Rethinking the training of intelligence analysts

Article *in* Intelligence & National Security · February 2016

DOI: 10.1080/02684527.2016.1147164

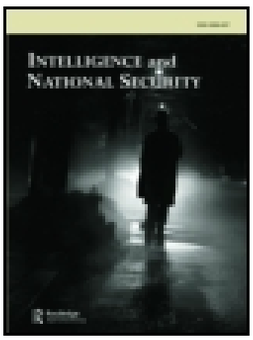
CITATIONS
14

READS
1,489

1 author:

Welton Chang
Johns Hopkins University
11 PUBLICATIONS **82** CITATIONS

SEE PROFILE



Intelligence and National Security

ISSN: 0268-4527 (Print) 1743-9019 (Online) Journal homepage: http://www.tandfonline.com/loi/fint20

Rethinking the training of intelligence analysts

Welton Chang & Philip E. Tetlock

To cite this article: Welton Chang & Philip E. Tetlock (2016): Rethinking the training of intelligence analysts, Intelligence and National Security, DOI: 10.1080/02684527.2016.1147164

To link to this article: http://dx.doi.org/10.1080/02684527.2016.1147164

Published online: 29 Feb 2016.

INTELLIGENCE AND NATIONAL SECURITY, 2016
http://dx.doi.org/10.1080/02684527.2016.1147164

Routledge
Taylor & Francis Group

Rethinking the training of intelligence analysts

Welton Chang and Philip E. Tetlock

ABSTRACT

Despite intense scrutiny and promised fixes resulting from intelligence 'transformation' efforts, erroneous analytic assessments persist and continue to dominate news coverage of the US intelligence community. Existing analytic training teaches analysts about common cognitive biases and then aims to correct them with structured analytic techniques. On its face, this approach is eminently reasonable; on close inspection, incomplete and imbalanced. Current training is anchored in a mid-twentieth century understanding of psychology that focuses on checking over-confidence and rigidity but ignores the problems of under-confidence and excessive volatility. Moreover it has never been validated against objective benchmarks of good judgment. We propose a new approach: (a) adopting scientifically validated content and regularly testing training to avoid institutionalizing new dogmas; (b) incentivizing analysts to view training guidelines as means to the end of improved accuracy, not an end in itself.

Whenever the intelligence community falters – such as under-connecting the 9/11 dots or over-connecting the WMD dots; under-estimating Russia's response to Ukrainian events or over-estimating the spread of democracy following the Arab Spring; under-rating the Islamic State or over-rating the Iraqi army – we hear demands to improve intelligence analysis.[1] The headline-grabbing reforms usually involve installing new leaders and redrawing organizational charts and budgets. Surprisingly little attention, though, has been directed at the essential but lower-profile task of improving training for line analysts, of giving them the cognitive skills and tools they need to assign realistic probability estimates to outcomes policymakers care about.[2] Training has remained remarkably static amidst the big structural changes proposed by Presidential Commissions and Congressional investigations over the last 20 years.[3]

The reasons for this inertia are many and most are beyond the scope of our study. But a core one is epistemological. The training programs adopted by organizations reflect deeply-ingrained assumptions about the knowledge their employees need to succeed and how to convert that knowledge into real world performance.[4] The United States intelligence community is no exception. The community invests in training systems that teach analysts about the dangers of biased thinking identified from past mistakes – and about how to deploy structured techniques to check those biases.[5]

CONTACT Welton Chang welton@sas.upenn.edu

[1]Lake and Dickey, 'US Spies Said No Russian Invasion'; Leibovich, 'Slam-Dunk'; Miller, 'Former CIA official'.
[2]Cooper, *Curing Analytic Pathologies,* 8; Rubin, Statement to Investigations Subcommittee.
[3]Kean and Hamilton, *The 9/11 Report*; IRTPA, 118; Duelfer, *Comprehensive Report of Iraq's WMD*. Richards Heuer's *Psychology of Intelligence Analysis* has not been revised since its publication, though training volumes on structured techniques are occasionally updated. We see these revisions as largely cosmetic, as the new techniques, like older ones, are untested and of unknown efficacy.
[4]March, 'Exploration and Exploitation,' 71–87; Tetlock, 'Cognitive Biases and Organizational Correctives,' 293–326; Weick, *Sensemaking in Organizations*.
[5]Heuer, *Psychology of Intelligence Analysis*; Marrin, 'Training and Educating Analysts,' 132–3.

Whether this training works is very much an open question. In the first half of this article, we discuss the logical and psychological shortcomings of the current training regimen. Current training draws on a skewed diagnosis of thinking errors and is grounded in dated premises. It focuses on the risks of over-confidence, over-assimilating new cases to historical analogies, and groupthink while ignoring the risks of activating opposing biases such as under-confidence, insensitivity to base rates and analysis-paralysis. Training fails to alert analysts to the dangers of overcorrection – and to offer guidance on error management.[6]

Current training is also methodologically imperfect. It encourages analysts to rely on vague verbiage judgments that are difficult – often impossible – to assess for accuracy. The consequences are far-reaching. The self-correcting feedback loops for linking theories, hypotheses and evidence operate less efficiently when there is ambiguity about both the original hypothesis and how the evidence could have fit alternative hypotheses. It is also harder for analysts to strike the right balance between under- and over-correcting biases without precise real-world feedback, the difference between telling an artilleryman that he is 'close' to the target instead of 100 meters off. Proponents of the training status quo are, however, skeptical of efforts both to quantify uncertainty and to validate training systems against their power to boost accuracy in real-world situations. The grounds for skepticism will be explored later, so it must suffice here to say that these grounds are both reasonable and wrong, a dangerous combination in any organization.[7] Psychological findings circa 1980 have left too deep an imprint, encouraging a narrow perspective on the flaws of intuition and a correspondingly narrow search for remedies. The result has been a defensive mindset aimed at avoiding mistakes, not a proactive mindset aimed at getting it right by reducing uncertainty as aggressively as possible.

Analytic training needs to be revamped.[8] In the second half of this article, we discuss how to do so. The revised program should acknowledge tensions among cognitive biases and focus on error-balancing. It should also include feedback systems that help analysts develop better-calibrated judgments and refine their skills at probabilistic reasoning.[9] And it should address the trade-offs between process and outcome accountability in organizations. Conventional wisdom is right to warn that holding analysts and training systems accountable for accuracy can lead to mistakes: rewarding people who were right because they were lucky, not because of their astute reasoning, and punishing those who were wrong because they were unlucky, not because of their flawed reasoning. But training aimed solely at improving reasoning, insulated from real-world accuracy metrics, can turn into purely scholastic exercises. The effectiveness of training needs to be regularly tested against both correspondence standards of good judgment (empirical accuracy) and coherence (logical rigor) standards. Drawing on the latest research, we detail how to improve training and institutionalize evidence-based norms for determining future training curricula.[10]

[6]For further reading on error management theory see Haselton and Buss, 'Error Management Theory: a new perspective on biases in cross-sex mind reading,' 81–91; McKay and Dennett, 'Evolution of Misbelief,' 493–510; and Haselton and Buss, 'Error Management Theory and the Evolution of Misbeliefs'. Additionally, the terms training and education are used interchangeably within academia. However, we note that the critique herein focuses on analytic training (i.e. the coursework that professional analysts undertake once in the community) as opposed to undergraduate or graduate courses, which typically form the foundational education of analysts.

[7]Another reasonable but wrong combination is the claim that right processes guarantee right outcomes. See Lowenthal, 'Towards a Reasonable Standard,' 308.

[8]Our critique of training requires distinguishing between aspects that focus on how to survive in an intelligence agency (i.e. 'Bureaucracy 101') and analysis (our focus). We do not address important training concepts such as proper sourcing, clear writing, and product timeliness. Our criticism focuses on community-wide analytic training, not single-source programs (such as those at NGA, NSA).

[9]Reiber, 'Intelligence Analysis and Judgmental Calibration,' 101.

[10]Change is difficult when analysts find themselves caught between trying to ensure that all threats are mitigated, which increases the risk of making false positive errors, and trying to ensure that they also do not unnecessarily cry wolf, which increases the risk of false negative errors. This incentivizes agencies, above all, to avoid repeating the most recent big error, which makes it difficult to develop better-calibrated political judgment and foresight.

Table 1. Balancing opposing errors.

Current training emphasizes avoiding…	Symptom	But de-emphasizes…	Symptom
Over-confidence	Exaggerating the strength of a conclusion (e.g. Iraq WMD)	Under-confidence	Minimizing the strength of evidence towards a conclusion (e.g. the rise of ISIL)
Under-stating chances of change	Minimizing status quo changes (e.g. Russian ground incursion into Eastern Europe)	Over-stating chances of change	Exaggerating status quo changes (e.g. collapse of North Korea after a leader's death)
Overweighting the consensus argument	Blindly following the group (e.g. optimism about democratization post-Arab Spring)	Underweighting the consensus argument	Adopting a contrarian stance pro forma (e.g. routine warnings of a major North Korean attack)

Debiasing our thinking about cognitive biases

The most influential training text is Richards Heuer's *Psychology of Intelligence Analysis*.[11] A National Academy of Sciences report on improving intelligence highlighted Heuer's seminal role in introducing intelligence analysts to the behavioral sciences.[12] Heuer's articles, written between 1978 and 1986, became the basis for the *Psychology of Intelligence Analysis*.[13] When the intelligence community consolidated its training systems post-9/11, analytic courses were based on Heuer's work.[14] Heuer urged agencies to embrace a more self-consciously rigorous approach to analysis. Consciousness-raising was deemed critical because the default mode of intuitive and unstructured information processing was so error prone.[15] Heuer made a strong case that the systematic structuring of analysis was the most effective way to 'reduce the frequency and severity of error'.[16] However, the heuristics-and-biases framework on which he drew so extensively was never intended to be a comprehensive theory of human cognition. Psychologists Daniel Kahneman and Amos Tversky were well aware from the outset that the heuristics they were uncovering were adaptive tools that only sometimes led people astray. Subsequent work has underscored the dangers of fixating on suppressing one type of error while downplaying or ignoring the mirror-image error.[17] The *Psychology of Intelligence Analysis* was groundbreaking for its time, but revisions are now necessary.

Intelligence analysis requires balancing opposing errors (see Table 1). This balancing act should play out at a cognitive level as each analyst asks: am I at greater risk of being over- vs. under-confident in my judgments, of over- vs. under-adjusting my beliefs in response to evidence, and of over- vs. under-estimating the uniqueness of the current problem in relation to possible historical precedents? The balancing act should also play out at an institutional level as senior managers aggregate the judgments of individual analysts: am I at greater risk of over-relying on consensus judgments (blindly following the crowd) or of under-relying on them (ignoring the wisdom of the crowd)? And then there is the trade-off that applies to everyone, high and low, in the hierarchy: are scarce analytic resources being allocated in optimal ways, given the relative risks and consequence of each error?

[11] *Psychology of Intelligence Analysis* is a key introductory reading for DIA's Critical Thinking and Structured Analysis Course, the NSA's critical thinking course, as well as ODNI's Analysis 101 course. See National Research Council, *Intelligence Analysis for Tomorrow*, 35; and National Security Agency, 'Critical Thinking Course Syllabus'.

[12] National Research Council, *Intelligence Analysis for Tomorrow,* 83.

[13] Heuer, *Psychology of Intelligence Analysis*, vii.

[14] Heuer and Pherson, *Structured Analytic Techniques*, 9–10.

[15] Heuer, *Psychology of Intelligence Analysis*, xx–xxi; Central Intelligence Agency, *Structured Analytic Techniques*, 1; Cooper, *Curing Analytic Pathologies,* 5–6.

[16] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 5; Immerman, *Transforming Analysis,* 169–70.

[17] Kahneman and Tversky, 'Reality of Cognitive Illusions,' 582–91; Gigerenzer, 'Fast and Frugal Heuristics,' 62–88; Wegener and Petty, 'The Flexible Correction Model,' 142–208; Johnson et al., 'The Evolution of Error,' 474–81. Some additional alternative perspectives to the error-and-bias model of cognition include Mishra, 'Decision-Making Under Risk,' 280–307; Griffiths et al., 'Bayesian Models of Cognition'; and Gigerenzer and Reinhard, 'Bounded Rationality: The Adaptive Toolbox'.

Over- vs. under-confidence

Unfortunately, this error-balancing theme is absent from current training. The focus is almost exclusively on one type of error, over-confidence in intelligence assessments, which arise from over-relying on existing mental models:

> 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.[18]

> Whenever people try to make sense of events, they begin with some body of experience or knowledge that gives them a certain perspective or viewpoint which we are calling a mental model. Intelligence specialists who are expert in their field have well developed mental models. Their mental model tells them, sometimes subconsciously, what to look for, what is important, and how to interpret what they see.[19]

These 'mental ruts' deepen as analysts find and discuss confirming evidence while discounting dissonant evidence.[20] Shared mental models can lead to overweighting the staying power of the status quo and prematurely cutting off valuable inquiry.[21] This results in a shared sense of over-confidence.

Although the over-confidence diagnosis illuminates a deep threat, it sheds no light on the flipside danger of under-confidence.[22] When the evidence supports a strong conclusion, understating it can be more damaging than overselling. The goal should be confidence calibrated to the diagnosticity and strength of the evidence in hand, not across-the-board corrections that cause under-confidence.[23]

Over- vs. under-reaction to new evidence

Along the same lines, traditional training stresses the dangers of belief perseverance and under reacting to new evidence. Closed-mindedness can obviously cause great harm, but the training guidance does not grapple with the danger of over-reacting to new evidence.[24] For example, behavioral economists have found that financial traders, in attempting to keep pace with the news and each other, often slip into self-reinforcing cycles of over-correction, causing excess volatility.[25] Odd though it sounds, one can be too open-minded. Focusing solely on suppressing cognitive conservatism can lead to erratic judgments. In Carl Sagan's words: 'keep an open mind, just not so open that your brains fall out'.[26] Inertia and countervailing forces often slow political change to a crawl.[27] Pushing analysts to imagine the possibility of change, without limitation, will inevitably over-inflate probability-of-change judgments.[28]

Over-vs. under-stating the uniqueness of events

Traditional training offers particularly confusing guidance on how to deal with historical analogies and assess the risks and opportunities posed by change. *Psychology of Intelligence Analysis* warns that

[18]Heuer, *Psychology of Intelligence Analysis*, 65.
[19]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 5.
[20]Heuer, *Psychology of Intelligence Analysis*, 5, 66; Heuer's interpretation of the psychological literature integrates a number of connected concepts including confirmation bias, biased assimilation and motivated reasoning. See Kunda, 'Case for Motivated Reasoning,' 480–98; Lord et al., 'Biased Assimilation and Attitude Polarization,' 2098–109; and Wason, *Psychology of Reasoning*.
[21]Heuer, *Psychology of Intelligence Analysis*, 10, 11, 15.
[22]David Mandel and Alan Barnes found that Canadian intelligence analysts were systematically underconfident in their forecasts. Mandel and Barnes, 'Accuracy of Forecasts in Intelligence,' 10984–9; Heuer, *Psychology of Intelligence Analysis*, 121; Erev et al., 'Simultaneous Over- and Under-confidence,' 519–27; Koriat et al., 'Comparing Objective and Subjective,' 147–62; Moore and Cain, 'Overconfidence and Underconfidence,' 197–213; Shynkaruk and Thompson, 'Confidence and Accuracy in Deductive Reasoning,' 619–32.
[23]Heuer, *Psychology of Intelligence Analysis*, 53–5.
[24]De Bondt and Thaler, 'Does the Stock Market Overreact,' 793–805; Odean, 'Do Investors Trade Too Much?,' 1279–98.
[25]See Howe, 'Evidence of Stock Market Overreaction,' 74–7; Granger, 'Forecasting Stock Market Prices,' 3–13; De Bondt and Thaler, 'Further Evidence on Investor Overreaction,' 557–81; Easterwood and Nutt, 'Inefficiency of Earnings Forecasts,' 1777–97.
[26]Sagan credits historian James Oberg for the quote. Sagan, *Demon-Haunted World*.
[27]Bueno De Mesquita, *The Predictioneer's Game*; Bueno De Mesquita and Smith, *The Logic of Political Survival*.
[28]Teigen, 'Overestimation of Subjective Probabilities,' 56–62; Teigen 'Subjective Sampling Distributions,' 50–5; Caroll, 'Effect of Imagining an Event,' 88–96.

over-reliance on existing mindsets cause analysts to over-weight analogies.[29] However, it does not mention how under-weighting similarities to salient precedents can lead us to exaggerate the uniqueness of the current case. As Carmen Reinhart and Kenneth Rogoff opine in their history of financial crises, 'this time is different' are among the most dangerous words investors can hear.[30]

The confusion deepens when we get to the penultimate chapter of *Psychology of Intelligence Analysis,* which discusses 'base-rates', a term for the relative frequency of events in a reference class. The question is, in judging the risk of a coup in an African country, how much weight should be given to how often coups occur in that region versus country-specific conditions? The guidance from research on base-rate neglect is that analysts should over-weight base-rates in estimating likelihoods. The net effect is to make analysts much more cautious about predicting near-term changes to the status quo.[31]

But this advice contradicts the warning against over-relying on mental models that blind us to looming change. Which bias – belief-perseverance or base-rate neglect – should analysts worry about more? Tension also exists between warnings against neglecting base rates and over-relying on precedents. Base-rates are simply sets of historical precedents. Pushing analysts to rely on base rates sets them up for errors of over-fitting historical data, the inductivist fallacy at the core of Nassim Taleb's parable of the Thanksgiving turkey, whose faith in human benevolence grew each day that humans protected and fed it so generously.[32]

In summary, traditional training fails to confront how biases can either reinforce or neutralize each other – and thus also misses the need to fashion prescriptions for managing these tensions.[33]

Lemmings or wise crowds: over vs. under-weighting the consensus

Traditional training encourages wariness of consensus judgments. As Heuer and Pherson note:

> There is a broad recognition in the Intelligence Community that failure to question a consensus judgment, or a long-established mental model, has been a consistent feature of most significant intelligence failures.[34]

The specter of groupthink is real, but such warnings should be balanced by acknowledging the wisdom-of-the-crowd effect. The consensus of an expert panel is generally more accurate than the majority of the individuals from whom the consensus was derived.[35] Whether the signal value of the crowd judgment outweighs the risk of irrational conformity hinges, among other things, on the independence of individual judgments.[36] The average is often surprisingly accurate because, when independence holds, the idiosyncratic errors of each individual tend to be uncorrelated and thus cancel each other out. Unquestioned acceptance of the consensus can be disastrous, but so too can reflexive skepticism. Any analyst can imagine myriad ways a North Korean dictator could fall or a resurgent Russia could undermine NATO allies. However, an unrestricted imagination generates ever more improbable scenarios, diminishing the value of alternative thinking by swamping the signal with cognitive noise. Scrutiny of each new possibility requires additional effort. Determining when to cut off inquiry and when to continue thinking is a recurring choice. Training should grapple with these time-effort trade-offs.

Structuring analysis: building on an imbalanced foundation

The balancing of effort-accuracy trade-offs should be an integral part of the community's approach to structured analysis. Intelligence instructors Jack Davis, Pherson and Heuer developed structured analytic

[29]Heuer, *Psychology of Intelligence Analysis*, 38–40.
[30]Reinhart and Rogoff, *This Time is Different*.
[31]Tversky and Kahneman, 'Availability,' 207–232.
[32]Ward et al., 'Perils of policy by p-value,' 363–75; Taleb, *The Black Swan*.
[33]Heuer, *Psychology of Intelligence Analysis*, 157–160.
[34]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 215.
[35]Surowiecki, *The Wisdom of Crowds*.
[36]Mellers et al., 'Psychological Strategies for Winning,' 1106–15; Mozer et al., 'Optimal Predictions in Everyday Cognition,' 1113–47; Hackman and Katz, 'Group Behavior and Performance,' 3:III:32.

techniques to reduce what they saw as natural distortions in free-flowing thinking.[37] Structured techniques form a 'set of principles and procedures for qualitative analysis' that exposes biased thought processes for correction.[38] Use of these techniques cannot guarantee accuracy but can 'guarantee an appropriate process of analysis'.[39] The risk of ineffectiveness is ever-present for any proposed cognitive fix – analysts can apply the wrong technique or misapply the right technique. Assuming that analysts can adequately execute the processes is risky because misapplications can easily trigger flipside biases. Structuring analysis also requires additional time and effort. Proponents of thinking tools implicitly come down on the side of maximization (i.e., look at all the options) and against satisficing (i.e., find a good enough answer given the constraints).

Eight categories of structured techniques exist.[40] One, scenario analysis, 'identifies the multiple ways in which a situation might evolve' and targets the status quo and confirmation biases.[41] Scenario analysis systematically explores how a complex situation might unfold. But doing so can distort probabilistic reasoning, causing people to assign too much credibility to too many possibilities – with the net result that the probabilities sum to more than 1.0 (a logical impossibility).[42] Consistent with Tversky's support theory, unpacking possible futures over-inflates probabilities, causing incoherent assessments of nested likelihoods.[43]

Each of the other seven methods can lead to equally serious side effects, which remain unacknowledged. Fighting the privileging of the status quo, with a technique like low-probability/high-impact analysis, raises the risk of over-predicting the occurrence of change.[44] Sensitizing analysts to the impact of vivid, low-probability events can cause over-estimation of those events.[45] Fight mirror-imaging of an adversary's actions with 'red team' analysis and risk over-emphasizing the alien inscrutability of the target.[46] Red-teaming an adversary's behavior can 'free analysts from … their own sense of rationality, cultural norm, and personal values', but it can lead to a dismissive attitude toward conclusions derived from rational cost-benefit analyses.[47] Fight the tendency to under-adjust to new information with 'change sign-posting' and risk over-inflating the chance of a status quo departure.[48] Is the repositioning of tanks and troops signaling invasion, or the start of an exercise, or redeployment under a new chain of command, or something else entirely?

Current training does not grapple with trade-offs between maximizing and satisficing. Analysts complain that time pressure makes structured techniques inaccessible, which leads to use of the techniques by a small minority within the community.[49] Heuer and Pherson counter:

> The experience of many analysts shows that this criticism is not justified. Many techniques take very little time. Anything new does take some time to learn, but, once learned, the use of structured analytic techniques often saves analysts time.[50]

When legitimate pushback from analysts is dismissed, it short-circuits the search for fixes, for spotting new techniques that offer the greatest accuracy-gains for effort-expended.[51] Techniques such as deception detection and alternative futures analysis can easily turn valuable analysis into over-analysis.

[37]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 8.
[38]Ibid., 4.
[39]Heuer, *Psychology of Intelligence Analysis*, 109; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 5.
[40]This number continues to grow. Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 24. There are three categories of techniques in the CIA Analytic Tradecraft Primer and Heuer presented five categories at a 2008 International Studies Association conference. See Heuer, 'Taxonomy of Structured Analytic Techniques.'
[41]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 120.
[42]Rottenstreich and Tversky, 'Unpacking, Repacking, and Anchoring,' 383–406.
[43]Tversky and Koehler note that 'this problem is especially severe in tasks that require the generation of new hypotheses or the construction of novel scenarios.' Tversky and Koehler, 'Support Theory,' 547–67.
[44]Central Intelligence Agency. *A Tradecraft Primer*; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis,* 220.
[45]Tversky and Kahneman, 'Availability,' 207–32.
[46]Ibid., 243.
[47]Ibid., 243–4.
[48]Ibid., 132–43.
[49]To the best of our knowledge, the use of structured techniques is not systematically tracked or evaluated by analytic offices. Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 7; Wastell, 'Cognitive Predispositions and Analyst Reasoning,' 449–60; Folker, *Intelligence Analysts in Intelligence Centers*.
[50]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 7.
[51]MacKay, 'The Problems of Flexibility,' 483–506.

Structured-analysis advocates must acknowledge when the 'barrier to entry' is so high that it discourages analysts from using the techniques.[52] The high cognitive cost of initial employment prevents jumpstarting virtuous learning cycles that reduce technique employment times. Advocates assume efficiency will arise naturally, if only analysts would just practice the official techniques long enough. We believe these competing views of how best to boost analytic performance deserve to be put to the test in realistic work settings.

Does it work? The inadequate epistemology of current training

Intelligence agencies do not know whether structured techniques work because they were designed to be free-standing, without feedback mechanisms to accuracy criteria. As Stephen Marrin argues, 'effort should be devoted to developing a capacity to evaluate the utility of these approaches rather than merely developing and teaching them'.[53] The intelligence community also does not systematically evaluate analytic judgments despite repeated recommendations to do so from both scholars and insiders.[54] Thus, no judgmental accuracy baseline exists to determine whether the techniques help.[55] Structured methods also have not been shown to be efficacious when scientifically evaluated in external unclassified studies.[56]

Some proponents of traditional training argue that measuring analytic accuracy is not possible because assessments are inherently probabilistic and also not traceable to individuals.[57] They argue that it is unfair to evaluate techniques because analysts face many environmental variables that resist experimental control.[58] Some proponents believe that it is impossible to empirically test structured techniques for effectiveness, so the only test these techniques need to pass is 'face-validity'.[59] The existing training program lacks scientific testing and thus bases its validity on an appeal to self-evident truth.[60] Structured techniques may be a net positive force on analytic quality, but without proper testing it is

[52]Heuer and Pherson state, 'satisficing is a common analytic shortcut that people use in making everyday decisions when there are multiple possible answers. It saves a lot of time when you are making judgments or decisions of little consequence, but it is ill-advised when making judgments or decisions with significant consequences for national security. It seems self-evident that an analyst who deliberately identifies and analyzes alternative hypotheses before reaching a conclusion is more likely to find a better answer than an analyst who does not.' Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 311.

[53]Marrin, 'Training and Educating Analysts,' 133.

[54]The NAS report on improving intelligence emphasized testing techniques for effectiveness. McClelland, 'Use of Signal Detection Theory,' 83–100.

[55]Friedman and Zeckhauser, 'Assessing Estimative Accuracy,' 1–23; Reiber, 'Intelligence Analysis and Judgmental Calibration,' 97–112; Tetlock and Mellers, 'Intelligent Management of Intelligence Agencies,' 542–54; Wheaton, 'Evaluating Intelligence,' 614–31; Chang, 'Getting it Right,' 99–108.

[56]Psychological studies using methods such as decomposition and scenario generation have not been shown to be effective. Analysis of Competing Hypotheses failed to correct for confirmation bias. Büyükkurt and Büyükkurt, 'Effectiveness of Three Debiasing Techniques,' 60–73; Byram, 'Cognitive and Motivational Factors,' 216–39; Johnston, 'Analytic Culture in the US Intelligence Community,' 40–1; Moore, *Sensemaking*, 100–1; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 312–3.

[57]Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 310–1.

[58]'Using empirical experiments to evaluate structured techniques is difficult because the outcome of any experiment is influenced by so many variables. Experiments conducted outside the Intelligence Community typically fail to replicate the important conditions that influence the outcome of analysis within the community.' Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 316.

[59]In a presentation to a National Research Council panel on intelligence reform, Heuer stated:

> 'I understand you are all concerned about evaluating whether these structured techniques actually work. So am I. I'd love to see our methods tested, especially the structured analytic techniques Randy [Pherson] and I have written about. The only testing the Intelligence Community has done is through the experience of using them, and I think we all agree that's not adequate … Some of you have emphasized the need to test the accuracy of these techniques. That would certainly be the ideal, but ideals are not always achievable … I see four reasons for this. Testing for accuracy is difficult because it assumes that the accuracy of intelligence judgments can be measured. There is a subset of analytic problems such as elections, when a definitive answer will be known in 6 or 12 months. Even in these cases there is a problem in measuring accuracy, because intelligence judgments are almost always probabilistic. A third reason why a major effort to evaluate the accuracy of structured analytic techniques may not be feasible stems from our experience that these techniques are most effective when used as part of a group process.' Heuer, 'The Evolution of Structured Analytic Techniques'; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 312–16.

[60]60 Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 312–16.

impossible to know if the community is devoting time and effort to techniques that are a net neutral or net negative on analysis.

The objections to evaluation rest on what we now know are misconceptions about the degree to which people can make meaningful probability judgments of real-world events and about the feasibility of scoring them. First, it is methodologically possible to assess the accuracy of probability judgments.[61] One such method, Brier scoring, quantifies how close an analytic assessment comes to capturing the realized outcome. Thus accuracy can be objectively measured and measured cumulatively.[62] Objections that invoke the non-replicability of history can never be fully refuted, but they can be rendered less plausible by well-run field studies, well-designed experiments, and the proper utilization of statistics. One objection is that analysts deal with unique events, making it impossible to come up with a statistical basis for odds estimation. True, history cannot be rerun like an experiment, and so the conditional probability of any unique event given antecedent conditions at time *X* is ultimately unknowable. However, it is possible to penetrate the ultimately unknowable by posing a different conditional-probability question: how likely is the event, given the most current estimates from forecasters with the best track records on similarly complex problems? Tracking forecasters enables understanding which conditions enable more accurate forecasting.[63]

Does it work? Training transference to real-world performance

There is also little published evidence that analytic training transfers from the classroom to the office. 'Transfer', a well-known problem in personnel psychology, refers to how well analysts can translate the knowledge and skills acquired in formal training into sustainable improvement.[64] The entity which oversees all US government training puts it succinctly: 'The goal of training is not simply to gain knowledge and skills, but to transfer learning into performance, which in turn leads to improvements in agency results'.[65] Existing training conveys key concepts didactically in a manner similar to learning to shoot a rifle only by watching videos. It is difficult to become a well-calibrated analyst only through reading second-hand accounts of psychological research. Analysts need to learn (ideally experientially): (1) how the biases impact their work; and (2) how to develop effective ways to check these biases. Similarly, tutorials on structured analysis fall short, in part, because they are so scripted and formulaic. Imagine a rower who practices on a rowing machine and excels but who now finds herself on open water in a storm. This rower is dangerously ill-equipped for navigating real-world conditions. As they say in the military, 'train as you fight', and current training falls short of this mantra.

When does one-sidedness make sense? Incentive structures in analysis

Before turning to solutions, one point must be made: organizations often pressure analysts to avoid certain mistakes over others, and current training could be defended as encouraging analysts to make the right sorts of mistakes.[66] Stressing avoiding false-negative errors (missing wolves on the prowl) and downplaying false-positive errors (crying wolf) looks, on its face, sensible if policymakers worry more about being blindsided than about being distracted by spurious warnings. But this argument must have bounds. Its *reductio ad absurdum* is the precautionary principle: intelligence agencies should warn against anything and everything for which analysts can construct a minimally plausible worst-case

[61] Murphy and Winkler, 'Reliability of Subjective Forecasts,' 41–7; Brier, 'Verification of Forecasts,' 1–3.
[62] Mellers et al., 'Psychological Strategies for Winning,' 1106–15. Tetlock, *Expert Political Judgment*.
[63] Atkins, 'Why Grexit odds are probably 99% wrong.'
[64] Baldwin and Ford, 'Transfer of Training,' 63–105; Blume et al., 'Transfer of Training,' 1065–05; Cheng and Ho, 'Review of Transfer of Training,' 102–18; Grossman and Salas, 'Transfer of Training,' 103–20; Yamnill and McLean, 'Theories Supporting Transfer of Training,' 195–208.
[65] Office of Personnel Management, 'Training and Development Policy.'
[66] Tetlock, *Expert Political Judgment,* 21, 143.

scenario.[67] This is an unsustainable position. The call on Iraq WMD was a very costly error – and it was a false positive.[68]

The 'right mistake' defense encourages us to do something that decision scientists have long warned against: confusing our probability judgments of making an error with our (dis)utility judgments of the error.[69] If the intelligence apparatus is committed to respecting the official division of labor between analysis and policy, it should concentrate on executing the delicate error-balancing acts that yield the truest estimates of probability, undistorted by fears of future recriminations for under- or over-estimating this or that outcome – and leave the value judgments of which errors are worse to the President and his or her proxies.

A delicate balancing act: designing a new training system

How should training be redesigned? Acknowledging the psychological truth that cognitive biases pull in opposing directions is a start (see Table 2). Good judgment requires deftly managing error-balancing and effort-accuracy trade-offs. Balancing must be learned through exercises with objectively correct logical or empirical answers, exercises that provide a more secure epistemological foundation for determining whether training works as intended.

By extension, this means revamped training would hold analysts accountable for both good process and accurate outcomes. It would recognize that analysts can be right for incorrect reasons and vice versa. But predictive accuracy is typically a strong Bayesian signal that at least some of the underlying reasoning was correct. Measurement and accountability together form the core of a self-correcting training system.

Mastering mental balancing

Learning error-balancing starts with improving probabilistic reasoning. Tracking the accuracy of probabilistic estimates enables measurement of over-under errors. A new system should start with the core tenets of probability and review common errors people make.[70] For starters, quantifying uncertainty becomes more natural with practice. Practice at quantifying uncertainty further sensitizes analysts to the boundary between fine-grained and pseudo-precise estimates. Learning the limits of precision across domains deepens our understanding of the appropriate effort to commit to quantifying probabilities (the concept of cognitive triage).[71]

Improving probabilistic reasoning can help analysts become better at mental-balancing. To do so may not even require a formal, lengthy course. Some researchers have demonstrated improved probabilistic thinking (e.g. through a simplified version of Bayes theorem) without teaching complicated math.[72] These cognitive developments can enable analysts to distinguish finer grained probabilities.[73] The community's standardized verbal probability scale asks analysts only to distinguish seven likelihoods (and not all are routinely used). Improved reasoning, coupled with moving from verbal probabilities to numeric ones, should enable analysts to communicate more meaningfully granular estimates to policymakers. For example, the chances of a large-scale terrorist attack in the US and that of a nuclear

[67]Cameron and Abouchar, 'The Precautionary Principle,' 1–27.
[68]Pricing the error is hard because no one knows the costs incurred in a no-invasion counterfactual world but current estimates put it at roughly US$4–6 trillion and thousands of American lives – leaving out the destabilizing regional ripple effects. Bilmes, 'The Financial Legacy of Iraq and Afghanistan'; Wehrey et al., *The Iraq Effect*.
[69]Bell et al., *Decision Making*.
[70]Kahneman and Tversky, 'Study of Statistical Intuitions,' 123–41.
[71]See Tetlock and Gardner, *Superforecasting*, 277–85.
[72]Mandel, 'Instruction in Information Structuring'; Jaynes, *Probability Theory*; Gigerenzer, 'Make Cognitive Illusions Disappear,' 83–115; Gigerenzer and Hoffrage, 'Improve Bayesian Reasoning Without Instruction,' 684–704; Sedlmeier and Gigerenzer, 'Teaching Bayesian Reasoning,' 380–400; Johnson-Laird, 'Mental Models and Probabilistic Thinking,' 189–209.
[73]See Friedman et al., 'The Value of Precision.'

Table 2. Balanced thinking through better training.

Proposed Change	How It Improves the Current System
Incorporating probabilistic reasoning	Equips analysts to improve calibration
Explore, test and add scientifically-validated debiasing techniques	Equips analysts with quick thinking tools that can improve their reasoning over time
Closing the feedback loop by measuring and holding analysts accountable for accuracy	Sets a baseline from which to improve judgments and evaluates training efficacy

war between the US and Russia may both be 'remote', but one is orders of magnitude less likely than the other. Analysts might describe the probability of security forces retaking a city or a leader winning re-election as about 'even chance'. Numerical probabilities enable analysts to express meaningful differences in likelihoods between the two scenarios (e.g. to distinguish 60/40 from 40/60 odds). Using numerical values also enables more frequent incremental belief updating (e.g. 70 per cent becomes 75 per cent, as opposed to 'likely' remaining 'likely'), –allowing analysts to get to the truth more quickly.

People also tend to process magnitude and duration effects improperly, which can bias estimates of the impact of time on event probabilities.[74] This flaw, known as scope insensitivity, causes anomalies like judging as equally likely the probability of a leader being removed from power within one year versus five years.[75] Scope-insensitive forecasters reported virtually identical probabilities for Syrian President Assad's removal from power at three (40 per cent) and six months (41 per cent) into the future. Scope-sensitive 'superforecasters' (the most accurate 2 per cent of the forecaster pool) discovered in a ground-breaking US Intelligence Advanced Research Projects Activity (IARPA) research program, the Aggregative Contingent Estimation (ACE) project, reported 15 per cent at three months and 24 per cent at six months, a significant difference.[76] These superforecasters were not perfectly scope sensitive but they show what is possible with gritty practice.[77]

Developing base-rate reasoning (a key component of probabilistic reasoning) also belongs in future programs. Base-rates offer a reasonable starting point for probability estimation. For example, assessing the probability of an African dictator losing power can be initially calibrated by considering the overall probability of dictators losing power in Africa annually (rather infrequent for those that are long-serving). Starting with the base-rate, an analyst can then adjust the probability dial up or down by considering the details of the case (e.g. a floundering economy, a disgruntled military). Determining the appropriate comparison classes is both science and art and improved through practice. Is the appropriate class the probability of leadership transition in the whole of South America or the Andean region or perhaps an even wider class? In thinking about the probability of a coup in a Southeast Asian country, is it more appropriate to start with the probability of any country anywhere experiencing a coup (approximately four per year between 2004 and 2012) or a narrower class?[78] Once the basics of base-rate thinking are understood – such as the tricky trade-offs between sample size and sample relevance – shifting between the details of a case and the overall event-class becomes second-nature, allowing more calibrated estimates.

Another way to teach balancing is through behavioral game theory. Imagine asking students to guess what number will be 2/3 of the average of all guesses in the class. The task requires balancing notions of hyper-rationality ('everyone is rational so they will quickly converge on zero') with bounded rationality ('some people won't get it so what is 2/3 of the number derived from the estimated number of recursive thought-loops students perform?'). Game theory enables analysts to spot the counteracting

[74]Frederick and Fischhoff, 'Scope (In)Sensitivity in Elicited Valuations,' 109–23. Kahneman and Frederick, 'Representativeness Revisited,' 49.
[75]Mellers et al., 'Identifying and Cultivating Superforecasters,' 267–81.
[76]The IARPA ACE project, a large-scale experiment with thousands of participants, tested the limits of forecasting accuracy via geopolitical forecasting tournament from 2011 to 2015.
[77]Mellers et al., 'Identifying and Cultivating Superforecasters,' 267–81.
[78]Statistics from data collected by the Center for Systemic Peace.

positions of key players and the causal forces that lock or unlock stalemates. Thinking this way might bring down errors of both under- and over-estimating the rationality of opponents.[79]

A better understanding of complex systems can also deepen our appreciation for the problems posed by random processes as well as missing data.[80] Appreciating randomness helps analysts to better understand probabilities that at first glance seem to be at the tail ends of probability distributions. Imagine two countries in a territorial dispute. Each patrols the area with troops, aircraft and ships. While neither side wants to start a war, as more forces are deployed, the possibility of accidental encounters increases. Accidental events are even likelier to occur if less-controllable proxy actors such as rebel militias are armed with advanced weaponry (e.g. Russian-backed rebels in Eastern Ukraine). Intentions matter but some events are beyond the control of the authorities. Additionally, when events previously assessed as low-probability tail risks come to fruition, it highlights the inadequacy of relying on past data to predict all future events. Just because something hasn't previously happened does not mean it never will.[81] Analysts should contribute judgments to long-term databases that give management a sense for how good or bad they are at spotting when sharp departures from the status quo (due to complexity or stochasticity) deserve more scrutiny than the extremely low base rates of such events would suggest warranted.

Another applicable training regimen is the one developed for the IARPA ACE program's forecasting tournament by the Good Judgment Project. This training combined probabilistic reasoning and key social science principles. Trained research subjects were, on average, more accurate across hundreds of political and economic questions (selected by the US intelligence community for their relevance to issues that analysts routinely grapple with) over four years.[82] Analysts deserve the best state-of-the-art training and the IARPA techniques are among the few that have been scientifically validated against real-world accuracy metrics.

One way to teach error-balancing is by using real-world scenarios that help analysts identify the relative strengths of countervailing biases in their own thinking.[83] This testing should be done under realistic time constraints and accountability pressures. For example, trainees could perform exercises in which they receive either strongly or weakly diagnostic reports from the field, thus exploring trainees' tendencies either to under- or over-adjust to evidence.[84] Another way of diagnosing adaptive or maladaptive cognitive habits is through forecasting tournaments which can stimulate analysts to build cognitive bridges between abstract thinking principles and analytical practice. Confronted by a tournament question about whether instability will follow a dictator's death pressures analysts to balance a wider range of causes than they might otherwise consider. A structured method such as scenario generation might well highlight the same factors but it does not relentlessly push analysts to distinguish more from less probable scenarios, in as granular a fashion as humanly possible.

Counterintuition: confronting effort-accuracy trade-offs

New training protocols should acknowledge that an effort-maximization perspective, which is uncompromising about the need to use rigorously structured techniques, is inadequate in a world of competing priorities and finite resources. Analysts need to develop frugal overrides to sometimes-erroneous intuitive thinking, overrides Michael Mauboussin has called 'counterintuition'.[85] Heuristics are two-edged swords.[86] And blaming heuristics for the myriad cognitive ills plaguing intuition, as existing training does,

[79]See Camerer, *Behavioral Game Theory*.
[80]Langer, 'The Illusion of Control,' 311–328; Taleb, *The Black Swan*; Taleb, *Fooled by Randomness*.
[81]Taleb, *Fooled by Randomness*, 92–3.
[82]Mellers et al., 'Psychological Strategies for Winning,' 1106–1115; Mellers et al., 'The Psychology of Intelligence Analysis,' 1–14; Chang et al., 'Developing Expert Political Judgment.'
[83]Ford and Weissbein, 'Transfer of Training,' 22–41.
[84]Massey and Wu, 'Detecting Regime Shifts,' 932–47.
[85]Mauboussin, *Think Twice*.
[86]Tversky and Kahneman, 'Judgment under Uncertainty,' 1124–31.

leads to overlooking useful tools.[87] For example, 'inside-outside' thinking helps an analyst balance the tendency to privilege the engrossing details of a case (e.g. all indicators point to a North Korean attack across the DMZ) with the overall base-rate probability of an attack (e.g. we've seen patterns like this before and they are testing us again).[88] Other useful approaches include David Mandel's heuristics-based approach to Bayesian reasoning using visualizations for more consistent probabilistic judgments.[89]

A regimen for honing intuitive mental processing should offer many opportunities for refresher training and more advanced modules as basic concepts are mastered. Retraining 'fast-processing' through exercises, repetition, and feedback can debias judgments at the intuitive level of thought that analysts rely on most – the 'sensemaking' concept that has been connected to intelligence analysis.[90]

Improving analysis through feedback mechanisms

Improved training inculcates sensitivity to over-under error-balancing acts. It does so by requiring fine-grained measures of judgments and fine-grained feedback on accuracy. Probabilistic thinking expressed quantitatively (0–100 per cent) permits precision that vague verbiage makes impossible.[91] Tracking accuracy empowers analysts to test whether training actually improves analytic outcomes.

Calibrating probability judgments is similar to fine-tuning a dial. Increasing the granularity of judgment is much harder when we are confined to a seven-word scale of estimative probability ('unlikely', 'likely', etc.).[92] Recent US intelligence official guidance codifies this scale and translates these words into probability bins of sizes from 5–15 per cent.[93] *Psychology of Intelligence Analysis* and the structured techniques manual do discuss probabilities, but lack actionable guidance.[94] Training also slights quantitative methods that are essential for gauging whether analysts are striking reasonable balances between under- versus over-prediction.[95] Probabilistic reasoning training and Bayesian inference principles, shown to boost predictive accuracy, are mostly missing from the current regimen.[96]

Checking errors like base-rate neglect and trend over-extrapolation requires understanding basic probability theory concepts. Quantitative probability assessments provide nuance that vague expressions of uncertainty cannot. In estimating the likelihood of regime collapse during the Arab Spring, it might have been tempting to answer 50 per cent or higher. But grounding that initial estimate in historical base rates 'how often are authoritarian regimes overthrown in any 12-month period' would have introduced some useful caution. The same is true of other event classes: how often do interstate wars occur (not very)? How many African countries experience coups each year (a few)? How often do peace agreements wrap up long-festering conflicts (fewer than one might hope)? Attaching numerical probabilities to judgments also makes it possible to assess accuracy in ways that would be impossible if analysts were restricted to terms like 'might' or 'could'.[97] The importance of probabilistic reasoning cannot be understated: because reasoning about uncertainty underlies all complex cognitive tasks,

[87]Gigerenzer and Todd, *Simple Heuristics*.
[88]Kahneman, *Thinking, Fast and Slow*.
[89]Mandel, 'Visual Representation of Rational Belief'; Mandel, 'The Psychology of Bayesian Reasoning.'
[90]Moore, *Sensemaking*; Puvathingal and Hantula, 'Revisiting the Psychology of Intelligence,' 199–210.
[91]Virtually all judgments can be reduced to subjective probabilities about past, present or future states of the world. It is meaningless to warn about over or under-confidence but never actually assess calibration.
[92]Traditional training departs from the guidance offered by Heuer who, based on the work of Sherman Kent, recognized that verbal probabilities are often misunderstood. Heuer encouraged analysts to use numerical probabilities when appropriate. Heuer, *Psychology of Intelligence Analysis*, 154–6.
[93]Office of the Director of National Intelligence, 'Intelligence Community Directive 203: Analytic Standards.'
[94]Heuer, *Psychology of Intelligence Analysis*, 147–60; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 236.
[95]Heuer and Pherson see quantitative methods as a completely separate category of analysis from structured techniques. Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 22–3.
[96]Mellers et al., 'Psychological Strategies for Winning,' 1106–15; Mellers et al., 'The Psychology of Intelligence Analysis,' 1–14.
[97]Alan Barnes points out how inconsistent mapping of expressions to quantitative probabilities impedes comprehension. Barnes, 'Making Intelligence Analysis More Intelligent.'

from finance to medicine, to intelligence analysis. It follows that probabilistic reasoning should be an essential tool in analysts' arsenal.[98]

Closing the loop: feedback, measurement, and accountability can improve thinking

A key feature of an improved training program is its capacity to give analysts high-quality performance feedback. Forecasting competitions allow analysts to turn rough qualitative judgments into scorable quantitative predictions.[99] Scoring is needed to help spot mistakes helps – and illuminate post-mortems. Breaking big issues into discrete parts and time periods can turn a low feedback environment (e.g. North Korean politics) into a high feedback one, a low granularity environment (e.g. Iranian nuclear ambitions) into a moderate granularity one. Error balancing requires measuring, figuring out how much we have over/under-shot in given situations and adjusting to feedback.

Current efforts to measure analytic quality give almost no weight to accuracy metrics, yet a recent study of analytic forecasting skill in Canada demonstrates the feasibility of measuring accuracy.[100] Declaring that 'face-validity' is enough makes it impossible to figure out how loosely or tightly coupled current metrics of good reasoning process are to correspondence-to-reality metrics. Absent such data, we will never learn how common it is for analysts who can develop great narratives are poor forecasters, or conversely how often the best forecasters prove quite maladroit at developing compelling narrative rationales.

The probability training proposed in the previous section not only enables more granular judgments; it enables measurement of accuracy or correspondence to reality.[101] Analytic judgments are currently evaluated solely on how well they adhere to process and coherence guidelines. Scoring accuracy alongside process measures bridges the gap between explanatory quality and prediction accuracy by tracking the correlation between the two metrics. Even if the main purpose of intelligence is generating context-framing explanations, not forecasts, adding correspondence measures reveals the extent to which the best explanations match what comes to pass.

Intelligence agencies should integrate the correspondence measures generated from evaluations with existing coherence measures, creating hybrid accountability systems that reward analysts for both well-reasoned arguments and accurate forecasts. Every accountability system rests on assumptions about the weaknesses it needs to fix, the 'best' ways to fix them, and the best ways to tell if the 'best' ways are the best. The intelligence community has long recognized the need to gauge the accuracy of its judgments and a hybrid accountability system is the best way to begin the process without compromising the quality of existing analytic products.

If the intelligence community decides to continue utilizing structured techniques, it should, at least, start testing them. One approach would be to pit the community's favored techniques against the best known psychological debiasing tools in long-term comparative validity studies, ideally using accuracy metrics from forecasting tournaments as well as logical-coherence indicators used in most debiasing work. Critics and proponents should agree ex ante on how long long-term comparative studies need to run before defenders of the training status quo would concede the need for change,– or before critics would concede that they overstated the need for change. Agencies should get in the habit of basing training on the results of even-playing-field tests, not untested intuition or institutional inertia.

Taking our own medicine: acknowledging base-rates of organizational change

There are no off-the-shelf solutions. Implementing our recommendations will require overcoming bureaucratic resistance and solving tricky problems, akin to the shift from folk remedies to science-based

[98]See Pearl, *Probabilistic Reasoning in Intelligence Systems*. The famed investor Charlie Munger sums it up in his trademark blunt fashion: 'if you don't get … elementary probability into your repertoire, then you go through life like a one legged man in an asskicking contest.' Munger, 'A Lesson On Elementary, Worldly Wisdom.'

[99]Tetlock et al., 'Forecasting Tournaments,' 290–5.

[100]Mandel and Barnes, 'Accuracy of Forecasts,' 10984–9.

[101]We are not advocating perfection as a standard, a point Stephen Marrin makes when discussing the pitfalls of evaluating intelligence accuracy. Marrin, 'Evaluating the Quality of Intelligence Analysis,' 911.

modern medicine.[102] A sober assessment of the probability of success would be low. Why? Look at the base rates: Most organizational-change initiatives fail.[103] Machiavelli made his prediction 500 years ago: 'there is nothing more difficult and dangerous, or more doubtful of success, than an attempt to introduce a new order of things … whenever the opponents of the new order of things have the opportunity to attack it, they will do it with the zeal of partisans.' What may be even more remarkable than reform is that intelligence agencies have maintained a training regimen that assumes it knows the magnitude and direction of errors that analysts will make and that it can fix them – all without any scientific measurement.[104] For too long the intelligence community has shackled itself to a system of training that it never tested – and that almost certainly does not deliver promised performance benefits. That said, the national security stakes are high enough to justify even a low probability of success in improving the training of intelligence analysts. That is why we wrote this article.

Acknowledgements

We thank the Wharton Global Initiatives program for its generous support of our research and Kelsie Piedra for assistance with manuscript preparation. We also thank Jeffrey Friedman, David Mandel, Mark Parker, Carrie Plummer, and three anonymous reviewers for their insightful comments and critiques.

Notes on contributors

Welton Chang, a PhD student at the University of Pennsylvania, is a psychologist who studies judgment and decision-making. From 2005–2014, he served as an intelligence officer in the US Army and at the Defense Intelligence Agency. His overseas experience includes 21 months in Iraq and a year in South Korea. He sits on the board of the John Sloan Dickey Center for International Understanding and is a Truman National Security Fellow. Welton earned a BA in Government, *cum laude* with departmental high honors, from Dartmouth College, an MA in Security Studies from Georgetown University, and an MA in Psychology from the University of Pennsylvania.

Philip E. Tetlock is the Annenberg University Professor at the University of Pennsylvania, with cross appointments in psychology, the Wharton school, and political science. He previously held the Mitchell Endowed Professorship at the University of California, Berkeley. Tetlock has published widely and is the author of *Superforecasting: The Art and Science of Prediction* and of *Expert Political Judgment: How Good Is It? How Can We Know?* and has published widely in peer-reviewed journals on the topics of expert judgment, judgmental biases, and methods of debiasing. He has received scientific awards and honors from a wide range of scholarly organizations, including the National Academy of Sciences, the American Academy of Arts and Sciences, the American Political Science Association, and the American Psychological Association.

ORCID

Welton Chang http://orcid.org/0000-0001-6344-8732

Disclaimer

The views presented are those of the authors and do not represent the views of the Department of Defense or any of its components or the United States Government.

References

Atkins, Ralph. 'Why Grexit odds are probably 99% wrong.' *Financial Times*, July 2, 2015. Accessed January 19, 2016. http://www.ft.com/intl/cms/s/0/bb18e6e6-1ff8-11e5-aa5a-398b2169cf79.html.

Baldwin, Timothy T., and Ford, J. Kevin, 'Transfer of Training: A Review and Directions for Future Research", *Personnel Psychology* 41/1 (1988): 63–105.

[102] Johnston compares medicine, also a blend of art, tradecraft, and science, to analysis, noting that intelligence had yet to go through a similar scientific revolution. Johnston, *Analytic Culture in the US Intelligence Community*, 43.
[103] Hannan and Freeman, 'Structural Inertia and Organizational Change,' 149–64.
[104] Cooper, *Curing Analytic Pathologies*, 38.

Barnes, Alan. 'Making Intelligence Analysis More Intelligent: Using Numeric Probabilities,' *Intelligence and National Security* (2015).

Bell, David E., Raiffa, Howard, Tversky, Amos. *Decision Making: Descriptive, Normative, and Prescriptive Interactions* (New York: Cambridge University Press 1988.

Bilmes, Linda. 'The financial legacy of Iraq and Afghanistan: How wartime spending decisions will constrain future national security budgets.' Harvard Kennedy School Faculty Research Working Paper Series, 2013. Accessed on January 19, 2016. https://research.hks.harvard.edu/publications/getFile.aspx?Id=923.

Blume, Brian D., Ford, J. Kevin, Baldwin, Timothy T., Huang, Jason L, 'Transfer of Training: A Meta-Analytic Review', *Journal of Management* 36/4 (2010): 1065–1105.

Brier, Glenn W., 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review* 78/1 (1950): 1–3.

Bueno De Mesquita, Bruce. *The Predictioneer's Game: Using the Logic of Brazen Self-Interest to See and Shape the Future*. New York, NY: Random House, 2009.

Bueno De Mesquita, Bruce and Smith, Alastair, *The Logic of Political Survival*. Cambridge: MIT Press, 2005.

Büyükkurt, B. Kemal, and Büyükkurt, Meral Demirbag. 'An Experimental Study of the Effectiveness of Three Debiasing Techniques', *Decision Sciences* 22/1 (1991): 60–73.

Byram, Stephanie J., 'Cognitive and motivational factors influencing time prediction', *Journal of Experimental Psychology: Applied* 3/3 (1997): 216–239.

Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press, 2003.

Cameron, James, and Abouchar, Juli, 'The Precautionary Principle: A Fundamental Principle of Law and Policy for the Protection of the Global Environment', *Boston College International and Comparative Law Review* 14/1 (1991): 1–27.

Caroll, John S. 'The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic', *Journal of Experimental Social Psychology* 14/1 (1978): 88–96.

Central Intelligence Agency. *A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis*. (Washington D.C.: CIA Sherman Kent School, 2009.

Chang, Welton. 'Getting it Right: Assessing the Intelligence Community's Analytic Performance', *American Intelligence Journal* 30/2 (2012) pp.99–108.

Chang, Welton, Chen, Eva, Mellers, Barbara, and Tetlock, Philip E. 'Developing Expert Political Judgment: The Impact of Training on Judgmental Accuracy in Geopolitical Forecasting Tournaments.' *in preparation*.

Cheng, Eddie WL, Ho, Danny CK. 'A Review of Transfer of Training Studies in the Past Decade.' *Personnel Review* 30/1 (2001): 102–118.

Cooper, Jeffrey R. *Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis*. Washington DC: Central Intelligence Agency, 2005.

De Bondt, Werner F.M. and Thaler, Richard H., 'Does the Stock Market Overreact?', *The Journal of Finance* 40/3 (1985): 793–805.

De Bondt, Werner F.M. and Thaler, Richard H. 'Further Evidence On Investor Overreaction and Stock Market Seasonality', *The Journal of Finance* 42/3 (1987): 557–581.

Charles Duelfer, *Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD* Washington D.C.: Central Intelligence Agency 2004.

Easterwood, John C. and Nutt, Stacey R., 'Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction or Systematic Optimism?', *The Journal of Finance* 54/5 (1999): 1777–1797.

Erev, Ido, Wallsten, Thomas S., Budescu, David V. 'Simultaneous Over- and Under-confidence: The Role of Error in Judgment Processes', *Psychological Review* 101/3 (1994): 519–527.

Folker, Robert D. *Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment In Applying Structured Methods*. Washington: D.C.; Joint Military Intelligence College, 2000.

Ford, J. Kevin and Weissbein, Daniel A., 'Transfer of Training: An Updated Review and Analysis", *Performance Improvement Quarterly* 10/2 (1997) pp.22–41.

Frederick, Shane, and Fischhoff, Baruch, 'Scope (in) sensitivity in Elicited Valuations', *Risk Decision and Policy* 3/2 (1998): 109–123.

Friedman, Jeffrey A., Baker, Joshua D., Mellers, Barbara A., Tetlock, Philip E., and Zeckhauser, Richard J. 'The Value of Precision in Geopolitical Forecasting: Empirical Foundations for Intelligence Analysis and Foreign Policy Decision Making.' Paper presented to the American Political Science Association, San Francisco, 2015.

Friedman, Jeffrey A. and Zeckhauser, Richard. 'Why Assessing Estimative Accuracy Is Feasible and Desirable.' *Intelligence and National Security* (2014): 1–23.

Gigerenzer, Gerd, 'Fast and frugal heuristics: The tools of bounded rationality', *Blackwell Handbook of Judgment and Decision Making* (2004): 62–88.

Gigerenzer, Gerd and Reinhard, Selten. *Bounded rationality: The adaptive toolbox*. Cambridge: MIT Press, 2002.

Gigerenzer, Gerd, and Todd, Peter M., *Simple Heuristics That Make Us Smart* (New York: Oxford University Press 1999.

Gigerenzer, Gerd, and Hoffrage, Ulrich, 'How to Improve Bayesian Reasoning Without Instruction: Frequency Formats', *Psychological Review* 102/4 (1995): 684–704.

Gigerenzer, Gerd, 'How to Make Cognitive Illusions Disappear: Beyond 'Heuristics and Biases', *European Review of Social Psychology* 2/1 (1991): 83–115.

Granger, Clive W.J. 'Forecasting Stock Market Prices: Lessons for Forecasters.' *International Journal of Forecasting* 8, (1992): 3–13.
Griffiths, Thomas L., Kemp, Charles, Tenenbaum, Joshua. 'Bayesian models of cognition.' *The Cambridge Handbook of Computational Psychology*. Cambridge: Cambridge University Press, 2008.
Grossman, Rebecca, and Salas, Eduardo, 'The Transfer of Training: What Really Matters', *International Journal of Training and Development* 15/2 (2011): 103–120.
Hackman, J. Richard, and Katz, Nancy. 'Group Behavior and Performance.' in the *Handbook of Social Psychology* 2010, 3:III:32.
Hannan, Michael T., and Freeman, John, 'Structural Inertia and Organizational Change', *American Sociological Review* 49/2 (1984): 149–164.
Haselton, Martie G., and Buss, David M., 'Error management theory: a new perspective on biases in cross-sex mind reading', *Journal of Personality and Social Psychology* 78/1 (2000): 81–91.
Haselton, Martie G., and Buss, David M., 'Error management theory and the evolution of misbeliefs', *Behavioral and Brain Sciences* 32/6 (2009): 522–523.
Heuer, Richards J. *Psychology of Intelligence Analysis*, Washington D.C.: Center for the Study of Intelligence, 1999.
Heuer, Richards J. 'Taxonomy of Structured Analytic Techniques.' Paper presented at the annual meeting for the International Studies Association, San Francisco, March 26-29, 2008. Accessed on January 19, 2016. http://www.pherson.org/wp-content/uploads/2013/06/03.-Taxonomy-of-Structured-Analytic-Techniques_FINAL.pdf.
Heuer, Richards J. 'The Evolution of Structured Analytic Techniques.' Presentation to the National Academy of Science, National Research Council Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security, Washington, D.C., December 8, 2009. Accessed on January 19, 2016. http://sourcesandmethods.blogspot.com/2010/01/heuer-how-to-fix-intelligence-analysis.html.
Heuer, Richards J. and Pherson, Randolph. *Structured Analytic Techniques for Intelligence Analysis*. Washington D.C.: CQ Press, 2010.
Howe, John S. 'Evidence on Stock Market Overreaction', *Financial Analysts Journal* 42/4 (1986): 74–77.
Immerman, Richard H. *Transforming Analysis: The Intelligence Community's Best Kept Secret. Intelligence and National Security* 26/2-3 (2011): 169–170.
Intelligence Reform and Terrorism Prevention Act of 2004 (IRTPA), Public Law 108–458: 118.
Jaynes, Edwin T. *Probability Theory: The Logic of Science*. New York: Cambridge University Press, 2003.
Johnson-Laird, Philip N., 'Mental Models and Probabilistic Thinking', *Cognition* 50/1 (1994): 189–209.
Johnson, Dominic P., Blumstein, Daniel T., Fowler, James H., Haselton, Martie G., 'The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases', *Trends in Ecology and Evolution* 28/8 (2013): 474–481.
Johnston, Rob, *Analytic Culture in the US Intelligence Community: An Ethnographic Study* Washington, DC: Center for the Study of Intelligence, 2005.
Kahneman, Daniel, *Thinking, Fast and Slow* New York: MacMillan 2011.
Kahneman, Daniel and Frederick, Shane. 'Representativeness Revisited: Attribute Substitution in Intuitive Judgment.' *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York, NY: Cambridge University Press, 2002.
Kahneman, Daniel, and Tversky, Amos, 'On the Reality of Cognitive Illusions', *Psychological Review* 103/3 (1996): 582–591.
Kahneman, Daniel, and Tversky, Amos, 'On the Study of Statistical Intuitions', *Cognition* 11/2 (1982): 123–141.
Kean, Thomas and Hamilton, Lee. *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States*. Washington D.C.: Government Printing Office, 2011.
Koriat, Asher, Sheffer, Limor, Ma'ayan, Hilit. 'Comparing Objective and Subjective Learning Curves: Judgments of Learning Exhibit Increased Underconfidence With Practice.' *Journal of Experimental Psychology: General* 131 / 2 (2002): 147–162.
Kunda, Ziva, 'The Case for Motivated Reasoning', *Psychological Bulletin* 108/3 (1990): 480–498.
Lake, Eli and Dickey, Christopher, 'US Spies Said No Russian Invasion of Ukraine- Putin Disagreed,' *The Daily Beast*, February 28, 2014, Accessed January 19, 2015. http://www.thedailybeast.com/articles/2014/02/28/u-s-spies-said-no-invasion-putin-disagreed.html.
Langer, Ellen J., 'The Illusion of Control', *Journal of Personality and Social Psychology* 32/3 (1975): 311–328.
Leibovich, Mark. 'George Tenet's 'Slam-Dunk' Into the History Books,' *The Washington Post*, June 4, 2004, Accessed January 19, 2015. http://www.washingtonpost.com/wp-dyn/articles/A14030-2004Jun3.html.
Lord, Charles G., Ross, Lee, and Lepper, Mark R. 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence', *Journal of Personality and Social Psychology* 37/11 (1979): 2098–2109.
Lowenthal, Mark M. 'Towards a Reasonable Standard for Analysis: How Right, How Often on Which Issues?' *Intelligence and National Security* 23/3 (2008): 308.
McClelland, Gary H. 'Use of signal detection theory as a tool for enhancing performance and evaluating tradecraft in intelligence analysis,' in *Intelligence Analysis: Behavioral and Social Scientific Foundations*, Washington, D.C.: National Academies Press, 2011, 83–100.
McKay, Ryan T., and Dennett, Daniel C., 'The evolution of misbelief', *Behavioral and Brain Sciences* 32/6 (2009): 493–510.
MacKay, Donald G., 'The Problems of Flexibility, Fluency, and Speed-Accuracy Trade-Off in Skilled Behavior', *Psychological Review* 89/5 (1982): 483–506.
Mandel, David. 'Instruction in information structuring improves Bayesian judgment in intelligence analysts.' *Frontiers in Psychology: Cognition*, (2015).

Mandel, David R. 'Visual Representation of Rational Belief Revision: Another Look at the Sleeping Beauty Problem,' *Frontiers in Psychology: Cognition* 5 (2014).
Mandel, David R. 'The Psychology of Bayesian Reasoning,' *Frontiers in Psychology: Cognition* 5 (2014).
Mandel, David R., and Barnes, Alan, 'Accuracy of Forecasts in Strategic Intelligence', *Proceedings of the National Academy of Sciences* 111/30 (2014): 10984–10989.
March, James G., 'Exploration and Exploitation in Organizational Learning', *Organization Science* 2/1 (1991): 71–87.
Marrin, Stephen. 'Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure?' *Intelligence and National Security* 27/6 (2012): 911.
Marrin, Stephen, 'Training and Educating US Intelligence Analysts', *International Journal of Intelligence and Counterintelligence* 22/1 (2008): 132–133.
Massey, Cade, and Wu, George, 'Detecting Regime Shifts: The Causes of Under-and Overreaction', *Management Science* 51/6 (2005): 932–947.
Mauboussin, Michael J. *Think Twice: Harnessing the Power of Counterintuition*. Cambridge: Harvard Business Press, 2009.
Mellers, Barbara, Unger, Lyle, Baron, Jonathan, Ramos, Jamie, Gurcay, Burcu, Fincher, Katrina, Scott, Sydney E., Moore, Don, Atanasov, Pavel, Swift, Samuel A., Murray, Terry, Stone, Eric, and Tetlock, Philip E., 'Psychological Strategies For Winning a Geopolitical Forecasting Tournament', *Psychological Science* 25/5 (2014): 1106–1115.
Mellers, Barbara, Stone, Eric, Atanasov, Pavel, Rohrbaugh, Nick, Metz, Sarah E., Ungar, Lyle, Bishop, Michael, Horowitz, Michael, Merkle, Ed, and Tetlock, Philip E., 'The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics', *Journal of Experimental Psychology: Applied* 21/1 (2015): 1–14.
Mellers, Barbara, Stone, Eric, Murray, Terry, Minster, Angela, Rohrbaugh, Nick, Bishop, Michael, Chen, Eva, Baker, Joshua, Hou, Yuan, Horowitz, Michael, Ungar, Lyle, and Tetlock Philip E.. 'Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions.' *Perspectives on Psychological Science*, 10/3 (2015): 267-281.
Mishra, Sandeep. 'Decision-Making Under Risk Integrating Perspectives From Biology, Economics, and Psychology.' *Personality and Social Psychology Review*, (18/3) (2014): 280-307.
Miller, Greg. 'Former CIA official cites agency's failure to see al-Qaeda's rebound,' The Washington Post, May 3, 2015, Accessed January 19, 2015. http://www.washingtonpost.com/world/national-security/former-cia-official-cites-agencys-failure-to-see-al-qaedas-rebound/2015/05/03/d68e7292-f028-11e4-8abc-d6aa3bad79dd_story.html.
Moore, David T. *Sensemaking: A Structure for an Intelligence Revolution*. Washington, DC: Government Printing Office, 2012.
Moore, Don A. and Cain, Daylian M. 'Overconfidence and Underconfidence: When and Why People Underestimate (and Overestimate) the Competition', *Organizational Behavior and Human Decision Processes* 103/2 (2007): 197–213.
Mozer, Michael, Pashler, Harold, and Homaei, Hadjar, 'Optimal Predictions in Everyday Cognition: The Wisdom of Individuals or Crowds?', *Cognitive Science* 32/7 (2010): 1133–1147.
Munger, Charlie. 'A Lesson on Elementary, Worldly Wisdom As It Relates To Investment Management & Business.' Address at the University of Southern California Business School, 1994. Accessed January 19, 2016. https://old.ycombinator.com/munger.html.
Murphy, Allan H. and Winkler, Robert L. 'Reliability of Subjective Probability Forecasts of Precipitation and Temperature.' *Journal of the Royal Statistical Society. Series C* (Applied Statistics), 26/1 (1977): 41-47.
National Research Council, *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences*. Washington, D.C.: National Academies Press, 2011.
National Security Agency. 'Critical Thinking Course Syllabus.' Accessed January 19, 2016. http://www.au.af.mil/au/awc/awcgate/dia/ndic_moore_crit_analysis_nsa_syllabus.pdf.
Odean, Terrance, 'Do Investors Trade Too Much?', *The American Economic Review* 89/5 (1999): 1279–1298.
Office of the Director of National Intelligence. 'Intelligence Community Directive 203: Analytic Standards.' Accessed January 19, 2016. http://fas.org/irp/dni/icd/icd-203.pdf.
Office of Personnel Management. 'Training and Development Policy Wiki.' Accessed January 19, 2016. http://www.opm.gov/WIKI/training/Training-Transfer.ashx.
Pearl, Judea, *Probabilistic Reasoning in Intelligence Systems: Networks of Plausible Inference*, San Francisco: Morgan Kaufmann Publishers, 1988.
Puvathingal, Bess J. and Hantula, Donald A. 'Revisiting the Psychology of Intelligence Analysis: From Rational Actors to Adaptive Thinkers', *American Psychologist* 67/3 (2012): 199–210.
Rottenstreich, Yuval, and Tversky, Amos. 'Unpacking, repacking, and anchoring: advances in support theory', *Psychological Review* 104/2 (1997): 383–406.
Reinhart, Carmen M. and Rogoff, Kenneth. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press, 2009.
Reiber, Steven. 'Intelligence Analysis and Judgmental Calibration', *International Journal of Intelligence and Counterintelligence* 17/1 (2004): 97–112.
Rubin, Philip E., Statement to the Subcommittee on Investigations and Oversight Committee on Science, Space, and Technology, US House of Representatives, April 6, 2011, Accessed January 19, 2015. https://science.house.gov/sites/republicans.science.house.gov/files/documents/hearings/2011%2004%2001%20RubinTestimony.pdf.
Sagan, Carl, *Demon-Haunted World: Science as a Candle in the Dark*. New York: Ballantine Books, 2011.

Sedlmeier, Peter and Gigerenzer, Gerd. 'Teaching Bayesian Reasoning in Less than Two Hours', *Journal of Experimental Psychology: General* 130/3 (2001): 380–400.
Shynkaruk, Jody M. and Thompson, Valerie A., 'Confidence and Accuracy in Deductive Reasoning', *Memory and Cognition* 34/3 (2006): 619–632.
Surowiecki, James, *The Wisdom of Crowds* (New York: Anchor, 2005.
Taleb, Nassim N., *The Black Swan: The Impact of the Highly Improbable*. New York: Random House 2010.
Taleb, Nassim N. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets* New York: Random House, 2005.
Karl H. Teigen, 'Overestimation of subjective probabilities', *Scandinavian Journal of Psychology*, 15/1 (1974): 56–62.
Karl H. Teigen, 'Subjective sampling distributions and the additivity of estimates', *Scandinavian Journal of Psychology*, 15/1 (1974): 50–55.
Tetlock, Philip E., *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press, 2005.
Tetlock, Philip E., Mellers, Barbara, Rohrbaugh, Nick, and Chen, Eva. 'Forecasting Tournaments Tools for Increasing Transparency and Improving the Quality of Debate', *Current Directions in Psychological Science* 23/4 (2014): 290–295.
Tetlock, Philip E.. 'Cognitive Biases and Organizational Correctives: Do Both Disease and Cure Depend on the Politics of the Beholder?', *Administrative Science Quarterly* 45/2 (2000): 293–326.
Tetlock, Philip E. and Gardner, Dan. *Superforecasting: The Art and Science of Prediction*, New York, NY: Crown, 2015.
Tetlock, Philip E.. and Mellers, Barbara A. 'Intelligent Management of Intelligence Agencies: Beyond Accountability Ping-Pong', *American Psychologist* 66/6 (2011): 542–554.
Tversky, Amos, and Kahneman, Daniel, 'Availability: A heuristic for judging frequency and probability', *Cognitive Psychology* 5/2 (1973): 207–232.
Tversky, Amos, Kahneman, Daniel. 'Judgment Under Uncertainty: Heuristics and Biases.' *Science* (1974) 1124-1131.
Tversky, Amos, and Koehler, Derek J., 'Support Theory: A Nonextensional Representation of Subjective Probability', *Psychological Review* 101/4 (1994): 547–567.
Ward, Michael D., Greenhill, Brian D., Bakke, Kristin M. 'The perils of policy by p-value: Predicting civil conflicts', *Journal of Peace Research* 47/4 (2010): 363–375.
Wason, Peter C. *Psychology of Reasoning: Structure and Content*. Cambridge: Harvard University Press, 1972.
Wastell, Colin A., 'Cognitive Predispositions and Intelligence Analyst Reasoning', *International Journal of Intelligence and CounterIntelligence* 23/3 (2010): 449–460.
Wegener, Duane T. and Petty, Richard E. 'The flexible correction model: The role of naive theories of bias in bias correction.' *Advances in Experimental Social Psychology* 29 (1997): 142-208.
Wehrey, Frederic, Kaye, Dalia Dassa, Watkins, Jessica, Martini, Jeffrey, Guffey, Robert A. *The Iraq Effect: The Middle East After the Iraq War*. Washington D.C.: RAND Corporation, 2010.
Weick, Karl E. *Sensemaking in Organizations*. Thousand Oaks: Sage, 1995.
Wheaton, Kristan J. 'Evaluating Intelligence: Answering Questions Asked and Not', *International Journal of Intelligence and Counterintelligence* 22/4 (2009): 614–631.
Yamnill, Siriporn, and McLean, Gary N., 'Theories Supporting Transfer of Training', *Human Resource Development Quarterly* 12/2 (2001): 195–208.