

# Operationalizing Air Force Critical Thinking

Lt Col James M. Davitch, USAF\*

Lt Col Robert D. Folker Jr., USAF

Disclaimer: The views and opinions expressed or implied in the Journal are those of the authors and should not be construed as carrying the official sanction of the Department of Defense, Air Force, Air Education and Training Command, Air University, or other agencies or departments of the US government. This article may be reproduced in whole or in part without permission. If it is reproduced, the Air and Space Power Journal requests a courtesy line.

## The Requirement for Critical Thinkers

Air Force senior leaders have stressed the importance of developing and maintaining critical thinking capability. The service has approached the requirement as an academic shortfall, failing to accord this important skill its place as a core combat capability. The 2015 Air Force Future Operating Concept (AFFOC) plainly states that the Air Force must “recruit [and] assess individuals with [the] demonstrated potential for critical thinking” to successfully fight and win in contested environments.<sup>1</sup> Although the Air Force articulated an ambitious end state to build and utilize Airmen of the future who can think critically about vexing issues, it is not properly identifying personnel who possess the necessary skills. The USAF has habitually relied on intuitive assessments regarding high-stakes outcomes in uncertain conditions. Individual judgment is typically plagued by overconfidence, cognitive biases, and other psychological factors that lead to poor decision making. The Air Force needs a more deliberate approach if it wants to improve critical thinking so that it can make better decisions across a range of areas including strategic planning, budgeting, human capital management, intelligence, medicine, and acquisition.

Implementing a forecasting program is one low-cost method which would allow the Air Force to measure critical thinking, provide accountability, and identify Airmen with the ability to demonstrate and improve critical thinking by mitigating cognitive errors. To start this process, critical thinking is defined as a mode of reasoning in which one improves the quality of their thought by skillfully analyzing, assessing, and reconstructing their thought processes. A forecasting program, as will be discussed in this article, will provide the best means to measure progress.

Before discussing a practical implementation plan, it is useful to identify why, given multiple requests to improve critical thinking, it has not yet occurred. This problem requires a different resourcing strategy than the typical Air Force acquisition response to meet requirements. Critical thinking is essential to waging modern warfare today, but its intangible nature complicates the service’s ability to resource it

---

\*In the Summer 2017 edition, *Air & Space Power Journal* published an article by Col Adam J. Stone, USAF, entitled, “Critical Thinking Skills in USAF Developmental Education.” The authors believe that the proposal within could be one way to quantitatively measure and develop critical thinkers within the Air Force to meet Colonel Stone’s objectives.

as compared to how it resources most other combat capabilities. For instance, to adequately meet operational plan requirements for defensive counterair, the Air Force understands it must purchase a certain amount of jets, radars, and air-to-air missiles.

Thus, the Air Force is able to measure this traditional combat capability by the number of aircraft, weapons, and qualified aircrew. The Air Force approach to develop critical thinking has primarily consisted of formal training classes, such as the Critical Thinking and Structured Analysis course at Goodfellow AFB, Texas. While such attempts may be helpful, no process exists to routinely measure the critical thinking capability within the Air Force. Accurately measuring critical thinking cannot be done by counting graduates from a course. Rather, the individual critical thinking skills of each Airman should be developed and measured throughout their careers.

Critical thinking skills should be measured over time in a way similar to how instructor pilots conduct periodic check rides for their students. In a cognitive check ride, the evaluator can host a sort of forecasting debriefing or a survey, from which data can be used to improve thought processes. Until we hold ourselves accountable for our assessments derived from critically analyzing problems, it is impossible to judge whether one's subjective opinion is worth anything.

### The Illusion of Expert Judgment

In the absence of a systematic effort to collect critical thinking metrics, the Air Force turns to those with experience. Considering experienced individuals have seen and often been instrumental in key decision-making events, this seems to make perfect sense. It is not unreasonable to assume that these individuals would be best suited to recommend future solutions simply by their experiences. Unless decisions are captured in a policy memorandum, most experienced individuals rarely have any documented history of making the best decisions. Indeed, multiple scientific studies have shown that individual judgment is habitually plagued by overconfidence, cognitive biases, and other psychological factors.<sup>2</sup> Oftentimes, those affected by decision makers' judgments do not know whether critical thought was applied to a problem, or if the best decision to solve it was made. In many cases, the thought processes behind decisions simply are not documented using a standard rubric.

In the absence of these things, the Air Force by and large resorts to considering qualifying details such as one's time in service or some outward signifier of experience, such as a weapons school graduate patch on the uniform that signifies some specialized training or experience, rather than solid evidence of critical thought and good decision making. Certainly, extensive experience carries a quality all its own, but experience by itself does not equate to skill in critical thinking. Individuals with unexamined records of success should not answer complex predictive questions based solely on their intuition. At the very least, these same individuals, when asked to provide critical thought, should first be held to an objective standard that measures the secondary and tertiary effects of a proposed course of action.

### Some Inconvenient Results

Col Adam "Mez" Stone was one of the first Air Force officers to measure critical thinking ability. He used a standardized exam called the Watson-Glaser Critical

Thinking Appraisal (WGCTA). The test consisted of 40 questions, measured five critical thinking skills, and provided a means for identifying critical thinking ability in comparison to a similar reference population.<sup>3</sup> He knew senior leaders were asking for better critical thinkers, but his first task was to establish a baseline of behavior and to answer the question, “Where do we stand, right now?” His results, which were published in the fall of 2008, became an indictment of the Air Force’s critical thinking skills at the time.<sup>4</sup> The group of 180 junior Air Force officers who were the test subjects scored well below average when compared to the graduate degree norm group.

While studying at the Air War College (AWC) in 2015, Colonel Stone used the WGCTA again for a similar study of officers’ critical thinking skills at Air Command and Staff College (ACSC), AWC, and the School of Advanced Air and Space Studies (SAASS). In his study, SAASS students scored in the 61st percentile. The ACSC and AWC students scored in the 36th percentile, which was below average in comparison to similar master’s level programs.<sup>5</sup> The 2015 study concluded with a condemnation of the Air Force’s failure to appropriately educate and train its personnel to develop critical thinking skills through professional military education programs. His assertion is coincident with demands at the highest levels of our leadership for better critical thinking skills.

Despite Colonel Stone’s efforts to measure the Air Force’s critical thinking capability, there are still no sustained, long-term measurements collected within it. Although measuring critical thinking will not in and of itself provide a complete picture, the mere fact of having individuals make verifiable assessments will improve their critical thinking skills. In short, performance will improve through measurement, feedback, and repetition. The Air Force should capitalize on Colonel Stone’s findings and begin to methodically gather data to measure and improve the critical thinking skills of Airmen.

## Ways and Means: Practicing and Measuring Critical Thinking

Participants who learn to overcome cognitive traps by measuring their performance and adjusting their approach based on reliable feedback will demonstrate a quantifiable ability to think critically, consistent with the definition proposed earlier in this article. Fortunately, there is evidence that one’s subjective judgment can be aided in several ways to avoid mental pitfalls. In so doing, we may identify critical thinkers like Colonel Stone did, build critical thinking capability, and adequately respond to our senior leaders’ stated request for critical thinkers.

Few areas are as fraught with cognitive pitfalls as forecasting. While we do not dispute there are many avenues to improve one’s critical thinking skills, attempting to anticipate future events provides unique opportunities for individuals to get unambiguous feedback, identify cognitive errors, and improve skills. Therefore, due to the proven success demonstrated by the Good Judgment Project, an Intelligence Advanced Research Projects Activity (IARPA)-funded geopolitical forecasting tournament and research study in which “thousands of people around the world predict global events,” this article recommends a long-term critical thinking program which uses forecasting as one measure of critical thinking ability.<sup>6</sup> The program should include a modest amount of training to deal with typical errors in reasoning, such as

overconfidence, bias, and base-rate neglect.<sup>7</sup> The program's participants would make predictive estimates based on numeric probabilities (that is, 40 or 60 percent), rather than possible or probable estimative language. Finally, the program should track performance over time.

Multiyear-long research studies funded by IARPA have shown impressive results with this approach. The first IARPA tournament began in 2011 and explored the potential of crowd-sourced forecasting. Participants made predictions about real-world events, which were then judged by their forecasts' precision. Perhaps the most important aspect of the IARPA forecasting events was that it measured participants' performance longitudinally. These measurements identified individuals who consistently improved and performed well over time. Dubbed super forecasters, they demonstrated the same critical thinking skills, such as bias mitigation and open-mindedness, the Air Force desires in its personnel.

Thus, the primary method to develop critical thinking is submitting regular forecasts in areas of specific interest to the Air Force. For example, since the DOD programming, budgeting, and acquisitions cycle takes years to produce a new weapons system, it is necessary to make the right decisions as to which weapon systems the Air Force should invest to counter a future adversary threat. When making these forecasts, individuals should characterize uncertainty and express that characterization in probabilistic terms through predictive analysis that drives good decision making. Social scientists have published empirical data showing the ability to improve one's forecasting accuracy can be cultivated.<sup>8</sup> They have identified characteristics that differentiate between those who are better and worse at accurately predicting the results of a course of action over a period. Those elements are not indicative of natural-born intelligence or aptitude, but rather a mental determination to exercise critical thought and learn from mistakes. Critical thinkers will take the feedback obtained by measuring their performance, critique the process they used to make a forecast, and improve their decision making.

The process of evaluating an individual's forecast might appear to generate subjective results, which is the reason that the forecasting questions and scoring should be done by an independent central authority, such as the Air Force Research Laboratory. Moreover, proper academic preparation can help minimize the influence of natural heuristics and biases yielding forecasts with remarkable precision.<sup>9</sup> Several studies in the field of decision theory show that a modest amount of preparation can radically improve cognitive performance compared to those who do not receive training.<sup>10</sup> Training is needed that helps identify certain cognitive errors including overconfidence, confirmation bias, and base rate neglect. Daniel Kahneman, Amos Tversky, and Philip Tetlock all explored critical thinking in great detail as it relates to forecasting and cognitive dissonance.<sup>11</sup>

## What's Your Brier Score? Operationalizing Critical Thinking

The results of repeated assessments should be graded using a Brier Score, which is a useful way to verify the accuracy of a probability forecast. Brier Scores provide a quantitative means to compare and improve critical thinking while also holding individuals accountable for their estimates. For instance, consider the following question,

“Will the ruler of Country X conduct a nuclear test by the end of 2017?” The outcome is binary, the leader will (100 percent) or will not (0 percent) test a nuclear device. Assume a predictive analyst forecasts a 60 percent chance the test occurs and a 40 percent chance that it does not. If Country X conducts the test, then the score for the assessment would be 0.16. If it does not, the score would be 0.36. Since the Brier Score measures error, the lower the number is, the better, like a golf score.<sup>12</sup>

If the Air Force commits to improving its personnel’s critical thinking skills through a forecasting program, it could prove both inexpensive and lucrative. This program would require an administrator function to manage enrollment, generate forecast questions, and score the results. But how might the program attract participants? One option is through monetary incentives. The Air Force already incentivizes individuals to gain and maintain foreign language capabilities. If they attain a high enough reading, writing, and speaking proficiency level on the Defense Language Proficiency Test, they then receive additional compensation. If the Air Force judges that critical thinking skill is as valuable as, or more so than, foreign language capability, then there is a precedent for such incentive pay.

Alternatively, individuals could opt into the program purely to better their cognitive capabilities and compete with peers. It may be possible that the pursuit of a better Brier Score might be incentive enough to improve cognition. Studies show that job satisfaction routinely eclipses financial incentives as primary drivers of personal fulfillment.<sup>13</sup> Since a Brier Score is an objective method to determine the accuracy of a forecast, it levels the playing field. This approach could spotlight a young, inexperienced Airman seeking a reputation for being a person whose thinking is objective and uncluttered by bias. It could also repel those who have established a reputation for fear that their lack of CT skill will be exposed. In short, because it provides accountability, some may avoid establishing a Brier Score if given a choice.

Just as the Air Force requires physical training (PT) culminating in regular tests, so should it mandate participation in a “cognitive PT” program. While coercive, this approach could maximize participation at the lowest cost. Over time, the Brier Score could become a part of the Air Force culture, and the benefits would become obvious to all. Results from multiple large-scale forecasting tournaments revealed, “Prediction accuracy is possible when people participate in a setup that rewards only accuracy—and not the novelty of the explanation, or loyalty to the party line.”<sup>14</sup> In other words, competition like this fosters critical thinking while sharpening skills on an individual level. Furthermore, a mandatory competitive program may lend itself to developing and asking questions that can be answered, measured, and scored.

Competitive events are not new for the military. For decades, fighter pilots have trained against rival squadrons during “turkey shoot” events. Winners receive accolades and the recognition of their peers. The Air Force would be well-served by a cognitive turkey shoot, challenging participants to form their conclusions based on openly available information, thereby granting agency to the individual and allowing motivated professionals to best demonstrate their analytic prowess. Ideally, to check a peer’s decision-making process, individuals might routinely ask each other, “So, what’s your Brier Score?”

Furthermore, the prediction tournament proposed in this article would be one way to quantitatively measure and develop critical thinkers within the Air Force to

meet Colonel Stone's objectives. For instance, the top forecasters in the prediction tournament should be measured for critical thinking skills according to Colonel Stone's method to test for a correlation between forecasters with above average Bri'er's scores and higher than average critical thinking skills. If a positive correlation exists, then the forecasting tournament may prove to be one of the most effective ways to measure and develop stronger critical thinking skills within the Air Force.

## Summary

In short, we must value critical thinking as a core combat capability and measure it. It requires the same degree of training, monitoring, and validation that flying qualification demands. The Air Force would never allow a nonqualified aviator to pilot an aircraft. The risk to individual life and equipment is too great. Similarly, we must ask, why would we be less stringent about larger situations of uncertainty that could introduce risk to thousands? In areas that demand verified critical thinking skill, why would we turn to one's intuitive judgment that may be susceptible to unmitigated cognitive error?

President John F. Kennedy once said, "Too often we . . . enjoy the comfort of opinion without the discomfort of thought."<sup>15</sup> As scientific studies have shown, intuitive judgment is flawed. Institutionalizing a culture of critical thinking will complement expert intuition by mitigating cognitive error and bias. In doing so, the Air Force will step toward a process that rewards true skill through measurement, accountability, feedback, and improvement. ✪

## Notes

1. USAF, *Air Force Future Operating Concept*, <http://www.af.mil/Portals/1/images/airpower/AFFOC.pdf>, 30 September 2015, 43.

2. Welton Chang, Eva Chen, Barbara Mellers, and Philip Tetlock, "Developing Expert Political Judgment: The Impact of Training and Practice on Judgmental Accuracy in Geopolitical Forecasting Tournaments," *Judgment and Decision Making* 11, no. 5, 509–26, September 2016, <http://journal.sjdm.org/16/16511/jdm16511.pdf>.

3. Goodwin Watson and Edwin Glaser delineated the five skills of critical thinking: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments.

4. Col Adam J. Stone, "Critical Thinking Skills of Air Force Intelligence Officers: Are We Developing Better Critical Thinkers?" (master's thesis, National Defense Intelligence College, 2008).

5. Col Adam J. Stone, *Critical Thinking Skills of US Air Force Senior and Intermediate Developmental Education Students* (Maxwell AFB, AL: Air War College, 2016).

6. Office of the Director of National Intelligence, Intelligence Advanced Research Projects Activity (IARPA), "The Good Judgment Project," accessed 26 July 2017, <https://www.iarpa.gov/index.php/newsroom/iarpa-in-the-news/2015/439-the-good-judgment-project>.

7. Base rate neglect, or base rate bias, is a formal fallacy. If presented with related general information, such as 85 percent of small businesses fail within the first 5 years, the base rate, and specific information, such as this particular business owner has special training or skill, the mind tends to ignore the former and focus on the latter.

8. Chang, Chen, Mellers, and Tetlock, "Developing Expert Political Judgment," 509–26.

9. Based on the research report generated after one large-scale predictive tournament, this training can be executed in 45 minutes.

10. Barbara Mellers et al., "Psychological Strategies for Winning Geopolitical Forecasting Tournaments," *Psychological Science* 25, no. 5, 2014, <http://journals.sagepub.com/doi/abs/10.1177/0956797614524255>.

11. Daniel Kahneman and Amos Tversky, two eminent psychologists, have written extensively on each.

12. A Brier Score measures the mean squared error of one's assessment. If Country X's leader is deposed, the Brier Score is calculated as  $[(1.00 - 0.60)^2 + (0.00 - 0.40)^2] / 2 = 0.16$ . If Country X conducts a nuclear test, the Brier Score is calculated as  $[(0.00 - 0.60)^2 + (1.00 - 0.40)^2] / 2 = 0.36$ .

13. Daniel J. Benjamin, Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones, "What Do You Think Would Make You Happier? What Do You Think You Would Choose?," *American Economic Review* 102, no. 5 (2012): 2083–110, <https://www.aeaweb.org/articles?id=10.1257/aer.102.5.2083>.

14. Angela Chen, "Philip Tetlock's Tomorrows," *The Chronicle of Higher Education*, 5 October 2015, <http://www.chronicle.com/article/Philip-Tetlock-s-Tomorrows/233507>.

15. John F. Kennedy Presidential Library and Museum, "Kennedy Library and Museum Rededication Film (1993): Source of Quotation, 'We Enjoy the Comfort of Opinion. . .,' Address by President Kennedy, 11 June 1962, Yale University Commencement," <https://www.jfklibrary.org/Research/Research-Aids/Ready-Reference/Kennedy-Library-Fast-Facts/Yale-University-Commencement-Address.aspx>.



#### **Lt Col James M. Davitch, USAF**

Lieutenant Colonel Davitch (BA, Penn State University; MA, University of Oklahoma; MS, Blue Horizons Fellowship) is the division chief of intelligence operations at Air Force Global Strike Command, Barksdale AFB, LA. He is the command lead for all intelligence unit support, formal training, and analytical functions. Previously, he completed the Blue Horizons Fellowship at Air University where he analyzed new methods to identify indicators and warning of conflict. Lieutenant Colonel Davitch is a career intelligence officer and has held intelligence assignments at the tactical level in the combat air forces, operational level at the air operations center, and strategic level at Headquarters, US Air Force. He is a 2007 graduate of the Air Force Weapons School, where he was also an instructor.



#### **Lt Col Robert D. Folker Jr., USAF**

Lieutenant Colonel Folker (BS, Excelsior College; MS, National Intelligence University) is an airpower strategist in the Checkmate Directorate, Headquarters US Air Force, Washington, DC. He leads concept and strategy development to integrate emerging USAF capabilities into combatant command operational planning; directly supports the deputy chief of staff (DCS) for air, space, and cyber operations; and provides best military advice to the chief of staff of the Air Force. Previously, he supported the DCS for intelligence, surveillance, and reconnaissance (ISR) as the program element monitor for the Air Force Distributed Common Ground System and intelligence training programs and oversaw the programming, budgeting, and execution of more than \$5B. Before his assignment at the Pentagon, he served as the director of operations for the 19th Weapons Squadron, US Air Force Weapons School (USAFWS). As an intelligence weapons instructor, Lieutenant Colonel Folker conducted sensitive reconnaissance operations across the globe and combat ISR operations in Operations Enduring Freedom and Iraqi Freedom.

Distribution A: Approved for public release; distribution unlimited.

<http://www.airuniversity.af.mil/ASPJ/>