Capturing the Potential of Outlier Ideas in the Intelligence Community

Clint Watts and John E. Brennan

***Outlier**:*

—A data point far outside the norm for a variable or population;

—An observation that "deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism";

—A value that is "dubious in the eyes of the researcher";

—A contaminant.

Source: J. Osborne, "The Power of outliers (and why researchers should ALWAYS check for them)," http://pareonline.net/htm/v9n6

> *In war you will generally find that the enemy has at any time three courses of action open to him. Of those three, he will invariably choose the fourth.*
>
> —Helmuth Von Moltke

With that quip, Von Moltke may have launched a spirited debate within his intelligence staff. The modern version of the debate can be said to exist in the cottage industry that has been built on the examination and explanation of intelligence failures, surprises, omissions, and shortcomings.[1] The contributions of notable scholars to the discussion span multiple analytic generations, and each expresses points with equal measures of regret, fervor, and hope. Their diagnoses and their prescriptions are sadly similar, however, suggesting that the lessons of the past are lost on each succeeding generation of analysts and managers or that the processes and culture of intelligence analysis are incapable of evolution. It is with the same regret, fervor, and hope that we offer our own observations on avoiding intelligence omissions and surprise. Our intent is to explore the ingrained bias against outliers, the potential utility of outliers, and strategies for deliberately considering them.

Of all the examinations of intelligence surprise and failure, Richards Heuer provides perhaps the most succinct characterization of the problem:

> *Major intelligence failures are usually caused by failures of analysis, not failures of collection. Relevant information is discounted, misinterpreted, ignored, rejected, or overlooked because it fails to fit a prevailing mental model or mind-set.*

In his construction, Heuer identifies three reasons information is omitted from consid-

[1] David Moore (2011), Rob Johnston (2005), Warren Fishbein and Gregory Treverton (2004), Jack Davis (2003, 2002), Richards Heuer (1999), Christopher Brady (1993), J. J. Wirtz (1991), Ephraim Kam (1988), Richard Betts (1982, 1978), Abraham Ben-Zvi (1979), Richard W. Shryock (1977), Avi Schlaim (1976), Michael Handel (1976), Charles Fisk (1972), and Klaus Knorr (1964) have explored some of the well-known ones. See source note for bibliographic information on these works.

Outliers are data and hypotheses that analysts may too quickly dismiss.

eration, one that leads to its misinterpretation, and one that leads to information being unavoidably overlooked, suggesting it is outside the analyst's view or access. In Heuer's analysis, 60 percent of the sources of the problem may be attributable to analysts choosing to throw out data. What kind of information is likely to have been lost in the discarded data? Johnston and others who see confirmation bias at play would describe it as information that does not conform to an analyst's expectations, fails to align with the Intelligence Community's (IC) preconceived notions of a problem, or directly challenges analytic lines in which managers have confidence. These discarded pieces of data may be viewed as outliers.

What Are Outliers and Why Do They Matter?

Outliers have an unfortunate reputation: they are suspect, different, error, deviation, fraudulent. Throughout life we are trained and encouraged to think of them negatively. If they do not fit the data of our normal distribution, we are often encouraged to ignore them, purge them, or delete them. This tendency is very powerful and very useful. It keeps us from pursuing many strange ideas.

What are outliers in the context of the intelligence profession? Outliers are data and hypotheses that analysts may too quickly dismiss. They may be the imaginative, even prescient analyses policymakers cannot bring themselves to believe. Intelligence analysts generally possess healthy doses of skepticism to help them avoid the pitfalls of hubris and self-delusion, but, sadly, this is insufficient, for the outliers that ultimately prove to be the seeds of surprise are outlandish, unthinkable, and wholly anomalous. For example:

- Russia would destabilize the balance of power by deploying tactical nuclear missiles in Cuba.[2]
- North Vietnam would invade South Vietnam in the spring of 1975, resulting in the complete collapse of the South Vietnamese government.[3]
- An Islamic cleric would distribute sermons via cassettes, and the Iranian people would then overthrow their government.[4]
- Yugoslavia would not remain intact through the 1990s.[5]
- A construction company owner from Saudi Arabia would declare war on the United States and destroy two US embassies, a US Navy destroyer, and conduct an attack on US soil that would kill thousands.
- Saddam Hussein would abandon his weapons of mass destruction (WMD) program.
- A fruit vendor's self-immolation in Tunisia would set off a firestorm of demonstrations for self-determination across the Near East.

Examples from the finance industry.

Not only do we find outliers impossible to take seriously, but we dismiss the accurate reader of outliers as someone with extrahuman powers. Take Warren Buffett. His ability to beat his peers and the markets on a consistent basis has earned him the moniker the "Oracle of

[2] The Special NIE on Cuba records the IC's unwillingness to support the hypothesis of nuclear missiles in Cuba. This required analysts to ignore eight refugee reports (outliers) out of thousands of reports as bad data. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol51no3/revisiting-sherman-kent2019s-defense-of-snie-85-3-62.html

[3] Interagency Intelligence Memorandum, "Response to National Security Study Memorandum 213--Part I: Intelligence Appraisal—Factors Influencing the Course of Events in the Republic of Vietnam over the Next Five Years," 18 November 1974. Accessed 6 December 2011 at http://gateway.proquest.com.openurl?url_ver=Z39.88-2004&res_dat=xri:dnsa&rtf_dat=xri:dnsa:article:CVW01271.

[4] NSC staffer Gary Sick later concluded, "The Iranian revolution…refused to conform to the conventional wisdom of the day, and contemporary analyses often had more to say about the prejudices and assumptions of the observer than about the new reality being created in the mosques and in the streets of Iran." Gary Sick, *All Fall Down: America's Tragic Encounter with Iran* (New York: Random House, 1985), 106.

[5] In this case, the Intelligence Community correctly estimated the situation, but was considered the outlier in a policy community unwilling to accept that forecast. (Based on interview with the NIE author, August 2011).

Omaha." Yet he possesses no oracular powers. Instead, he deeply investigates the company he is considering investing in. He visits it and gets to know the leaders and the customers. Then he conducts a detailed financial analysis of the firm to determine its worth, and he patiently waits for the markets to distort the price to a low enough level that he is willing to buy in. It is nothing more than the wisdom of the outlier trouncing the wisdom of the crowds.

In hindsight, Buffett looks like a genius, but he looked the fool when he put his entire life savings into a single company for his first investment, with no diversification and no hedging, something his peers could have easily perceived as blind recklessness. He defied all the basic conventions and accepted norms of investing. It appears that making the first outlier call requires not only deep conviction backed up by solid analysis, but also professional courage. Not everyone demonstrates this courage. Our personal relations and our desire for group cohesion often stifle dissent.[6] For example,

> *An economist at Yale University,* [Robert] *Shiller is a leading scholar, a tenured professor, an innovator, and the author of the 2000 book* Irrational Exuberance, *which warned the boom in the tech stocks was really a bubble set to burst. He wrote in a 2005 edition of* Irrational Exuberance *that there could be "a substantial increase in the rate of personal bankruptcies, which could lead to a secondary string of bankruptcies of financial institutions as well." A recession would follow, perhaps even "worldwide." Thus, Robert Shiller can reasonably claim to be one of the very few economists who predicted the disaster of 2008. Unlike anyone else, he was a member of a panel that advises the president of the Federal Reserve Bank of New York. But when the advisory panel met in 2002 and 2003, Shiller did not shout and jump up and down on the table. "I felt the need to use restraint," he recalled. The consensus in the group was that there was no bubble and no need to raise interest rates. To suggest otherwise was distinctly uncomfortable. Shiller did make this point, but "I did so very gently, and felt vulnerable expressing such quirky views. Deviating too far from consensus leaves one feeling potentially ostracized from the group, with the risk that one may be terminated."*[7]

This example demonstrates how uncomfortable it can sometimes be to hold a contrarian or outlier idea.

This example demonstrates how uncomfortable it can sometimes be to hold a contrarian or outlier idea.

Examples from within the IC.

Consider the "heretics" of the IC who have dabbled in the dark arts of open-mindedness and radical skepticism. One of our earliest cases can be traced to the autumn of 1952. That year a group of open-source translators in the Foreign Documents Division and analysts at CIA noticed differences in the way Russian and Chinese propaganda treated common subjects, especially in their descriptions of communism. The analysts focused on specific omissions in the Chinese—a striking example was absence of the term "Stalinism"—even while propaganda continued to embrace Marxism-Leninism. The group made other observations in 1953 and in 1954 after the death of Josef Stalin. These observations, however, failed to convince their colleagues, managers, and policy-makers who were convinced that communism was an indivisible bloc. They remained unconvinced a Sino-Soviet split was happening until Russia and China fought each other briefly in 1969.[8]

[6] There are strong incentives for analysts to be loyal to their issue managers and group chiefs.
[7] D. Gardner, *Future Babble* (New York: Dutton, 2011), 106–107.

More surprising, both the "at-large" and "professional" responses selected the "Status Quo/No Change" option at high rates.

It turns out that our analysts were not alone in this conviction. An even greater heretic than the analysts on China drew the same conclusion almost half a year before the IC analysts first gave credence to the idea. In February 1952, Franz Borkenau, a "student of history, Comintern organization man, freelance journalist and historian, father of Kremlinology, and philosopher of history" wrote an analysis for the US Department of State, which began:

> *In the view of this writer a profound conflict between the Communist regimes of Russia and of China is in the long run as certain as anything predictable in politics. Its necessity can be demonstrated by a very simple formula. Totalitarian regimes live by an inherent urge to establish their absolute, "totalitarian" control as far as they can. A totalitarian regime, and more especially the Russian regime, is striving for absolute world domination. It therefore cannot have genuine allies, but must try to subjugate everything within its reach. This is incompatible with the obvious Chinese quest for national independence.*[9]

How might US foreign policy toward China have been different if the IC and the White House had come to accept the feasibility of this outlandish idea in 1953 or 1954? Perhaps it would not have changed our commitment to the Chinese Nationalists and Taiwan, and it is unlikely to have colored our thinking toward China's entry into the Korean War. Yet imagine if the United States had taken *before* 1967 Richard Nixon's advice of that year, "We simply cannot afford to leave China forever outside the family of nations, there to nurture its fantasies, cherish its hates, and threaten its neighbors."[10] Would the "Domino Theory" have carried the same weight? Would we have intervened in Vietnam to the degree we did?

The Quest for the Wisdom of Crowds

The first phase of collection and the insight.

Our interest in outliers was born out of Internet-based surveying that Clint Watts undertook on 2 January 2011. The purpose of that survey was to test the ability of crowds to make accurate hypotheses about future counterterrorism issues in the event of Usama Bin Laden's death. The survey attracted a little over 30 voters (we will call this the "at large" sample). An additional 30 respondents (professionals known by the author to have significant terrorism and counterterrorism expertise) were queried separately, resulting in two samples of 30 or more voters each.[11]

As we reviewed the survey design and sample test results, we noted how frequently respondents selected what appeared to be safe or conventional ?nswers. The "at-large" sample made selections that would be expected from people fed a steady diet of mainstream media. The professionals' selections suggested they had some unique knowledge that steered them away from popular sentiment—although they too herded or clustered together around certain answers. More surprising, both the "at-large" and "professional" responses selected the "Status Quo/No Change" option at high rates across all questions (see graph on facing page).

The results of this initial survey prompted us to consider an entirely different hypothesis: outliers in the survey might provide more important insights into a post–Bin Laden, non–status quo world. Those that responded outside the "typical" responses of their professional group and especially

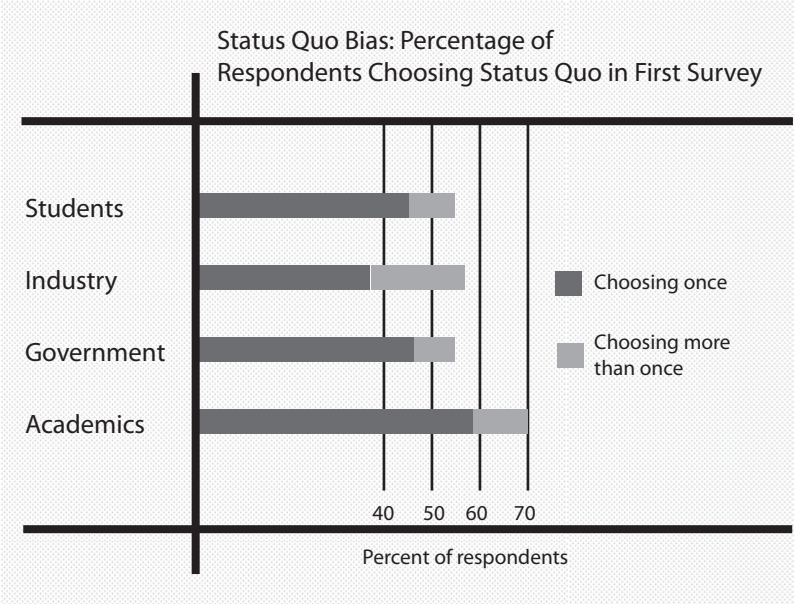
[8] Harold Ford. "Calling the Sino-Soviet Split: The CIA and Double Demonology." *Studies in Intelligence.* Winter 98-99. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/winter98_99/art05.html

[9] J. Tashjean, "The Sino-Soviet Split: Borkenau's Predictive Analysis of 1952." *China Quarterly,* No. 94 (1983): 342–61.

[10] "Asia After Vietnam," *Foreign Affairs,* Vol. XLVI (October 1967): 121.

[11] The authors do not believe that a sample of 30 adequately represents a "crowd." However, such a sample size could easily reflect the approximate size of a team of community analysts grappling with an intelligence problem.

for an option other than the status quo became particularly interesting to us. Those providing outlying responses also often provided their reasoning in the question's comment section. This suggested that they anticipated their answer broke from mainstream views. Their responses and comments provided the richest insights and prompted us to alter the intended purpose of the surveys. This insight prompted us to ask, "How can we find the most insightful outlier opinions in a crowd of responses?"



The second phase of collection.

From March to April 2011, we crafted a more exhaustive survey designed to evaluate several key dimensions of al Qaeda's future. The "Al Qaeda's Strategy 2011–2012" survey queried visitors to the Watts blog, SelectedWisdom.com, and personal contacts of ours beginning on 27 April 2011. The strategy poll asked respondents to answer 11 questions on the future of al Qaeda in a post–Bin Laden world. The survey concluded by asking respondents to estimate their confidence level and to provide demographic background on their profession, education, international experience, and information sources. Altogether, 325 respondents answered this survey to some degree (82 percent completed all questions).[12] Little did we know how prescient the survey would become.

Fortuitous events

US Special Operations forces killed Usama Bin Laden on 2 May 2011, only five days after we had initiated the "Al Qaeda's Strategy 2011–2012" survey. Bin Laden's death provided a unique opportunity to compare perspectives immediately before and immediately after the elimination of the organization's key leader. On the morning of 2 May 2011, we conducted a third web-based poll, including again the questions from the initial "post–Bin Laden" poll of 2 January 2011 and some of the same questions queried the week before in the "Al Qaeda's Strategy 2011–2012" survey. From 2 May 2011 through 20 May 2011, 160 voters participated, answering 11 questions about the implications of Bin Laden's death.[13] Like the previous week's poll, this survey asked respondents to rate their confidence and provide demographic information on their profession and education. Combined, the two polls engaged just under 400 respondents, who answered completely. The survey provided raw material with which to evaluate the notion of outliers.

Interpreting the results.

In the second and third surveys we deliberately sought outliers. This required a much different survey design. First, the poll consisted of cognition problems, and most of the questions were highly complex counterterrorism issues truly requiring some expertise. For example, the question, "What will be the chief consequence of Usama bin Laden's death?" provided voters 12 options, all of

[12] The AQ Strategy poll collected responses from 325 unique voters. However, only 82 percent of the respondents completed all of the survey's questions. The responses were collected between 27 April 2011 and 17 May 2011.

[13] The post–Bin Laden poll executed on 2 May 2011 included some of the same voters from the "Al Qaeda's Strategy 2011–2012" poll from the week of 27 April 2011 and the first post–Bin Laden poll conducted on 2 January 2011. For this third overall poll and second "post–Bin Laden" poll, there were 160 total respondents and 85 percent completed all questions asked.

which were interrelated and required respondents to think through the dynamics of each response in picking their best estimate. This complexity (not unlike real life) appeared in most cases to push voters to the status quo.

Second, we wanted to understand the relationship between the opinions of respondents and the information sources upon which they based their opinions. Our crowdsourced sample predominately consisted of English-speaking Americans with limited international travel. For the most part, these respondents receive terrorism-related information via mainstream television and newspapers, with some social media commentary on mainstream media content. This limited Western information stream contributed to herding around common Western media perspectives. For example, the "Post-UBL Survey" (2 May) asked, "Which al Qaeda leader has the necessary attributes to become al Qaeda's global leader?" The majority selected "Ayman al-Zawahiri"—a commonly cited figure in global media usually referred to as "the number-two man in al Qaeda." Selecting this response was logical and expected. However, the first person identified as interim leader of al Qaeda and potential successor to Bin Laden was instead Sayf al-Adel, a longtime al Qaeda veteran not well known to international media. Of 130 respondents to this question, only five selected Sayf al-Adel, and only one respondent pointed to *al Jazeera* as an information source.[14]

While Bin Laden's successor ultimately turned out to be Ayman al-Zawahiri, the crowd demonstrated the potential to be swayed by popular sentiment or media reporting. For example, the academics in the sample clustered around one answer before Bin Laden's death and then shifted en masse to another answer after his death. This raises important questions about whether to use outside academic experts to fill knowledge gaps in government communities or for assistance with estimating the future course of events.

Third, we noted from Philip Tetlock's findings in *Expert Political Judgment* and Dan Gardner's commentary on them in *Future Babble* that confidence levels may be, at best, immaterial and at worst, deceptive. Even so, we asked respondents to estimate their confidence in the responses they provided in the survey.

Tetlock conducted an experiment over many years collecting more than 27,000 expert judgments. Tetlock found most expert predictions were no more accurate than random guessing. Of particular interest to the question of confidence, Tetlock looked specifically at the accuracy of media pundits, concluding that the bigger the media profile of the expert, the lower the expert's accuracy. Gardner argued that media pundits share a common characteristic: confidence. A talking head who hedges or appears dubious does not attract the same ratings that a bold and confident one does. We prefer confidence, but are we good judges of confidence?

Admittedly our sample of "experts" differed from Tetlock's "experts." While some readers may argue that Tetlock's experts do not resemble our experts in the way they consider problems, what we found in our suvey is consistent with Tetlock's findings that accuracy and confidence levels do not necessarily go together. We found that respondents with master's degrees were slightly more confident, on average, than respondents with PhDs. Even more interesting, these respondents were more confident than individuals with bachelor's degrees, but so were respondents with associate degrees or only high school diplomas!

The Theory of the Wisdom of Outliers: Hunting for Red and Brown Foxes

To describe the kinds of cognitive processes he saw demonstrated in his experiments, Tetlock borrowed from a Greek saying, "The fox knows many things, but the hedgehog knows one big thing," that was popularized in Isaiah Berlin's 1953 essay *The Hedgehog and the*

[14] Poll results #3, http://selectedwisdom.com/?p=277

Fox.[15] In his research, Tetlock determined that it was *how* experts think, not *what* they think, that resulted in accurate future forecasts, and he characterized his experts into two categories of thinker. The better-than-average prophets he labeled "foxes."[16] Foxes

> *thought very differently…they had no template. Instead, they drew information and ideas from multiple sources and sought to synthesize it. They were self-critical, always questioning whether what they believed to be true really was. And when they were shown they had made mistakes, they didn't try to minimize, hedge, or evade. They simply acknowledged they were wrong and adjusted their thinking accordingly. Most of all, these experts were comfortable seeing the world as complex and uncertain—so comfortable that they tended to doubt the ability of anyone to predict the future. That resulted in a paradox: the experts who were more accurate than others tended to be much less confident they were right.*[17]

We wanted to attempt to determine if education, profession, and information sources mattered [in the creation of outlying ideas].

The other class Tetlock called hedgehogs. These were individuals who were

> *not comfortable with complexity or uncertainty …they sought to reduce the problem to some core theoretical theme…they used that theme over and over, like a template, to stamp out predictions …these experts were also more confident than others that their predictions were accurate…why wouldn't they be? They were sure their One Big Idea was right and so the predictions they stamped out with that idea must be too*[18]

Where Tetlock's labels accounted only for attributes of thought he saw in his experts, we also wanted to account for the demographic attributes of individuals we surveyed, in order to determine if those qualities (i.e., education, profession, information sources) are influential. To characterize the participants in our surveys we adopted Tetlock's labels and created variations of our own. If respondents' answers to the demographic questions were within the 95 percentile of responses, they were considered to have typical demographic characteristics. If a respondent's demographic response represented fewer than 5 percent of our sample, then the respondent was regarded as atypical demographically. We applied the same concept to respondents' answers to forecasting questions, identifying those that were atypical substantively.

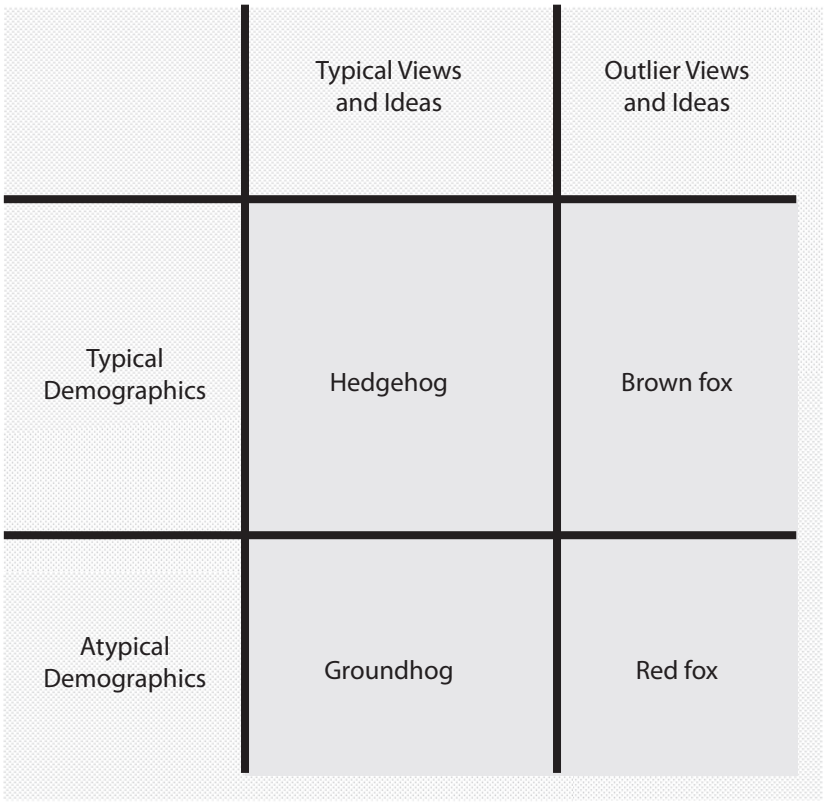
With these two measures, we created four categories of respondents:

- Hedgehogs—those with typical demographic characteristics who offered typical substantive responses
- Groundhogs—those with atypical demographics but who provided typical substantive responses
- Brown Foxes—those with typical demographics but who provided outlier substantive responses
- Red Foxes—those with outlier demographic and outlier substantive responses.

We suspect that analysts could be characterized in one of these four ways, though it is likely that their characteristics will change over the course of a

[15] Gardner, 27.
[16] Tetlock believes that we are not permanently typecast in these categories; rather, each time we make an analytic judgment, we could be acting as a fox or a hedgehog.
[17] Gardner, 26–27
[18] Ibid., 26.

career, depending on specific assignments, the demographics of an analytic shop, and the knowledge and experience the analyst has on a particular problem. For instance, according to an analyst working on the CIA analytic team researching Iraq's WMD program in 2003, the team contained analysts who could be described as hedgehogs and one groundhog. The groundhog acknowledged never going back and reviewing the earliest Curveball reporting, relying rather on the synopsis provided by the hedgehogs on the team.[19] When the groundhog went back and read the original reports, that analyst started to demonstrate fox-like qualities.



Armed with these definitions, we began our analysis by recoding all of the responses into choices that were consistent with the sample and those that were outliers. For this recoding, we gave full outlier credit to any respondent who picked a choice that 5 percent of the population or less selected. We gave partial credit to respondents who were right at the outlier dividing line, especially when the outliers made up 5 to 7 percent of the responses and it was impossible to distinguish them. Some of our questions did not result in a clear outlier minority; instead, the respondents might have split 60:40 on a question, so we gave those in the 40-percent group partial outlier credit.

Next we looked for those respondents who provided atypical demographic responses and outlying responses, i.e., substantive opinions outside of the norm. Then we rank-ordered them based on the degrees to which responses were outliers and then filtered the respondents into our four categories. Out of 260 respondents, 73 percent fit the "typical" demographics. However, this same group also provided outlying ideas, so it was necessary to focus on the respondents with the highest degree of outlier ideas. We selected 13 red foxes (4 percent) and 14 brown foxes (4.3 percent). The 21 groundhogs were easily identifiable and formed only 6.5 percent of the sample.

Finally, we went back to the original survey and isolated the responses from the red and brown foxes to compare their ideas with those of the overall population. As expected, we found that these foxes had ideas different from the main population. For example, when asked what the most likely strategy for al Qaeda was in a post-UBL environment, the foxes' selections spread across the spectrum of choices. In this case, the "irregular warfare attacks inside Pakistan to erode Pakistani-US cooperation" choice received the highest votes (22 percent) by the population and the strongest concentration of interest by the foxes.

In one experiment we cannot begin to understand whether attention should be given to the outlier ideas of brown foxes, red foxes, or areas of significant overlap between them. Additional experimentation across a

[19] John E. Brennan interview with the analyst, 2010.

number of scenarios inside and outside this setting (on the scale of Professor Tetlock's work) is required to determine if there is a definitive rule. Hopefully, crowdsourcing research, which is being sponsered by Intelligence Advanced Research Projects Activity, will shed more light on this question. Therefore, we will conclude our discussion of the experiment and move to suggest a potential application of the theory.

A Potential Application of the Theory

Imagine back to the summer and fall of 2002. Intelligence analysts across the US government were assessing Saddam Hussein and Iraq's WMD capability. Most inside the government were convinced Saddam maintained and continued to develop a major capability representing an existential threat to the United States. Some outlying opinions from intelligence officers and analysts suggested that may not have been the case. However, those opinions were few and spread throughout the government. This dispersed, dissenting perspective ultimately proved true, but it was overwhelmed by the majority opinion, led by confident experts, hedgehogs.

Though the IC has several methods devoted to countering groupthink (e.g., analysis of competing hypotheses), it is still a human process, as we saw with Dr. Shiller's unwillingness to voice a strong opinion in the Federal Reserve

Internal polling designed to identify outliers can provide a systematic method to analytically hedge against the potential for groupthink.

meetings, and it is subject to human frailties. These methods are applied according to the predilections of the individual managers and analytic teams involved. Few of these teams are likely able to pattern-match and repeat the process of their best peers. The level of rigor is no doubt subjective and variable.

Internal polling designed to identify outliers can provide a systematic method to analytically hedge against the potential for groupthink. If executed properly, analytic managers can consciously explore alternative perspectives from otherwise overlooked foxes. Survey populations that overlap between the at-large population and the professional population can have interesting results, especially if the outside population includes members from the intelligence target group, region, country, or acceptable proxies. These polls could serve as preludes to analyses of competing hypotheses, helping identify the alternatives analysts are still worried about.

Here is how the process might work.

- First, the organization would survey all relevant members on a particular topic.
- Second, an independent arbiter (perhaps an ombudsman) could identify and vet potential outliers and their alternative analyses.
- Third, a leader would evaluate the entire analytical portfolio on the issue queried and compare it with the results of the internal poll. In the case of WMD for example, the organizational leader might note that almost 100 percent of the organization's analytical horsepower is pursuing the hypothesis that Iraq sustains a massive WMD program. These results are compared with the internal poll results where 5 percent of respondents believe Saddam has no WMD capability.
- Fourth, the leader decides to make an analytical hedge based on the imbalance of resources committed to hypotheses. The leader decides to move a higher percentage of the organization's analytical effort from the majority hypothesis that Iraq had WMD and then dedicate these resources to exploring an alternative hypothesis; Iraq does not have WMD.
- Fifth, the leader empowers this alternative analytical group by staffing it with the very outliers that selected the alternative possibility during the internal poll (rather than staffing the endeavor with members of the majority opinion who are inherently primed to discount alternative perspectives.)[20] The leader must

also provide the alternative analysis team with time and dedicated collection to explore their hypothesis.[21]

The result is an organization responsibly leveraging its people and resources to identify and examine outlier ideas.

A possible by-product of this exercise might be the insight that comes from looking at trends in personal confidence levels. Another useful insight might come in the form of seeing how concentrated analysts' information sources are and whether at-large, academic researchers and independent analysts use something new that the inside analysts might benefit from. In either case, these data are intended for mission-management purposes only, not something that would be shared with intelligence consumers. Intelligence managers would look at the responses and decide which potential hypotheses to hedge against by exploring them more deliberately.

Conclusion

It seems axiomatic: surprising outcomes were outliers until they occurred. If the IC wants to deliberately and systematically counter groupthink and reduce the potential for surprise, it should consider standard methods, like surveys, to elicit and then identify outlier ideas. The nascent theoretical method described in this article requires additional scrutiny and experimentation.

The concept of confidence also deserves fresh scrutiny. Policymakers may feel comfortable in receiving such barometric readings, but the evidence not only from our experiment but also from Tetlock's work suggests that effort put into measuring confidence levels may be futile, with the time better spent on clearly identifying and explaining facts and underlying assumptions and dynamics.

The process of selecting and managing the inputs of outside advisers and experts also warrants further review. The groupthink seen among academics, not only in this study but also in the studies related to Tunisia and Egypt, is troublesome.[22] Furthermore, the evidence from our survey demonstrating how quickly academics moved from one option to another following Bin Laden's death, prompts many questions.

❖ ❖ ❖

[20] Some readers might feel this isolates groupthink inside the alternative analysis team, and this may be true. In the lead author's opinion, the outlier ideas are often spread thinly around the analyst population; concentrating them, temporarily, in one team brings a set of "doubters" together who may convince themselves otherwise.

[21] As "dedicated" as any collection is.

[22] Bush School Study, 2011 for the Deputy DNI for Intelligence Integration and ODNI I-21 Study, 2011.

Source Note:

Examples of the literature on failure and remediation of intelligence analysis include: Richards Heuer, *Psychology of Intelligence Analysis* (Washington, DC: CIA, 1999). Heuer also reviewed Christopher Brady, "Intelligence Failures: Plus Ca Change…" in *Intelligence and National Security* 8, No. 4 (October 1993); N. Cigar, "Iraq's Strategic Mindset and the Gulf War: Blueprint for Defeat," *The Journal of Strategic Studies* 15, No. 1 (March 1992); J. J. Wirtz, *The Tet Offensive: Intelligence Failure in War* (Ithaca: Cornell University Press, 1991); Ephraim Kam, *Surprise Attack* (Cambridge, MA: Harvard University Press, 1988); Richard Betts, *Surprise Attack: Lessons for Defense Planning* (Washington, DC: Brookings, 1982). Abraham Ben-Zvi, "The Study of Surprise Attacks," *British Journal of International Studies* 5 (1979). *Iran: Evaluation of Intelligence Performance Prior to November 1978* (Staff Report, Subcommittee on Evaluation, Permanent Select Committee on Intelligence, US House of Representatives, January 1979); Richard Betts, "Analysis, War and Decision: Why Intelligence Failures Are Inevitable," *World Politics* 31, No. 1 (October 1978); Richard W. Shryock, "The Intelligence Community Post-Mortem Program, 1973–1975," *Studies in Intelligence* 21, No. 1 (Fall 1977); Avi Schlaim, "Failures in National Intelligence Estimates: The Case of the Yom Kippur War," *World Politics* 28 (April 1976); Michael Handel, *Perception, Deception, and Surprise: The Case of the Yom Kippur War* (Jerusalem: Leonard Davis Institute of International Relations, Jerusalem Paper No. 19, 1976); Klaus Knorr, "Failures in National Intelligence Estimates: The Case of the Cuban Missiles," *World Politics* 16 (1964).