1

Pathologies of Probability Assessment

This chapter describes how foreign policy officials often avoid assessing uncertainty in a manner that supports sound decision making. This problem extends beyond the fact that foreign policy analysts are generally unwilling to articulate their judgments precisely. Instead, we will see how foreign policy officials often appear reluctant to assess basic elements of uncertainty at all. For example, the chapter shows that foreign policy officials frequently orient debates around determining what policies offer the best prospects for success or what actions are necessary to achieve strategic objectives, without analyzing the chances that these measures will actually work. This behavior is ingrained throughout foreign policy discourse, from day-to-day intelligence reports to the highest levels of national security decision making.

The chapter describes this problem in two ways. First, it reviews official guidelines for assessing uncertainty in the U.S. military, Intelligence Community, and other national security agencies. These guidelines demonstrate a widespread aversion to probabilistic reasoning. The chapter then presents a case study that describes how U.S. officials debated their strategic prospects during the Vietnam War between 1961 and 1965. Through qualitative and quantitative analysis of key documents from this period, we will see that senior leaders consistently avoided assessing the uncertainty that surrounded their decisions to escalate the war.

The Vietnam case holds special significance considering that Secretary of Defense Robert McNamara and his staff were some of the most analytically inclined decision makers in the history of U.S. foreign policy. Conventional narratives of Vietnam War decision making portray these so-called Whiz Kids as being overly committed to methodological rigor, especially when it came to conducting the kinds of quantitative, cost-benefit analyses that Secretary McNamara had pioneered as an executive at Ford Motors. This narrative may accurately characterize the immense analytic firepower that the Pentagon devoted to measuring tactical progress during the war. Yet we will also see that

100% Certainty		
The General Area of Possibility		
93%	give or take about 6%	Almost certain
75%	give or take about 12%	Probable
50%	give or take about 10%	Chances about even
30%	give or take about 10%	Probably not
7%	give or take about 5%	Almost certainly not
0% Impossibility		

Figure 1.1 Sherman Kent's odds table. Source: Kent, "Words of Estimative Probability."

U.S. officials put virtually no systematic effort into assessing the chances that their overall strategy would succeed.

Seen from this vantage point, the McNamara Pentagon did not suffer from an excess of methodological rigor when guiding the Vietnam War, but rather from an almost complete absence of attention to the logical foundations of their most important choices. The fact that even the Whiz Kids could not bring themselves to address these matters in clear and structured ways highlights a basic divergence between the theory and the practice of foreign policy decision making. Although virtually every theory of decision making in international politics depends on assessing uncertainty in one form or another, this chapter shows that foreign policy officials can place lives and resources at risk without carefully evaluating what those risks entail.

"Words of Estimative Probability"

In 1950, an intelligence official named Sherman Kent supervised the production of a report that warned of a "serious possibility" that the Soviet Union would invade Yugoslavia within a year. Kent later asked his analysts to explain how they interpreted the phrase "a serious possibility." Their answers ranged from twenty percent to eighty percent. Surely, Kent reasoned, intelligence reports meant little if their own authors disagreed so strongly about what their judgments meant. Kent described this experience in a 1964 essay, titled "Words of Estimative Probability," in which he recommended that intelligence analysts articulate assessments of uncertainty using the odds table shown in Figure 1.1.[1]

[1] Sherman Kent, "Words of Estimative Probability," *Studies in Intelligence*, Vol. 8, No. 4 (1964), pp. 49–65. Kent defined the numeric ranges in the odds table by surveying how foreign policy officials intuitively quantified probabilistic phrases. This method of eliciting "inferred probabilities" is now

Sherman Kent was one of the most influential figures in the early years of the United States Intelligence Community. A former history professor at Yale, Kent became the director of the Board of National Estimates, an organization similar to today's National Intelligence Council. Kent is known as the "founding father of intelligence estimation" in the United States, and the Central Intelligence Agency training center is named in his honor. But despite his position of authority, Kent could not convince his colleagues to use the odds table or to clarify their assessments of uncertainty in any other systematic way.

In his essay, Kent described receiving pushback from a group of analysts he called the "poets" (in contrast to the "mathematicians," who supported making assessments of uncertainty more precise). Kent wrote that the poets "appear to believe the most a writer can achieve when working in a speculative area of human affairs is communication in only the broadest general sense. If he gets the wrong message across or no message at all—well, that is life." To the extent that the poets saw the odds table as reflecting "bogus precision," Kent argued, it represented a "fundamentally defeatist" attitude. The poets nevertheless retained the upper hand in this debate for four decades, defeating subsequent attempts by other high-ranking officials, including National Intelligence Council chairman Joseph Nye, to clarify assessments of uncertainty in intelligence.[2]

The poets were ultimately forced to cede ground following mistaken assessments of Saddam Hussein's weapons of mass destruction (WMD) programs, which President George W. Bush used to justify the 2003 invasion of Iraq. A National Intelligence Estimate (NIE) on this subject opened by stating, "We judge that Iraq has continued its weapons of mass destruction programs in defiance of UN resolutions and restrictions."[3] Phrases like "we judge" are what

a standard tool in the decision sciences. See Ruth Beyth-Marom, "How Probable Is Probable?" *Journal of Forecasting*, Vol. 1 (1982), pp. 257–269; Frederick Mosteller and Cleo Youtz, "Quantifying Probabilistic Expressions," *Statistical Science*, Vol. 5, No. 1 (1990), pp. 2–12; and David R. Mandel, "Accuracy of Intelligence Forecasts from the Intelligence Consumer's Perspective," *Policy Insights from the Behavioral and Brain Sciences*, Vol. 2, No. 1 (2015), pp. 111–120.

[2] Joseph S. Nye, Jr., "Peering into the Future," *Foreign Affairs*, Vol. 73, No. 4 (1994), pp. 82–93. Reviewing a corpus of 379 NIEs written between 1964 and 1994, Richard Zeckhauser and I found that only 4 percent of key judgments assessed uncertainty using quantitative indicators, whereas 18 percent of key judgments did not convey even a qualitative sense of the chances that a statement was true. Jeffrey A. Friedman and Richard Zeckhauser, "Assessing Uncertainty in Intelligence," *Intelligence and National Security*, Vol. 27, No. 6 (2012), p. 837. For similar descriptions of the aversion to transparent probabilistic reasoning in Israeli and Canadian intelligence, respectively, see Zvi Lanir and Daniel Kahneman, "An Experiment in Decision Analysis in Israel in 1975," *Studies in Intelligence*, Vol. 50, No. 4 (2006), pp. 11–19; and Alan Barnes, "Making Intelligence Analysis More Intelligent: Using Numeric Probabilities," *Intelligence and National Security*, Vol. 31, No. 1 (2016), pp. 327–344.

[3] NIE 2002-16HC, *Iraq's Continuing Programs for Weapons of Mass Destruction* (Oct. 2002).

intelligence analysts call "estimative verbs." In principle, estimative verbs are meant to emphasize the presence of uncertainty. A statement that analysts *judge* to be true is thus distinct from a statement that analysts *know* to be true. Yet most people read the NIE's key judgments as though they conveyed facts, not inferences. It was widely believed that intelligence analysts had presented the evidence on Iraq's weapons programs as a "slam dunk," and almost every postmortem analysis of the episode emphasized that the Iraq NIE was unclear on this point.[4]

This lack of clarity ran much deeper than confusing semantics. Several prominent postmortem analyses of the Iraq NIE argued that the document's lack of linguistic clarity reflected a more fundamental failure to grapple with the uncertainty surrounding key judgments. For instance, Robert Jervis describes how intelligence analysts' perceptions in this case were heavily conditioned by assumptions drawn from Saddam Hussein's intentions and past behavior. Because these assumptions seemed reasonable and were widely shared, it was easy to overlook how heavily they relied on circumstantial evidence. Jervis thus concludes that "the central analytic error [with the Iraq NIE] was not that inferences were driven by their plausibility [instead of direct evidence] . . . but that the analysts did not make this clear and probably did not even understand it."[5] Former deputy director of intelligence Michael Morell has similarly argued that "By far the biggest mistake made by the analysts . . . was not that they came to the wrong conclusion about Iraq's WMD program, but rather that they did not rigorously ask themselves how confident they were in their judgments."[6] Morell specifically singled out a line in the NIE that assigned "high confidence" to the judgment that Saddam Hussein was developing chemical, biological, and nuclear weapons, explaining that

> the analysts did not really think about that statement before making it. It was a reflection of their gut view. It did not reflect a thorough

[4] See, for example, Michael Schrage, "What Percent Is Slam Dunk?" *Washington Post*, February 20, 2005, p. B01. Chapter 5 will return to this episode, describing how the "slam dunk" term was mistakenly (but foreseeably) used to characterize the Iraq NIE. Richard Betts similarly argues that the principal error in intelligence on Iraqi weapons programs was a "failure to make clear how weak the direct evidence was for reaching any conclusion." Richard K. Betts, *Enemies of Intelligence: Knowledge and Power in American National Security* (New York: Columbia University Press, 2006), p. 116.

[5] Robert Jervis, "Reports, Politics, and Intelligence Failures: The Case of Iraq," *Journal of Strategic Studies*, Vol. 29, No. 1 (2006), p. 44.

[6] Michael Morell, *The Great War of Our Time: The CIA's Fight against Terrorism from Al Qa'ida to ISIS* (New York: Twelve, 2014), p. 102.

> assessment of the question of confidence levels. Such a rigorous assessment was missing. It was simply not part of an analyst's toolkit in those days.[7]

The Iraq WMD controversy provoked widespread pressure to improve analytic standards in intelligence reporting. Congress even translated this pressure into law, requiring intelligence analysts to "properly caveat and express uncertainties" in their published work.[8] Yet because Congress did not explain what "properly" assessing uncertainty meant, intelligence agencies responded by creating idiosyncratic standards as opposed to common doctrine. Figure 1.2 presents four examples.[9] Nodding to Sherman Kent, scholars call these expressions "words of estimative probability," though it is worth noting that this phrase is usually a misnomer, as most such spectrums do not include the kinds of explicit definitions that Kent had originally recommended.

While these guidelines all possess important advantages over unstructured communication, they also reveal how intelligence agencies remain reluctant to provide policymakers with clear assessments of uncertainty. Indeed, these guidelines show that intelligence agencies are willing to take special steps to avoid transparent probabilistic reasoning. The Defense Intelligence Agency's doctrine is especially outspoken on this matter, stating—emphasis in the original—that "*DIA does not condone the use of probability percentages in its products to portray likelihood.*"[10] But each of the guidelines shown in Figure 1.2 conveys a similar sentiment. To apply these guidelines properly, analysts must determine where their probability estimates fall along the number line, after which analysts are instructed to coarsen their beliefs into broader categories, most of which lack clear definitions.

The application of these analytic standards also remains inconsistent. For example, the guidelines shown at the bottom of Figure 1.2 appear in an Intelligence

[7] Ibid. Note that in this and the previous quotation, Morell uses the term "confidence" as a synonym for probability. The Iraq NIE conflates these concepts in similar ways—or, at least, never draws a clear distinction between the chances that a statement is true and the basis on which analysts have drawn those conclusions. Chapter 2 returns to the distinction between probability and confidence.

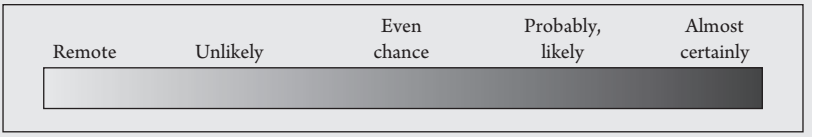
[8] 2004 Intelligence Reform and Terrorism Prevention Act, Section 1019(b)(2)(A).

[9] The Defense Intelligence Agency guidelines in Figure 1.1 have been redrawn for clarity; each of the other guidelines in Figure 1.2 is a facsimile based on original documents.

[10] Defense Intelligence Agency Tradecraft Note 01-15, *Expressing Analytic Certainty* (January 2015), p. 1. The Defense Intelligence Agency (DIA) guidelines contain other peculiarities. For example, they do not give analysts any clear option for communicating probability estimates that are near fifty percent. Terms in the middle of the DIA's spectrum, such as "undetermined" and "perhaps," suggest lack of knowledge as opposed to considered judgments, and the guidelines state that these phrases should only be used to reflect "purely exploratory analysis."

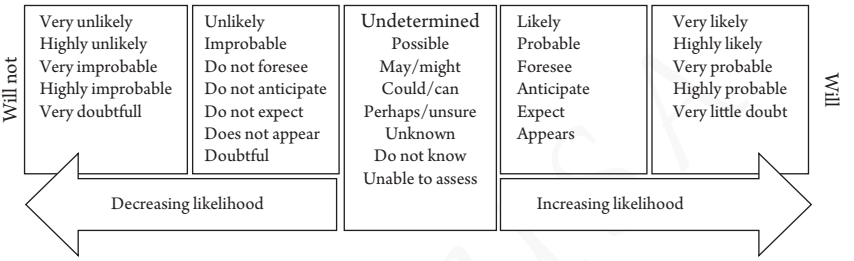
From the 2007 National Intelligence Estimate, "Prospects for Iraq's Stability":

Intelligence judgments pertaining to likelihood are intended to reflect the community's sense of the probability of a development or event. Assigning precise numerical ratings to such judgments would imply more rigor than we intend. The chart below provides a rough idea of the relationship of terms to each other.



From the Defense Intelligence Agency's 2015 Tradecraft Standards:

Likelihood expresses the probability that an event or development will **or** will not happen.



From the 2015 Intelligence Community Directive on "Analytic Standards":

(a) For expressions of likelihood or probability, an analytic product must use one of the following sets of terms:

almost no chance	very unlikely	unlikely	roughly even chance	likely	very likely	almost certain(ly)
remote	highly improbable	improbable (improbably)	roughly even odds	probable (probably)	highly probable	nearly certain
01–05%	05–20%	20–45%	45–55%	55–80%	80–95%	95–99%

From the 2017 National Intelligence Community Assessment on Russian Hacking:

Percent

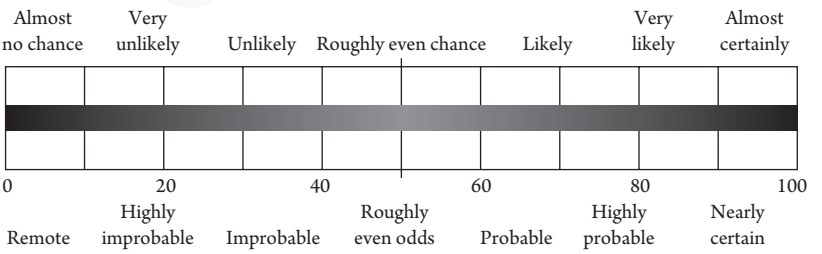


Figure 1.2 Words of estimative probability in intelligence analysis.

Community Assessment describing Russia's attempts to influence the 2016 U.S. presidential election. These guidelines indicate that intelligence analysts should communicate probabilistic judgments using fourteen "words of estimative probability" grouped into seven segments along the number line. One could argue about whether or not this is the best way to communicate important assessments of uncertainty. But that argument would be largely irrelevant in this case, given that the report's key judgments did not actually use those terms to convey key judgments. Instead, the report described uncertainty about Russian interference in the 2016 election using estimative verbs and confidence levels, as in the following statements:

> *We assess* Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential election. Russia's goals were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency. *We further assess* Putin and the Russian Government developed a clear preference for President-elect Trump. *We have high confidence in these judgments.*
>
> *We also assess* Putin and the Russian Government aspired to help President-elect Trump's election chances when possible by discrediting Secretary Clinton and publicly contrasting her unfavorably to him. All three agencies agree with this judgment. *CIA and FBI have high confidence in this judgment; NSA has moderate confidence.*[11]

It is hard to say exactly what these judgments mean. One plausible interpretation is that analysts conflated assessments of probability (which reflect the chances that a statement is true) with assessments of confidence (which reflect the extent to which analysts believe they have a sound basis for drawing conclusions).[12] This would raise further questions about what probabilities a "high confidence" or "moderate confidence" judgment might reflect. If analysts *did* intend to distinguish between probability and confidence, then their conclusions would be even harder to interpret, because the only other probabilistic term these judgments use is the estimative verb, "we assess." Either way, these judgments are not consistent with the analytic standards for assessing

[11] ICA 2017-01D, *Assessing Russian Activities and Intentions in Recent U.S. Elections* (January 2017). Emphasis added. See https://www.dni.gov/files/documents/ICA_2017_01.pdf.

[12] Chapter 2 expands on this conceptual distinction. Friedman and Zeckhauser, "Assessing Uncertainty in Intelligence," pp. 834–841, show that intelligence analysts appear to regularly conflate the language of probability and the language of confidence, particularly when they are dealing with uncertainty about questions to which the answers are knowable. Intelligence analysts are thus more likely to use the language of confidence to describe uncertainty about events that have occurred in the past; they are more likely to use the language of probability to describe uncertainty about events that will occur in the future; and they rarely employ both concepts when offering a single judgment.

RISK ASSESSMENT MATRIX						
		Probability				
Severity		Frequent A	Likely B	Occasional C	Seldom D	Unlikely E
Catastrophic	I	E	E	H	H	M
Critical	II	E	H	H	M	L
Marginal	III	H	M	M	L	L
Negligible	IV	M	L	L	L	L
E - Extremely High		H-High		M - Moderate		L- Low

Assess Each Hazard on the Probability of the Event or Occurrence

1-23. Probability is the likelihood of an event. This is your estimate, given what information you know and what others have experienced. The probability levels estimated for each hazard are based on the mission, COA, or frequency of a similar event. For the purpose of CRM, there are five levels of probability—frequent, likely, occasional, seldom, and unlikely:

- Frequent – Occurs very often, known to happen regularly. In illustration, given 500 or so exposures to the hazard, expect that it will definitely happen to someone. Examples of frequent occurrences are vehicle rollovers, rear-end collisions, and heat injury during a battalion physical training run with hot weather or nonacclimated Soldiers.
- Likely – Occurs several times, a common occurrence. In illustration, given 1000 or so exposures without proper controls, it will occur at some point. Examples might include improvised explosive devices (IEDs), wire strikes for aircraft, controlled flights into terrain, and unintentional weapons discharges.
- Occasional – Occurs sporadically, but is not uncommon. You may or may not get through your deployment without it happening. Some examples might include unexploded ordnance (UXO) and fratricide.
- Seldom - Remotely possible, could occur at some time. Usually several things must go wrong for it to happen. Examples might include things like heat-related death or electrocution.
- Unlikely – Can assume will not occur, but not impossible. Examples might include detonation of containerized ammunition during transport.

Figure 1.3 U.S. Army guidelines for risk assessment. *Source:* U.S. Army Field Manual 5-19, *Composite Risk Management*.

uncertainty that the document itself put forth. This discrepancy highlights how intelligence analysts remain ambivalent about the importance of assessing uncertainty in clear and structured ways, despite decades of effort spent promoting that goal.

Words of Estimative Probability Outside Intelligence Studies

Official guidelines for assessing uncertainty are most commonly debated in intelligence studies, but similar controversies appear in other areas of foreign policy analysis. For example, Figure 1.3 presents current U.S. Army doctrine on

risk assessment. This doctrine instructs planners to describe probability using five terms. These terms superficially resemble the words of estimative probability spectrums shown in Figure 1.2, but they convey very different meanings. For example, the term *frequent* describes a hazard that "occurs very often, [and is] known to happen regularly. In illustration, given 500 exposures to the hazard, expect that it will definitely happen to someone." The word "likely" means that a hazard "will occur at some point" given "1000 or so exposures without proper controls."[13]

Even though the guidelines shown in Figure 1.3 define key terms using numbers, the resulting definitions remain extremely vague. What does it mean to "expect that something will definitely happen"? Does an "exposure" refer to a single action, so that we would need to take that action 1,000 times before a "likely" risk would occur? Or is every person exposed to the risk separately, so that the risk would occur if a unit with 1,000 people took that action once? And why is a "likely" risk only defined with respect to actions undertaken "without proper controls"? How would planners express a risk that would occur one time in 1,000 trials *with* proper controls?

Figure 1.4 presents a different set of guidelines, used by the Joint Chiefs of Staff to assess the risks surrounding military strategies.[14] This scale involves four categories of risk: low, moderate, significant, and high. Each designation reflects a complex array of vaguely defined factors. Moreover, none of the terms allows the Joint Chiefs to indicate that a military strategy is unlikely to succeed—the category of "high risk" indicates only that achieving desired objectives "requires extraordinary measures." Later in the chapter, we will see how U.S. officials sent combat forces to Vietnam thinking that, even if this strategy was unlikely to work, its chances of success were still large enough to justify a high-stakes gamble. As of this writing, the U.S. military is committed to "destroying" the Islamic State, an outcome that most senior leaders presumably see as unlikely, at least in the short run. The Joint Chiefs of Staff's risk assessment system makes it impossible to describe that logic, let alone to grapple in a meaningful way with the uncertainty it entails.

Figure 1.5 presents the analytic standards that U.S. military interrogators use when describing the reliability of human sources.[15] Whenever foreign policy analysts consider information from a human source, it is crucial to assess the probability that the source could be lying. One of the principal critiques of

[13] U.S. Army Field Manual 5-19, *Composite Risk Management* (2006), paras. 1–23. U.S. Army doctrine can be found online at https://usacac.army.mil/cac2/Doctrine.asp.

[14] "CJCS Joint Risk Assessment System" (2004), accessed October 5, 2016, http://dde.carlisle.army.mil/LLL/DSC/ppt/L14_CRA.pdf.

[15] U.S. Army Field Manual 2-22.3, *Human Intelligence Collector Operations* (2006), p. B-1.

	Low	Moderate	Significant	High
Strategic Objectives	Strategic objective: Assured	Strategic objective: Very likely	Strategic objective: Likely	Strategic objective: Requires extraordinary measures
Authorities	Full authorities provided to achieve all strategic objectives	Authorities provided to achieve most strategic objectives	Authorities are insufficient to achieve key strategic objectives	Critical authorities are not provided; ability to achieve strategic objectives is compromised
Plans	SCG/CPG/JSCP direct advanced planning: OPLANS (Level IV) or CONPLANS (Level III)	SCG/CPG/JSCP direct preliminary planning: Base Plans (Level II) or CDR's Estimate (Level I)	SCG/CPG/JSCP do not direct planning but local plans exist or are being developed	SCG/CPG/ JSCP do not direct plans and planning is not in progress
Resources	As planned	Additional resources from other plans and operations	Additional resources from other plans & operations; some significant capability shortfalls	Significant resources from other operations; some resources severely deficient or absent altogether
Resources: Timelines	As planned	Extended	Significant adjustments	Significant adjustments; may not achieve desired end-states
Resources: Unanticipated Requirements	Easily managed, minimal impact	May necessitate adjustments to plans	Will necessitate significant adjustments to plans	Unable to manage
Resources: Force Provider	Full capacity to source COCOM rqmts	Can source all rqmts. Worldwide force allocation solutions may result in limited duration capability gaps	Can source priority COCOM rqmts. Worldwide force allocation solutions may result in extended duration capability gaps	Require full mobilization to sustain sourcing solutions to achieve strategic objectives
Resources: Services Functions, Force Management, Institutional Capacity	Full capacity to source COCOM rqmts	Requires intra-Service adjustments to source COCOM rqmts	Requires joint source solutions and force substitutions to source COCOM rqmts	COCOM rqmts exceed Joint Force capacity to substitute capabilities

Figure 1.4 Joint Chiefs of Staff risk assessment system.

A	**Reliable**	**No doubt** of authenticity, trustworthiness, or competency; has a history of complete reliability
B	**Usually Reliable**	**Minor doubt** about authenticity, trustworthiness, or competency; has a history of valid information most of the time
C	**Fairly Reliable**	**Doubt** of authenticity, trustworthiness, or competency but has provided valid information in the past
D	**Not Usually Reliable**	**Significant doubt** about authenticity, trustworthiness, or competency but has provided valid information in the past
E	**Unreliable**	**Lacking** in authenticity, trustworthiness, and competency; history of invalid information
F	**Cannot Be Judged**	**No basis** exists for evaluating the reliability of the source

Figure 1.5 U. S. Army guidelines for source reliability assessment. *Source:* U.S. Army Field Manual 2-22.3, Human Intelligence Collector Operations.

prewar intelligence on Iraq, for instance, was that analysts should have been more skeptical of the source known as "Curveball," who had fabricated information about biological weapons. Even though several intelligence officials doubted Curveball's reliability, that skepticism was not clearly conveyed in reports that drew on Curveball's interviews.[16]

To help avoid such problems, collectors of human intelligence generally attach source reliability assessments to their reports. The source assessment doctrine shown in Figure 1.5 encourages analysts to communicate these evaluations using six qualitative terms. These assessments of reliability are defined in relation to another set of qualitative terms describing doubt. The doctrine then provides no indication of how either reliability or doubt relates to the probability that a source is telling the truth.

The guidelines described in this section reflect just a handful of relevant examples of how intelligence agencies and military organizations encourage analysts to leave probability assessments vague.[17] Nevertheless, the guidelines

[16] Senate Select Committee on Intelligence, *Report on the U.S. Intelligence Community's Prewar Intelligence Assessments on Iraq* (2004), ch. IV.

[17] This behavior is not limited to the United States. The current "NATO Common Standard" for assessing probability in intelligence analysis involves five terms: *highly unlikely* (less than 10 percent), *unlikely* (10 to 40 percent), *even chance* (40 to 60 percent), *likely* (60 to 90 percent), and *highly likely* (more than 90 percent). NATO Standard AJP 2.1, *Allied Joint Doctrine for Intelligence Procedures*, pp. 3–15. For NATO standards on source reliability assessment that are similar to the doctrine shown in Figure 1.5, see Dan Irwin and David Mandel, "Standards for Evaluating Source Reliability

described in this section cover a vast range of consequential analyses. In principle, they shape every key judgment in every intelligence report, every risk assessment in every Army strategy and plan, and every source assessment provided by military interrogators.

Chapters 3 and 4 will show that this imprecision discards information that can improve the accuracy of foreign policy analysis and the effectiveness of foreign policy decision making. For the moment, however, the purpose of describing these standards is simply to demonstrate the lengths to which foreign policy agencies are willing to go to in order to avoid transparent probabilistic reasoning: accepting the risk of miscommunication, creating divergent definitions of key words, and requiring analysts and decision makers to memorize a bewildering array of phrases—the analytic standards described in this section alone contain more than eighty distinct terms. Yet the chapter's next section explains why imprecision is actually the least concerning form of vagueness that surrounds assessments of uncertainty in foreign policy discourse. Two other problems, which I call *relative probability* and *conditioning*, pose even greater obstacles to making sound decisions.

Relative Probability and Conditioning

As the Taliban gained ground throughout Afghanistan in 2009, President Obama appointed General Stanley McChrystal to command allied forces in the country. Two months later, General McChrystal submitted an assessment in which he recommended a new counterinsurgency strategy to be bolstered by forty thousand additional troops. McChrystal argued that his proposal would "improve effectiveness" and that that it offered "the best prospect for success in this important mission."[18] When the document leaked to the press, it helped to build political momentum behind the policy known as the Afghan Surge.[19]

and Information Credibility in Intelligence Production," in David Mandel, ed., *Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making* (Brussels, Belgium: NATO Science and Technology Organization, in press). For additional examples drawn from Britain, Canada, and Israel, see Mandeep K. Dhami, *Understanding and Communicating Uncertainty in Intelligence Analysis* (London, U.K.: Report Prepared for H.M. Government, 2013); Barnes, "Making Intelligence Analysis More Intelligent"; and Lanir and Kahneman, "Experiment in Decision Analysis in Israel."

[18] Commander, International Security Assistance Force-Afghanistan, *COMISAF Initial Assessment*, August 2009, pp. 1–3, 1–4, and 2–22. See http://www.washingtonpost.com/wp-dyn/content/article/2009/09/21/AR2009092100110.html.

[19] Kevin Marsh, "Obama's Surge: A Bureaucratic Politics Analysis of the Decision to Order a Troop Surge in the Afghanistan War," *Foreign Policy Analysis*, Vol. 10, No. 3 (2014), pp. 265–288.

General McChrystal's report presented a candid, detailed argument for strategic change. Yet it was essentially trivial to argue that deploying forty thousand additional troops would improve effectiveness. The key question was *how much* those forces would improve effectiveness, and whether their expected benefit justified their expected cost. Similarly, saying that the Afghan Surge represented the best prospect for success did not imply that the strategy was worth pursuing, especially given growing concerns about whether the United States had the resources and political will to defeat the Taliban. Logically speaking, there was no way to evaluate General McChrystal's proposal without estimating the extent to which his recommendations would raise the chances of strategic success. Yet the proposal itself left that judgment to the reader.

McChrystal's report exemplifies the use of *relative probability*: a probability assessment that conveys an analyst's beliefs with respect to an unspecified baseline. Relative probability provides limited information, and it can bias assessments of expected value. Pharmaceutical companies, for example, often encourage customers to purchase drugs that reduce risks even if those risks are already vanishingly small.[20] Regardless of the extent to which foreign policy officials consciously seek to fool their colleagues in this manner, it is reasonable to expect their use of relative probability to produce a similar result. Decision makers may thus find it easier to accept a strategy billed as their best chance of success over a strategy whose odds of succeeding are roughly one in five, even if those strategies are the same.

Of course, there is no reason why decision makers should not seek to minimize risks or to maximize their chances of success, so long as they are cognizant of what those risks and chances entail. Yet the analytic standards that guide military planning often revolve exclusively around relational judgments. U.S. Army operations doctrine thus states that when commanders evaluate potential courses of action, they should define the problem, gather information, develop possible solutions, and "select the best solution."[21] To do this,

> the staff compares feasible COAs [courses of action] to identify the one with the *highest probability* of success . . . . The selected COA should also pose the *minimum risk* to the force and mission accomplishment; place the force in the *best posture* for future operations; provide *maximum*

[20] Jonathan Baron, "Confusion of Relative and Absolute Risk in Valuation," *Journal of Risk and Uncertainty*, Vol. 14 (1997), pp. 301–309; and Gerd Gigerenzer, *Calculated Risks: How to Know When Numbers Deceive You* (New York: Simon and Schuster, 2002). On how governments likewise spend inordinate resources reducing risks that are already very small, see Cass R. Sunstein, *Worst-Case Scenarios* (Cambridge, Mass.: Harvard University Press, 2007); and John Mueller and Mark G. Stewart, *Terror, Security, and Money: Balancing the Risks, Benefits, and Costs of Homeland Security* (New York: Oxford University Press, 2011).

[21] Emphasis added. U.S. Army Field Manual 5-0, *The Operations Process* (2009), para. 2–19.

> *latitude* for initiative by subordinates; provide the *most flexibility* to meet unexpected threats and opportunities; provide the *most secure* and stable environment for civilians in the AO; [and] *best facilitate* initial information themes and messages.[22]

Army doctrine on staff organization similarly states that planners should "compare feasible courses of action to identify the one that has the highest probability of success."[23] Army doctrine on tactical intelligence recommends a series of exercises that help to "identify the [course of action] that has the highest probability of success."[24] Official guidelines for strategic assessment instruct commanders to "determine which COA (or combination) hold the greatest promise of success."[25] These analyses demonstrate that General McChrystal's report on the Afghan Surge was not unusual in presenting a recommendation based on minimizing risk and maximizing strategic prospects rather than evaluating what those risks and prospects were. This was, in fact, consistent with the instructions that military commanders are asked to follow when making high-stakes decisions.

Part of what makes this doctrine problematic is that choosing the course of action with the highest probability of success is almost never the correct decision to make. It is usually possible to increase the probability of success by devoting more resources to a task. Yet short of all-out existential warfare, decision makers must at some point determine where additional resource commitments are no longer worth the expense.[26]

One could attempt to skirt this issue by redefining "success" as an outcome for which an action's benefits exceed its costs. Perhaps Army planning doctrine aims to channel that idea by instructing commanders to select the strategy with the highest chances of success among all "feasible" courses of action. But that would not present a valid basis for decision making either. If a policy has a very small chance of becoming a very large disaster, then it might be wise to pursue an alternative course of action that trades a smaller chance of success for less downside risk.[27] The fact of the matter is that it is impossible to judge a policy's

[22] Ibid., para. B-173.

[23] U.S. Army Field Manual 101-5, *Staff Organization and Operations* (1997), para. 5–24.

[24] U.S. Army Field Manual 34-130, *Intelligence Preparation of the Battlefield* (1994), p. A-7.

[25] U.S. Joint Forces Command, *Commander's Handbook for an Effects-Based Approach to Joint Operations* (2006), pp. 3–14.

[26] Alain C. Enthoven and K. Wayne Smith, *How Much Is Enough? Shaping the Defense Program, 1961–1969* (Santa Monica, Calif.: Rand, 2005).

[27] George W. Downs and David M. Rocke, "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War," *American Journal of Political Science*, Vol. 38, No. 2 (1994), pp. 362–380.

expected value without estimating its chances of success in absolute terms. Any heuristic that avoids doing this will generate systematic biases.

The terrorism advisory system of the Department of Homeland Security (DHS) offers another example of how relative probability provides limited and potentially misleading information. Starting in 2003, the system involved five color-coded threat levels: red (severe risk), orange (high risk), yellow (elevated/significant risk), blue (guarded/general risk), and green (low risk). Since none of these terms received a clear definition, the primary way in which DHS used the system to communicate information was by shifting the color-coded threat levels over time. These shifts essentially instructed citizens to adjust their prior assumptions about the risk of terrorism up or down. But since most people already hold dramatically inflated perceptions about what the risk of terrorism entails, raising official threat levels may have only created unnecessary alarm.[28] Faced with mounting critiques, the DHS scrapped this framework in 2011. The department's new National Terrorism Advisory System provides information about specific threats, whose risk levels are classified into three tiers ("elevated," "intermediate," and "imminent"), none of which receives a definition that maps to a discernible segment of the number line.[29]

Relative probability plays a prominent role in intelligence analysis, too. For example, the most widely taught "structured analytic technique" for assessing uncertainty in intelligence is called analysis of competing hypotheses (ACH).[30] ACH instructs analysts to create a matrix in which the columns represent plausible conclusions and the rows represent pieces of available evidence. Analysts populate this matrix by indicating the extent to which each piece of evidence is inconsistent with each hypothesis. When the analysis is complete, the hypothesis with the lowest inconsistency score is considered to be the most likely conclusion.

ACH combats confirmation bias by ensuring that analysts consider alternative judgments and at the same time fosters a falsificationist mindset that prevents jumping to conclusions. These goals are undeniably important. Yet identifying

[28] Rose McDermott and Philip G. Zimbardo, "The Psychological Consequences of Terrorist Alerts," in Bruce Bongar et al., eds., *Psychology of Terrorism* (New York: Oxford University Press, 2007), pp. 357–370; and Jacob N. Shapiro and Dara Kay Cohen, "Color Blind: Lessons from the Failed Homeland Security Advisory System," *International Security*, Vol. 32, No. 2 (Fall 2007), pp. 121–154.

[29] For example, an "elevated" alert "warns of a credible terrorism threat," and an "imminent" alert "warns of a credible, specific, and impending terrorism threat." See Department of Homeland Security, "National Terrorism Advisory System (NTAS)," https://www.dhs.gov/national-terrorism-advisory-system, accessed August 6, 2018.

[30] Richards Heuer and Randolph Pherson, *Structured Analytic Techniques for Intelligence Analysis* (Washington, D.C.: CQ Press, 2010), pp. 160–169.

the most probable conclusion says little about what that probability entails. Just as a strategy that offers the best chances of success might not be worth pursuing, the most likely conclusion is not necessarily worth betting on.[31]

This was indeed the heart of the critiques we encountered earlier regarding intelligence on Iraq's weapons of mass destruction programs. Though the National Intelligence Estimate may have reflected the most plausible interpretation of the available evidence, it was still important to assess the probability that this interpretation was true. Highlighting the presence of this uncertainty could then have provoked important questions about why existing evidence about Iraq's weapons programs could just sustain more conclusive judgments. As Morell describes it,

> The way analysts talked and wrote about their judgments would have led anyone to think that it was a slam dunk—that is, that Saddam definitively had active WMD programs. No one ever said to me, [CIA Deputy Director for Intelligence Jami] Miscik, [Deputy Director of Central Intelligence John] McLaughlin, [CIA Director George] Tenet, [Secretary of State Condoleezza] Rice, or the president, "You know, there is a chance he might not have them." Such a statement would have gotten everyone's attention.[32]

Conditioning

A third way that assessments of uncertainty can be too vague to support sound foreign policy decision involves a practice that I call *conditioning*. This occurs when analysts describe the conditions that must hold in order for a statement to be true. Conditioning is the least informative approach to assessing uncertainty described in this chapter. The logic of conditioning states that *if* a statement is true, *then* an assumption must hold. Note, however, that this logic does not flow in the opposite direction. If the assumption holds, this implies nothing about the probability that the statement is true.

[31] Indeed, there is remarkably little systematic evidence that proves that ACH improves judgmental accuracy. See David R. Mandel, Christopher W. Karvetski, and Mandeep K. Dhami, "Boosting Intelligence Analysts' Judgmental Accuracy: What Works, What Fails?" *Judgment and Decision Making*, forthcoming; Welton Chang and Philip Tetlock, "Rethinking the Training of Intelligence Analysts," *Intelligence and National Security*, Vol. 31, No. 6 (2016), pp. 903–920.

[32] Morell, *Great War of Our Time*, p. 103. Note how this statement implies that the proper way to assess the impact of the Iraq NIE is not to speculate about what would have happened if the NIE had never been published, but rather to explore what might have happened if the NIE had been researched and written in a manner that gave higher priority to the importance of assessing and communicating uncertainty.

One vivid example of conditioning involves recent debates over force sizing for counterinsurgency. In 2006, the U.S. military released a new counterinsurgency doctrine known as Field Manual 3-24 (FM 3-24).[33] Here is the guidance that FM 3-24 offered for sizing forces in counterinsurgency wars:

> Most [troop] density recommendations fall within a range of 20 to 25 counterinsurgents for every 1,000 residents in an AO [area of operations]. Twenty counterinsurgents per 1,000 residents is often considered the minimum troop density required for effective COIN [counterinsurgency] operations; however, as with any fixed ratio, such calculations remain very dependent upon the situation.[34]

At a glance, this paragraph appears to offer clear guidance, saying that if counterinsurgents wish to succeed, then they should deploy at least twenty troops per thousand residents in an area of operations. But even if that claim had been correct,[35] FM 3-24 still said nothing about the chances that a counterinsurgent who met this threshold would achieve its goals. FM 3-24 only claimed that if counterinsurgents deployed twenty troops per thousand inhabitants in an area of operations, then their probability of success might be greater than zero. Even *that* judgment was hedged by the caveat that troop-density thresholds vary across cases.

Taken literally, FM 3-24 thus stated that if counterinsurgents deploy a particular number of forces, then it is possible that success would then be possible. And though that insight clearly fell short of a rigorous basis for sending large numbers of soldiers into harm's way, FM 3-24's guidance on force sizing played a major role in building support for the eventual troop surges in both Iraq and Afghanistan.[36] Part of the problem with these arguments was that their flaws were subtle: on a first read, FM 3-24 appeared to offer clear instructions for conducting successful counterinsurgency operations, even though the doctrine's actual empirical claim was far weaker than that. It is also easy to see how these kinds of judgments can bias policy evaluations. Warning decision makers that they will fail without taking some action surely provides more encouragement for committing resources than saying that this action has a negligible chance of succeeding, even though both statements can be true simultaneously.

[33] The manual's formal title was U.S. Army Field Manual 3-24/Marine Corps Warfighting Publication 3-33.5, *Counterinsurgency*.

[34] Field Manual 3-24, *Counterinsurgency*, para. 1–67.

[35] This is highly dubious: see Jeffrey A. Friedman, "Manpower and Counterinsurgency: Empirical Foundations for Theory and Doctrine," *Security Studies*, Vol. 20, No. 4 (2011), pp. 556–591.

[36] Peter Mansoor, *Surge* (New Haven, Conn.: Yale University Press, 2013), ch. 2.

In principle, one could offer a clear probability assessment *and* identify the conditions that must hold for a statement to be true. In practice, however, many scholars and practitioners argue that foreign policy analysts should limit their focus to the latter of those tasks. Intelligence scholars call this the "linchpins and drivers" approach to foreign policy analysis.[37] The basic idea behind this approach is that analysts should lay out the key assumptions that are required to sustain a conclusion under uncertainty, but then leave it to readers to assess those assumptions' collective plausibility. Some intelligence estimates explicitly define their scope along these lines. For example, the National Intelligence Council's *Global Trends 2025* report opened by explaining, "The study as a whole is more a description of the factors likely to shape events than a prediction of what will actually happen. By examining a small number of variables that we judge probably will have a disproportionate influence on future events and possibilities, the study seeks to help readers to recognize signposts indicating where events are headed." *Global Trends 2030* similarly describes its aim as being to identify "megatrends" and "game-changers" that could potentially lead to "alternative worlds," but not to make predictions about the chances that those worlds will actually materialize.

Yet analysts cannot speak about the future without engaging in some form of probabilistic reasoning. There is little practical value in analyzing events that truly have no chance of taking place. And since analysts cannot devote equal text to describing every global trend or possible future that they can imagine, those judgments depend—however implicitly—on the premise that these statements are sufficiently likely to deserve the reader's attention. Leaving these judgments vague only defers the challenge of assessing uncertainty from analysts to decision makers, who generally have much less time, expertise, and incentive to make accurate judgments.[38]

Relative probability and conditioning are important because they demonstrate how foreign policy officials' aversion to probabilistic reasoning runs much deeper than semantics. The problem is not just that analysts and decision makers are reluctant to assess uncertainty in clear and structured ways. Indeed,

[37] Douglas J. MacEachin, "The Tradecraft of Analysis," in Roy Godson, Ernest R. May, and Gary Schmitt eds., *U.S. Intelligence at the Crossroads: Agendas for Reform* (Washington, D.C.: Brassey's, 1995); Jack Davis, *A Compendium of Analytic Tradecraft Notes* (Washington, D.C.: Central Intelligence Agency, 1997); and Thomas Fingar, *Reducing Uncertainty: Intelligence Analysis and National Security* (Stanford, Calif.: Stanford Security Studies, 2011), pp. 53–74.

[38] As Richard Betts puts it, "The greater the ambiguity, the greater the impact of preconceptions." Richard K. Betts, *Surprise Attack: Lessons for Defense Planning* (Washington, D.C.: Brookings, 1982), p. 103. Joshua Rovner, *Fixing the Facts: National Security and the Politics of Intelligence* (Ithaca, N.Y.: Cornell University Press, 2011), further explains how decision makers can exploit judgmental ambiguity for political ends.

this section has shown how intelligence agencies and military organizations deliberately develop heuristics that allow analysts and decision makers to avoid grappling with crucial probabilities entirely.

It is also important to note that while all of the examples described in this chapter come from official government sources, similar problems recur throughout broader foreign policy discourse. Recent works by Daniel Gardner, Chaim Kaufmann, John Mueller, Nate Silver, Mark Stewart, and Philip Tetlock, among others, document a range of strategies that foreign policy commentators use to appear insightful without actually making falsifiable claims.[39] These strategies include focusing excessive attention on hypothetical scenarios or worst-case outcomes, defending policies based on desirability rather than feasibility, and lacing statements with vague caveats and contingencies. In many cases, it is virtually impossible to define the outcomes that analysts are assessing, let alone to understand how probable analysts think these outcomes might be. Here, too, the problem is not just that foreign policy analysts use imprecise language to express their beliefs, but that they avoid addressing the uncertainty that surrounds high-stakes choices.

The remainder of the chapter shows how these problems surround the highest levels of foreign policy decision making. To do this, I present a mixed-method case study that examines Vietnam War decision making, focusing on how senior leaders assessed the risks of escalating U.S. involvement in Vietnam between 1961 and 1965. I devote special attention to this case for four reasons.

The first of these reasons is that, if any group of leaders should have been expected to evaluate the logical foundations of high-stakes choices in clear and structured ways, it would have been the group that shaped Vietnam War policy from 1961 to 1965. President Kennedy appointed Robert McNamara as secretary of defense based on his reputation for conducting rigorous cost-benefit analyses as an executive at Ford Motor Company. McNamara staffed his office with the so-called Whiz Kids, who possessed an unusual zeal for quantitative analysis.[40] The Whiz Kids found a receptive audience among the Kennedy administration's roster of public intellectuals, including National Security Advisor McGeorge Bundy, a political scientist who had previously served as

[39] Daniel Gardner, *Future Babble* (New York: Plume, 2011); Chaim Kaufmann, "Threat Inflation and the Failure of the Marketplace of Ideas," *International Security*, Vol. 29, No. 1 (Summer 2004), pp. 5–48; Nate Silver, *The Signal and the Noise* (New York: Penguin, 2012); Philip Tetlock and Daniel Gardner, *Superforecasting* (New York: Crown, 2015); John Mueller, *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them* (New York: Free Press 2006); John Mueller and Mark G. Stewart, *Chasing Ghosts: The Policing of Terrorism* (New York: Oxford University Press, 2016).

[40] David Halberstam, *The Best and the Brightest* (New York: Penguin, 1983); John A. Byrne, *The Whiz Kids* (New York: Doubleday, 1993).

dean of Harvard's Faculty of Arts and Sciences. If even these leaders assessed uncertainty in the problematic ways that this chapter describes, that would lend credibility to the notion that these are consistent flaws in foreign policy decision making.[41]

Debates about escalating the Vietnam War also provide exactly the kind of context in which we would expect rational decision makers to assess uncertainty as carefully as possible. U.S. officials knew that all available strategies in Vietnam presented serious risks. They had genuine doubts about whether those risks were worth accepting, and several years to develop opinions on this issue. In this sense, the escalation of U.S. involvement in Vietnam presents a case where leaders should have had both the motivation and the opportunity to study their strategic options in detail.[42]

The documentary record of Vietnam War decision making is, furthermore, unusually thorough. Collections such as the *Foreign Relations of the United States*, the *Pentagon Papers*, and the National Intelligence Council's *Estimative Products on Vietnam* provide scholars with a broad base of well-known, widely available, trusted primary sources. We will see how these documents contain candid discussions of U.S. strategy, as public officials did not yet anticipate that sensitive opinions voiced in private would consistently leak to the press.[43] Debates over Vietnam War strategy thus present a case where assessments of uncertainty were especially important, where decision makers were especially capable of confronting this challenge directly, and where transparent discussions of those issues should be especially accessible to scholars.

Finally, Vietnam War decision making is important because it relates so closely to the theoretical frameworks that international relations scholars use to study coercion. During this period, international relations theorists developed the contemporary view that the primary purpose of coercion is to influence how opponents perceive their strategic prospects.[44] Secretary McNamara himself wrote in 1965 that the basic goal of military operations in Vietnam was "to create

[41] For a similar logic of selecting cases as a window into the (ir)rationality of foreign policy decision making, see Brian Rathbun, "The Rarity of Realpolitik: What Bismarck's Rationality Reveals about International Politics," *International Security*, Vol. 43, No. 1 (2018), pp. 7–55.

[42] Leslie Gelb and Richard K. Betts, *The Irony of Vietnam: The System Worked* (Washington, D.C.: Brookings, 1979); Fredrik Logevall, *Choosing War: The Lost Chance for Peace and the Escalation of War in Vietnam* (Berkeley, Calif.: University of California Press, 1999).

[43] Yuen Foong Khong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton, N.J.: Princeton University Press, 1992).

[44] Prominent articulations of this view include Thomas Schelling, *Arms and Influence* (New Haven, Conn.: Yale University Press, 1966); Robert A. Pape, *Bombing to Win: Air Power, Coercion, and War* (Ithaca, N.Y.: Cornell University Press, 1996); and Branislav Slantchev, "The Principle of Convergence in Wartime Negotiations," *American Political Science Review*, Vol. 97, No. 4 (2003), pp. 621–632.

conditions for a favorable settlement by demonstrating [to the Communists] that the odds are against their winning."[45] Yet we will see how McNamara and his senior colleagues consistently avoided analyzing their own odds of winning in a meaningful way. This presents a clear divergence between the theory and practice of strategic decision making, even according to U.S. officials' own standard of rational leadership.

The case study presented here involves a combination of qualitative and quantitative evidence. The analysis begins by describing how U.S. officials assessed their chances of success in documents that historians have identified as shaping President Kennedy's decision to expand the United States' advisory role in Vietnam in 1961, as well as President Johnson's decisions to escalate the air and ground campaigns in 1964–1965. Then it looks at the broader patterns in how U.S. officials assessed uncertainty when debating Vietnam War strategy, using a database of 1,757 probabilistic statements gathered from primary sources.

This mixed-method approach demonstrates how the pathologies of probability assessment described in this chapter appeared at key junctures of Vietnam War decision making, while confirming that these examples reflect systematic trends in how U.S. officials assessed uncertainty across a fifty-four-month period. The empirical record shows that officials were overwhelmingly vague in assessing key elements of uncertainty surrounding the war, and that their judgments were least informative on the issue that mattered most—namely, the chances that U.S. strategy would succeed. This account departs sharply from the conventional idea that the Whiz Kids were excessively devoted to methodological rigor. Instead, we will see how the Kennedy and Johnson administrations progressively committed the United States to war without carefully considering the risks their decisions involved.

The Probability of Success in Vietnam, 1961–1965

When John F. Kennedy became president in 1961, Vietnam was not a top priority for U.S. foreign policy. There were fewer than seven hundred U.S. soldiers in the country at the time, employed in noncombat roles as "advisers." Yet the situation in Vietnam was deteriorating. The Communist insurgency known as the Viet Cong was gaining strength, and the regime in Saigon led by South

[45] McNamara to Johnson, *Foreign Relations of the United States, 1964–1968*, Vol. III, Doc 38 (1 July 1965). Hereafter cited as *FRUS*. On U.S. officials' conception of coercion in Vietnam, see Wallace J. Thies, *When Governments Collide: Coercion and Diplomacy in the Vietnam Conflict, 1964–1968* (Berkeley, Calif.: University of California Press, 1980).

Vietnamese President Ngo Dinh Diem was rapidly losing public support. President Kennedy thus asked Secretary McNamara to write a report analyzing how the United States could respond to this growing crisis.

Secretary McNamara presented his views to the president in November 1961. He wrote, "The chances are against, probably sharply against, preventing [the fall of South Vietnam] by any measures short of the introduction of U.S. forces on a substantial scale." McNamara rejected proposals for limited escalation, saying that if the Kennedy administration introduced just eight to ten thousand additional soldiers, then they would be "almost certain to get increasingly mired down in an inconclusive struggle." McNamara recommended sending up to 220,000 troops to Vietnam. But he never assessed the chances that the United States would succeed in this effort, writing only that large-scale commitments were necessary to keep Saigon from falling to the Communists.[46]

Kennedy had solicited a second report on the growing crisis in Vietnam, which was coauthored by General Maxwell Taylor and the deputy national security adviser, Walt Rostow. They argued that "morale in Viet-Nam will rapidly crumble" without "a hard U.S. commitment to the ground." But when Taylor and Rostow recommended that the United States commit combat forces to stop the insurgency's momentum, they indicated only that this initiative was more likely to succeed than the alternatives, explaining that "intervention under SEATO [Southeast Asian Treaty Organization] or U.S. plans is the best means of saving SVN [South Vietnam] and indeed, all of Southeast Asia."[47] Employing a similar emphasis on relative probability, the Joint Chiefs of Staff argued that defeating the Viet Cong insurgency required expanding U.S. involvement and that "the over-all objective could best be served" by attacking North Vietnam directly.[48]

These documents established a pattern that recurred throughout debates about Vietnam War strategy from 1961 to 1965. When U.S. officials evaluated existing strategies, they candidly acknowledged that the war's current trajectory offered little chance of success. But when they turned their focus to recommending new strategies, their probabilistic reasoning shifted as well. Instead of directly assessing the odds that their recommendations would work, U.S. officials typically justified new measures through relative probability or

[46] *FRUS, 1961–1963*, Vol. I, Doc 214 (5 November 1961). Also see ibid., Doc 222, indicating concurrence from the secretary of state, Dean Rusk, and the Joint Chiefs of Staff.

[47] *FRUS, 1961–1963*, Vol. I, Doc 210 (3 November 1961); and *Pentagon Papers*, Vol. V.B.4, pp. 331–342. Taylor and Rostow called for 6,000–8,000 combat troops to be sent to Vietnam, while acknowledging that there would be "no limit to our possible commitment" once U.S. forces were committed in this way.

[48] JCS to McNamara, *Pentagon Papers*, Vol. V.B.4, p. 288 (9 October 1961).

conditioning, saying that new policies offered the best chance of success or that they were necessary to prevent Saigon's collapse. These forms of probabilistic reasoning were not just semantically imprecise—they reflected a deeper lack of engagement with the uncertainty surrounding strategic choices that would come to dominate U.S. foreign policy for the next decade.

This method of analysis was not limited to proponents of military escalation.[49] In several letters to President Kennedy, for instance, foreign policy adviser John Kenneth Galbraith denounced military measures for their "high risk and limited promise," arguing that "there is scarcely the slightest practical chance" that President Diem would implement the needed reforms. There was "no chance of success," Galbraith concluded, if the United States continued to back Diem and his family. Galbraith thus recommended that Washington shift its support to other Vietnamese leaders without expanding its military commitments. But beyond arguing that this course was "not hopeless," Galbraith did not evaluate the odds that his proposal would work—he simply wrote that it was "the only solution" worth considering.[50] The chairman of the Joint Chiefs of Staff, Lyman Lemnitzer, rebutted Galbraith using the same style of argumentation, writing that "the President's policy of supporting the Diem regime . . . appears to be the only practicable alternative at this time," without explaining how practicable the policy actually was.[51]

Again, the problem with this debate was not just that U.S. officials articulated their proposals in an imprecise manner that made it difficult to determine which argument was more likely to be correct. It is indeed plausible that *both* sides of this debate offered valid arguments: the United States may have had little chance of creating a stable noncommunist South Vietnam through political measures or through military measures, with or without the Diem regime. Focusing on which measure was the most likely to succeed concealed deeper uncertainty about whether the United States was pursuing feasible objectives.[52] Secretary McNamara himself later wrote that failing to address this uncertainty was one of the key factors that led the United States into the Vietnam quagmire. In his memoirs, McNamara acknowledged that "we never carefully debated what

[49] See Logevall, *Choosing War*, on how Ball and other opponents of Americanization rarely went into depth in articulating alternative proposals.

[50] *FRUS, 1961–1963*, Vol. I, Doc 209 (3 November 1961); Galbraith to Kennedy, *Pentagon Papers*, Vol. V.B.4, pp. 406–407 (20 November 1961); Galbraith to Kennedy, ibid., pp. 414–416 (21 November 1961).

[51] Lemnitzer to McNamara, *Pentagon Papers*, Vol. V.B.4, p. 465 (13 April 1962).

[52] Aaron Rapport, *Waging War, Planning Peace: U.S. Noncombat Operations in Major Wars* (Ithaca, N.Y.: Cornell University Press, 2015), argues that this reflects a broader tendency for foreign policy decision makers to base high-stakes decisions on the desirability of their objectives rather than the feasibility of their methods.

U.S. force would ultimately be required, what our chances of success would be, or what the political, military, financial, and human costs would be if we provided it. Indeed, these basic questions went unexamined. We were at the beginning of a slide down a tragic and slippery slope."[53]

Expanding the War, 1964–1965

President Kennedy responded to the crisis in Vietnam by sending fifteen thousand additional soldiers to the country, but he also refused to expand their mission beyond an advisory role. This limited investment was unable to halt the growth of the Viet Cong insurgency or to stop Saigon's political decay. In November 1964, a group of Vietnamese military officers assassinated President Diem. The coup plotters then failed to consolidate a stable regime. When Lyndon Johnson became president in December 1964, he thus faced renewed questions about the basic tenets of U.S. strategy in Vietnam. As Johnson's advisers debated those questions, they fell back on their previous style of argumentation, offering clear judgments that the current strategy was failing without assessing the chances that alternative proposals would succeed.

National Security Advisor McGeorge Bundy led the push to expand air strikes against Hanoi. "The situation in Vietnam is deteriorating," he wrote, "and without new U.S. action defeat appears inevitable." Bundy then identified strategic bombing as "the most promising course available . . . the best available way of increasing our chance of success in Vietnam." But Bundy also wrote that "we cannot estimate the odds of success with any accuracy—they may be somewhere between 25% and 75%."[54] (This statement was even less informative in context, because Bundy had defined success in terms of whether strategic bombing would "change the course of the contest in Vietnam," not whether the United States would ultimately achieve its strategic objectives.) Maxwell Taylor, now the U.S. ambassador to Saigon, concurred that an escalated bombing program "offers the best available means of exerting increasing pressure on the DRV [Democratic Republic of Vietnam] leaders to induce them to cease their intervention in SVN," without indicating how that policy was likely to fare.[55]

[53] Robert S. McNamara with Brian VanDeMark, *In Retrospect: The Tragedy and Lessons of Vietnam* (New York: Times Books, 1995), p. 107.

[54] McG. Bundy to Johnson, *FRUS, 1964–1968*, Vol. II, Doc 84 (7 February 1965).

[55] Taylor to State Department, *FRUS, 1964–1968*, Vol. II, Doc 93 (9 February 1964).

The Joint Chiefs of Staff played an ambivalent role in this debate.[56] They agreed that the United States could not succeed in Vietnam without committing additional forces. In an October 1964 memorandum, the chairman of the Joint Chiefs of Staff Earle Wheeler stated, "Unless we move now to alter the present evolution of events, there is great likelihood of VC [Viet Cong] victory."[57] However, the Joint Chiefs believed that civilian leaders underestimated the importance of directly attacking Hanoi. Wheeler thus explained: "The military course of action which would contribute most to defeating insurgencies in Southeast Asia remains the destruction of the DRV will and capabilities as necessary to compel the DRV to cease providing support to those insurgencies." Wheeler repeatedly backed this position with relational judgments, arguing that bombing North Vietnam offered "the best probability of success," the "greatest assurance of success," and the "best probability of achieving our objectives."[58] Yet despite warning that existing U.S. policy would almost surely fail, Wheeler and the Chiefs did not describe the chances that their preferred strategy might succeed.

As in 1961, opponents of escalating the war also presented recommendations without assessing their chances of success. The most prominent internal critic of escalation during this period was the undersecretary of state, George Ball. In a sixty-seven-page memorandum, Ball argued that his colleagues offered implausible justifications for expanding the war.[59] Ball instead recommended that the White House attempt to negotiate a neutral political status for South Vietnam. But though Ball wrote that this strategy's prospects were "at least as good as the chance of success through military action alone," he did not describe what those chances were.[60] In his memoirs, Ball admitted that he never tried to "do more than outline the possibilities of such a settlement."[61] Thus, though Ball advocated a different strategy than most of the other high-ranking members of the Johnson administration, he mirrored his colleagues' style of argumentation, identifying the risks and logical flaws in other proposals without evaluating the chances that his own recommendations would succeed.

[56] Herbert R. McMaster, *Dereliction of Duty: Lyndon Johnson, Robert McNamara, the Joint Chiefs of Staff, and the Lies That Led to Vietnam* (New York: HarperCollins, 1997).

[57] Wheeler to McNamara, *FRUS, 1964–1968*, Vol. I, Doc 388 (27 October 1964). The JCS also wrote to McNamara saying that "if present trends are not reversed, the counterinsurgency campaign in South Vietnam will be lost" and that U.S. combat troops were required "to turn the tide of the war." JCS to McNamara, *FRUS, 1964–1968*, Vol. II, Doc 208 (18 March 1965).

[58] Wheeler to McNamara, *FRUS, 1964–1968*, Vol. I, Doc 420 (7 February 1964).

[59] George Ball, "How Valid Are the Assumptions Underlying Our Viet-Nam Policies?" 5 October 1964. Lyndon B. Johnson Presidential Library, National Security File, Vietnam Country File, Box 222.

[60] *FRUS, 1964–1968*, Vol. II, Doc 300 (13 May 1965).

[61] George Ball, *The Past Has Another Pattern* (New York: Norton, 1982), p. 383.

Measuring Tactical Progress without Assessing Strategic Prospects

The manner in which high-ranking U.S. officials avoided assessing their chances of strategic success throughout this period struck a sharp contrast with the unprecedented analytic firepower they devoted to measuring tactical progress.[62] This elaborate effort created the impression that Secretary McNamara and his colleagues were unusually committed to methodological rigor. The McNamara Pentagon's failures in Vietnam have thus generated lasting mistrust among scholars and practitioners regarding the extent to which sophisticated analytic tools can provide value for assessing military strategy.[63]

Both impressions are misplaced. Regardless of how extensively the McNamara Pentagon analyzed tactical progress in Vietnam, this was a separate matter from analyzing the chances that U.S. strategy would ultimately succeed. Even though U.S. Army tactical reporting requirements alone used up to fourteen thousand pounds of paper per day,[64] the Pentagon undertook just one formal study of whether the United States had the capacity to ultimately achieve its central objectives.[65] And even that report, written by Andrew Goodpaster, ducked the central question by debating whether the United States *could* win the war instead of assessing the chances that the United States *would* win the war: "There appears to be no reason we cannot win if such is our will," Goodpaster explained. Goodpaster even wrote in this document that he did not believe it was necessary to offer any "assessment of the assurance the U.S. can have of winning."[66]

U.S. decision makers regularly discussed the probability of making progress in Vietnam as though progress itself were the strategic objective. A Defense Department memorandum of discussion from fall 1964 stated that the "guiding principle" for bombing Hanoi was that "the situation in Southeast Asia can be improved over what it would otherwise be if pressure is brought to bear on North Vietnam."[67] The memorandum then discusses which measures had the

[62] Thomas C. Thayer, *War without Fronts: The American Experience in Vietnam* (Boulder, Colo.: Westview Press, 1985); Gregory Daddis, *No Sure Victory: Measuring U.S. Army Effectiveness in the Vietnam War* (New York: Oxford University Press, 2011).

[63] On the Vietnam War's legacy for debates about methodological rigor and military strategy, see Halberstam, *The Best and the Brightest*; Ben Connable, *Embracing the Fog of War* (Santa Monica, Calif.: Rand, 2012).

[64] Daddis, *No Sure Victory*, p. 121.

[65] Gelb and Betts, *Irony of Vietnam*, p. 125.

[66] *FRUS, 1964–1968*, Vol. III, Doc 69 (21 July 1965).

[67] *FRUS, 1964–1968*, Vol. I, Doc 361 (25 September 1964).

potential to make these improvements, without presenting any conclusions about the extent to which those measures would influence the overall chances of winning the war. In an August 1964 memorandum to President Johnson that recommended deploying major combat forces to Vietnam, McGeorge Bundy similarly wrote that "the larger question is whether there is any course of action that can improve the chances in this weakening situation."[68] Of course, that question remained substantially narrower than asking whether the United States could actually achieve the goals for which it was preparing to sacrifice so much blood and treasure.

Estimating the chances of making progress is an inherently indirect way to evaluate policy. For example, many historians use a January 1961 report from Brigadier General Edward Lansdale to mark the start of serious debate about escalating U.S. commitment to Vietnam. Lansdale explained to President Kennedy that although the Viet Cong were making headway, "We still have a chance of beating them if we can give the people some fighting chance of gaining security and some political basis of action."[69] Assistant Secretary of State William Bundy used similarly convoluted language that fall in writing: "An early and hard-hitting operation has a good chance (70% would be my guess) of arresting things and giving Diem a chance to do better and clean up."[70] In other words, Bundy argued that expanding the war had a chance of giving Diem a chance to do something that would, in turn, improve the chances of success in Vietnam. Part of what makes this statement remarkable is that even though Bundy attempted to clarify part of his reasoning with a numeric estimate, the broader sentence conveys virtually no useful information about the strategic value of placing U.S. forces in harm's way.

None of this is to say that U.S. officials were unaware of the uncertainty surrounding the decision to escalate the Vietnam War. Most historians agree that senior officials in the Kennedy and Johnson administrations realized that the odds were against them as they deepened their commitment to Saigon. If the stakes are high enough, it can be rational for decision makers to accept low-probability gambles. And when debating Vietnam War strategy, senior leaders like Secretary McNamara, National Security Advisor Bundy, and President Johnson appear to have genuinely believed that losing Vietnam would represent a major blow to U.S. interests. It furthermore seemed implausible that a relatively small and economically undeveloped country would not eventually reach

[68] Ibid., Doc 335 (31 August 1964).

[69] Lansdale to McNamara, *Pentagon Papers*, Vol. V.B.4, p. 11 (17 January 1961).

[70] W. Bundy to Rusk, ibid., p. 312 (10 October 1961).

its breaking point when confronting the might of United States military.[71] Leslie Gelb and Richard Betts thus titled their famous book on decision making during this period *The Irony of Vietnam: The System Worked*. Even though U.S. strategy in Vietnam ultimately failed, Gelb and Betts argued, it might have still represented a reasonable, calculated risk.[72]

Yet to conclude that escalating the Vietnam War reflected reasonable logic, one must show that decision makers actually attempted to analyze their choices in a logical, reasonable manner. If McNamara, Bundy, and their colleagues had carefully considered whether the chances of success were high enough to justify escalating the war, then critics might have to accept that this was a subjective matter over which reasonable people could disagree. Instead, we have seen how senior U.S. officials repeatedly avoided assessing the uncertainty that surrounded their most important strategic choices and focused instead on identifying which measures were necessary or which strategies had the highest probability of success. As Gelb and Betts describe it themselves, debates about U.S. strategy in Vietnam "revolved around how to do things better, and whether they could be done, not whether they were worth doing."[73]

Seen from this perspective, the McNamara Pentagon's approach to decision making during the Vietnam War was not characterized by an excess of methodological rigor, but rather by a nearly complete absence of debate about the logical foundations of U.S. strategy. In his memoirs, Robert McNamara reflected on how he of all people should have understood this. "I had spent twenty years as a manager," McNamara wrote, "identifying problems and forcing organizations—often against their will—to think deeply and realistically about alternative courses of action and their consequences. I doubt I will ever fully understand why I did not do so here."[74]

[71] John Mueller, "The Search for the 'Breaking Point' in Vietnam: The Statistics of a Deadly Quarrel," *International Studies Quarterly*, Vol. 24, No. 4 (1980), pp. 497–519; Stanley Hoffmann et al., "Vietnam Reappraised," *International Security*, Vol. 6, No. 1 (1981), p. 9.

[72] Gelb and Betts, *Irony of Vietnam*.

[73] Leslie Gelb and Richard K. Betts, "Vietnam: The System Worked," *Foreign Policy*, No. 3 (1971), p. 146. Furthermore, saying that these officials shared a sense of pessimism does not mean that further analysis was unwarranted. There are situations in which decision makers might accept a thirty percent chance of success, but not a five percent chance of success, and certainly not a zero percent chance of success, even though all of these assessments could be described as pessimistic. When decision makers are pessimistic about their prospects, this is in fact exactly when it should be most important to parse probability estimates carefully, because a strategy's chances of success may be too low to justify moving forward.

[74] McNamara, *In Retrospect*, pp. 202–203.

Patterns of Probability Assessment across *Foreign Relations* Documents

The previous section used qualitative analysis of key documents to show how senior foreign policy officials were consistently vague when analyzing their strategic prospects in Vietnam, and how their judgments were especially uninformative when it came to evaluating proposals for escalating the war. One might nevertheless question the extent to which the individual documents described in the last section reflected general patterns. To substantiate an argument that senior leaders systematically avoided debating their chances of success in Vietnam, I need to analyze the documentary record systematically myself.

To do that, I constructed a database of probabilistic statements drawn from documents in the Foreign Relations of the United States (FRUS) series. FRUS is an anthology of primary sources compiled by the U.S. State Department's Office of the Historian; it constitutes "the official documentary historical record of major U.S. foreign policy decisions."[75] The six FRUS volumes that cover Vietnam War decision making from January 1961 through June 1965 contain 2,199 documents, drawn from sources that include presidential libraries, the Departments of State and Defense, the National Security Council, the Central Intelligence Agency, and private papers of individuals involved with formulating U.S. policy.

The editors of the FRUS series explain that they "choose documentation that illuminates policy formulation and major aspects and repercussions of its execution."[76] These documents thus tend to focus on the views of the high-ranking officials who set policy in Washington, as opposed to the experiences of personnel who implement those decisions in the field. Moreover, written reports and memoranda of discussions are bound to offer oversimplified versions of foreign policy officials' actual beliefs. Although these are significant drawbacks for many areas of Vietnam War scholarship, the main aim of this case study is to describe how senior leaders like Robert McNamara and McGeorge Bundy assessed their strategic prospects when they were debating escalation in Vietnam. The fact that these debates often excluded critical information is exactly why one should be concerned with how this discourse shaped the highest levels of foreign policy decision making.

[75] "About the *Foreign Relations of the United States* Series," Historical Documents, *Office of the Historian* website, https://history.state.gov/historicaldocuments/about-frus, accessed May 18, 2017.

[76] Ibid.

To analyze these documents systematically, I composed a list of twenty-one probabilistic terms based on previous scholarship exploring the meanings of qualitative expressions of uncertainty.[77] Then I collected every statement in the FRUS documents that contained any of these terms.[78] These statements made it possible to build a database containing 1,757 assessments of uncertainty, spanning 600 separate documents.

I coded each assessment as representing one of four categories. *Precise* assessments of probability establish unambiguous meaning. I defined this category liberally to include qualitative expressions such as "even chance" and "almost certain." *Imprecise* assessments of probability use qualitative expressions or numeric ranges to give a vague sense of what an analyst's belief entails. *Relative probability* assessments indicate that some conclusion is more likely than another without specifying the relevant reference point. Statements that involve *conditioning* indicate that a conclusion would be impossible to sustain unless some logical condition were met, without assessing the probability that the conclusion is actually true.

To examine how these forms of probability assessment varied across different kinds of judgments, I divided statements in the database into nine categories. First, I identified 94 statements that involved U.S. officials describing the chances that their strategic recommendations would succeed. Next, I identified 58 statements that involved U.S. officials describing the probability of success in Vietnam *without* making a strategic recommendation, usually in the context of critiquing other people's proposals or explaining why the status quo was not viable. The remaining seven categories dealt with elements of uncertainty besides the probability of strategic success. These categories, which were neither mutually exclusive nor exhaustive, included predictions about coups or Government of Vietnam (GVN) leadership succession; assessments of the chances that other states would turn

[77] The terms were "almost certain," "cannot rule out," "chance," "conceivable," "doubt," "doubtful," "fifty-fifty," "improbable," "inconceivable," "liable," "likelihood," "likely," "odds," "percent," "probability," "probable," "probably," "prospect," "remote," "risk," and "unlikely." I chose these search terms because they appear in other scholars' empirical studies of verbal uncertainty expressions, specifically, Kent, "Words of Estimative Probability"; Beyth-Marom, "How Probable Is Probable?"; and Mosteller and Youtz, "Quantifying Probabilistic Expressions."

[78] I applied three criteria for inclusion in this database. First, I only retained statements that expressed the beliefs of the document's author or that were recorded in a memorandum of discussion. Second, I only retained statements that reflected the beliefs of U.S. government officials; though independent scholars or members of the press appear frequently in FRUS documents, the goal of this study is to analyze how U.S. officials debated uncertainty. Third, I only included documents and memoranda intended for an internal audience, on the assumption that public speeches, letters to President Diem, or other public statements would not reliably reflect officials' true beliefs.

Table 1.1. **Patterns of Probability Assessment across *Foreign Relations* Documents**

	Precise assessments	*Imprecise assessments*	*Relative probability or conditioning*
Probability of a proposed strategy succeeding (N = 94)	0%**	9%***	91%***
Probability of any other strategy succeeding (58)	2%	88%	10%
All statements (1,759)	*8%*	*78%*	*14%*
Coups/South Vietnamese succession (213)	9%	74%	17%
Other states turning Communist (30)	20%*	67%	13%
GVN domestic policies (432)	5%*	88%***	7%***
GVN economic/fiscal policies (39)	0%	95%**	5%
Communist strategy/perceptions (472)	12%***	76%	12%
Foreign policies of noncombatants (106)	10%	85%	36%***
Chances of tactical policy success (136)	7%	87%**	25%**

Asterisks reflect statistical significance levels when comparing each category to baseline patterns across the data as a whole, as follows: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < .0.001$.

or lean Communist; assessments of GVN domestic policies; assessments of GVN economic or fiscal policies; assessments of Communist combatants' strategies or intentions; assessments of noncombatant countries' foreign policies or intentions; and the chances that specific policies would succeed in meeting tactical objectives.

The database strongly confirms the patterns shown in the previous section's qualitative analysis. As shown by the top row of Table 1.1, U.S. officials were overwhelmingly vague when describing the chances that their recommendations would succeed. Ninety-one percent of those statements involved either relative probability or conditioning. Most of the remaining statements were not especially informative either. Some involved unspecified logic, such as Ambassador Frederick Nolting's claim that "we have better than [a] 50-50 chance of winning on this policy line provided the border with Laos is reasonably well protected," in which Nolting neither explained what it meant to "reasonably" protect the

border nor assessed the probability that this could be accomplished.[79] Other statements convey extremely vague sentiments, such as the argument of Deputy National Security Advisor Rostow that escalation would produce "a fair chance of avoiding war," General Wheeler's assessment that "victory is now a hopeful prospect," or Director of Central Intelligence John McCone's reporting to President Johnson that winning the war would be "formidable and difficult, but not impossible."[80]

The database also confirms that U.S. officials consistently provided more informative probability assessments when critiquing the strategies they opposed. Fifty-eight statements in the database involve U.S. officials describing the current strategy's prospects or critiquing other officials' recommendations. Although only one of these statements could reasonably be considered precise,[81] just six (ten percent) involved relative probability or conditioning. Most of these statements offer at least some rough way to gauge U.S. officials' pessimism: for example, stating that "the chances of a turn-around in South Vietnam remain less than even," explaining that current U.S. strategy "would probably not save South Viet-Nam from eventual loss," or warning that "our present course of action . . . is very likely to fail."[82]

The database also confirms that the manner in which U.S. officials discussed their strategic prospects contrasted with how they assessed other aspects of uncertainty about the war.[83] Though these officials almost never made their probability assessments explicit, just fourteen percent of the statements across the database as a whole involved either relative probability or conditioning. No other category of statements employed these forms of probabilistic reasoning nearly as frequently as when officials described the chances

[79] Frederick Nolting, "Telegram from the Embassy in Vietnam to the Department of State," *FRUS, 1961–1963*, Vol. 1, Doc 147 (6 October 1961). Similarly, Major General Charles Bonesteel argued in a meeting of the Presidential Task Force on Vietnam that "if we had a reasonably workable settlement in Laos, bolstered by the barrier between Laos and Viet-Nam and along the 17th parallel, we could probably succeed." *FRUS, 1961–1963*, Vol. 1, Doc 43 (4 May 1961).

[80] Rostow to Johnson, *FRUS, 1961–1963*, Vol. 1, Doc 251 (14 November 1961); JCS Team Report on South Vietnam, *FRUS, 1961–1963*, Vol. II, Doc 26 (January 1963); McCone to Johnson, *FRUS, 1961–1963*, Vol. IV, Doc 375 (23 December 1963).

[81] This is a memorandum written by the Defense Department's Office of International Security Affairs, which argues that continuing U.S. policy toward President Diem would "almost certainly seal the fate of Vietnam." *FRUS, 1961–1963*, Vol. II, doc. 35 (26 January 1962).

[82] McG. Bundy to Johnson, *FRUS, 1964–1968*, Vol. II, Doc 183 (6 March 1965); Maxwell Taylor, "The Current Situation in South Vietnam," *FRUS, 1964–1968*, Vol. I, Doc 426 (November 1964); Chester Bowles to Kennedy, "Recommendations for a Fresh Approach to the Vietnam Impasse," *FRUS, 1961–1963*, Vol. III, Doc 52 (7 March 1963).

[83] The statistical significance tests reported in Table 1.1 reflect two-way t-tests comparing differences in means.

that their favored strategies would succeed. Altogether, these data support the chapter's qualitative description of how U.S. officials were overwhelmingly vague when assessing uncertainty in Vietnam; of how this problem ran much deeper than the use of imprecise language; and of how assessments of uncertainty were least informative when it came to proposals for escalating the war.

Beyond "Words of Estimative Probability"

This chapter began by describing the famous 1964 essay by Sherman Kent, titled "Words of Estimative Probability." In that essay, Kent criticized intelligence analysts for making imprecise assessments of uncertainty. More than a half-century later, Kent's writing on this topic remains an effective vehicle for sparking discussion about the nature and limits of foreign policy analysis. Yet the chapter described three ways in which Kent did not go far enough in describing the degree to which foreign policy analysts avoid probabilistic reasoning.

First, Kent's essay suggests that vague assessments of uncertainty are mainly a problem for intelligence analysts. The chapter has explained, by contrast, how similar issues recur across a broad range of foreign policy agencies, in public debates among scholars and pundits, and among decision makers forming policy at the highest levels.

Second, when Kent offered his critiques in 1964, he was raising an issue to which foreign policy analysts had previously devoted little attention. In recounting his experience writing a report describing a "serious possibility" that the Soviet Union would invade Yugoslavia, Kent admitted that he had not given this phrase much thought. This chapter has shown, by contrast, that aversion to probabilistic reasoning in foreign policy discourse is now entirely deliberate, guided by official doctrine that encourages intelligence analysts and military planners to avoid assessing uncertainty in clear and structured ways.

Third, and most importantly, this chapter has demonstrated how problems with assessing uncertainty in international politics run much deeper than semantics. It is even more worrisome to see that foreign policy analysts often avoid assessing crucial probabilities at all, especially by engaging in the practices that this chapter called relative probability and conditioning. These are not just inefficient methods for assessing uncertainty: these kinds of analysis are fundamentally unsuited to supporting sound decisions. Nor do these problems appear solely among the so-called poets who are uncomfortable with advanced analytics. Even Robert McNamara and the Whiz Kids systematically avoided assessing the uncertainty surrounding their most important foreign policy choices.

This experience motivates a broader concern that will run throughout the remainder of the book. In most areas of policy and scholarship, it is generally assumed that rigor, objectivity, and analytic precision go together.[84] Yet that is not the case when it comes to probabilistic reasoning in international politics. Any attempt to assess uncertainty in this domain is inherently subjective, and that is indeed one of the main reasons why foreign policy analysts are typically reluctant to describe their judgments explicitly. The Whiz Kids' experience in Vietnam highlights the dangers of failing to grapple with this challenge—of privileging objectivity over rigor in foreign policy analysis. But if the most important elements of uncertainty in international politics are inherently subjective, then what would rigorously analyzing these issues look like? How well can fallible individuals handle this challenge? And how can decision makers incorporate these judgments into real foreign policy decisions? These are the questions that occupy the rest of the book.

OUP-USA

[84] For a history of debates about rigor, objectivity, and quantification in public policy analysis, see Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, N.J.: Princeton University Press, 1995).

2

Subjective Probability and International Politics

This chapter examines the theoretical foundations of assessing uncertainty in international politics. The chapter's first section explains that all major foreign policy decisions depend on perceptions of subjective probability. Even when foreign policy analysts base their judgments on statistical analyses or mathematical models, their assessments still reflect personal convictions rather than objective truths. The chapter's second section explains why it is logically acceptable to believe that subjective probabilities are meaningless, but shows that this premise has consequences that no foreign policy analyst can accept. The chapter's third section demonstrates how, conditional on believing that assessments of subjective probability contain any meaningful insight, it is always possible to form and communicate that insight in clear and structured ways. The subjective nature of probabilistic reasoning in international politics is, in fact, exactly what makes it possible to debate that reasoning directly, because foreign policy analysts can always describe their personal convictions however precisely they like.

This theoretical analysis has several practical implications. For instance, the chapter explains how it is always possible to resolve the kind of confusion that President Obama encountered in reconciling different views about the chances that Osama bin Laden was living at Abbottabad. We will see that there is no situation in which foreign policy analysts can communicate their beliefs using qualitative language but not numeric percentages. And I will argue that the impulse that leads foreign policy analysts to prize objectivity is better served by drawing clear distinctions between the concepts of probability and confidence. Altogether, the chapter shows how foreign policy analysts always possess a coherent theoretical basis for assessing uncertainty in international politics.

Theoretical Foundations of Probability Assessment

What did it mean to estimate a "sixty percent chance" that Osama bin Laden was living in Abbottabad? Ultimately, bin Laden was there or he wasn't—how could any other judgment be considered meaningful? And if foreign policy analysts cannot accurately model phenomena as complex as international politics and armed conflict, then how can they form coherent estimates of the uncertainty that these phenomena contain? These are nontrivial questions that raise long-standing debates about the conceptual foundations of probabilistic reasoning and the nature of foreign policy analysis.[1]

Broadly speaking, there are three ways to define what probability assessments represent. These frameworks are called the *frequency*, *propensity*, and *subjective* theories of probability.[2] This section explains how the subjective theory of probability is the only viable foundation for assessing uncertainty in international politics.

Much of the discussion in this section revolves around the distinction between the concepts of *aleatory* and *epistemic* uncertainty. Aleatory uncertainty applies to situations in which outcomes are unknown because they reflect processes involving genuine randomness. Epistemic uncertainty applies to situations in which outcomes are unknown because analysts possess incomplete information.[3] Uncertainty about Osama bin Laden's location or the status of Saddam Hussein's nuclear weapons programs are both examples of epistemic uncertainty. Epistemic uncertainty is always subjective, because it depends on the state of an analyst's personal knowledge. In brief, this section argues that all probabilistic reasoning in international politics involves epistemic uncertainty in one form or another.[4]

[1] Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference* (New York: Cambridge University Press, 1975); Donald Gillies, *Philosophical Theories of Probability* (New York: Routledge, 2000); Gerd Gigerenzer et al., *The Empire of Chance: How Probability Changed Science and Everyday Life* (New York: Cambridge University Press, 1989); and David Howie, *Interpreting Probability: Controversies and Developments in the Early Twentieth Century* (New York: Cambridge University Press, 2002).

[2] This typology omits two paradigms: the classical theory of probability of Bernoulli and Laplace, and Keynes's theory of logical probability. These theories played an important role in the history of probabilistic thought, but they are no longer considered viable. See Gillies, *Philosophical Theories of Probability*, for an overview of these concepts.

[3] This "dual nature" of probability assessment is a central theme in Hacking, *Emergence of Probability*.

[4] Note that the contrast between aleatory and epistemic uncertainty is different from the way in which intelligence scholars distinguish between *puzzles* and *mysteries*. In intelligence studies, puzzles represent questions that have knowable right answers, whereas mysteries represent questions about which it is impossible to provide conclusive judgments. Some kinds of epistemic uncertainty (like Osama bin Laden's location in 2011) are indeed puzzles, but other forms of epistemic uncertainty

The Frequency Theory of Probability

The first way of defining what probability assessments represent is called the *frequency theory* of probability. According to the frequency theory, probability assessments reflect inferences drawn from analyzing a series of previous identical events.[5] When you say that a coin flip has a fifty percent chance of turning up heads, the frequency theory's interpretation for this statement is that you have previously witnessed many such coins being flipped, and half of those coins have landed facing up. Probability assessments based on the frequency theory are considered to be "correct" to the extent that they accurately characterize the patterns observed across these previous, identical events.

The frequency theory serves as the basis for many statistical techniques. Nevertheless, the frequency theory does not provide a coherent foundation for probabilistic reasoning in international politics. It cannot, for example, resolve uncertainty over factual matters. When intelligence analysts debated the chances that Saddam Hussein was pursuing nuclear weapons, their judgments reflected incomplete information, and not inferences drawn from a stream of prior cases. And though foreign policy analysts can often use statistical patterns to inform their expectations about the likely outcomes of a policy, those patterns almost never reflect a series of previous, identical events. Richard von Mises, the German mathematician who made seminal contributions to the frequency theory, was clear on this point. "'The probability of winning a battle,'" he explained, "has no place in our theory of probability, because we cannot think of a collective to which it belongs."[6]

This does not mean that statistical analysis is useless for informing foreign policy debates. Yet the moment statistical analysts can no longer consider data points to be identical to the specific issue they are assessing, they can no longer defend their conclusions on the basis of the frequency theory alone. This is what von Mises meant by saying that battles do not naturally belong to "collectives": constructing a statistical model of battle outcomes requires determining which data points belong in the analysis, assessing how similar the cases are, and speculating about what kinds of confounds a statistical model might

(such as judging a foreign leader's intentions) are essentially impossible to resolve, and should thus be considered mysteries. Meanwhile, some kinds of aleatory uncertainty (such as estimating the probability that a coin flip will come up heads) do yield "correct" answers, and should therefore be considered puzzles. On the distinction between mysteries and puzzles, see Wilhelm Agrell and Gregory F. Treverton, *National Intelligence and Science: Beyond the Great Divide in Analysis and Policy* (New York: Oxford University Press, 2015), pp. 32–35.

[5] Richard von Mises, *Probability, Statistics and Truth* (Mineola, N.Y.: Dover Publications, 1981 [1928]).

[6] von Mises, *Probability, Statistics, and Truth*, pp. 11–15.

be unable to measure.[7] These are all subjective judgments that reflect epistemic uncertainty about the quality of a statistical model's assumptions, rather than objective properties of randomness.[8] In this sense, even if the statistical methods used in foreign policy analysis are often described in terms of "frequentist statistics," they are properly viewed as analogies to the frequentist paradigm, rather than strict applications of that conceptual framework.[9]

The Propensity Theory of Probability

The second way to describe what probability assessments represent is called the *propensity theory* of probability. According to the propensity theory, probability assessments reflect an analyst's knowledge of the manner in which physical systems generate random output.[10] When you say that a coin flip has a fifty percent chance of turning up heads, the propensity theory's interpretation of this statement is that you have studied the coin's physical properties, and you have concluded that they will cause the coin to land face up half of the time. Probability assessments based on the propensity theory are considered to be "correct" to the extent that they accurately characterize the properties of mechanical systems.

The propensity theory plays a major role in the natural sciences, particularly when it comes to grounding quantum mechanics. But like frequency theory, propensity theory is inappropriate for describing probability assessments in international politics. Once again, the theory cannot capture what analysts are saying when they describe uncertainty surrounding factual matters. When Michael Morell said there was a sixty percent chance that bin Laden was living

[7] The subjective nature of determining what it means for cases to be "comparable" in this fashion is known as the *reference class problem*. See John Venn, *The Logic of Chance* (London, U.K.: Macmillan, 1888); Hans Reichenbach, *The Theory of Probability*, tr. Ernest H. Hutten and Maria Reichenbach (Berkeley, Calif.: University of California Press, 1949).

[8] Charles F. Manski, *Public Policy in an Uncertain World: Analysis and Decision* (Cambridge, Mass.: Harvard University Press, 2013), explains that statistical probability estimates present an incomplete (and usually narrower) picture of the overall uncertainty that surrounds empirical inference.

[9] In other words, the use of frequentist statistics in foreign policy analysis is properly understood as a tool for generating subjective probability estimates. It is, then, an empirical question as to whether those statistical methods are more reliable than other forms of subjective judgment. For more on how frequentist statistics rarely adhere to the frequency theory of probability itself, see Harold Jeffreys, "Probability and Scientific Method," *Proceedings of the Royal Society of London A*, Vol. 146, No. 856 (1933), pp. 9–16; E. T. Jaynes, *Probability Theory: The Logic of Science*, ed. G. Larry Brethorst (New York: Cambridge University Press, 2003).

[10] The propensity theory originates with Karl Popper, "The Propensity Interpretation of Probability," *British Journal for the Philosophy of Science*, Vol. 10, No. 37 (1959), pp. 25–52.

at Abbottabad, he was not implying that some mechanical process had randomly assigned bin Laden to this location six times out of ten. And though foreign policy analysts can use physical models and simulation techniques to guide some elements of policy design, no one would claim that these analyses depict the actual properties of social systems. Even the most sophisticated models of armed conflict and international politics reflect extreme simplifications of reality. Interpreting these models' results thus requires making subjective judgments about the appropriateness of simplifying assumptions.[11]

Moreover, even if foreign policy analysts could create a deterministic model of international politics, that would *still* be insufficient to make objective predictions, because many of the model's key parameters would remain uncertain. When waging wars or bargaining over contentious issues, leaders have incentives to conceal their intentions, capabilities, and resolve. Since any model of conflict or cooperation should account for these variables, those models' output will reflect epistemic uncertainty about underlying conditions.[12] This is another reason why even the most sophisticated piece of foreign policy analysis could never meet the propensity theory's standards of objectivity.[13]

Of course, there is no need to invoke probability theory just to understand the difficulties of predicting international politics. The purpose of this discussion is

[11] On the limits to objectivity in military modeling in particular, see Joshua M. Epstein, *Strategy and Force Planning* (Washington, D.C.: Brookings, 1987); Thomas J. Czerwinski, *Coping with the Bounds: Speculations on Nonlinearity in Military Affairs* (Washington, D.C.: National Defense University Press, 1998); Michael E. O'Hanlon, *The Science of War* (Princeton, N.J.: Princeton University Press, 2009).

[12] On the pervasiveness of this uncertainty, see Robert Jervis, *System Effects: Complexity in Political and Social Life* (Princeton, N.J.: Princeton University Press, 1997); John Mearsheimer, *The Tragedy of Great Power Politics* (New York: Norton, 2001), pp. 29–54; and Kristopher W. Ramsey, "Information, Uncertainty, and War," *Annual Reviews of Political Science*, Vol. 20 (2017), pp. 505–527. For practitioners' views of this subject, see James N. Mattis, "USJFCOM Commander's Guidance for Effects-Based Operations," *Parameters*, Vol. 38, No. 3 (2008), pp. 18–25; and Paul K. Van Riper, "EBO: There Was No Baby in the Bathwater," *Joint Forces Quarterly*, No. 52 (2009), pp. 82–85. On the limits of academic forecasting in international politics, see Gerald Schneider, Nils Petter Gleditsch, and Sabine Carey, "Forecasting in International Relations," *Conflict Management and Peace Science*, Vol. 20, No. 1 (2011), pp. 5–14; and Michael D. Ward, "Can We Predict Politics?" *Journal of Global Security Studies*, Vol. 1, No. 1 (2016), pp. 80–91.

[13] Furthermore, even if foreign policy analysts *could* somehow develop a purely objective probability assessment, most decision makers would be unable to know that this was the case. Decision makers do not possess the time or expertise to scrutinize every data point and assumption behind the estimates that analysts give them. Indeed, the purpose of delegating analytic tasks is so that decision makers do *not* need to scrutinize these details. But since these minutiae matter, any time decision makers solicit input in forming judgments under uncertainty, their reactions to that input depend on the subjective trust that they place in their analysts for the purposes of understanding a specific issue.

instead to place those difficulties into proper theoretical perspective. When foreign policy analysts attempt to assess the uncertainty that surrounds high-stakes decisions, their judgments should not be viewed as unreliable attempts to estimate frequencies or propensities. Logically speaking, these judgments do not reflect frequencies or propensities at all—even if foreign policy analysts base their assessments of uncertainty on the output of simulations or statistical analyses, their conclusions always reflect some measure of subjective belief. The next part of this section explains how the logic of translating subjective beliefs into probability assessments is entirely different from the manner in which analysts typically interpret output from statistical analyses or mechanical models.

The Subjective Theory of Probability

The subjective theory of probability defines probability assessments with respect to an analyst's personal convictions.[14] The standard way to represent these convictions is to describe an analyst's willingness to bet. Thus, when you say that a coin flip has a fifty percent chance of coming up heads, the subjective theory's interpretation for this statement is that you would be indifferent between betting on the coin's landing heads versus the coin's landing tails.[15] According to the subjective probability framework, this judgment is considered to be "correct" to the extent that it accurately characterizes an analyst's personal beliefs.[16]

The logic of describing an analyst's personal beliefs can sharply diverge from the logic of describing objective properties of randomness. For example, consider flipping a coin that has an unknown bias. In other words, you know that this coin is unevenly weighted, but you have no information about whether the imbalance inclines the coin toward landing heads versus landing tails. Since you should be indifferent between betting on either outcome, the subjective theory of probability indicates that you should estimate the probability of this coin landing heads as fifty percent—even though you know this is the only estimate that does *not* describe the coin's actual, physical properties. The same

[14] The Italian economist Bruno de Finetti and the British philosopher Frank P. Ramsey developed the theory of subjective probability independently. See Frank P. Ramsey, "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays* (New York: Harcourt, Brace, 1931); and Bruno de Finetti, "Probabilism: A Critical Essay on the Theory of Probability and the Value of Science," *Logos*, Vol. 14 (1931), pp. 163–219; reprinted in *Erkenntnis*, Vol. 31, Vol. 2/3 (1989), pp. 169–223. The American mathematician Leonard J. Savage then extended the formal logic behind this framework in *The Foundations of Statistics* (New York: Wiley, 1954).

[15] And that you would be unwilling to bet on any other outcome, regardless of the payoff. Gillies, *Philosophical Theories of Probability*, pp. 55–58, explores conceptual nuances of the comparison of lotteries (or "betting quotient") method for defining what probability assessments represent.

[16] Frank Lad, *Operational Subjective Statistical Methods* (New York: Wiley, 1996), pp. 8–9.

logic applies to interpreting the "chances" that Osama bin Laden was living at Abbottabad. When President Obama's advisers estimated those chances, their degrees of personal conviction reportedly ranged from thirty percent to ninety-five percent, even though everyone knew that the "real" probability of bin Laden living at Abbottabad was either zero percent or one hundred percent.

Many people find it counterintuitive to equate convictions with probabilities, and laboratory experiments confirm that subjects tend to treat these concepts differently. For example, if given the opportunity to bet on calling the outcome of a fair coin versus calling the outcome of a coin with an unknown bias, most people will choose the first of these gambles. Some people will even pay nontrivial amounts to ensure that they can bet on less-ambiguous probabilities.

This behavior is known as *ambiguity aversion*, and it is easy to demonstrate that ambiguity aversion is misguided.[17] For example, you could simply randomize whether you predict that the coin will land heads or tails. In that case, you would guarantee yourself a fifty percent chance of winning the bet regardless of whether the coin you flip is fair or imbalanced. If anything, it is actually in your interest to flip the coin with an unknown bias, because this would provide potentially valuable information about how that coin is weighted. If you ever have a chance to take this gamble again, you could pick the outcome that you observed from the last time you flipped this coin, and your expected probability of winning the bet would then be greater than fifty percent.[18]

Another source of discomfort with the subjective probability framework is that it conflicts with the desire to base high-stakes choices on objective judgments.[19] Objectivity is in fact a formal requirement in many areas of foreign policy analysis. For example, the very first requirement laid out in the U.S. Intelligence Community's analytic standards is that "analysts must perform their functions with objectivity."[20] U.S. Army planning doctrine similarly revolves

[17] The term *ambiguity aversion* was coined by Daniel Ellsberg, the economist who later went on to leak the Pentagon Papers. See Daniel Ellsberg, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, Vol. 75, No. 4 (1961) pp. 643–669. In some situations ambiguity can serve as a signal of other decision-relevant factors, but this does not change the fact that ambiguity regarding probability estimates is not, by itself, a component of rational decision making. See Nabil al-Najjar and Jonathan Weinstein, "The Ambiguity Aversion Literature: A Critical Assessment," *Economics and Philosophy*, Vol. 25, No. 3 (2009), pp. 249–284.

[18] Stefan T. Trautmann and Richard Zeckhauser, "Shunning Uncertainty: The Neglect of Learning Opportunities," *Games and Economic Behavior*, Vol. 79 (2013), pp. 44–55.

[19] For a history of debates about rigor, objectivity, and quantification in public policy analysis, see Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, N.J.: Princeton University Press, 1995).

[20] Intelligence Community Directive 203 (ICD-203), *Analytic Standards*, p. 2. U.S. law actually mandates that the Intelligence Community develop institutional mechanisms to "safeguard objectivity." See the 2004 Intelligence Reform and Terrorism Prevention Act, Section 1020A. Recent

around "safeguards to protect objectivity in the assessment process."[21] This is one reason why many scholars and practitioners believe that foreign policy analysts should avoid making their assessments of uncertainty explicit. Thus, in what is arguably the most important contemporary textbook on intelligence studies, Mark Lowenthal writes that numeric probability assessments "run the risk of conveying to the policy client a degree of precision that does not exist. What is the difference between a 6-in-10 chance and a 7-in-10 chance, beyond greater conviction? In reality, the analyst is back to relying on gut feeling."[22]

Yet leaving assessments of uncertainty vague does not change their conceptual foundations. These judgments do not become subjective once they pass some threshold of specificity. Regardless of how foreign policy analysts express their beliefs about uncertainty, there is no coherent way to say what these beliefs reflect that does not rely on personal conviction. If foreign policy analysts are truly intended to "perform their functions with objectivity," then they would have to reject *all* assessments of uncertainty, not just those expressed using numbers instead of words. The next section explains why this is a position that no foreign policy analyst can accept.

Probability, Confidence, and Analytic Rigor

To say that a judgment is subjective does not mean that it is careless or uninformative. For example, Carl von Clausewitz argued that the key to assessing uncertainty in war lay with cultivating an intuitive "genius" that did not "yield to academic wisdom."[23] In Clausewitz's view, the truly careless move would be to base strategic decisions on a hidebound use of formal logic that left no room to

survey research indicates that intelligence analysts indeed prioritize "objectivity" over ICD-203's other analytic standards: see David R. Mandel and Tonya Hendriks, "ODNI's ICD-203 Standards on Analytic Integrity," Paper presented at the 7th Biannual Meeting of the NATO Systems Analysis and Studies Panel Research Technical Group on the Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making (Madrid, Spain, 2018).

[21] U.S. Army Field Manual 5-0, *The Operations Process* (2010), p. H-4.

[22] Mark Lowenthal, *Intelligence: From Secrets to Policy*, 3rd edition (Washington, D.C.: CQ Press, 2006), p. 129. In a similar vein, the former director of national intelligence James Clapper reflected on the debate about the chances that Osama bin Laden was living at Abbottabad, saying: "We put a lot of discussion [into] percentages of confidence, which to me is not particularly meaningful. In the end it's all subjective judgment anyway." *CNN*, "The Axe Files," Podcast Ep. 247 (31 May 2018).

[23] Carl von Clausewitz, *On War*, tr. Michael Howard and Peter Paret (Princeton, N.J.: Princeton University Press, 1984); and Clifford J. Rogers, "Clausewitz, Genius, and the Rules," *Journal of Military History*, Vol. 66, No. 4 (2002), pp. 1167–1176.

accommodate a decision maker's personal insight.[24] And in the next chapter, we will also see how fine-grained variations in subjective probability assessments can capture meaningful information about international politics.

It is nevertheless impossible to demonstrate that a single subjective probability assessment is meaningful—in other words, there is no way to demonstrate that an analyst's personal convictions provide a useful guide to assessing uncertainty in any one case.[25] This has led many prominent scholars to question whether subjective probabilities can ever be trusted as a basis for making high-stakes decisions. The philosopher John Stuart Mill expressed this view when he argued that probability estimates "are of no real value" unless analysts derive them from large volumes of reliable data.[26] The economist John Maynard Keynes expressed similar disdain for subjective probabilities, writing that "about these matters, there is no scientific basis on which to form any calculable probability whatsoever. We simply do not know."[27]

Although this position is logically coherent, its consequences are untenable for anyone who seeks to contribute to foreign policy debates. In order to believe that it is possible to distinguish between policies that are acceptable and those that are not, one must implicitly assume that subjective probability assessments are meaningful. For example, consider General Stanley McChrystal's 2009 recommendation that President Obama deploy forty thousand additional troops to Afghanistan. We saw in chapter 1 that General McChrystal argued that a new strategy backed by forty thousand additional troops represented "the best prospect for success in this important mission." Elsewhere in his report, General McChrystal wrote that his proposed troop

[24] The rejection of Enlightenment rationality was indeed crucial to Clausewitz's intellectual project. On this point, see Azar Gat, *The Origins of Military Thought: From the Enlightenment to Clausewitz* (Oxford, U.K.: Clarendon Press, 1989). On the distinction between subjectivity and deliberativeness in foreign policy analysis more generally, see Philip E. Tetlock, "Second Thoughts about *Expert Political Judgment*," *Critical Review*, Vol. 22, No. 4 (2010), pp. 467–488. In a similar vein, psychologists have produced a large volume of research showing that unconscious reasoning can improve decision outcomes. See Ap Dijksterhuis et al., "On Making the Right Choice: The Deliberation-Without-Attention Effect," *Science*, Vol. 311 (2006), pp. 1005–1007.

[25] Thus one of the principal critiques of Clausewitz's theory of decision making: if two commanders disagree on the chances that a strategy will succeed, how do we know which one possesses greater "genius"?

[26] John Stuart Mill, *A System of Logic, Ratiocinative and Inductive*, 8th ed. (New York, N.Y.: Harper and Brothers, 1882), p. 539.

[27] John Maynard Keynes, "The General Theory of Employment," *Quarterly Journal of Economics*, Vol. 51, No. 2 (1937), pp. 213–214; see also John Maynard Keynes, *A Treatise on Probability* (London: Macmillan, 1921). For reviews of this argument in international affairs, see Richard K. Betts, "Is Strategy an Illusion?" *International Security*, Vol. 25, No. 2 (2000), pp. 5–50; and Ben Connable, *Embracing the Fog of War* (Santa Monica, Calif.: Rand, 2012).

surge represented "the minimum force levels to accomplish the mission with an acceptable level of risk."

Both statements reflected the premise that sending fewer than forty thousand reinforcements to Afghanistan would produce lower chances of success.[28] It is impossible to believe that General McChrystal could coherently argue for sending additional troops to Afghanistan while also claiming that he had no basis for assessing the chances that the deployment would succeed. If General McChrystal had no basis for assessing those chances, then President Obama would have had no reason to expect that sending forty thousand additional forces to Afghanistan would produce a better outcome than deploying thirty thousand more troops. Iterating this logic, it would become unjustifiable to send any additional soldiers to Afghanistan, or even to keep any remaining U.S. forces in the country at all. Who could have said that withdrawal would lower the chances of defeating the Taliban? What is the purpose of expending lives and resources if there is no basis for believing that this raises the chances of obtaining valuable outcomes?

Few observers would adopt such an extreme viewpoint.[29] Regardless of how complex or subjective a decision might be, it is ultimately necessarily to draw a line somewhere between resource allocations that are acceptable and those that are not. Any deployment of resources thus depends on the notion that it is, in fact, possible to form coherent perceptions of subjective probability.[30]

Of course, this does not mean that anything goes in foreign policy analysis. If a team of subject-matter experts conducts a rigorous, mixed-method analysis to estimate the chances than some event will occur, their judgment clearly deserves

[28] This statement also implies that General McChrystal could judge that the *magnitude* of this shift in probability was large enough to justify the additional investment of resources. Similar assumptions are also evident in McChrystal's claim that deploying more than 40,000 additional troops "could achieve low risk, but this would be excessive in the final analysis." Commander, International Security Assistance Force-Afghanistan, *COMISAF Initial Assessment*, August 2009, pp. 2–20 to 2–21.

[29] In fact, if foreign policy analysts cannot make valid probability assessments, then the only ethical move would be to eliminate all defense expenditures. There would be no reason to think that those resources actually make the country safer, and thus the money would be better spent elsewhere (or else returned to taxpayers). As Betts puts it, if there is no hope of predicting the results of military strategy, then "analysts as well as politicians and generals should all quit and go fishing" ("Is Strategy an Illusion," p. 20). See also Robert A. Pape, "The Air Force Strikes Back," *Security Studies*, Vol. 7, No. 2 (1997/98), pp. 191–214.

[30] Note how this discussion inverts the traditional relationship between objectivity and analytic precision. In many areas of public life, scholars and practitioners seek to quantify important parameters as a way of pursuing objective judgment. Yet, when dealing with probability assessments in international politics, many scholars and practitioners seek to *avoid* making precise judgments, on the grounds that these judgments would be excessively subjective. See Porter, *Trust in Numbers*, for more detail on the first of these viewpoints.

more weight than a pundit's off-the-cuff opinion, even if both views reflect subjective probabilities. But how should decision makers determine how much weight these judgments deserve? If objectivity is not the proper standard for assigning credibility to foreign policy analysts' judgments, then what criterion should we use?

Answering these questions requires distinguishing between the concepts of "probability" and "confidence." As mentioned in the introduction, *probability* reflects analysts' beliefs about the chances that a statement is true, while *confidence* reflects the extent to which analysts believe they possess a sound basis for assessing uncertainty. Elsewhere, I have argued for further parsing analytic confidence into at least three components: the reliability of available evidence supporting a judgment, the range of reasonable opinion surrounding a judgment, and the degree to which analysts believe their judgment could change in response to new information. These concepts can vary independently, and each carries distinct implications for high-stakes decision making.[31]

In U.S. criminal trials, for example, jurors are only allowed to convict defendants based on reliable evidence. Inferences based on hearsay, stereotypes, or coerced confessions are barred from jurors' consideration on the grounds that these inferences do not provide a valid basis for revoking defendants' civil liberties. Most criminal trials furthermore require that jurors reach a guilty verdict unanimously, reflecting the premise that a single dissenting view creates enough reasonable doubt to preclude conviction. In this respect, the application of criminal justice is not just a matter of estimating the probability of defendants' guilt. Estimating a high probability of guilt is necessary, but not sufficient, to sustain a guilty verdict.[32]

Controversy over the U.S. invasion of Iraq in 2003 shows how similar factors shape the politics and ethics of foreign policy decision making. As described in chapter 1, most intelligence analysts believed that Iraq was pursuing weapons of mass destruction (WMD), and the Bush administration invoked that belief to justify going to war. However, inferences about Iraq's WMD programs relied heavily on circumstantial evidence and questionable informants. The U.S. State Department had, furthermore, lodged an explicit dissent regarding assessments of Iraq's nuclear program, and the International Atomic Energy Agency had

[31] Jeffrey A. Friedman and Richard Zeckhauser, "Analytic Confidence and Political Decision Making: Theoretical Principles and Experimental Evidence from National Security Professionals," *Political Psychology*, Vol. 39, No. 5 (2018), pp. 1069–1087.

[32] On how the standard of "guilt beyond a reasonable doubt" combines assessments of probability and confidence in this way, see Peter Tillers and Jonathan Gottfried, "Case Comment—*United States* v. *Copeland*, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond a Reasonable Doubt Is Unquantifiable?" *Law, Probability, and Risk*, Vol. 5, No. 2 (2006), pp. 135–157.

warned that there was no factual basis for concluding that the program was operative. Many of the Bush administration's critics argued that the lack of reliable evidence and the presence of dissenting views undermined the case for war, or at least justified delaying the invasion in order to gather more reliable information and to build greater consensus.[33] Note that this is distinct from critiquing how the Bush administration perceived the *probability* that Iraq was pursuing WMDs. The argument is instead that analysts and decision makers paid insufficient attention to the *confidence* surrounding this judgment.[34]

The degree to which probability assessments might change in response to new information plays an especially important role when it comes to timing high-stakes decisions. For example, when the Central Intelligence Agency (CIA) initially approached President Obama with indications that Osama bin Laden might be living in Abbottabad, Pakistan, Obama decided that the available information did not yet warrant taking military action. Instead, he asked intelligence analysts to conduct additional surveillance of the compound. Over the next eight months, the CIA's collection efforts became increasingly creative. When none of these efforts produced significant insights, President Obama eventually decided that he could no longer afford to delay the decision. What swayed Obama's thought process over time was not that he had gained new insight about the probability that bin Laden was living in Abbottabad, but rather that it had become increasingly clear that perceptions of this probability were unlikely to change in response to new information.[35]

These examples emphasize how assessments of probability provide only part of what decision makers need to know about the uncertainty surrounding high-stakes choices. Reliability of available information, range of reasonable opinion, and responsiveness to new information can also shape decision makers'

[33] The lack of direct evidence regarding Iraq's WMD programs was particularly relevant to judging the Bush administration's attempt to justify the invasion using the doctrine of "preemption," which is generally reserved for cases of clear and present danger. See Michael B. Doyle, *Striking First: Preemption and Prevention in International Conflict* (Princeton, N.J.: Princeton University Press, 2008). For broader arguments about how the ethics of using force depend on the reliability of the evidence decision makers possess, see Jeff McMahan, *The Ethics of Killing* (New York: Oxford University Press, 2002); and Michael Walzer, *Just and Unjust Wars*, 4th ed. (New York: Basic Books, 2006).

[34] For expressions of similar views, see Richard K. Betts, *Enemies of Intelligence: Knowledge and Power in American National Security* (New York: Columbia University Press, 2006), p. 116; Robert Jervis, "Reports, Politics, and Intelligence Failures: The Case of Iraq," Vol. 29, No. 1 (2006), p. 44; and Michael Morell, *The Great War of Our Time: the CIA's Fight against Terrorism from Al Qa'ida to ISIS* (New York: Twelve, 2014), p. 102.

[35] Peter Bergen, *Manhunt: The Ten-Year Search for Bin Laden from 9/11 to Abbottabad* (New York: Crown 2012); and Mark Bowden, *The Finish: The Killing of Osama bin Laden* (New York: Atlantic, 2012); Morell, *Great War of Our Time*, ch. 7.

willingness to take risks.[36] These attributes are far more relevant to foreign policy decision making than objectivity—which is, in any case, an impossible standard to apply when assessing uncertainty in international politics. This discussion highlights the importance of drawing clear distinctions between the probability and the confidence that foreign policy analysts assign to their judgments. And in the next section, we will see how foreign policy analysts can always assess probability in clear and structured ways.

Practical Implications of the Subjective Probability Framework

One of the main drawbacks with the frequency and propensity theories of probability assessment is that both frameworks generally preclude analysts from making precise judgments. Strictly speaking, these approaches can only indicate that the correct probability assessment falls within a plausible range of values.[37] And when analysts lack large volumes of well-behaved data, those ranges can be very broad. This is one of the clearest places where the subjective theory of probability departs from alternative frameworks. Conditional on believing that subjective probability assessments convey any insight whatsoever, it is always possible to summarize that insight with respect to a single, numeric point estimate. The subjectivity of these judgments is in fact the very thing that makes it possible to deal with them precisely, as analysts are always in a position to articulate their personal convictions in clear and structured ways.[38]

[36] Friedman and Zeckhauser, "Analytic Confidence and Political Decision Making," provide experimental evidence showing that these factors do, in fact, shape the way that national security professionals evaluate risky decisions.

[37] These ranges are often called *confidence intervals* (or credible intervals), but note that this only reflects one of the three elements of analytic confidence described earlier. One reason to shift existing discussions of analytic confidence toward more precise terminology is simply to avoid the confusion that the word "confidence" can generate—including, as we have seen in several places throughout the book, conflating the concepts of confidence and probability.

[38] It is important to note that this logic holds for estimating subjective probabilities, but not for estimating material costs and benefits. The difference stems from the fact that individuals can assign different utilities to these costs and benefits. These utility functions create "risk preferences" that shape the manner in which rational actors resolve ambiguity. For example, a gamble that offers a fifty percent chance of winning $0 and a fifty percent chance of $100 would have an expected monetary value of $50. But the expected *utility* of taking this gamble is not necessarily the utility that gamblers assign to a certain $50 payoff. Risk-averse individuals would value this bet at less than $50, while risk-acceptant individuals would value this bet at more than $50. For more on the contrast between resolving ambiguity about outcomes versus probabilities, see Ellsberg, "Risk, Ambiguity, and the Savage Axioms."

One way to demonstrate this point is with a thought experiment involving the comparison of lotteries. For example, consider the chances that a Republican candidate will win the next U.S. presidential election. Now consider which of the following two gambles you prefer. In the first gamble, you receive $1,000 if a Republican wins the next presidential election, and nothing otherwise. In the second gamble, we will flip a coin at the same time as the election results are announced. If the coin lands heads you will receive $1,000, and if the coin lands tails you will win nothing. If you prefer to bet on the election of a Republican presidential candidate, this indicates that you believe that outcome has a probability of at least fifty percent; otherwise you would have thought that the coin flip was a better bet. If you prefer to bet on the coin flip, it indicates that you believe the chances of a Republican candidate's election are no greater than fifty percent; otherwise you would have bet on that outcome. In principle, we can toggle this comparison of lotteries until you become indifferent between the two gambles. This method will reveal your personal conviction regarding the chances that the next president of the United States will be a Republican.[39]

Rational actors should have just one indifference point when comparing lotteries, or else they can be exploited by a betting strategy known as the Dutch Book. For example, imagine that you believe the chances of a Republican nominee winning the next U.S. presidential election are anywhere between forty and sixty percent, and you are truly indifferent among gambles within this range. This means that you would be willing to offer other gamblers favorable odds that a Republican will be elected, and that you would *also* be willing to offer other gamblers favorable odds that a Republican will *not* be elected. Smart people would take you up on both of these bets, thereby guaranteeing that you will lose money.[40]

Of course, you can always choose to avoid betting on an issue when you think that your probability of success is ambiguous, and thus people are rarely exploited by Dutch Books in real life. But the Dutch Book argument is not necessary for explaining why political analysts can always assess subjective probabilities with a single point estimate. As long as these analysts are willing to scrutinize their own beliefs carefully, it is impossible to be indifferent about what subjective probabilities should entail. For example, imagine that you believe the chances of a Republican candidate winning the next presidential election are somewhere between forty percent and sixty percent, and you believe

[39] For a review of similar techniques, see Paul H. Garthwaite, Joseph B. Kadane, and Anthony O'Hagan, "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, Vol. 100, No. 470 (2005), pp. 680–700.

[40] On the logic of the Dutch Book, see Ken Binmore, *Rational Decisions* (Princeton, N.J.: Princeton University Press, 2009), pp. 123–126.

that all estimates within this range are equally credible. The only sensible way to resolve that kind of ambiguity is to take the range's midpoint of fifty percent. Any estimate higher than fifty percent would implicitly give more weight to possibilities above that value; any estimate lower than fifty percent would implicitly give more weight to possibilities below that value; and either of those options would violate your own personal belief that all estimates within this range are equally credible.

Similar logic allows decision makers to resolve ambiguity when analysts offer diverging views. When analysts study the same information and offer different assessments, a decision maker should consider those assessments' relative credibility.[41] If the available estimates appear equally credible—and this includes situations in which decision makers have no basis for evaluating analysts' credibility at all—then the only reasonable way to resolve the disagreement is to average the available estimates together. Any other approach would violate a decision maker's own assumptions by implicitly assigning some estimates more weight than others. In cases where a decision maker believes that probability assessments are not equally credible, the only logical solution is to take a weighted average.[42]

Viewed in light of the subjective probability framework, seemingly intractable debates about assessing uncertainty become conceptually straightforward, as long as analysts and decision makers are willing to approach these debates in a principled manner. For example, we saw in the book's introduction that President Obama's advisers voiced a broad range of opinions about the chances that Osama bin Laden was living in Abbottabad. Red Teams assigned to draw skeptical conclusions placed these chances around forty percent. CIA Deputy Director Michael Morell offered an estimate of sixty percent. Most advisers placed their judgments around eighty percent, while the leader of the CIA unit assigned to track bin Laden said there was a ninety-five percent chance that he was living in the suspected compound. Multiple sources describe how President Obama found this discussion confusing, and how he concluded that the odds

[41] The weighted averaging approach described here is only valid when analysts base their assessments on the same body of information. When analysts form their estimates based on nonidentical information, then resolving their disagreements becomes more complex. See Richard Zeckhauser, "Combining Overlapping Information," *Journal of the American Statistical Association*, Vol. 66, No. 333 (1971), pp. 91–92 and Robert T. Clemen, "Combining Overlapping Information," *Management Science*, Vol. 33, No. 3 (1987), pp. 373–380.

[42] Thus, if one analyst estimates a probability as sixty percent, another analyst estimates that probability as ninety percent, and the latter judgment seems twice as credible, then the only coherent way for a decision maker to resolve the ambiguity is to believe that the probability is eighty percent. For more on the weighted averaging issue, see Jeffrey A. Friedman and Richard Zeckhauser, "Handling and Mishandling Estimative Probability," *Intelligence and National Security*, Vol. 30, No. 1 (2015), pp. 77–99.

were fifty-fifty. "Look guys, this is a flip of the coin," President Obama said according to one account. "I can't base this decision on the notion that we have any greater certainty than that."[43]

Two aspects of President Obama's conclusion are worth noting. First, a fifty-fifty estimate was remarkably conservative given the viewpoints that the president's advisers presented. The simple average of the four estimates described in the previous paragraph is sixty-nine percent. President Obama's coin flip analogy thus implicitly afforded special weight to the lowest estimates he received.[44] Yet it is not at all clear that these were the most credible positions on the table. The Red Teams were developed to offer deliberately pessimistic conclusions, not unbiased reflections of any analysts' actual views. Even Morell indicated that his sixty percent estimate involved intentionally low-balling the issue given how intelligence analysts had overestimated the chances that Iraq was pursuing nuclear weapons a decade earlier.

The point of this discussion is not to second-guess President Obama's judgment, but to show that no matter how much his advisers disagreed over the chances that bin Laden was at Abbottabad, there was no reason to believe that resolving this disagreement represented an intractable problem. As long as analysts are willing to assess uncertainty in transparent ways, and as long as decision makers are willing to consider the relative credibility of those judgments, then it is always possible to reconcile diverging opinions. Greater ambiguity actually makes it easier, not harder, to resolve these disagreements. In the extreme case where a decision maker has no reason to believe that any estimate is more credible than the others, then the only valid approach is simply to average the available assessments together.[45]

This reasoning also implies that any time foreign policy analysts formulate judgments using words of estimative probability, they also possess a coherent basis for converting those judgments into quantitative expressions. Chapter 1, for example, showed how U.S. Intelligence Community standards currently

[43] Bowden, *The Finish*, p. 163.

[44] This estimate may have reflected the well-established "fifty-fifty bias" in which individuals use this phrase to refer to situations in which they feel they have no idea about the right answer and not to indicate that the chances of a particular outcome are actually fifty percent. See Baruch Fischhoff and Wandi Bruine de Bruin, "Fifty-Fifty = 50%?" *Journal of Behavioral Decision Making*, Vol. 12, No. 2 (1999), pp. 149–163.

[45] As mentioned earlier, the weighted averaging approach depends on the assumption that analysts have based their estimates on similar information. This assumption appears safe in most high-level national security debates such as the discussion of whether Osama bin Laden was at Abbottabad. Deputy CIA Director Morell reportedly told President Obama during the meeting: "People don't have differences [in their probability estimates] because they have different intel. . . . We are all looking at the same things." Bowden, *The Finish*, p. 161.

define probabilistic terms with respect to numeric ranges. If analysts have no reason to believe that any value within the range they selected is more credible than others, then the appropriate way to resolve this ambiguity is to state the range's midpoint. If analysts instead believe that estimates toward one end of the relevant range are more plausible, then they can resolve this ambiguity using the logic described earlier in this section.

Though several of the words of estimative probability guidelines shown in chapter 1 entail more ambiguity, it nevertheless remains straightforward to convert these phrases into point estimates. For example, we saw in chapter 1 that the Defense Intelligence Agency currently recommends that analysts express uncertainty by dividing probability assessments into five equally spaced segments. All else being equal, one can assume that each segment spans twenty percentage points. From here, one can employ the previous logic for quantifying ambiguous estimates, interpreting the phrase "highly improbable" as meaning ten percent, "doubtful" as meaning thirty percent, and so forth. Of course, analysts can always shift these interpretations up or down if they believe that other judgments would be more credible on a case-specific basis.

This discussion shows that there is no instance in which foreign policy analysts can coherently offer a qualitative probability estimate but not a quantitative probability estimate. While the search for a middle ground between Sherman Kent's "mathematicians" and "poets" may make sense from an organizational or cultural standpoint, there is no reason to seek this middle ground based on logical principles alone. Scholars, practitioners, and pundits who prefer to leave their assessments of uncertainty vague cannot justify that practice on theoretical grounds. The fact that these probability assessments are subjective is, in fact, exactly what makes it possible to describe those judgments explicitly.

Conclusion

This chapter has argued that aversion to assessing uncertainty in international politics rests, at least in part, on misperceptions of what probabilistic reasoning entails. Much of the existing debate about this subject revolves around the expectation that assessments of uncertainty in international politics should obey the same kind of logic that characterizes statistics or quantum mechanics. Many observers appear to believe that if foreign policy analysts cannot meet those objective standards, then their beliefs cannot provide a credible basis for high-stakes decision making. This emphasis on objectivity is enshrined throughout official analytic standards for intelligence analysis and military planning, and it is often used to justify leaving controversial judgments deliberately vague.

These arguments are misguided. The chapter explained that assessments of uncertainty in international politics reflect a logic that is distinct from statistical frequencies or mechanical propensities. We have seen that objectivity is an unrealistic standard for foreign policy analysis, and that no foreign policy analyst could ever accept the implications of pursuing that goal. We have seen that even when assessments of uncertainty are subjective, they can still vary in terms of credibility, and that explicitly distinguishing between probability and confidence can help to convey that variation to policymakers.[46] And the chapter explained that subjectivity is ultimately the main reason why foreign policy analysts can articulate their assessments of uncrtainty in clear and structured ways. It is, in fact, the objective theories of probability that usually force analysts to accept irreducible ambiguity.

Of course, just because some approach makes sense in principle does not mean that it also works in practice. The remainder of the book will thus explore a range of psychological and political obstacles that may prevent foreign policy analysts from implementing the logic that this chapter has described. It is nevertheless important to understand that the only valid objections to pursuing this approach are empirical, not theoretical. To the extent that scholars, practitioners, and pundits wish to avoid assessing the uncertainty surrounding foreign policy debates, it is not enough to say that those assessments are subjective, and it is wrong to say that those judgments are incoherent.

Moreover, the chapter has argued that understanding the theoretical foundations of subjective probability can help foreign policy analysts and decision makers to resolve confusion that otherwise seems intractable. The difficulty that President Obama encountered when debating intelligence during the search for Osama bin Laden is a prime example. Based on the evidence available at the time, there may have been no "right answer" when it came to estimating the chances that bin Laden was living in Abbottabad. But there was also no reason to get stuck addressing this issue. There were, in fact, logically valid and reasonably straightforward ways for President Obama and his advisers to resolve the ambiguity surrounding one of their seminal decisions. And in the next two chapters, we will see how foreign policy analysts and decision makers are remarkably capable of handling this challenge.

[46] Friedman and Zeckhauser, "Analytic Confidence and Political Decision Making," provide experimental evidence showing that national security professionals interpret and react to these distinctions in sensible ways.