currently living?' These situations can be addressed by decomposing the issue into binary components, such as 'Is bin Laden living in location A?', 'Is bin Laden living in location B?', and so on.

not one but *two* urns. If you draw from one urn, there is a 40 per cent chance of picking a red marble; and if you draw from the other urn, there is an 80 per cent chance of picking a red marble. You can select either urn, but you do not know which is which and you cannot see inside the urns to inform your decision.

This situation represents a *compound lottery*, because it involves multiple points of uncertainty: first, the odds of selecting a particular urn and, second, the odds of picking a red marble from the chosen urn. This situation is analogous to presenting a decision maker with the notion that bin Laden is either 40 or 80 per cent likely to be living at Abbottabad, without providing any information as to which of these assessments is more plausible.¹⁶

In this case, you should conceive of the *expected probability* of drawing a red marble as being the odds of selecting each urn combined with the odds of drawing a red marble from each urn, respectively. If you cannot distinguish the urns, there is a one-half chance that the odds of picking a red marble are 80 per cent and a one-half chance that the odds of picking a red marble are 40 per cent. The overall expected probability of drawing a red marble is thus 60 per cent: this is simply the average of the relevant possibilities, which makes sense given that we are combining them under the assumption that they deserve equal weight.¹⁷

Many people do not immediately accept that one should act on expected probabilities derived in this fashion. The economist Daniel Ellsberg, who became famous for leaking the *Pentagon Papers*, initially earned academic renown for pointing out that most people prefer to bet on known probabilities.¹⁸ Scholars call this behavior ‘ambiguity aversion’. In general, people have a tendency to tilt their probabilistic estimates towards caution so as to avoid being over-optimistic when faced with uncertainty.¹⁹

¹⁶By extension, this arrangement covers situations where the best estimate could lie anywhere between 40 and 80 per cent, and you do not believe that any options within this range are more likely than others, since the contents of the urn could be combined to construct any intermediate mix of colors.

¹⁷ $1/2 \times 0.40 + 1/2 \times 0.80 = 0.60$, or 60 per cent.

¹⁸Ellsberg’s seminal example involved deciding whether to choose from an urn with exactly 50 red marbles and 50 black marbles to win a prize versus an urn where the mix of colors was unknown. Subjects will often pay non-trivial amounts to select from the urn with known risk, which makes no rational sense. If people are willing to pay more in order to gamble on drawing a red marble from the urn with a 50/50 distribution, this implies that they believe there is less than a 50 per cent chance of drawing a red marble from the ambiguous urn. But by the same logic, they will also be willing to pay more in order to gamble on drawing a *black* marble from the 50/50 urn, which implies they believe there is less than a 50 per cent chance of drawing a *black* marble from the ambiguous urn. These two statements cannot be true simultaneously. This is known as the ‘Ellsberg paradox’. See Daniel Ellsberg, ‘Risk, Ambiguity, and the Savage Axioms’, *Quarterly Journal of Economics* 75/4 (1961) pp.643–69.

¹⁹See Stefan T. Trautmann and Gijs van de Kuilen, ‘Ambiguity Attitudes’ in Gideon Keren and George Wu (eds.) *The Blackwell Handbook of Judgment and Decision Making* (Malden, MA: Blackwell forthcoming).

Yet, in the example given here, interpreting the likelihood of bin Laden's being in Abbottabad as anything other than 60 per cent would lead to a contradiction. If you assessed these chances as being below 60 per cent, then you would be giving the lower-bound estimate extra weight. If you assessed those chances as being higher than 60 per cent, then you would be giving the upper-bound estimate extra weight. Either way, you would be violating your own beliefs that neither one of those possibilities is more plausible than the other.

Moreover, any time analysts tilt their estimates in favor of caution, they are conflating information about likelihoods with information about outcomes. When predicting the odds of a favorable outcome (such as the United States having successfully located bin Laden), the cautionary tendency would be to shift probabilities downward, so as to avoid being disappointed by the results of a decision (such as ordering a raid on the complex and discovering that bin Laden was not in fact there). But when predicting the odds of an unfavorable outcome (such as a possible impending terrorist attack), the cautionary tendency would be to shift probabilities upward in order to reduce the chances of underestimating a potential threat.²⁰ This is not to deny that decision makers should be risk averse, but to make clear that risk aversion appropriately applies to evaluating outcomes and not their probabilities. How the president should have chosen to deal with the chances of bin Laden's being in Abbottabad was logically distinct from understanding what those chances were.²¹

In many (if not most) cases, analysts will have cogent reasons for thinking that some point estimates are more credible than others. For instance, it is said that intelligence operatives initially found the Abbottabad complex when they tailed a person suspected of being bin Laden's former courier, who had told someone by phone that he was 'with the same people as before'. There was a good chance that this indicated the courier was once again working for Al Qaeda, in which case the odds that the Abbottabad compound belonged to bin Laden might be relatively high (perhaps about 80 per cent). But there was a chance that intelligence analysts had misinterpreted this information, in which case the chances that bin Laden was living in the compound would have been much lower (perhaps about 40 per cent). This is just one of many examples of how analysts must wrestle with the fact that their overall inferences are based on information that can point in different

²⁰See W. Kip Viscusi and Richard J. Zeckhauser, 'The Less Than Rational Regulation of Ambiguous Risks', University of Chicago Law School Conference, 26 April 2013. These contrasting examples show how risk preferences depend on a decision maker's views of whether it would be worse to make an error of commission or omission – in decision theory more generally, a key concept is balancing the risks of Type I and Type II errors.

²¹It is especially important to disentangle these questions when it comes to intelligence analysis, a field in which one of the cardinal rules is that it is the analyst's role to provide information that facilitates decision making, but not to interfere with making the decision itself. Massaging probabilistic estimates in light of potential policy responses inherently blurs this line.

directions. Imagine you believed that, on balance, the high-end estimate given in this example is about twice as credible as the low-end estimate. How then should you assess the chances of bin Laden living in Abbottabad?

Once again, the compound lottery framework tells us how this information implies an expected probability. Continuing the analogy to marbles and urns, we could represent this situation in terms of having two urns each with 80 red marbles, and a third urn with 40 red marbles. In choosing an urn now, you are twice as likely to be picking from the one where the odds of drawing red are relatively high, so the expected probability of picking a red marble is 67 per cent.²² As before, you would contradict your own assumptions by interpreting the situation any other way.

Even readers not familiar with the compound lottery framework are certainly acquainted with the principles supporting it, which most people employ in everyday life. When you consider whether it might rain tomorrow, for instance, or the odds that a sports team will win the championship, you presumably begin by considering a number of answers based on various possible scenarios. Then you weigh the plausibility of these scenarios against each other and determine what seems like a reasonable viewpoint, on balance. You will presumably think through how a number of different possible scenarios could affect the overall outcome (what if the team's star player gets injured?) along with the chances that those scenarios might occur (how injury-prone does that player tend to be and how healthy does he seem at the moment?) Compound lotteries are structured ways to manage this balance without violating your own beliefs.

Examining this logic makes clear not merely that subjective probabilities *can* be expressed using weighted averages, but that, in an important sense, this is the only logical approach. Analysts who do not try to give more weight to the more plausible possibilities would be leaving important information out of their estimates. And if they attempted to hedge those estimates by providing a range of possible values instead of a single subjective probability, then the only changes to the message would be semantic as, in the absence of other information, rational decision makers should simply act as though they have been given that range's midpoint. Thus while there are many situations in which analysts might wish to avoid providing decision makers with an estimated probability – whether because they do not feel they have enough information to make a reliable guess, or because they do not want to be criticized for making a prediction that seems wrong in hindsight, or for any other reason – there is actually not much of a logical alternative. Even when analysts avoid explicitly stating a point estimate, they cannot avoid implying one.

Interpreting Estimates from Groups

The same basic logic holds for interpreting estimates provided by different analysts, as in the debate about bin Laden's location. We have seen that President Obama received several different point estimates of the chances that

²² $1/3 \times 0.80 + 1/3 \times 0.80 + 1/3 \times 0.40 = 0.67$, or 67 per cent.

bin Laden was living in Abbottabad, including figures of 40 per cent, 60 per cent, 80 per cent, and 95 per cent. The president was unsure of how to handle this information, but the task of combining probabilities is largely the same as before.²³ For instance, if these estimates seemed equally plausible, then, all else being equal, the appropriate move would simply be to average them together, which using these particular numbers would come to 69 per cent.²⁴

Note how this figure is substantially above President Obama's determination that the odds of bin Laden's living in Abbottabad were 50/50. Given existing reports of this debate, it thus appears that the president implicitly gave the lower-bound estimates disproportionate influence in forming his perceptions. This is consistent with the notion that people often tilt probability estimates towards caution. It is also consistent with a common bias in which individuals default to 50/50 assessments when presented with complex situations, regardless of how well that stance fits with the evidence at hand.²⁵ Both of these tendencies can be exacerbated by presenting a range of predicted probabilities rather than a single point estimate.

Once again, we can usually improve our assessments by accounting for the idea that some estimates will be more plausible than others, and this was an important subtext of the Abbottabad debate. According to Mark Bowden's account summarized above, the high estimate of 95 per cent came from the CIA team leader assigned to the bin Laden pursuit. It might be reasonable to expect that this person's involvement with intelligence collection may have influenced his or her beliefs about that information's reliability.²⁶ Meanwhile, the lowest estimate offered to the president (which secondary sources have reported as being either 30 or 40 per cent) was the product of a 'Red Team' of analysts specifically charged to offer a 'devil's advocate' position questioning the evidence; this estimate was explicitly biased downward, and for that reason it should have been discounted.²⁷ As this section demonstrates, anyone confronted with multiple estimates should combine them in a manner that weights each one in proportion to its credibility. Contrary to the confusion that

²³Throughout this discussion, the information available to one analyst is assumed to be available to all, hence analysts are basing their estimates on the same evidence. The importance of this point is discussed below.

²⁴ $1/4 \times 0.40 + 1/4 \times 0.60 + 1/4 \times 0.80 + 1/4 \times 0.95 = 0.69$, or 69 per cent.

²⁵See Baruch Fischhoff and Wandi Bruine de Bruin, 'Fifty-Fifty = 50%?', *Journal of Behavioral Decisionmaking* 12/2 (1999) pp.149–63. Consistent with this idea, Paul Lehner et al., 'Using Inferred Probabilities to Measure the Accuracy of Imprecise Forecasts' (MITRE 2012) Case #12–4439, show that intelligence estimates predicting that an event will occur with 50 per cent odds tend to be especially inaccurate.

²⁶In this case, some people believed the analyst fell prey to confirmation bias and thus interpreted the evidence over-optimistically. In other cases, of course, analysts involved with intelligence collection may deserve extra credibility given their knowledge of the relevant material.

²⁷This is not to deny that there may be instances where the views of 'Red Teams' or 'devil's advocates' will turn out to be more accurate than the consensus opinion (and more generally, that these tools are useful for helping to challenge and refine other estimates). It is to say that, on balance, one should expect an analysis to be less credible if it is biased, and Red Teams are explicitly tasked to slant their views.

clouded the debate about bin Laden's location (and that befogs intelligence matters in general), there is in fact an objective way to proceed.

This discussion naturally leads to the question of how to assess analyst credibility. In performing this task, it is important to keep in mind that estimates of subjective probability depend on three analyst-specific factors: prior assumptions, information, and analytic technique. Prior assumptions are intuitive judgments people make that shape their inferences. Deputy Director Morell, for instance, reportedly discussed how he thought the evidence about Abbottabad compared to the evidence on Iraqi Weapons of Mass Destruction nine years earlier, a debate in which he had participated and which had made him cautious about extrapolating from partial information.²⁸ One might see Morell's view here as being unduly influenced by an experience unrelated to the intelligence on bin Laden, but one might also see his skepticism as representing a useful corrective to natural overconfidence.²⁹ Either way, Morell's estimate was clearly influenced by his prior assumptions. Again, the goal of decision theory is not to pretend that these kinds of subjective assumptions do not (or should not) matter, but rather to take them into account in explicit, structured ways.³⁰

Analysts employing different information will also arrive at different estimates. This makes it difficult to determine what their assessments imply in the aggregate, because it forces decision makers to synthesize information that analysts did not combine themselves.³¹ But this situation is avoidable

²⁸As Morell reflected: 'People don't have differences [over estimative probabilities] because they have different intel ... We are all looking at the same things. I think it depends more on your past experience' (Bowden, *The Finish*, p.161).

²⁹On overconfidence in national security decision making, see Dominic D. P. Johnson, *Overconfidence in War: The Havoc and Glory of Positive Illusions* (Cambridge, MA: Harvard University Press 2004).

³⁰It is important to note that prior assumptions do not always lead analysts towards the right conclusions. Robert Jervis argues, for instance, that flawed assessments of Iraq's potential weapons of mass destruction programs were driven by the assumed plausibility that Saddam Hussein would pursue such capabilities. The point of assessing priors is thus not to reify them, but rather to make their role in the analysis explicit (and, where possible, to submit such assumptions to structured critique). See Robert Jervis, 'Reports, Politics, and Intelligence Failures: The Case of Iraq', *Journal of Strategic Studies* 29/1 (2006) pp.3–52.

³¹For example, imagine that two analysts are assigned to assess the likelihood that a certain state will attack its neighbor by the end of the year. These analysts share the prior assumption that the odds of this happening are relatively low (say, about 5 per cent). They independently encounter different pieces of information suggesting that these chances are higher than they originally anticipated. Analyst A learns that the country has been secretly importing massive quantities of armaments, and Analyst B learns that the country has been conducting large, unannounced training exercises for its air force. Based on this information, our analysts respectively estimate a 30 per cent (A) and a 40 per cent (B) chance of war breaking out by the end of the year. In this instance, it would be problematic to think that the odds of war are somewhere between 30 and 40 per cent, because if the analysts had been exposed to and properly incorporated each others' information, both their respective estimates would presumably have been higher. On such processes, see Richard Zeckhauser, 'Combining Overlapping Information', *Journal of the American Statistical Association* 66/333 (1971) pp.91–2.

when analysts are briefed on the relevant material and, in high-level discussions like the Abbottabad debate, it is probably reasonable to assume that key participants were familiar with the basic body of evidence.³²

Even if analysts have equal access to intelligence, informational biases can persist. Some analysts may be predisposed to skepticism of information gathered by others, and thus rely predominantly on information they have developed themselves.³³ Analysts might also differ on the kinds of information from which they feel comfortable drawing inferences. An analyst who has spent her whole career working with human sources may assign this kind of information more value than someone who has spent a career working with signals intelligence.³⁴

Finally, despite shared prior assumptions and identical bodies of information, analysts can still arrive at varying estimates if they employ different analytic techniques. Sometimes analytic disparities are clear. As mentioned earlier, the ‘Red Team’ in the Abbottabad debate was explicitly instructed to interpret information skeptically. However useful this approach may have been for stimulating discussion, it also meant that the Red Team’s estimate should have carried less weight than more objective assessments. Thus, while there is rarely a single, ‘right way’ to assess uncertainty, it is still often possible to get a rough sense of the analytic foundations of different estimates and of the biases that may influence them. Prior assumptions, information, and analytic approaches should be considered in determining the credibility of a given estimate.³⁵

³²See the quote in Footnote 28.

³³Scholars have shown this to be the case in several fields including medicine, law, and climate change. See, for instance, Elke U. Weber et al., ‘Determinants of Diagnostic Hypothesis Generation: Effects of Information, Base Rates, and Experience’, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19/5 (1993) pp.1151–64; Jonas Jacobson et al., ‘Predicting Civil Jury Verdicts: How Attorneys Use (and Misuse) a Second Opinion’, *Journal of Empirical Legal Studies* 8/S1 (2011) pp.99–119.

³⁴This tendency follows from the availability heuristic, one of the most thoroughly-documented biases in behavioral decision, which shows that people inflate the probability of events that are easier to bring to mind; see Amos Tversky and Daniel Kahneman, ‘Availability: A Heuristic for Judging Frequency and Probability’, *Cognitive Psychology* 5 (1973) pp.207–32. In the context of intelligence analysis, this implies that analysts may conflate the predictive value of a piece of information with how easily they are able to interpret it.

³⁵An analyst’s past record of making successful predictions may also inform the credibility of her estimates. Recent research demonstrates that some people are systematically better than others at political forecasting. However, contemporary systems for evaluating analyst performance are relatively underdeveloped, especially since analysts tend not to specify probabilistic estimates in a manner that can be rated objectively. Moreover, when decision makers consider the past performance of their advisors, they may be tempted to extrapolate from a handful of experiences that offer little basis for judging analytic skill (or from personal qualities that are not relevant for estimating likelihood). Determining how to draw sound inferences about an analyst’s credibility from their past performance is thus an area where further research can have practical benefit. In the interim, we focus on evaluating the logic of each analyst’s assessment per se. See Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press 2005) and Lyle Ungar

The substantial literature on group decision making lays out a range of methods for assessing expert credibility, while emphasizing that any recommended approach is bound to be subjective and contentious.³⁶ This makes it all the more important that analysts address this task and not leave it to decision makers. In order to interpret a group of point estimates from different sources, it is essential to draw conclusions about the credibility of those sources. This task is unavoidable, yet it is often left to be performed implicitly by decision makers who are neither trained in this exercise, nor specially informed about available intelligence, and often are not even familiar with people advising them. Analysts should take on this task. Had they done so in discussions about Abbottabad, President Obama almost surely would have had much clearer and more helpful information than what he received. This section sought to demonstrate that, since a range of estimative probabilities always implies a single point estimate, then *not* giving a single point estimate may only confuse a decision maker, while simultaneously preventing analysts from refining and controlling the message they convey.

Expressing Confidence: The Value of Additional Information

Point estimates convey likelihood but not confidence. As explained earlier, these concepts should be kept separate. And if providing a range of estimates such as those President Obama's advisors offered in debates about Abbottabad is not an effective way to convey information about likelihood, it is no better in articulating confidence. This section argues that instead of expressing confidence in order to 'hedge' estimates by saying how ambiguous they are *right now*, it is better to express confidence in terms of how much those estimates might shift as a result of gaining new information moving forward. This idea would allow analysts to express the ambiguity surrounding their estimates in a manner that ties into decision making much more directly than the way these assessments are typically conveyed.

Why Hedging does not Help

One of the intended functions of expressing confidence is to inform decision makers about the extent of the ambiguity that estimates involve. Standard procedure is to express confidence levels as being 'low', 'medium', or 'high' in

et al., 'The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions', AAI Technical Report FS-12-06 (2012).

³⁶Some relevant techniques include: asking individuals to rate the credibility of their own predictions and using these ratings as weights; asking individuals to debate among themselves whose estimates seem most credible and thereby determine appropriate weighting by consensus; and denoting a member of the team to assign weights to each estimate after evaluating the reasoning that analysts present. For a review, see Detlof von Winterfeldt and Ward Edwards, *Decision Analysis and Behavioral Research* (NY: Cambridge University Press 1986) pp.133–6.

a fashion that ‘reflect[s] the scope and quality of the information’ behind a given estimate.³⁷ Though official intelligence tradecraft explicitly distinguishes between the confidence levels assigned to an estimate of likelihood and the statement of likelihood itself, these concepts routinely get conflated in practice.³⁸ When President Obama’s team offered several different point estimates about the chances of bin Laden’s being in Abbottabad, for instance, this conveyed estimates and ambiguity simultaneously, rather than distinguishing these characteristics from one another.

Some aspects of intelligence tradecraft encourage analysts to merge likelihood and confidence. Take, for example, the ‘Words of Estimative Probability’ (WEPs) employed in recent National Intelligence Estimates. [Figure 1](#) shows how the National Intelligence Council identified seven terms that can be used for this purpose, evenly spaced on a spectrum that ranges from ‘remote’ to ‘almost certainly’. The intent of using these WEPs instead of reporting numerical odds is to hedge estimates of likelihood, accepting the ambiguity associated with estimative probabilities, and avoiding the impression that these estimates entail scientific precision.³⁹

Yet this kind of hedging does not enjoy the logical force that its advocates intend. WEPs essentially offer ranges of subjective probabilities and, as we have seen, these can always be condensed to single points. Absent additional information to say whether any parts of a range are more plausible than others, decision makers should treat an estimate that some event is ‘between 40 and 80 per cent likely to occur’ just the same as an estimate that the event is 60 per cent likely to occur: these statements are logically equivalent for decision makers trying to establish expected probabilities.

Similarly, the use of WEPs or other devices to hedge estimative probabilities obfuscates content without actually changing it. In [Figure 1](#), for instance, the term ‘even chance’ straddles the middle of the spectrum. It is hard to know exactly how wide a range of possibilities falls within an acceptable definition of this term – but absent additional information, *any* range of estimative probabilities centered on 50 per cent can be interpreted as simply meaning ‘50 per cent’.⁴⁰

³⁷Office of the Director of National Intelligence [ODNI], *US National Intelligence: An Overview* (Washington, DC: ODNI 2011) p.60.

³⁸For example, Bergen prints a quote from Director of National Intelligence James Clapper referring to estimates of the likelihood that bin Laden was living in Abbottabad as ‘percentage [s] of confidence’ (*Manhunt*, p.197). On the dangers of conflating likelihood and confidence in intelligence analysis more generally, see Jeffrey A. Friedman and Richard Zeckhauser, ‘Assessing Uncertainty in Intelligence’, *Intelligence and National Security* 27/6 (2012) pp.835–41.

³⁹On the WEPs and their origins, see Wheaton, ‘Revolution Begins on Page Five’, pp.333–5.

⁴⁰To repeat, these statements are equivalent only from the standpoint of how decision makers should choose among options *right now*. In reality, decision makers must weigh immediate action against the potential costs and benefits of delaying to gather new information. In this respect, the estimates ‘between 0 and 100 per cent’ and ‘50 per cent’ may indeed have different interpretations, as the former literally relays no information (and thus a decision maker might be inclined to search for additional intelligence) while an estimate of ‘50 per cent’ could

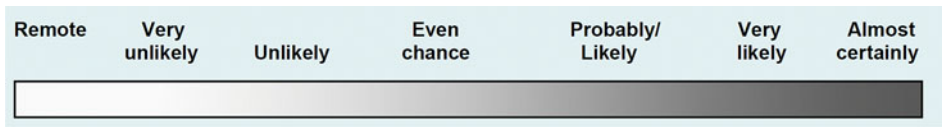


Figure 1. Words of Estimative Probability. *Source:* Graphic displayed in the 2007 National Intelligence Estimate, *Iran: Nuclear Intentions and Capabilities*, as well as in the front matter of several other recent intelligence products.

Some of the other words on this spectrum are harder to interpret. The range of plausible definitions for the word ‘remote’, for instance, clearly starts at 0 per cent, but then might extend up to 10 or 15 per cent. This introduces a problem of its own, as different people could think that the word has different meanings. But there is another relevant issue here, which is that if we interpret ‘remote’ as meaning anything between 0 and 10 per cent, with no special emphasis on any value within that range, then this is logically equivalent to saying ‘5 per cent’. This may often be a gross exaggeration of the odds analysts intend to convey. The short-term risks of terrorists capturing Pakistani nuclear weapons or of North Korea launching a nuclear attack on another state are presumably very small, just as were the risks of a nuclear exchange in most years during the Cold War. Saying that those probabilities are ‘remote’ does not even offer decision makers an order of magnitude for what the odds entail. They could be one in ten, one in a hundred, or one in a million, and official estimative language provides no way to tell these estimates apart.

Seen from this perspective, the Words of Estimative Probability displayed in Figure 1 constrict potential ways of expressing uncertainty. It is difficult to imagine that the intelligence community would accept a proposal stating that analysts could only report estimative probabilities using one of seven numbers; but this is the effective result of employing standard terminology. These words attempt to alter estimates’ meanings by building ambiguity into expressions of likelihood, but this function is simply semantic. If analysts wish to convey information about ambiguity and confidence, that information should be offered separately and explicitly.

Assessing Responsiveness

Another drawback to the way that confidence is typically expressed in intelligence analysis is that those expressions are not directly relevant to decision making. The issue is not just that terms like ‘low’, ‘medium’, or ‘high confidence’ are vague. The more important conceptual problem is that

represent a careful weighing of evidence that is unlikely to shift much moving forward. This distinction shows why it is important to assess both likelihood *and* confidence, and why estimates of confidence should be tied directly to questions about whether decision makers should find it worthwhile to gather additional information. This is the subject of discussion below.

ambiguity itself is not what matters when it comes to making hard choices. As shown earlier, if a decision maker is considering whether to take a gamble – whether drawing a red marble from an urn or counting on bin Laden’s being in Abbottabad – it makes no difference whether the odds of this gamble’s paying off are expressed with a range or a point estimate. If a range is given, it does not matter whether it is relatively small (50 to 70 per cent) or relatively large (20 to 100 per cent): absent additional information, these statements imply identical expected probabilities. By the same token, if analysts make predictions with ‘low’, ‘medium’, or ‘high’ confidence, this should not, in itself, affect the way that decision makers view the chances that a gamble will pay off.

What *does* often make a difference for decision makers, however, is thinking about how much estimates might shift as a result of gathering more information. Securing more information is almost always an option, though the benefits reaped must be balanced against the costs of delay and of collecting new intelligence.⁴¹ This trade-off was reportedly at the front of President Obama’s mind when debating the intelligence about Abbottabad. As time wore on, the odds only increased that bin Laden might learn that the compound was being watched, and yet the president did not want to commit to a decision before relevant avenues for collecting intelligence had been exploited.

Confidence levels, as traditionally expressed, fail to facilitate this crucial element of decision making. For instance, few analysts seem to have believed that they could interpret the intelligence on bin Laden’s location with ‘high confidence’. But it might also have been fair to say that, by April 2011, the intelligence community had exploited virtually all available means of learning about the Abbottabad compound. Even if analysts had ‘low’ or ‘medium’ confidence in estimating whether bin Laden was living in that compound, they might *also* have assessed that their estimates were unlikely to change appreciably in light of any new information that might be gathered within a reasonable period of time. Thus, the relevant question is not necessarily where the ‘information content’ of assessments stands at the moment, but how much that information might change and how far those assessments might shift moving forward. Existing intelligence tradecraft does not express this information directly.

A straightforward solution is to have analysts not merely state estimative probabilities, but also to explain how much those assessments might change in light of further intelligence. This attribute might be termed an estimate’s *responsiveness*. What follows is a brief discussion of how responsiveness can be described.

The first step is to establish the relevant time period for which responsiveness should be assessed. On an urgent matter like the Abbottabad debate, for instance, decision makers might want to know how much analysts’ views might shift within a month, or after the conclusion of specific,

⁴¹This is one of the central subjects of decision theory. See Winkler, *Introduction to Bayesian Inference and Decision*, ch.6, and Raiffa, *Decision Analysis*, ch.7.

near-term collection efforts. On other issues, it might be appropriate to assess responsiveness more broadly (for example, how much an estimate might change within six months or a year).

Once this benchmark is established, analysts should think through how much their estimates could plausibly shift in the interim, and why that might be the case. In order to show what form these assessments might take, here are three different scenarios that analysts might have conveyed to President Obama in April 2011: in each case, analysts start with the same perception about the chances that bin Laden is living in the suspected compound, but they have very different view about how much that estimate might change in the near future.

Scenario #1: No information forthcoming. There is currently a 70 per cent chance that bin Laden is living in Abbottabad. All avenues for gathering additional information have been exhausted. Based on continuing study of available evidence, it is equally likely that, at the end of the month, analysts will believe that the odds of bin Laden living in Abbottabad are as low as 60 per cent, as high as 80 per cent, or that the current estimate of 70 per cent will remain unchanged.

Scenario #2: Small-scale shifting. There is currently a 70 per cent chance that bin Laden is living in Abbottabad. By the end of the month, there is a relatively small chance (about 5 per cent) of learning that bin Laden is definitely not there. There is a relatively high chance (about 75 per cent) that additional information will reinforce existing assessments, and the odds of bin Laden being in Abbottabad might climb to 80 per cent. There is a smaller chance (about 20 per cent) that additional intelligence will reduce the plausibility of existing assessments to 50 per cent.

Scenario #3: Potentially dispositive information forthcoming. There is currently a 70 per cent chance that bin Laden is living in Abbottabad. Ongoing collection efforts may resolve this uncertainty within the next month. Analysts believe that there is roughly a 15 per cent chance of determining that bin Laden is definitely not living in the suspected compound; there is roughly a 35 per cent chance of determining that bin Laden definitely *is* living in the suspected compound; and there is about a 50 per cent chance that existing estimates will remain unchanged.

These assessments are all hypothetical, of course, but they help to motivate some basic points. First, all of these assessments imply that at the end of the month analysts will have the same expected probability of thinking that bin Laden is living in Abbottabad as they did at the beginning.⁴² The expected probability of projected future estimates *must* be the same as the expected probability of current estimates, because if analysts genuinely expect that their

⁴²Scenario 1: $\frac{1}{3} \times 0.60 + \frac{1}{3} \times 0.70 + \frac{1}{3} \times 0.80 = 0.70$. Scenario 2: $0.15 \times 0.00 + 0.35 \times 1.00 + 0.50 \times 0.70 = 0.70$. Scenario 3: $0.05 \times 0.00 + 0.75 \times 0.80 + 0.20 \times 0.50 = 0.70$.

subjective probabilities will go up in the future, then this should be reflected in their present estimates (just as if investors expect a stock price to go up in the future, they will buy more of it until current and future expected values align).⁴³ Thus, explicitly assessing an estimate's responsiveness in the manner described here is actually a good way to diagnose whether analysts' current views make sense, or whether they instead suffer from logical contradictions.

Second, it is worth noting that even though these scenarios are described in concise paragraphs, they may have radically different consequences for the actions taken by decision makers. Scenario 1 clearly offers the least expected benefit of delay, Scenario 3 clearly offers the most expected benefit of delay, and Scenario 2 falls somewhere in the middle. Without knowing President Obama's preferences and beliefs in the Abbottabad debate, it is not possible to say how these distinctions would have influenced his decision making in this particular case. What is clear, however, is that these kinds of considerations *can* make a major difference, even though existing tradecraft provides no structured manner for conveying this information directly.

Moreover, the ideas suggested in this article impose demands on analysts that again are less about how individuals should form their views about ambiguous information than they are about how to *express* those views in ways that decision makers will find useful. Decision makers already have to form opinions about the potential value of gathering more information, and in doing so they already must think through different scenarios for how much existing assessments might respond as new material comes to light. That determination will always involve subjective factors, but those factors can at least be identified and conveyed in a structured fashion.

Implications and Critiques

Motivated by the debate about Osama bin Laden's location while speaking to the literature on intelligence analysis more generally, this article advanced two main arguments about expressing and interpreting estimative probability. First, analysts should convey likelihood using point estimates rather than ranges of predicted probabilities such as those offered by President Obama's advisors in the Abbottabad debate. This article showed that, even if analysts do not state a point estimate, they cannot avoid at least implying one; therefore, they might as well present a point estimate explicitly so as to avoid unnecessary confusion. Second, analysts should articulate confidence by assessing the 'responsiveness' of their estimates in a manner that explains how much their views might shift in light of new intelligence.

⁴³This holds constant the idea that the situation on the ground might change within the next month: bin Laden might have learned that he was being watched and might have fled, for instance, and that was something which the president reportedly worried about. This, however, is a matter of how much the state of the world might change, which is different from thinking about how assessments of the current situation might respond to additional information.

This should help decision makers to understand the benefit of securing more information, an issue that standard estimative language does not address.

How might these suggestions be implemented? While the scope of this article is more conceptual than procedural, here are some practical steps.

First, it is critical that analysts have access to the same body of evidence. Analysts are bound to assess uncertainty in different ways given their differing techniques and prior assumptions, but disagreements stemming from asymmetric information can be avoided.

Second, analysts should determine personal point estimates ('How likely is it that bin Laden is living in Abbottabad?'), and should assess how responsive those estimates might be within a given time frame ('How high or low might your estimate shift in response to newly gathered information within the next month?'). As the previous section showed, this need not require undue complexity, and this material can be summarized in a concise paragraph if necessary.

Third, analysts should decide whether they are in a position to evaluate each other's credibility. This can be done through self-reporting, or by asking analysts to debate among themselves which estimates seem more or less credible, or by having team leaders distill these conclusions based on analysts' argumentation and backgrounds. A potential measure that may be useful is to designate a member of the team as a 'synthesizer', one whose job is not to analyze the intelligence per se, but rather to assess the plausibility of different estimates.

Finally, when reporting to a decision maker, analysts' assessments should be combined into a single point estimate and a single judgment of responsiveness. In the case of the Abbottabad debate, this would have led to a very different way of expressing uncertainty about bin Laden's location. Rather than being confronted with a range of predictions, the president would have received a clear assessment of the odds that bin Laden was living in Abbottabad in a manner that logically reflected his advisors' opinions. Rather than having to disentangle the concepts of likelihood and confidence from the same set of figures, the president would have had separate and explicit characterizations of each. And when analysts articulated ambiguity, they would have done so in a manner that more directly helped the president to determine whether to act immediately or to wait for more information.

If followed, these recommendations would not add undue complexity to intelligence analysis. The main departures from existing practice are less about how to conduct intelligence analysis than about how to express and interpret that analysis in ways that are clear, structured, and directly useful for decision makers. The most difficult part of this process – judging analysts' relative credibility – is an issue that was already front-and-center in the Abbottabad debate, and which the president was forced to address himself, even though analysts were almost surely in a better position to do this.

Having laid out these arguments, it may be useful to close by presenting likely critiques and brief responses. This article does not seek the final word on long-standing debates about expressing uncertainty in intelligence, but rather to indicate how those debates might be extended. In that spirit, here are four lines of questioning that might apply to the ideas presented here.

1. Intelligence analysts have to take decision makers' irrational impulses into account. If decision makers are likely to interpret point estimates as being overly scientific, then is it not important to hedge them, even if this is purely a matter of semantics?

This question builds from the valid premise that most decision makers suffer biases, and analysts in any profession must understand how to convey information in light of this reality. At the same time, it is possible to overstate the drawbacks of expressing estimative probabilities directly, to ignore the way that current methods also encourage irrational responses, and to underestimate the possibility that analysts and decision makers could adjust to new tradecraft.

Saying that analysts should present decision makers with a point estimate is not the same as saying that they should ignore ambiguity. In fact, this article advocated for expressing ambiguity explicitly, separating assessments of likelihood from assessments of confidence, and articulating the latter in a fashion that directly addresses decision makers' concerns about whether or not it is worth acting now on current intelligence. By all accounts, such articulation was absent during debates about Abbottabad. This fact, among others, led President Obama to profess both confusion and frustration in trying to interpret the information his advisors presented.

2. If analysts know that their estimates will be combined into weighted averages, doesn't this give them an incentive to exaggerate their predictions in order to sway the group product?

No process can eliminate the incentive to exaggerate when analysts disagree, though analysts generally pride themselves on professionalism and, if they detect any strategic behavior from their colleagues, this can (and should) affect assessments of credibility. It is also important once again to recognize that both motivated and unmotivated biases already influence the production of estimative intelligence. In the Abbottabad debate, for instance, we know that the CIA team leader gave an extremely high estimate of the chances that bin Laden had been found. That may have been a strategic move intended to sway the president into acting, with the expectation that others would be giving much lower estimates. Also, the 'Red Team' gave a low-end estimate based on explicit instructions to be skeptical. Decision makers already have to counterbalance these kinds of biases when they sort through conflicting predictions. This article argued for incorporating credibility assessments directly into the analytic process so as to mitigate this problem.

The concepts advanced here may also help to reduce another problematic kind of strategic behavior, which involves intentionally hedging estimates in order to avoid accountability for mistaken projections. In this view, the broader the range of estimative probabilities that analysts assert, the less of a chance there is for predictions to appear mistaken. There is debate as to whether and how extensively analysts actually do this, and an empirical assessment is well beyond the scope of this article. Yet one of this article's

main points is that there is no logic for why smudging estimative probabilities should avoid criticism or blame if decision makers and intelligence officials know how to interpret probabilistic language properly. Even if analysts avoid explicitly stating a point estimate, they cannot avoid implying one. Thus, the strategic use of hedging to avoid accountability is at best a semantic trick.

3. Is providing a point estimate not the same as single-outcome forecasting? If a core goal of intelligence analysis is to present decision makers with a range of possibilities, then wouldn't providing a single point estimate be counterproductive?

The answer to both questions is 'no', because there is an important distinction between predicting outcomes and assessing probabilities. When it comes to discussing different scenarios – such as what the current state of Iran's nuclear program might be, or who might win Afghanistan's next presidential election, or what the results of a prospective military intervention overseas might entail – analysts should articulate a wide range of possible 'states of the world', and should then assess each on its own terms.⁴⁴ By this logic, it would be inappropriate to condense a range of possible outcomes (say, the number of nuclear weapons another state might possess, or the number of casualties that might result from some military operation) into a single point estimate.

It is only when assessing the *probabilities* attached to different outcomes that different point estimates should be combined into a single expected value. As this article demonstrated, a decision maker should be no more or less likely to take a gamble with a known risk than to take a gamble where the probabilities are ambiguous, all else being equal. To the extent that some people are 'ambiguity averse' such that they *do* treat these gambles differently in practice, they are falling prey to a common behavioral fallacy that the ideas in this article may help to combat.

4. Most people seem to think that the president made the right decision regarding the 2011 raid on Abbottabad. Should we not see this as an episode worth emulating, rather than as a case study that motivates changing established tradecraft?

This critique helps to emphasize that while the Abbottabad debate was in many ways confusing and frustrating (as the president himself observed), those problems did not ultimately prevent decision makers from achieving what most US citizens saw as a favorable outcome. The arguments made here are thus not motivated by the same sense of emergency that spawned a broad literature on perceived intelligence failures in the wake of the 9/11 terrorist attacks, or mistaken assessments of Iraqi weapons of mass destruction.

Yet the outcome of the Abbottabad debate by no means implies that the process was sound, just as it is a mistake to conclude that if a decision turns

⁴⁴For broader discussions of this point, see Willis C. Armstrong et al., 'The Hazards of Single-Outcome Forecasting', *Studies in Intelligence* 28/3 (1984) pp.57–70; and Friedman and Zeckhauser, 'Assessing Uncertainty in Intelligence', pp.829–34.

out poorly, then this alone necessitates blame or reform.⁴⁵ Sometimes bad decisions work out well, while even the most logically rigorous methods of analysis will fall short much of the time. It is thus often useful to focus on the conceptual foundations of intelligence tradecraft in their own right, and the discussions of bin Laden's location suggest important flaws in the way that analysts and decision makers deal with estimative probability. The president is the intelligence community's most important client. If he struggles to make sense of the way that analysts express uncertainty, this suggests an area where scholars and practitioners should seek improvement.

Whether and how extensively to implement the concepts addressed in this article is ultimately a question of cost-benefit analysis. Analysts could be trained to recognize and overcome ambiguity aversion and to understand that a range of probabilities always implies a single point estimate. Expressing confidence by articulating responsiveness is a technique that can be taught and incorporated into analytic standards. Doing this would require training in advance and effort at the time, but these ideas are fairly straightforward, and the methods recommended here could offer real advantages for expressing uncertainty in a manner that facilitates decision making.

After all, similar situations are sure to arise in the future. As President Obama stated in an interview about the Abbottabad debate: 'One of the things you learn as president is that you're always dealing with probabilities'.⁴⁶ Improving concepts for expressing and interpreting uncertainty will raise the likelihood that intelligence analysts and decision makers successfully meet this challenge.

Acknowledgements

For helpful comments, the authors thank Carol Atkinson, Welton Chang, Miron Varouhakis, S. C. Zidek, participants at the 2013 Midwest Political Science Association's annual conference, and four anonymous reviewers.

⁴⁵Ronald Howard argues that this is perhaps *the* fundamental takeaway from decision theory: 'I tell my students that if they learn nothing else about decision analysis from their studies, this distinction will have been worth the price of admission. A good outcome is a future state of the world that we prize relative to other possibilities. A good decision is an action we take that is logically consistent with the alternatives we perceive, the information we have, and the preferences we feel. In an uncertain world, good decisions can lead to bad outcomes, and vice versa. If you listen carefully to ordinary speech, you will see that this distinction is usually not observed. If a bad outcome follows an action, people say that they made a bad decision. Making the distinction allows us to separate action from consequence and hence improve the quality of action.' Ronald Howard, 'Decision Analysis: Practice and Promise', *Management Science* 34/6 (1988) p.682. On the danger of pursuing intelligence reform without drawing this distinction, see Richard K. Betts, *Enemies of Intelligence: Knowledge and Power in American National Security* (NY: Columbia University Press 2007) and Paul D. Pillar, *Intelligence and US Foreign Policy: Iraq, 9/11, and Misguided Reform* (NY: Columbia University Press 2011), among others.

⁴⁶Bowden, *The Finish*, p.161.

Notes on Contributors

Jeffrey A. Friedman is a postdoctoral fellow at Dartmouth's Dickey Center for International Understanding.

Richard Zeckhauser is Frank P. Ramsey Professor of Political Economy at the Harvard Kennedy School.