25 September 1984

MEMORANDUM FOR: Editor, Studies in Intelligence

25X1 FROM: [redacted]
Chief, Classification Review Division
Office of Information Services, DA

SUBJECT: [redacted] Review of Two Articles 25X1
25X1

REFERENCE: Memo Under CSI 84-0902, dated 19 September 1984 to Chairman, Editorial Board from Editor, Studies in Intelligence, Subject: [redacted] 25X1

We reviewed the two articles from Studies in Intelligence written by Richards J. Heuer, Jr. and found no classified information in either. The unclassified articles are titled "Do You Really Need More Information?" and "Cognitive Biases: Problems in Hindsight Analysis."

25X1 [redacted]

Attachments:
Memo CSI 84-0902
Copies of Two Articles

25X1 WARNING NOTICE
INTELLIGENCE SOURCES
AND METHODS INVOLVED [redacted]

25X1

DDA/OIS/CRD/ [redacted] (25 Sep 84)

Distribution:
Orig - Addressee w/atts
1 - CRD Liaison w/Misc w/atts
1 - CRD Chrono w/o atts

Next 1 Page(s) In Document Denied

Sanitized Copy Approved for Release 2010/03/15 : CIA-RDP87-00181R000200440006-3

More collection may not be the way to get better analysis.

DO YOU REALLY NEED MORE INFORMATION?

Richards J. Heuer, Jr.

The difficulties associated with intelligence analysis are often attributed to the inadequacy of available information. Thus the intelligence community has invested heavily in improved collection systems while analysts lament the comparatively small sums devoted to enhancing analytical resources, improving analytical methods, or gaining better understanding of the cognitive processes involved in making analytical judgments.

This article challenges the often implicit assumption that lack of information is the principal obstacle to accurate intelligence estimates. It describes psychological experiments that examine the relationship between amount of information, accuracy of estimates based on this information, and analysts' confidence in their estimates. In order to interpret the disturbing but not surprising findings from these experiments, it identifies four different types of information and discusses their relative value in contributing to the accuracy of analytical judgments. It also distinguishes analysis whose results are driven by the data from analysis that is driven by the conceptual framework employed to interpret the data. Finally, it outlines a strategy for improving intelligence analysis.

The key findings from the relevant psychological experiments are:

- **Once an experienced analyst has the minimum information necessary to make an informed judgment, obtaining additional information generally does not improve the accuracy of his estimates. Additional information does, however, lead the analyst to become more confident in his judgment, to the point of overconfidence.**
- **Experienced analysts have an imperfect understanding of what information they actually use in making judgments. They are unaware of the extent to which their judgments are determined by a few dominant factors, rather than by the systematic integration of all available information. Analysts use much less of the available information than they think they do.**

As will be noted in further detail below, these experimental findings should not necessarily be accepted at face value. There are, for example, circumstances when additional information does contribute to more accurate analysis. There are also circumstances when additional information—particularly contradictory information—decreases rather than increases an analyst's confidence in his judgment. But the experiments highlight important relationships between the amount of information an analyst has available, judgmental accuracy, and analyst confidence. An understanding of these relationships has implications for both the management and conduct of intelligence analysis. Such an understanding suggests analytical procedures and management initiatives that may indeed contribute to more accurate analytical judgments. It also suggests that resources needed to attain a better understanding of

the entire analytical process might profitably be diverted from some of the more massive and costly collection programs.

Betting on the Horses

Intelligence analysts have much in common with doctors diagnosing illness, psychologists identifying behavioral traits, stockbrokers predicting stock market performance, college admissions officers estimating future academic performance, weather forecasters, and horserace handicappers. All accumulate and interpret a large volume of information to make judgments about the future. All are playing an "information game," and all have been the subject of psychological research to determine how this game gets played.

Experts in these and similar professions analyze a finite number of identifiable and classifiable kinds of information to make judgments or estimates that can subsequently be checked for accuracy. The stock market analyst, for example, commonly works with information relating to price/earnings ratios, profit margins, earnings per share, market volume, and resistance and support levels. By controlling the information made available to a number of experts and then checking the accuracy of judgments based on this information, it has been possible to conduct experiments concerning how people use information to arrive at analytical judgments.

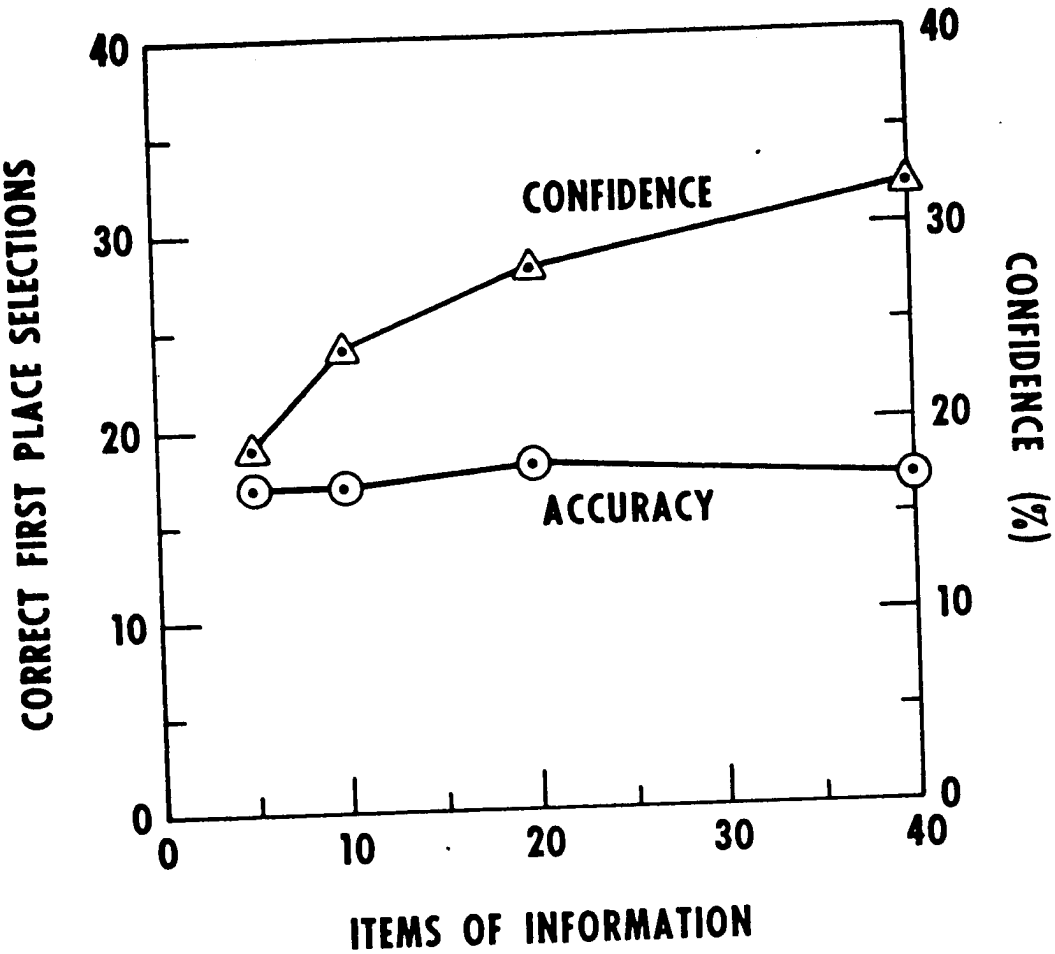
In one experiment,[1] eight experienced horserace handicappers were shown a list of 88 variables found on a typical past-performance chart—for example, weight to be carried; percentage of races in which horse finished first, second, or third during the previous year; jockey's record; number of days since horse's last race. Each handicapper was asked to identify, first, what he considered to be the five most important items of information—those he would wish to use to handicap a race if he were limited to only five items of information per horse. Each was then asked to select the 10, 20, and 40 most important variables he would use if limited to those levels of information.

The handicappers were at this point given true data (sterilized so that horses and actual races could not be identified) for 40 past races and were asked to rank the top five horses in each race in order of expected finish. Each handicapper was given the data in increments of the 5, 10, 20 and 40 variables he had judged to be most useful. Thus, he predicted each race four times—once with each of the four different levels of information. For each prediction, each handicapper assigned a value from 0 to 100 percent to indicate his degree of confidence in the accuracy of his prediction.

When the handicappers' predictions were compared with the actual outcomes of these 40 races, it was clear that average accuracy of predictions remained the same regardless of how much information the handicappers had available. Three of the handicappers actually showed less accuracy as the amount of information increased, two improved their accuracy, and three were unchanged. All, however, expressed steadily increasing confidence in their judgments as more information was received. This relationship between amount of information, accuracy of the handicappers' prediction of the first place winners, and the handicappers' confidence in their predictions is shown graphically in Figure 2. Note that with only five items of information, the handicappers' confidence was well calibrated with their accuracy, but that as additional information was received, they became overconfident.

The same relationship between amount of information, accuracy, and analyst confidence has been confirmed by similar experiments in other fields, especially

[1] Paul Slovic, "Behavioral Problems of Adhering to a Decision Policy," Mimeo, 1973.



clinical psychology.[2] In one experiment, a psychological case file was divided into four sections representing successive chronological periods in the life of a relatively normal individual. Thirty-two psychologists with varying levels of experience were asked to make judgments on the basis of this information. After reading each section of the case file, the psychologists answered 25 questions (for which there were known answers) about the personality of the subject of the file. As in other experiments, increasing information resulted in a strong increase in confidence but a negligible increase in accuracy.[3]

A series of experiments to examine the mental processes of medical doctors diagnosing illness found little relationship between thoroughness of data collection and accuracy of diagnosis. Medical students whose self-described research strategy stressed thorough collection of information (as opposed to formation and testing of hypotheses) were significantly below average in the accuracy of their diagnoses. It seems that the explicit formulation of hypotheses directs a more efficient and effective search for information.[4]

---

[2] For a list of references, see Lewis R. Goldberg, "Simple Models or Simple Processes? Some Research on Clinical Judgments," *American Psychologist*, 23 (1968), p. 484.

[3] Stuart Oskamp, "Overconfidence in Case-Study Judgments," *Journal of Consulting Psychology*, 29 (1965), pp. 261-265.

[4] Arthur S. Elstein *et al.*, *Medical Problem Solving: An Analysis of Clinical Reasoning* (Harvard University Press, Cambridge, Mass. and London), 1978; pp. 270 and 295.

Modeling Expert Judgment

Another significant question concerns the extent to which analysts possess an accurate understanding of their own mental processes. How good is our insight into how we actually weigh evidence in making judgments? For each situation we analyze, we have an implicit "mental model" consisting of beliefs and assumptions about which variables are most important and how they are related to each other. If we have good insight into our own mental model, we should be able to describe accurately which variables we have considered most important in making our judgments.

There is strong experimental evidence, however, that such self-insight is faulty. The expert perceives his own judgmental process, the number of different kinds of information he takes into account, as being considerably more complex than is in fact the case. He overestimates the importance he attributes to factors that have only a minor impact on his judgment, and underestimates the extent to which his decisions are based on a very few major variables. In short, our mental models are far simpler than we think, and the analyst is typically unaware not only of which variables *should* have the greatest influence on his judgments, but also of which variables actually are having the greatest influence.

This has been shown by a number of experiments in which analysts were asked to make quantitative estimates concerning a relatively large number of cases in their area of expertise, with each case defined by a number of quantifiable factors. In one experiment, stock market analysts were asked to predict long-term price appreciation for each of 50 securities, with each security being described in such terms as price/earnings ratio, corporate earnings growth trend, and dividend yield.[5] After completing this task, the analysts were instructed to explain how they reached their conclusions, including a description of how much weight they attached to each of the variables. They were told to be sufficiently explicit so that another person going through the same information could apply the same judgmental rules and arrive at the same conclusions.

In order to compare the analyst's verbal rationalization with the judgmental policy reflected in his actual decisions, multiple regression analysis or some similar statistical procedure can be used to develop a mathematical model of how each analyst actually weighed and combined information on the relevant variables.[6] There have been at least eight studies of this type in diverse fields,[7] including one involving prediction of future socioeconomic growth of underdeveloped nations.[8] The mathematical model based on the analyst's actual decisions is invariably a better predictor of that analyst's past and future decisions than his own verbal description of how he makes his judgments.

Although the existence of this phenomenon has been amply demonstrated in many experiments, its causes are not well understood. The literature on these experiments contains only the following speculative explanation:

> Possibly our feeling that we can take into account a host of different factors comes about because although we remember that at some time or other we have attended to each of the different factors, we fail to notice that it is seldom more than one or two that we consider at any one time.[9]

[5] Paul Slovic, Dan Fleissner, and W. Scott Bauman, "Analyzing the Use of Information in Investment Decision Making: A Methodological Proposal," *The Journal of Business*, 45 (1972), pp. 283-301.

[6] For a discussion of the methodology, see Slovic, Fleissner, and Bauman, *loc. cit.*

[7] For a list of references, see Paul Slovic and Sarah Lichtenstein, "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior and Human Performance*, 6 (1971), p. 684.

[8] David A. Summers, J. Dale Taliaferro, and Donna J. Fletcher, "Subjective vs. Objective Description of Judgment Policy," *Psychonomic Science*, 18 (1970) pp. 249-250.

[9] R. N. Shepard, "On Subjectively Optimum Selection Among Multiattribute Alternatives," in M. W. Shelly, II and G. L. Bryan, eds., *Human Judgments and Optimality* (New York: Wiley, 1964), p. 266.

How Can This Happen To Smart People Like Us?

In order to evaluate the relevance and significance of these experimental findings in the context of our own experience as intelligence analysts, it is necessary to distinguish four types of additional information that an analyst might receive:

1. **Additional detail about variables already included in our analysis.** Much raw intelligence reporting falls into this category. We would not expect such supplementary information to affect the over-all accuracy of our judgment, and it is readily understandable that further detail which is consistent with previous information increases our confidence. Analyses for which considerable depth of detail is available to support the conclusions tend to be more persuasive to their authors as well as to their readers.

2. **Information on additional variables.** Such information permits the analyst to take into account other factors that may affect the situation. This is the kind of additional information used in the horserace handicapper experiment. Other experiments have employed some combination of additional variables and additional detail on the same variables. The finding that our judgments are based on a very few critical variables rather than on the entire spectrum of evidence helps to explain why information on additional variables does not normally improve predictive accuracy. Occasionally, in situations when there are known gaps in our understanding, a single report concerning some new and previously unconsidered factor—for example, an authoritative report on some policy initiative or planned coup d'etat—will have a major impact on our judgments. Such a report would fall into either of the next two categories of new information.

3. **Information concerning the level or value attributed to variables already included in the analysis.** An example of such information would be the horserace handicapper learning that a horse he thought would carry 110 pounds will actually carry only 106. Current intelligence reporting tends to deal with this kind of information—for example, the analyst learning that coup planning was far more advanced than he had anticipated. New facts clearly affect the accuracy of our judgments when they deal with changes in variables that are critical to our estimates. Our confidence in judgments based on such information is influenced by our confidence in the accuracy of the information, as well as by the amount of information.

4. **Information concerning which variables are most important and how they relate to each other.** Knowledge and assumptions concerning which variables are most important and how they are interrelated comprise our mental model that tells us *how* to analyze the data we receive. Explicit investigation of such relationships is one factor that distinguishes systematic research from current intelligence reporting and raw intelligence. In the context of the horserace handicapper experiment, for example, handicappers had to select which variables to include in their analysis. Is weight carried by a horse more, or less, important than several other variables that affect a horse's performance? Any information that affects this judgment affects how the handicapper analyzes the available

data; that is, it affects his mental model. Events in Iran in late 1978 have probably had a permanent impact on the mental models not only of the Iran analysts, but of analysts dealing with internal politics in any of the Muslim countries. As a consequence of Iranian developments, analysts will consciously or subconsciously pay more attention and attribute increased importance to conservative religious opposition movements throughout the Muslim world.

The accuracy of our judgment depends upon both the accuracy of our mental model (the fourth type of information discussed above) and the accuracy of the values attributed to the key variables in the model (the third type of information discussed above). Additional detail on the variables in our model and information on other variables that do not in fact have a significant influence on our judgment (the first and second types of information) have a negligible impact on accuracy, but form the bulk of the raw material we work with. These kinds of information increase confidence because our conclusions seem to be supported by such a large body of data.

Important characteristics of the mental models analysts use vary substantially according to the type of intelligence problem faced. In particular, information is accorded a different role in different types of problems. In analyzing the readiness of a military division, for example, there are certain rules or procedures to be followed. The totality of these procedures comprise our mental model that influences our perception of the overhead photography of the unit and guides our judgment concerning what information is important and how this information should be analyzed to arrive at judgments conerning readiness. Most elements of the mental model can be made explicit so that other analysts may be taught to understand and follow the same analytical procedures and arrive at the same or very similar results. There is broad though not necessarily universal agreement on what the best model is. There are relatively objective standards for judging the quality of analysis, for the conclusions follow logically from the application of the agreed upon model to the available data.

Most important in the context of this discussion is that the accuracy of the estimate depends primarily upon the accuracy and completeness of the available data. If one makes the reasonable assumption that the analytical model is correct, and the further assumption that the analyst properly applies this model to the data, then the accuracy of the analytical judgment depends entirely upon the accuracy and completeness of the data. Because the analytical results are so heavily determined by the data, this may be called *data-driven analysis*.

At the opposite end of this spectrum is *conceptually-driven analysis*. For example, in most political analysis the questions to be answered do not have neat boundaries and there are many unknowns. The number of potentially relevant variables, and the diverse and imperfectly understood relationships between these variables, involve the analyst in enormous complexity and uncertainty. There is little tested theory to inform the analyst concerning which of the myriad pieces of information are most important, and how they should be combined to arrive at estimative judgments. In the absence of any agreed upon analytical schema, the analyst is left to his own devices. He interprets information with the aid of mental models which are largely implicit rather than explicit. The assumptions he is making concerning political forces and processes in the subject country may not be apparent even to the analyst himself. Such models are not representative of an analytical consensus. Other analysts examining the same data may well reach different

conclusions, or reach the same conclusions for different reasons. This analysis is conceptually driven because the outcome depends at least as much upon the conceptual framework employed to analyze the data as it does upon the data itself.

Not all military analysis is data-driven, and not all political analysis is concept-driven. In citing military and political analysis as the opposite ends of this spectrum, we are making a broad generalization that permits many exceptions. In comparing economic and political analysis, we note that economic models are usually more explicit, and that they represent a consensus of at least broad factions within the discipline.

In the light of this distinction between data-driven and conceptually driven analysis, it is instructive to look at the function of the analyst responsible for current intelligence, especially current political intelligence as distinct from longer-term research. His daily work is driven by the incoming reporting from overseas which he must interpret for dissemination to consumers, but this is not what is meant by data-driven analysis. The current intelligence analyst must provide immediate interpretation of the latest, often unexpected events. Apart from his store of background information, he may have no data other than the initial, usually incomplete report. Under these circumstances, his interpretation is based upon his implicit mental model of how and why events normally transpire in the country for which he is responsible. The accuracy of his judgment depends almost exclusively upon the accuracy of his mental model, for he has virtually no other basis for judgment.

If the accuracy of our mental model is the key to accurate judgment, it is necessary to consider how this mental model gets tested against reality and how it can be changed so that we can improve the accuracy of our judgment. There are two reasons that make it hard to change one's mental model. The first relates to the nature of human perception and information processing. The second concerns the difficulty, in many fields, of learning what truly is the best model.

Partly because of the nature of human perception and information processing, beliefs of all types tend to resist change. This is especially true of the implicit assumptions and "self-evident truths" that play an important role in determining our mental models.[10] Information that is consistent with our existing mindset is perceived and processed easily. However, since our mind strives instinctively for consistency, information that is inconsistent with our existing mental image tends to be overlooked, perceived in a distorted manner, or rationalized to fit existing assumptions and beliefs.[11] Thus, new information tends to be perceived and interpreted in a way that reinforces existing beliefs.

A second difficulty in revising our mental models arises because of the nature of the learning process. Learning to make better judgments through experience assumes systematic feedback concerning the accuracy of previous judgments and an ability to link the accuracy of a judgment with the particular configuration of variables that promoted an analyst to make the judgment. In practice, however, we get little

---

[10] We are often surprised to learn that what are to us self-evident truths are by no means self-evident to others, or that self-evident truth at one point in time may be commonly regarded as naive assumption 10 years later.

[11] We are, of course, referring to subconscious processes; no analyst is consciously going to distort information that does not fit his preconceived beliefs. Important aspects of the perception and processing of new information occur prior to and independently of any conscious direction, and the tendencies described here are largely the result of these subconscious or preconscious processes.

systematic feedback, and even when we know a predicted event has occurred or failed to occur, we typically do not know for certain whether this happened for the reasons we had foreseen. Thus, an analyst's personal experience may be a poor guide to revision of his mental model.[12]

Improving Intelligence Analysis

To the intelligence policy maker seeking an improved intelligence product, our findings offer a reminder that this can be achieved by improving analysis as well as by improving collection. There are, of course, many traditional ways to seek to improve analysis—language and area training, revising employee selection and retention criteria, manipulating incentives, improving management, and increasing the number of analysts. Any of these measures may play an important role, but we ought not to overlook the self-evident fact that intelligence analysis is principally a cognitive process. If we are to penetrate to the heart and soul of the problem of improving analysis, we must somehow penetrate and affect the mental processes of the individuals who do the analysis. The findings in this article suggest a central strategy for pursuing that goal: **this strategy is to focus on improving the mental models employed by the analyst to interpret his data.** While this will be very difficult to achieve, it is so critical to effective intelligence analysis that even small improvement could have large benefits.

There are a number of concrete actions to implement this strategy of improving mental models that can be undertaken by individual analysts and middle managers as well as by organizational policy makers. All involve confronting the analyst with alternative ways of thinking. The objective is to identify the most fundamental analytical assumptions, then to make these assumptions explicit so that they may be critiqued and re-evaluated.

The basic responsibility for proper analysis rests, of course, with the individual analyst. To guide his information search and analysis, the analyst should first seek to identify and examine alternative models or conceptual frameworks for interpreting the already available information. Because people have very limited capacity for simultaneously considering multiple hypotheses, the alternatives should be written down and evidence compared against them in a systematic manner. This permits the analyst to focus on the degree to which the evidence is diagnostic in helping him select the best among competing models, rather than simply the degree to which it supports or undermines his own previous belief. This helps overcome the tendency to ignore the possibility that evidence consistent with one's own belief is equally consistent with other hypotheses. **The analyst must, from time to time, attempt to suspend his own beliefs and develop alternative viewpoints,** to determine if some alternative—when given a fair chance—might not be as compelling as one's own previous view. Systematic development of an alternative scenario generally increases the perceived likelihood of that scenario.

The analyst should then try to disprove, rather than prove, each of the alternatives. He or she should try to rebut rather than confirm hypotheses, and actively seek information that permits this rather than review passively

---

[12] A similar point has been made in rebutting the belief in the accumulated wisdom of the classroom teacher. "It is actually very difficult for teachers to profit from experience. They almost never learn about their long-term successes or failures, and their short-term effects are not easily traced to the practices from which they presumably arose." B. F. Skinner, *The Technology of Teaching* (Appleton-Century-Crofts, New York, 1968), pp. 112-113.

information flowing through the in box. It is especially important for the analyst to seek information that, if found, would disprove rather than bolster his own arguments. One key to identifying the kinds of information that are potentially most valuable is for the analyst to ask himself what it is that could make him change his mind. Adoption of this simple tactic would do much to avoid intelligence surprises.

Management can play a role by fostering research on the mental models of analysts. Since these models serve as a "screen" or "lens" through which we perceive foreign developments, research to identify the impact of our mental models on our analysis may contribute as much to accurate estimates as research focused more directly on the foreign areas themselves. When the mental models are identified, further research is in order to test the assumptions of these models. To what extend can one determine, empirically, what are the key variables and how these variables relate to each other in determining an estimated outcome?

Management should insist on far more frequent and systematic retrospective evaluation of analytical performance. One ought not to generalize from any single instance of a correct or incorrect estimate, but a series of related judgments that are, or are not, borne out by subsequent events can be very diagnostic in revealing the accuracy or inaccuracy of or mental model. Obtaining systematic feedback on the accuracy of our past judgments is frequently difficult or impossible, especially in the political analysis field.

Political estimates are normally couched in vague and imprecise terms (to say that something "could" happen conveys no information that can be disproven by subsequent events) and are normally conditional upon other developments. Even in retrospect, there are no objective criteria for evaluating the accuracy of most political estimates as they are presently written. In the economic and military fields, however, where estimates are frequently concerned with numerical quantities, systematic feedback on analytical performance is feasible. Retrospective evaluation should be standard procedure in those fields where estimates are routinely updated at periodic intervals. It should be strongly encouraged in all areas as long as it can be accomplished as part of an objective search for improved understanding, rather than to identify scapegoats or assess blame. This requirement suggests that retrospective evaluation ought to be done within the organizational unit and perhaps by the same analysts that prepared the initial evaluation, even if this results in some loss of objectivity.

The pre-publication review and approval process is another point at which management can impact on the quality of analysis. Such review generally considers whether a draft publication is properly focused to meet the perceived need for that publication. Are the key judgments properly highlighted for the consumer who scans but does not read in depth? Are the conclusions well supported? Is the draft well written? Review procedures should also explicitly examine the mental model employed by the analyst in searching for and examining his evidence. What assumptions has the analyst made that are not discussed in the draft itself, but that underlie his principal judgments? What alternative hypotheses have been considered but rejected? What could cause the analyst to change his mind? These kinds of questions should be a part of the review process. Management should also consider the advisability of assigning another analyst to play the role of devil's advocate.

One common weakness in the pre-publication review process is that an analyst's immediate colleagues and supervisor are likely to share a common mindset, hence these are the individuals least likely to raise fundamental issues challenging the

validity of the analysis. Peer review by analysts handling other countries or issues and with no specialized knowledge of the subject under review may be the best way to identify assumptions and alternative explanations. Such non-specialist review has in the past been a formal part of the review process, but it is not now common practice.

At the policy making level, CIA directors since 1973 have been moving the agency in directions that ensure CIA analysts are increasingly exposed to alternative mental models. The realities of bureaucratic life still produce pressures for conformity, but efforts are made to ensure that competing views have the opportunity to surface. There is less formal inter-agency coordination than there used to be, and increased use of informal coordination aimed more at surfacing areas of disagreement and the reasons therefore than at enforcing consensus.

Sharply increased publication of CIA analyses in unclassified form has stimulated challenge and peer review by knowledgeable analysts in academia and industry. The public debate that followed publication of several CIA oil estimates in 1977 is the most noteworthy case in point. Such debate can only sharpen the perception and judgment of the participating CIA analysts. The 1976 Team A–Team B experiment in competitive analysis of the strategic balance with the Soviet Union, on the other hand, was a miscarriage. Confrontation of alternative mental models is a critical element of the analytical process, but this confrontation must take place in an environment that promotes attitude change rather than hardening of positions.

The most recent development has been the formal establishment in December 1978 of the Review Panel within the National Foreign Assessment Center. The panel, which presently consists of three senior officials from the State Department, the military and academia, is designed to bring outside perspectives to bear on the review of major analytical products.

Conclusion

The function of intelligence is frequently described by analogy to a mosaic. Intelligence services collect small pieces of information which, when put together like a mosaic or a jigsaw puzzle, eventually enable us to see a clear picture of reality. The analogy suggests that accurate estimates depend primarily upon having all the pieces, that is, upon accurate and relatively complete information. It is important to collect and store the small pieces of information, as these are the raw material from which the picture is made; we never know when it will be possible to fit a piece into the puzzle. Much of the rationale for large, technical collection systems is rooted in this mosaic analogy.

The mosaic theory of intelligence is an oversimplification that has distorted perception of the analytical process for many years. It is properly applied only to what has been described as data-driven analysis. A broader theory of intelligence encompassing conceptually driven as well as data-driven analysis ought to be based on insights from cognitive psychology. Such insights suggest that the picture formed by the so-called mosaic is not a picture of reality, but only our self-constructed mental image of a reality we can never perceive directly. We form the picture first, and only then do we fit in the pieces. Accurate estimates depend at least as much upon the mental model we use in forming that picture as upon the accuracy and completeness of the information itself.

The mosaic theory of intelligence has focused attention on collection, the gathering together of as many pieces as possible for the analyst to work with. A more

psychologically oriented view would direct our concern to problems of analysis, and especially to the importance of mental models that determine what we collect and how we perceive and interpret the collected data. To the extent that this is the more appropriate guide to comprehending the analytical process, there are important implications for the management of intelligence resources. There seem to be inherent practical and theoretical limits on how much can be gained by efforts to improve collection, but an open and fertile field for imaginative efforts to improve analysis.

(This entire article is UNCLASSIFIED.)

Intelligence Vignette

ON ECONOMIC INTELLIGENCE

(from the Historical Intelligence Collection)

Perhaps the first record of economic reporting by an American foreign intelligence agent is that of William Carmichael, who had been dispatched to Holland in late 1776 by Silas Deane, in the guise of a merchant, served as the secret agent in France for the Committee of Secret Correspondence—the foreign intelligence directorate of the Continental Congress. He had tasked Carmichael with a number of economic intelligence requirements, partially reported to the Committee of Secret Correspondence in Carmichael's dispatch from Amsterdam of November 2, 1776. In the report, which went by way of a secret mail facility on St. Eustatia Island, Carmichael reported:

"You have been threatened that the Ukraine would supply Europe with tobacco. It must be long before that time can arrive. I have seen some of its tobacco here, and the best of it is worse than the worst of our ground leaf. Four hundred thousand pounds have been sent here this year."

Our past intelligence judgments were neither as good as we think they were, nor as bad as others believe.

COGNITIVE BIASES: PROBLEMS IN HINDSIGHT ANALYSIS

Richards J. Heuer, Jr.

Psychologists observe that limitations in man's mental machinery (memory, attention span, reasoning capability, etc.,) affect his ability to process information to arrive at judgemental decisions. In order to cope with the complexity of our environment, these limitations force us to employ various simplifying strategies for perception, comprehension, inference, and decision. Many psychological experiments demonstrate that our mental processes often lead to erroneous judgments. When such mental errors are not random, but are consistently and predictably in the same direction, they are known as *cognitive biases*.

This article discusses three cognitive biases affecting how we evaluate ourselves and how others evaluate us as intelligence analysts.

- The analyst who thinks back about how good his past judgments have been will normally overestimate their accuracy.
- The intelligence consumer who thinks about how much he learned from our reports will normally underestimate their true value to him.
- The overseer of intelligence production who conducts a postmortem of an intelligence failure to evaluate what we should have concluded from the information that was available will normally judge that events were more readily foreseeable than was in fact the case.

Evidence supporting the existence of these biases is presented in detail in the second part of this article. None of the biases is surprising. We have all observed these tendencies in others—although probably not in ourselves. What may be unexpected is that these biases are not solely the product of self-interest and lack of objectivity. They are specific examples of a broader phenomenon that seems to be built into our mental processes and that cannot be overcome by the simple admonition to be more objective. In the experimental situations described below, conscious efforts to overcome these biases were ineffective. Experimental subjects with no vested interest in the results were briefed on the biases and encouraged to avoid them or compensate for them, but there was little or no improvement in their estimates. While self-interest and lack of objectivity will doubtless aggravate the situation, bias is also cause by mental processes unrelated to these baser instincts.

The analyst, consumer, and overseer evaluating estimative performance all have one thing in common: they are exercising hindsight. They take their current state of knowledge and compare it with what they or others did or could or should have known before the current knowledge was received. Intelligence estimation, on the other hand, is an exercise in foresight, and it is the difference between these two kinds of thought—hindsight and foresight—that seems to be the source of the bias.

The amount of good information that is available obviously is greater in hindsight than in foresight. There are several possible explanations of how this affects mental processes. One is that the additional information available for hindsight apparently

changes our perceptions of a situation so naturally and so immediately that we are largely unaware of the change. When new information is received, it is immediately and unconsciously assimilated into our prior knowledge. If this new information adds significantly to our knowledge—if it tells us the outcome of a situation or the answer to a question about which we were previously uncertain—our mental images are restructured to take the new information into account. With the benefit of hindsight, for example, factors previously considered relevant may become irrelevant, and factors previously thought to have little relevance may be seen as determinative.

Once our view has been restructured to assimilate the new information, there is virtually no way we can accurately reconstruct our prior mental set. We may *recall* our previous estimates if not much time has elapsed and they were precisely articulated, but we apparently cannot *reconstruct* them accurately. The effort to reconstruct what we previously thought about a given situation, or what we would have thought about it, is inevitably influenced by our current thought patterns. Knowing the outcome of a situation makes it harder to imagine other outcomes that we might have considered. Simply understanding that our mind works in this fashion, however, does little to help us overcome the limitation.

The overall message we should learn from an understanding of these biases is that our intelligence judgments are not as good as we think they are, or as bad as others seem to believe. Since the biases generally cannot be overcome, they would appear to be facts of life that need to be taken into account in evaluating our own performance and in determining what evaluations to expect from others. This suggests the need for a more systematic effort to:

- Define what should be expected from intelligence analysis.
- Develop an institutionalized procedure for comparing intelligence judgments and estimates with actual outcomes.
- Measure how well we live up to the defined expectations.

Discussion of Experiments

The experiments that demonstrated the existence of these biases and their resistance to corrective action were conducted as part of a research program in decision analysis funded by the Defense Advanced Research Projects Agency. Before examining these experiments, it is appropriate to consider the nature of experimental evidence *per se*, and the extent to which one can generalize from these experiments to conclude that the same biases are prevalent in the intelligence community.

When we say that psychological experiments demonstrate the existence of a bias, we do not mean the bias will be found in every judgment by every individual. We mean that in any group of people, the bias will exist to a greater or lesser degree in most of the judgments made by a large percentage of the group. On the basis of the kind of experimental evidence discussed here, we can only generalize about the tendencies of groups of people, not make statements about individual analysts, consumers, or overseers.

All the experiments described below used students, not members of the intelligence community, as test subjects. There is, nonetheless, ample reason to believe the results can be generalized to apply to the intelligence community. The experiments deal with basic mental processes common to everyone, and the results do seem consistent with our personal experience. In similar psychological test using various experts (including intelligence analysts) as test subjects, the experts showed the same pattern of responses as students.

Our own imperfect efforts to repeat one of these experiments using CIA analysts support the validity of the findings. In order to test the assertion that intelligence analysts normally overestimate the accuracy of their past judgments, there are two necessary preconditions. First, analysts must make a series of estimates in quantitative terms—they must say not just that a given occurrence is probable, but that there is, for example, a 75-percent chance of its occurrence. Second, it must be possible to make an unambiguous determination whether the estimated event did or did not occur. When these two preconditions are present, one can then go back and test the analyst's recollection of his or her earlier estimate. Because CIA estimates are rarely stated in terms of quantitative probability, and because the occurrence of an estimated event within a specified time period often cannot be determined unambiguously, these preconditions are rarely met.

We did, however, identify several analysts in CIA's Office of Regional and Political Analysis who on two widely differing subjects had made quantitative estimates of the likelihood of events that we now know either did or did not occur. We went to these analysts and asked them to recall their earlier estimates. The conditions for this miniexperiment were far from ideal, and the results were not clear-cut, but they did tend to support the conclusions drawn from the more extensive and systematic experiments described below.

These reasons lead us to conclude that the three biases are found in intelligence community personnel as well as in the specific test subjects. In fact, one would expect the biases to be even greater in foreign affairs professionals whose careers and self-esteem depend upon the presumed accuracy of their judgments. We can now turn to more detailed discussion of the experimental evidence demonstrating these biases from the perspective of the analyst, consumer, and overseer.

The Analyst's Perspective [1]

Analysts interested in improving their own performance need to evaluate their past estimates in the light of subsequent developments. To do this, an analyst must either recall (or be able to refer to) his past estimates, or he must reconstruct his past estimates on the basis of what he remembers having known about the situation at the time the estimates were made. The effectiveness of the evaluation process, and of the learning process to which it gives impetus, depends in part upon the accuracy of these recalled or reconstructed estimates.

Experimental evidence suggests, however, a systematic tendency toward faulty memory of our past estimates. That is, when events occur, we tend to overestimate the extent to which we had previously expected them to occur. And conversely, when events do not occur, we tend to underestimate the probability we had previously assigned to their occurrence. In short, events generally seem less surprising than they should on the basis of past estimates. This experimental evidence accords with our intuitive experience; analysts, in fact, rarely seem very surprised by the course of events they are following.

In experiments to test the bias in memory of past estimates, 119 subjects were asked to estimate the probability that a number of events would or would not occur during President Nixon's trips to Peking and Moscow in 1972. Fifteen possible outcomes were identified for each trip, and each subject assigned a probability to each of these outcomes. The outcomes were selected to cover the range of possible developments and to elicit a wide range of probability values.

[1] This section is based on research reported by Baruch Fischhoff and Ruth Beyth in "'I Knew It Would Happen': Remembered Probabilities of Once-Future Things," *Organizational Behavior and Human Performance*, 13 (1975), pp. 1-16.

At varying time periods after the trips, the same subjects were asked to recall or reconstruct their predictions as accurately as possible. (No mention was made of the memory task at the time of the original prediction.) Then the subjects were asked to indicate whether they thought each event had or had not occurred during these trips.

When three to six months were allowed to elapse between the subjects' estimates and their recollection of these estimates, 84 percent of the subjects exhibited the bias when dealing with events they believed actually happened. That is, the probabilities they remembered having estimated were higher than their actual estimates of events they believed actually occurred. Similarly, for events they believed did not occur, the probabilities they remembered having estimated were lower than their actual estimates, although here the bias was not as great. For both kinds of events, the bias was more pronounced after three to six months had elapsed than when subjects were asked to recall estimates they had given only two weeks earlier.

In sum, knowledge of the outcomes somehow affected most test subjects' memory of their previous estimates of these outcomes, and the more time was allowed for memories to fade, the greater was the effect of the bias. The developments during the President's trips were perceived as less surprising than they would have been if actual estimates were compared with actual outcomes. For the 84 percent of the subjects who showed the anticipated bias, their retrospective evaluation of their estimative performance was clearly more favorable than was warranted by the facts.

The Consumer's Perspective[2]

When the consumer of intelligence reports evaluates the quality of the intelligence product, he asks himself the question, "How much did I learn from these reports that I did not already know?" In answering this question, there is a consistent tendency for most people to underestimate the contribution made by new information. This kind of "I knew it all along" bias causes consumers to undervalue the intelligence product.

That people do in fact commonly react to new information in this manner was confirmed in a series of experiments involving some 320 people, each of whom answered the same set of 75 factual questions taken from almanacs and encyclopedias. They were then asked to indicate how confident they were in the correctness of each answer by assigning to it a probability percentage ranging from 50 (no confidence) to 100 (absolute certainty).

As a second step in the experiment, subjects were divided into three groups. The first group was given 25 of the previously asked questions and instructed to respond to them exactly as they had previously. This simply tested the subjects' ability to remember their previous answers. The second group was given the same set of 25 questions but with the correct answers circled "for your [the subjects'] general information." They, too, were asked to respond by reproducing their previous answers. This tested the extent to which learning the correct answers distorted the subjects' memory of their previous answers, thus measuring the same bias in recollection of previous estimates that was discussed above from the analyst's perspective.

The third group was given a different set of 25 questions that they had not previously seen, but of similar difficulty so that results would be comparable with the

[2] The experiments described in this section are reported in Baruch Fischhoff, *The Perceived Informativeness of Factual Information*, Technical Report DDI-1 (Oregon Research Institute, Eugene, Ore., 1976).

other two groups. The correct answers were indicated, and the subjects were asked to respond to the questions as they would have had they not been told the answer. This tested the subjects' ability to remember accurately how much they had known before they learned the correct answer. The situation is comparable to that of the intelligence consumer who is asked to evaluate how much he learned from a report and who can do this only by trying to reconstruct the extent of his knowledge before he read the report.

The most significant results came from this third group of subjects. The group clearly overestimated what they had known originally and underestimated how much they learned from having been told the answer. For 19 of 25 items in one test and 20 of 25 items in another, this group assigned higher probabilities to the correct alternatives than it is reasonable to expect they would have assigned had they not already known the correct answers.

The bias was stronger for deceptive questions than for easier questions. For example, one of the deceptive questions was:

Aladdin's nationality was:

(a) Persian
(b) Chinese

The correct answer, which is surprising to most people, is Chinese. The average probabilities assigned to each answer by the three groups varied as follows:

- When subjects recalled their previous response without having been told the correct answer, the average of the probabilities they assigned to the two possible responses was:

  (a) .838
  (b) .134

  As these subjects did not know the correct answer, they had no opportunity to exhibit the bias. Therefore, the above figures are the base against which to compare the answers of the other two groups that were aware of the correct answer.

- When subjects tried to recall their previous response after having been told the correct answer, their average responses were:

  (a) .793
  (b) .247

- When subjects not previously exposed to the question were given the correct answer but asked to respond as they would have responded before being told the answer, their average responses were:

  (a) .542
  (b) .321

In sum, the experiment confirms the results of the previous experiment showing that people exposed to an answer tend to remember having known more than they actually did, and it demonstrates that people tend even more to exaggerate the likelihood that they would have known the correct answer if they had not been informed of it. In other words, *people tend to underestimate how much they learn from new information*. To the extent that this bias affects the judgments of intelligence consumers—and there is every reason to expect that it does—these consumers will tend to underrate the value of intelligence estimates.

The Overseer's Perspective[3]

An overseer, as the term is used here, is one who investigates intelligence performance by conducting a postmortem examination, for example, of why we failed to foresee the 1973 Yom Kippur War. Such investigations are carried out by Congress and by our own management, and independent judgments are also made by the press and others. For those outside the executive branch who do not regularly read the intelligence product, this sort of retrospective evaluation in cases of known intelligence failure is a principal basis for judgments about the quality of our intelligence analysis.

A fundamental question posed in any postmortem investigation of intelligence failure is: Given the information that was available at the time, should we have been able to foresee what was going to happen? Unbiased evaluation of intelligence performance depends upon the ability to provide an unbiased answer to this question.

Once an event has occurred, it is impossible to erase from our mind the knowledge of that event and reconstruct what our thought processes would have been at an earlier point in time. In reconstructing the past, there is a tendency toward determinism, toward thinking that what occurred was inevitable under the circumstances and therefore predictable. In short, there is a tendency to believe we should have foreseen events that were in fact unforeseeable on the basis of the available information.

The experiments reported here tested the hypotheses that knowledge of an outcome increases the perceived probability of that outcome, and that people who are informed of the outcome are largely unaware that this information has changed their perceptions in this manner.

A series of sub-experiments used brief (150-word) summaries of several events for which four possible outcomes were identified. One of these events was the struggle between the British and the Gurkhas in India in 1814. The four possible outcomes for this event were (1) British victory, (2) Gurkha victory, (3) military stalemate with no peace settlement, and (4) military stalemate with a peace settlement. Five groups of 20 subjects each participated in each sub-experiment. One group received the 150-word description of the struggle between the British and the Gurkhas with no indication of the outcome. The other four groups received the identical description but with one sentence added to indicate the outcome of the struggle—a different outcome for each group.

The subjects in all five groups were asked to estimate the likelihood of each of the four possible outcomes and to evaluate the relevance to their judgment of each fact in the description of the event. Those subjects who were informed of an outcome were placed in the same position as our overseer who, although knowing what happened, seeks to estimate the probability of that outcome without the benefit of hindsight. The results are shown in the table below.

Table

Experimental Groups	Average Probabilities Assigned to Outcomes			
	1	2	3	4
Not Told Outcome	33.8	21.3	32.3	12.3
Told Outcome 1	57.2	14.3	15.3	13.4
Told Outcome 2	30.3	38.4	20.4	10.5
Told Outcome 3	25.7	17.0	48.0	0.9
Told Outcome 4	33.0	15.8	24.3	27.0

[3] The experiments described in this section are reported in Baruch Fischhoff, "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty," *Journal of Experimental Psychology: Human Perception and Performance*, I, 3 (1975), pp. 288-299.

The group not informed of any outcome judged the probability of Outcome 1 as 33.8 percent, while the group told that Outcome 1 was the actual outcome perceived the probability of this outcome as 57.2 percent. The estimated probability was clearly influenced by knowledge of the actual outcome. Similarly, those informed that Outcome 2 was the actual outcome perceived this outcome as having a 38.4 percent probability, as compared with a judgment of only 21.3 percent for the control group with no outcome knowledge. An average of all estimated outcomes in six sub-experiments (a total of 2,188 estimates by 547 subjects) indicates that the knowledge or belief that an outcome has occurred approximately doubles the perceived probability that that outcome will occur.

The relevance that subjects attributed to any fact was also strongly influenced by which outcome, if any, they had been told was true. As Wohlstetter has indicated, "It is much easier after the fact to sort the relevant from the irrelevant signals. After the event, of course, a signal is always crystal clear. We can now see what disaster it was signaling since the disaster has occurred, but before the event it is obscure and pregnant with conflicting meanings."[4] The fact that knowledge of the outcome automatically restructures our judgments on the relevance of available data is probably one reason it is so difficult to reconstruct what our thought processes were or would have been without this outcome knowledge.

In several variations of this experiment, subjects were asked to respond as though they did not know the outcome, or as others would respond if they did not know the outcome. The results were little different, indicating that subjects were largely unaware of how knowledge of the outcome affected their own perceptions. The experiment showed that subjects were unable to empathize with how others would judge these situations. Estimates of how others would interpret the data were virtually the same as the subjects' own retrospective interpretations.

These results indicate that overseers conducting postmortem evaluations of what CIA should have been able to foresee in any given situation will tend to perceive the outcome of that situation as having been more predictable than it in fact was. Because they are unable to reconstruct a state of mind that views the situation only with foresight, not hindsight, overseers will tend to be more critical of intelligence performance than is warranted.

Can We Overcome These Biases?

We tend to blame biased evaluations of intelligence performance at best on ignorance, at worst on self-interest and lack of objectivity. These factors may also be at work, but the experiments described above suggest that the nature of our mental processes is a principal culprit. This is a more intractable cause than either ignorance or lack of objectivity.

The self-interest of the experimental subjects was not at stake; yet they showed the same kinds of bias with which we are familiar. Moreover, in these experimental situations the biases were highly resistant to efforts to overcome them. Subjects were instructed to make estimates as if they did not already know the answer, but they were unable to do so. In the experiments using 75 almanac and encyclopedia questions, one set of subjects was specifically briefed on the bias, citing the results of previous experiments; this group was instructed to try to compensate for the bias, but it too was unable to do so. Despite maximum information and the best intentions, the bias persisted.

---

[4] Roberta Wohlstetter, *Pearl Harbor: Warning and Decision* (Stanford University Press, Stanford Calif., 1962), p. 387.

This intractability suggests that the bias does indeed have its roots in the nature of our mental processes. The analyst who tries to recall his previous estimate after learning the actual outcome of events, the consumer who thinks how much a report has added to his prior knowledge, and the overseer who judges whether our analysts should have been able to avoid an intelligence failure, all have one thing in common. They are engaged in a mental process involving hindsight. They are trying to erase the impact of knowledge, so as to recall, reconstruct, or imagine the uncertainties they had or would have had about a subject prior to receiving more or less definitive information on that subject.

It appears, however, that the receipt of what is accepted as definitive or authoritative information causes an immediate but unconscious restructuring of our mental images to make them consistent with the new information. Once our past perceptions have been restructured, it seems very difficult, at best, to reconstruct accurately what our thought processes were or would have been before this restructuring.

There is one procedure that may help to overcome these biases. It is to pose such questions as the following. The analyst should ask himself, "If the opposite outcome had occurred, would I have been surprised?" The consumer should ask, "If this report had told me the opposite, would I have believed it?" And the overseer should ask, "If the opposite outcome had occurred, would it have been predictable given the information that was available?" These questions may help us to recall the degree of uncertainty we had prior to learning the content of a report or the outcome of a situation. They may help us remember the reasons we had for supporting the opposite answer, which we now know to be wrong.

This method of overcoming the bias can be tested by readers of this article, especially those who believe it failed to tell them much they had not already known. If this article had reported that psychological experiments show no consistent pattern of analysts overestimating the accuracy of their estimates, and of consumers underestimating the value of our product, would you have believed it? (Answer: Probably not.) If it had reported that psychological experiments show these biases to be caused only by self-interest and lack of objectivity, would you have believed this? (Answer: Probably yes.) And would you have believed it if the article had reported that these biases can be overcome by a conscientious effort at objective evaluation? (Answer: Probably yes.)

These questions may lead the reader to recall the state of his knowledge or beliefs before reading this article, and thus to highlight what he has learned from it—namely, that significant biases in the evaluation of intelligence estimates are attributable to the nature of human mental processes, not just to self-interest and lack of objectivity, and that they are, therefore, exceedingly difficult to overcome.