Evaluating Intelligence

By Kristan Wheaton, Mercyhurst College
and Diane Chido, McManis & Monsalve Associates

Hand-wringing, brow-beating, and finger-pointing have all followed the events of September 11, 2001, and the National Intelligence Estimate on Iraq's Continuing Programs for Weapons of Mass Destruction (WMD) of 2002. The "failure of analysis" in the U.S. intelligence community has become a common catchphrase in American popular culture. Congress insisted on and the administration approved two new institutions to correct this problem: the Department of Homeland Security and the Office of the Director of National Intelligence.

Six years and billions of dollars after 9/11 and the WMD estimate, how do we know if these and other changes have improved the quality of intelligence analysis? More important, how do we evaluate intelligence generally?

THE NATURE OF INTELLIGENCE

The intelligence process seems straightforward. Under the traditional view:

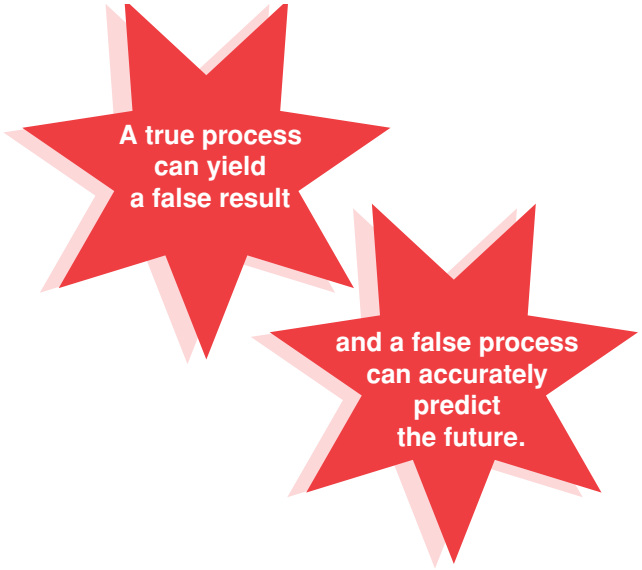
- Decision makers state requirements.
- Collectors go out and gather relevant and preferably unique information.
- Analysts figure out what it all means.
- Analysts give a neat and tidy answer to the decision makers, who are, in turn, grateful for these insights.

Those dependent on such estimates may sleep better imagining that the analytic process is as exacting and methodical as science, but this is a dangerous myth. (In fact,

the traditional description of the process — the intelligence cycle — is under increasing attack. See Clark, 2004.)

Decision makers are often unclear on their requirements, and even if they are clear, layers of bureaucracy may water down or alter the original requests by the time they get to the intelligence professionals tasked to produce the answers. Collection has never really been a discrete process. It is ongoing, and today's collectors can provide relevant information at any point in the process. Furthermore, the explosion of secondary source (print) information has given the power of collection to anyone with an internet connection, while it increasingly calls into question the costs and benefits of traditional methods of acquiring unique information from individuals.

Intelligence analysis itself is still both a science and an art. Analysts are engaged in estimating the likelihood of future events — something essentially unknowable no matter how good the current evidence. Even with sophisticated tools and methodologies, analysts must make a variety of decisions concerning the reliability of the sources and their own analytic confidence, which, in turn, effectively weights, or, as some say, skews, the evidence in one direction or another.



Ultimately, intelligence is an externally focused process designed to reduce the level of uncertainty for a decision maker, using information from all sources (see Wheaton and Beerbower, 2006). One way analysts can clearly produce assessments indicating a degree of certainty is to use standardized words of estimative probability (see Kent, 1962). Clearly indicating the likelihood of particular events so that a decision maker can determine appropriate action is the analyst's most important role. However, even when analysts convey a clear degree of probability, they can still get the wrong answer.

One of the most famous instances of this phenomenon was the Battle of Waterloo. Before marching into battle on the afternoon of June 18, 1815, Napoleon assessed his odds in achieving victory over British forces as 9 in 10. Napoleon's army at Waterloo "was generally considered to be much superior on paper, in terms of quality, experience, homogeneity and even motivation." (Roberts 2001). To this day, historians debate how and why Napoleon did not win, attributing the British victory to factors from the rainy weather delaying the French attack, to the size of the field, to the superior choice of terrain by Wellington. This list is by no means exhaustive, but it points to the heart of the matter. Napoleon, an experienced and successful general, adopted his tactics for this battle from his distinctive and successful fighting methods, including the use of massed infantry assaults and prolonged artillery bombardments. Napoleon's strategy worked in previous engagements, but for whatever reason, was unsuccessful at Waterloo (Roberts 2001).

How, then, does a decision maker go about evaluating the analyst's work?

THE WEATHERMAN ANALOGY

In physics, chemistry, or even psychology or sociology, a scientist confronted by a particularly difficult problem will first attempt to solve the simplest possible case. This method is designed to get at the most basic elements of a phenomenon and often yields deep insights. How might such an approach work in intelligence analysis? Clearly, the problem would have to have its answers couched in probabilities. In addition, the problem should avoid historical examples, as the multitude of facts and the opinions of historians could easily obscure any insights that might be derived from the experiment.

It is in this context that we introduce the weatherman analogy. Weathermen, like intelligence analysts, deal daily with complex data and, again like intelligence analysts, they voice their forecasts in terms of probabilities. Furthermore, while no one expects the weatherman to be always correct, there is a general expectation of competence, and there are penalties for a weatherman (or an intelligence analyst) who is consistently incorrect.

In addition, while historical examples might prove contentious for all the wrong reasons, discussions of the weather are generally benign, allowing, therefore, the possibility of some deep revelation. (As stated by Kin Hubbard, American humorist, "Don't knock the weather; nine-tenths of the people couldn't start a conversation if it didn't change once in a while.") Finally, weather can be closely associated with intelligence, as it can often be presented alongside the intelligence brief during operational planning.

Consider, then, the example of a weatherman who just started a new job at a local news station. He only knows one "rule" (he is still in weatherman school): If it is sunny today, then it is likely to be sunny tomorrow. Later that day, the boss asks him what the weather will be tomorrow. The weatherman looks out the window at the sun shining and replies that it is likely to be sunny tomorrow. When the boss wakes up the next day, he sees that it is sunny. Was the weatherman right? Suppose the boss woke up and it was raining. Was the weatherman's forecast wrong? Most people would say that, under these circumstances, the weatherman was wrong if it rained and right if it was sunny.

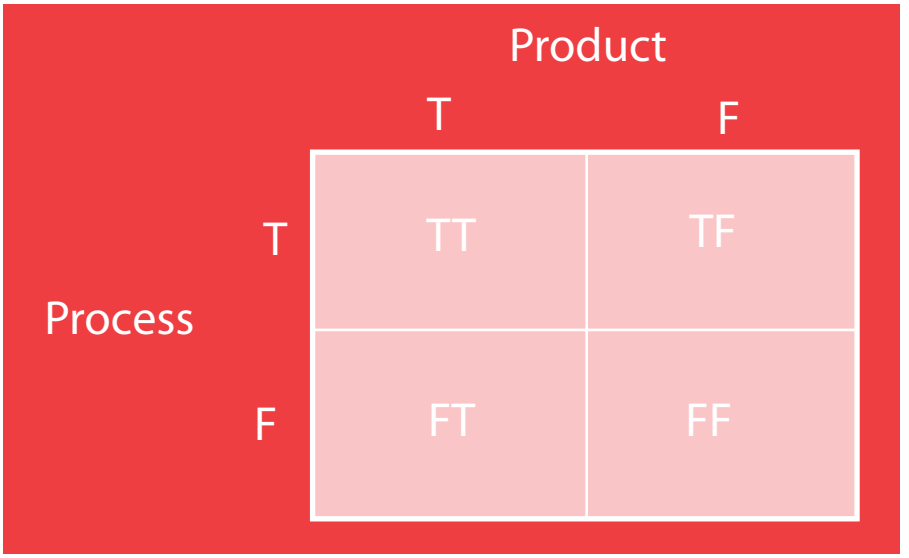


Figure 1: Results of looking at product and process individually

Let's push the analogy a bit further. Suppose the weatherman had predicted sun and it had rained. The weatherman walks into the office the next morning and is yelled at by his boss for making the wrong forecast. After the boss calms down a bit, the weatherman, humbled, asks the boss, "How can I improve?" The boss instructs him to learn more rules, to which the weatherman readily agrees. "I know I should learn more rules," the weatherman goes on, "but, based on what I knew yesterday, what should I have done differently?" Most people, if put in the boss' shoes would say, "Nothing." If the answer is that he did what he could, applied his one rule correctly and made his best estimate based on the outcome of that process, how can we then say that the weatherman was "wrong"?

Now, consider a bad weatherman. He also knows only one rule. However, this weatherman learned the rule incorrectly — he thinks that if it is sunny today, it is likely to rain tomorrow. His boss asks what the forecast is for tomorrow, so the weatherman looks out the window, sees that it is sunny, and predicts that it will rain tomorrow. The next day, the boss wakes up to a rainy day. Was this weatherman right? What if the boss wakes up to a sunny day? Was the weatherman wrong?

Depending on how you look at the questions, it is possible to make the case that both weathermen were both right and wrong. One implication of this disconcerting result is to assume that there is no way to evaluate inherently probabilistic statements, like those made by weathermen and analysts. Another possibility, and one that we intend to explore, is that "right" and "wrong" (or "true" and "false") are inappropriate labels when it comes to evaluating an analytic estimate.

PROCESS AND PRODUCT

"Right" and "wrong" don't work because there is both a product and a process to be evaluated in any intelligence (or weather) report. Intelligence products are final estimates derived from a process. (The intelligence practitioner also produces nonestimative products. The estimate, however, is the "highest form of the art" and as such, deserves the most consideration in developing a theory of evaluation.) Analysts create such products in the form of analytic reports and presentations for decision makers. Analytic products come in a variety of forms, ranging from technological trends and product development plans to financial forecasts and stock assessments.

Process, on the other hand, is the application of rules, methodologies, and other analytic tools to the information available on a specific target. One can draw parallels between processes used in intelligence and in the medical profession, as Richards Heuer reasoned:

"The doctor observes indicators (symptoms) of what is happening, uses his or her specialized knowledge of how the body works to develop hypotheses that might explain these observations, conducts tests to collect additional information to evaluate the hypotheses, then makes a diagnosis. This medical analogy focuses attention on the ability to identify and evaluate all plausible hypotheses. Collection is focused narrowly on information that will help to discriminate the relative probability of alternate hypothesis [sic]" (Heuer, 1999).

In the weatherman analogy,

the prediction was the product and the application of the single rule was the process used to get that product. The failure to separate the evaluation of the process from the evaluation of the product is what likely caused the sense of confusion at the end of the previous section. What, then, are "right" (or "true") processes and products and what are "wrong" (or "false") products? (Note that right and wrong are used synonymously with true and false in this context. True and false are probably better terms, however, and we will use them throughout the remainder of the discussion.)

Here are some examples:

- True processes are those that produce generally reliable results and include statistics, tested methodologies, and perhaps even intuition based on experience — for example, a company completing its balance sheet each quarter with quantitative, verifiable data (unless it is Enron, of course), using a variety of clearly reliable, repeatable, mathematical methods.
- False processes are those that produce generally unreliable results and include invalidated methods, such as reading horoscopes, tea leaves, or goat entrails.
- True products are those that turn out to be correct in their major conclusions (using 20-20 hindsight). The Value Line Investment Survey has ranked companies on "timeliness" for decades. Most of their top-ranked stocks have done very well, allowing Value Line to state, "Value Line #1 Ranked stocks outperformed the Dow Jones Industrial Average by 20 to 1."
- False products are those that turn out to be incorrect in their major conclusions (using 20-20 hindsight). Faulty assessments, such as 2002 and early 2003 reports of large stockpiles of weapons of mass destruction in Iraq, are good examples (Commission, 2005).

Evaluating intelligence by looking at product and process individually, rather than collectively, yields four possible results. Figure 1 shows the four different combinations of product and process, with T standing for true and F for false.

Once an analyst gives an intelligence product to a decision maker, it immediately becomes open to evaluation. Problems rarely arise from cases in which the product and process are both true. Likewise, it is not surprising when obviously false processes yield false results. Of more interest are those times when a true process yields a false result or when a false process accurately predicts the future.

WHICH IS MORE IMPORTANT TO INTELLIGENCE?

Given the probabilistic nature of intelligence products, evaluators of intelligence must look at both product and process. However, is one of these factors more important than the other? Should evaluators give additional weight to either the product or the process in the course of an evaluation?

The answer depends entirely on the time frame of the evaluation. Returning to the two weathermen, the bad weatherman who incorrectly used the rule to forecast the weather clearly misapplied the process. Although he might manage to forecast the weather correctly some of the time, the product is less likely to be reliable over time, and his estimates will fail more often than not due to this false process. That said, when the bad weatherman correctly forecasts the weather, the estimate is unlikely to invite evaluation from his boss, who does not have the time to constantly monitor employees who appear to be doing their jobs correctly.

Eventually, however, the bad weatherman is bound to establish a poor track record and to attract negative attention. This situation parallels that of intelligence analysts and decision makers. When intelligence is "wrong," senior management will inevitably conduct a review and attempt to correct the flaws using recommendations that are based on postmortem analyses. In these cases, decision makers will likely focus their concerns on the failure of the product, which, in turn, encourages ad hoc solutions. At the other end of the spectrum, the intelligence professionals will attempt to focus the debate on the process, often claiming that even the best of processes, like the best of weathermen, occasionally get it wrong. As these two extremes circle around each other, the fact that they are only examining half of the possible outcomes is lost.

The truth is that product and process cannot be separated. Because of the probabilistic nature of intelligence, decision makers and intelligence professionals alike must consider both process and product in the evaluation of intelligence. While the focus may shift from one to the other, they are completely intertwined. Scientists refer to this as "complementarity," while the Chinese have a symbol for it: the familiar picture of yin and yang.

Evaluating intelligence, then, is about more than merely determining whether it was right or wrong. It is about examining both the product and the processes of intelligence in every case and determining whether either was incorrect. Decision makers have a legitimate role in fixing processes that are broken, whether the analysis was ultimately proved to be right or wrong. Similarly, decision makers also have to be prepared to write off — in much the same way we write off failed weather forecasts — properly prepared analyses that get the future wrong as bad luck, not bad intelligence.

The question really becomes this: Can we afford not to have a well-considered and robust evaluative procedure? Our current strategy of spending millions on evaluating only product failures is much like evaluating chess matches in which only brown-haired people win — it is a strategy that will miss a large number of important cases. Likewise, without some sort of effort to identify and promote true

processes, the intelligence community is doomed to repeat its failures, thus increasingly calling into question the wisdom of spending even a penny on intelligence.

Intelligence, in one respect, is intended to warn, to give decision makers time to take action, to change the course of events and, in turn, change an unfavorable potential future into a more favorable actual one. Should responsible decision makers receive such a warning (born, of course, from a true process), they should take action to prevent the unfavorable potential future. Should this action be successful, the warning that spurred the action will, of course, turn out to be incorrect. Clearly, this will be a true process that, ultimately, yields a false product, but the decision makers will congratulate their intelligence professionals for their fine work in warning them in time to take action.

In some, perhaps many cases, the causal chain will not be so clear. The outcome might be the result of many small actions, rather than one or two decisive ones. There may be multiple analyses over many years, so that it becomes increasingly difficult to determine where a process went wrong or where it was performed adequately. In the end, the precise course of events may be so unclear that evaluation of the process may be extremely difficult.

The fact that evaluation is difficult and in some select cases may be impossible is no reason not to institute such assessments. This is a common problem in all quality control systems. The question becomes "What is our tolerance for error?" In industry, some machines can create parts that have millimeter tolerances, while other machines are far more or far less exacting. Likewise, in intelligence, some products and processes will require evaluation in greater or lesser detail than others.

Some products and processes might not lend themselves to after-the-fact evaluation at all. Determining where these limits are (or should be) is the appropriate work of both the intelligence practitioners and their clients. Deciding not to have the debate because some cases might not lend themselves to oversight seems foolish and shortsighted.

REFERENCES

Clark, Robert M. (2004). *Intelligence Analysis: A Target-Centric Approach*. CQ Press, Washington, DC.

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction Report to the President (March 31, 2005). Accessed March 31, 2006, at www.wmd.gov/report/.

Heuer, Richards J. (1999). *Psychology of Intelligence Analysis*. CIA Center For the Study of Intelligence, Washington, DC. Accessed March 31, 2006, at www.cia.gov/csi/Psychology/PsychofIntelNew.pdf.

Heuer, Richards J., Jr. (2005). "Limits of Intelligence Analysis," *Orbis*, Winter, pp. 75–94.

Hubbard, Kin. "Don't knock the weather; nine-tenths of the people couldn't start a conversation if it didn't change once in a while." Accessed 22 April 22, 2006, at www.nonstopenglish.com/reading/quotations/k_Weather.asp.

Kent, Sherman (1962). "Words of Estimative Probability," *Studies In Intelligence*, Fall. Accessed April 2, 2006, at www.cia.gov/csi/books/shermankent/6words.html.

Lowenthal, Mark M. (2003). *Intelligence: From Secrets to Policy*. CQ Press, Washington, DC, p20.

Roberts, Andrew (2001). *Napoleon & Waterloo*. Simon & Schuster, New York, p. 168.

Wheaton, Kristan J., and Beerbower, Michael T. (2006). "Toward a New Definition of Intelligence," *Stanford Law & Policy Review*, (17)1.

Diane Chido is a senior associate at McManis & Monsalve Associates using Intelligence Led Management (ILM), a trademarked approach integrating intelligence analysis with decision-maker-focused management consulting for a range of public and private sector organizations. She is a graduate of the Mercyhurst College Applied Intelligence program, and is co-author of Structured Analysis of Competing Hypotheses: Theory and Application, and co-editor of the Analysts' Cookbook. She can be reached at: DChido@McManis-Monsalve.com

Kristan J. Wheaton is an assistant professor of intelligence studies at Mercyhurst College in Erie, Pennsylvania. He is the author of The Warning Solution: Intelligent Analysis in the Age of Information Overload_and co-author of Structured Analysis of Competing Hypotheses: Theory and Application. He also teaches and provides materials for short courses in structured analysis of competing hypotheses to the U.S. Department of Defense. He can be reached at kwheaton@mercyhurst.edu.