rates are not well known but must be inferred or researched, they are even less likely to be used.[149]

The so-called planning fallacy, to which I personally plead guilty, is an example of a problem in which base rates are not given in numerical terms but must be abstracted from experience. In planning a research project, I may estimate being able to complete it in four weeks. This estimate is based on relevant case-specific evidence: desired length of report, availability of source materials, difficulty of the subject matter, allowance for both predictable and unforeseeable interruptions, and so on. I also possess a body of experience with similar estimates I have made in the past. Like many others, I almost never complete a research project within the initially estimated time frame! But I am seduced by the immediacy and persuasiveness of the case-specific evidence. All the causally relevant evidence about the project indicates I should be able to complete the work in the time allotted for it. Even though I know from experience that this never happens, I do not learn from this experience. I continue to ignore the non-causal, probabilistic evidence based on many similar projects in the past, and to estimate completion dates that I hardly ever meet. (Preparation of this book took twice as long as I had anticipated. These biases are, indeed, difficult to avoid!)

149. Many examples from everyday life are cited in Robyn M. Dawes, *Rational Choice in an Uncertain World* (Harcourt Brace Jovanovich College Publishers, 1988), Chapter 5.

Chapter 13

Hindsight Biases in Evaluation of Intelligence Reporting

Evaluations of intelligence analysis—analysts' own evaluations of their judgments as well as others' evaluations of intelligence products—are distorted by systematic biases. As a result, analysts overestimate the quality of their analytical performance, and others underestimate the value and quality of their efforts. These biases are not simply the product of self-interest and lack of objectivity. They stem from the nature of human mental processes and are difficult and perhaps impossible to overcome.[150]

* * * * * * * * * * * * * * * * * * *

Hindsight biases influence the evaluation of intelligence reporting in three ways:

- Analysts normally overestimate the accuracy of their past judgments.
- Intelligence consumers normally underestimate how much they learned from intelligence reports.
- Overseers of intelligence production who conduct postmortem analyses of an intelligence failure normally judge that events were more readily foreseeable than was in fact the case.

None of the biases is surprising. Analysts have observed these tendencies in others, although probably not in themselves. What may be

150. This chapter was first published as an unclassified article in *Studies in Intelligence*, Vol. 22, No. 2 (Summer 1978), under the title "Cognitive Biases: Problems in Hindsight Analysis." It was later published in H. Bradford Westerfield, editor, *Inside CIA's Private World: Declassified Articles from the Agency's Internal Journal, 1955-1992* (New Haven: Yale University Press, 1995.)

unexpected is that these biases are not only the product of self-interest and lack of objectivity. They are examples of a broader phenomenon that is built into human mental processes and that cannot be overcome by the simple admonition to be more objective.

Psychologists who conducted the experiments described below tried to teach test subjects to overcome these biases. Experimental subjects with no vested interest in the results were briefed on the biases and encouraged to avoid them or compensate for them, but could not do so. Like optical illusions, cognitive biases remain compelling even after we become aware of them.

The analyst, consumer, and overseer evaluating analytical 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. This is in sharp contrast with intelligence estimation, which is an exercise in foresight, and it is the difference between these two modes of thought—hindsight and foresight—that seems to be a source of 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 changes perceptions of a situation so naturally and so immediately that people are largely unaware of the change. When new information is received, it is immediately and unconsciously assimilated into our pre-existing knowledge. If this new information adds significantly to our knowledge—that is, if it tells 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.

After a view has been restructured to assimilate the new information, there is virtually no way to accurately reconstruct the pre-existing mental set. Once the bell has rung, it cannot be unrung. A person may *remember* his or her previous judgments if not much time has elapsed and the judgments were precisely articulated, but apparently people cannot accurately *reconstruct* their previous thinking. The effort to reconstruct what we previously thought about a given situation, or what we would

porate the prior probability into their reasoning because it does not seem relevant. It does not seem relevant because there is no causal relationship between the background information on the percentages of jet fighters in the area and the pilot's observation.[148] The fact that 85 percent of the fighters in the area were Vietnamese and 15 percent Cambodian did not *cause* the attack to be made by a Cambodian rather than a Vietnamese.

To appreciate the different impact made by causally relevant background information, consider this alternative formulation of the same problem. In paragraph (b) of the problem, substitute the following:

(b) Although the fighter forces of the two countries are roughly equal in number in this area, 85 percent of all harassment incidents involve Vietnamese fighters, while 15 percent involve Cambodian fighters.

The problem remains mathematically and structurally the same. Experiments with many test subjects, however, show it is quite different psychologically because it readily elicits a causal explanation relating the prior probabilities to the pilot's observation. If the Vietnamese have a propensity to harass and the Cambodians do not, the prior probability that Vietnamese harassment is more likely than Cambodian is no longer ignored. Linking the prior probability to a cause and effect relationship immediately raises the possibility that the pilot's observation was in error.

With this revised formulation of the problem, most people are likely to reason as follows: We know from past experience in cases such as this that the harassment is usually done by Vietnamese aircraft. Yet, we have a fairly reliable report from our pilot that it was a Cambodian fighter. These two conflicting pieces of evidence cancel each other out. Therefore, we do not know—it is roughly 50-50 whether it was Cambodian or Vietnamese. In employing this reasoning, we use the prior probability information, integrate it with the case-specific information, and arrive at a conclusion that is about as close to the optimal answer (still 41 percent) as one is going to get without doing a mathematical calculation.

There are, of course, few problems in which base rates are given as explicitly as in the Vietnamese/Cambodian aircraft example. When base

148. Maya Bar-Hillel, "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica*, 1980.

with a sample of fighters (half with Vietnamese markings and half with Cambodian) the pilot made correct identifications 80 percent of the time and erred 20 percent of the time.

(b) Base rate data: 85 percent of the jet fighters in that area are Vietnamese; 15 percent are Cambodian.

Question: What is the probability that the fighter was Cambodian rather than Vietnamese?

A common procedure in answering this question is to reason as follows: We know the pilot identified the aircraft as Cambodian. We also know the pilot's identifications are correct 80 percent of the time; therefore, there is an 80 percent probability the fighter was Cambodian. This reasoning appears plausible but is incorrect. It ignores the base rate—that 85 percent of the fighters in that area are Vietnamese. The base rate, or prior probability, is what you can say about any hostile fighter in that area before you learn anything about the specific sighting.

It is actually more likely that the plane was Vietnamese than Cambodian despite the pilot's "probably correct" identification. Readers who are unfamiliar with probabilistic reasoning and do not grasp this point should imagine 100 cases in which the pilot has a similar encounter. Based on paragraph (a), we know that 80 percent or 68 of the 85 Vietnamese aircraft will be correctly identified as Vietnamese, while 20 percent or 17 will be incorrectly identified as Cambodian. Based on paragraph (b), we know that 85 of these encounters will be with Vietnamese aircraft, 15 with Cambodian.

Similarly, 80 percent or 12 of the 15 Cambodian aircraft will be correctly identified as Cambodian, while 20 percent or three will be incorrectly identified as Vietnamese. This makes a total of 71 Vietnamese and 29 Cambodian sightings, of which only 12 of the 29 Cambodian sightings are correct; the other 17 are incorrect sightings of Vietnamese aircraft. Therefore, when the pilot claims the attack was by a Cambodian fighter, the probability that the craft was actually Cambodian is only 12/29ths or 41 percent, despite the fact that the pilot's identifications are correct 80 percent of the time.

This may seem like a mathematical trick, but it is not. The difference stems from the strong prior probability of the pilot observing a Vietnamese aircraft. The difficulty in understanding this arises because untrained intuitive judgment does not incorporate some of the basic statistical principles of probabilistic reasoning. Most people do not incor-

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 might have been considered. Unfortunately, simply understanding that the mind works in this fashion does little to help overcome the limitation.

The overall message to be learned from an understanding of these biases, as shown in the experiments described below, is that an analyst's intelligence judgments are not as good as analysts think they are, or as bad as others seem to believe. Because the biases generally cannot be overcome, they would appear to be facts of life that analysts need to take into account in evaluating their 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 analysts.
- Develop an institutionalized procedure for comparing intelligence judgments and estimates with actual outcomes.
- Measure how well analysts live up to the defined expectations.

The discussion now turns to the experimental evidence demonstrating these biases from the perspective of the analyst, consumer, and overseer of intelligence.

The Analyst's Perspective

Analysts interested in improving their own performance need to evaluate their past estimates in the light of subsequent developments. To do this, analysts must either remember (or be able to refer to) their past estimates or must reconstruct their past estimates on the basis of what they remember 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 remembered or reconstructed estimates.

Experimental evidence suggests a systematic tendency toward faulty memory of past estimates.[151] That is, when events occur, people tend to overestimate the extent to which they had previously expected them to occur. And conversely, when events do not occur, people tend to underestimate the probability they 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 analysts' intuitive experience. Analysts rarely appear—or allow themselves to appear—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.

At varying time periods after the trips, the same subjects were asked to remember or reconstruct their own 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 did happen. That is, the probabilities they remembered having estimated were higher than their actual estimates of events they believed actually did occur. 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 summary, knowledge of the outcomes somehow affected most test subjects' memory of their previous estimates of these outcomes, and the more time that was allowed for memories to fade, the greater the

151. This section is based on research reported by Baruch Fischoff 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.

x .70 or slightly over 34 percent. Adding a fourth probable (70 percent) event to the scenario would reduce its probability to 24 percent.

Most people do not have a good intuitive grasp of probabilistic reasoning. One approach to simplifying such problems is to assume (or think as though) one or more probable events have already occurred. This eliminates some of the uncertainty from the judgment. Another way to simplify the problem is to base judgment on a rough average of the probabilities of each event. In the above example, the averaging procedure gives an estimated probability of 70 percent for the entire scenario. Thus, the scenario appears far more likely than is in fact the case.

When the averaging strategy is employed, highly probable events in the scenario tend to offset less probable events. This violates the principle that a chain cannot be stronger than its weakest link. Mathematically, the least probable event in a scenario sets the upper limit on the probability of the scenario as a whole. If the averaging strategy is employed, additional details may be added to the scenario that are so plausible they increase the perceived probability of the scenario, while, mathematically, additional events must necessarily reduce its probability.[146]

Base-Rate Fallacy

In assessing a situation, an analyst sometimes has two kinds of evidence available—specific evidence about the individual case at hand, and numerical data that summarize information about many similar cases. This type of numerical information is called a base rate or prior probability. The base-rate fallacy is that the numerical data are commonly ignored unless they illuminate a causal relationship. This is illustrated by the following experiment.[147]

During the Vietnam War, a fighter plane made a non-fatal strafing attack on a US aerial reconnaissance mission at twilight. Both Cambodian and Vietnamese jets operate in the area. You know the following facts:

(a) Specific case information: The US pilot identified the fighter as Cambodian. The pilot's aircraft recognition capabilities were tested under appropriate visibility and flight conditions. When presented

146. Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein, "Cognitive Processes and Societal Risk Taking," in J. S. Carroll and J.W. Payne, eds., *Cognition and Social Behavior* (Potomac, MD: Lawrence Erlbaum Associates, 1976), pp. 177-78.
147. This is a modified version, developed by Frank J. Stech, of the blue and green taxicab question used by Kahneman and Tversky, "On Prediction and Judgment," Oregon Research Institute Research Bulletin, 12, 14, 1972.

ambiguity can be especially troubling when dealing with low-probability, high-impact dangers against which policymakers may wish to make contingency plans.

Consider, for example, a report that there is little chance of a terrorist attack against the American Embassy in Cairo at this time. If the Ambassador's preconception is that there is no more than a one-in-a-hundred chance, he may elect to not do very much. If the Ambassador's preconception is that there may be as much as a one-in-four chance of an attack, he may decide to do quite a bit. The term "little chance" is consistent with either of those interpretations, and there is no way to know what the report writer meant.

Another potential ambiguity is the phrase "at this time." Shortening the time frame for prediction lowers the probability, but may not decrease the need for preventive measures or contingency planning. An event for which the timing is unpredictable may "at this time" have only a 5-percent probability of occurring during the coming month, but a 60-percent probability if the time frame is extended to one year (5 percent per month for 12 months).

How can analysts express uncertainty without being unclear about how certain they are? Putting a numerical qualifier in parentheses after the phrase expressing degree of uncertainty is an appropriate means of avoiding misinterpretation. This may be an odds ratio (less than a one-in-four chance) or a percentage range (5 to 20 percent) or (less than 20 percent). Odds ratios are often preferable, as most people have a better intuitive understanding of odds than of percentages.

Assessing Probability of a Scenario

Intelligence analysts sometimes present judgments in the form of a scenario—a series of events leading to an anticipated outcome. There is evidence that judgments concerning the probability of a scenario are influenced by amount and nature of detail in the scenario in a way that is unrelated to actual likelihood of the scenario.

A scenario consists of several events linked together in a narrative description. To calculate mathematically the probability of a scenario, the proper procedure is to multiply the probabilities of each individual event. Thus, for a scenario with three events, each of which will probably (70 percent certainty) occur, the probability of the scenario is .70 x .70

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 subjects who showed the anticipated bias, their retrospective evaluation of their estimative performance was clearly more favorable than warranted by the facts.

The Consumer's Perspective

When consumers of intelligence reports evaluate the quality of the intelligence product, they ask themselves 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 "I knew it all along" bias causes consumers to undervalue the intelligence product.[152]

That people do in fact commonly react to new information in this manner was tested 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. As a measure of their confidence in their answers, the subjects assigned to each question a number ranging from 50 percent to 100 percent, indicating their estimate of the probability that they had chosen the correct answer.

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' memories of their own 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 they had not previously seen, but which were of similar difficulty so that results

152. Experiments described in this section are reported in Baruch Fischhoff, *The Perceived Informativeness of Factual Information*, Technical Report DDI- I (Eugene, OR: Oregon Research Institute, 1976).

would be comparable with the other two groups. The correct answers were marked on the questionnaire, and the subjects were asked to respond to the questions as they would have responded had they not been told the answer. This tested their ability to recall accurately how much they had known before they learned the correct answer. The situation is comparable to that of intelligence consumers who are asked to evaluate how much they learned from a report, and who can do this only by trying to recollect the extent of their knowledge before they 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 running of the experiment and 20 of 25 items in another running, 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.

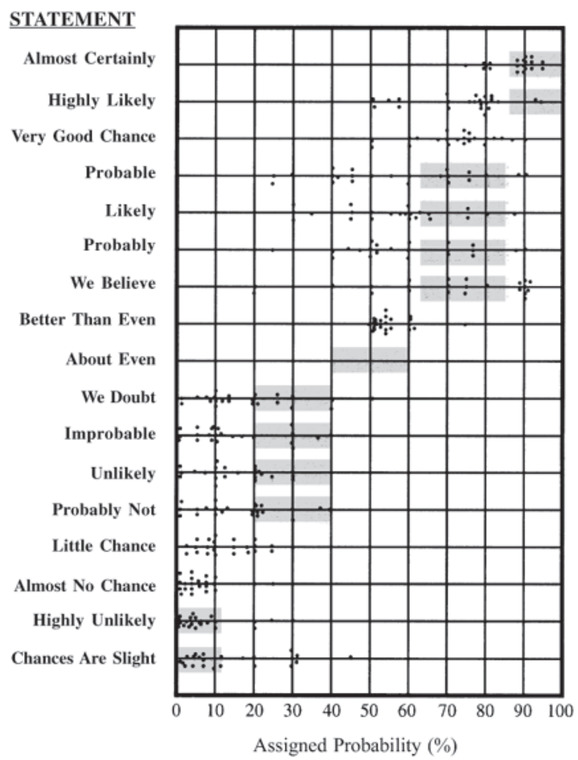
In summary, the experiment confirmed the results of the previous experiment showing that people exposed to an answer tend to remember having known more than they actually did. It also demonstrates that people have an even greater tendency to exaggerate the likelihood that they would have known the correct answer if they had not been informed of it. In other words, people tend to underestimate both how much they learn from new information and the extent to which new information permits them to make correct judgments with greater confidence. To the extent that intelligence consumers manifest these same biases, they will tend to underrate the value to them of intelligence reporting.

The Overseer's Perspective

An overseer, as the term is used here, is one who investigates intelligence performance by conducting apostmortemexamination of a high-profile intelligence failure. Such investigations are carried out by Congress, the Intelligence Community staff, and CIA or DI management. For those outside the executive branch who do not regularly read the intelligence product, this sort of retrospective evaluation of known intelligence failures is a principal basis for judgments about the quality of intelligence analysis.

A fundamental question posed in anypostmorteminvestigation of intelligence failure is this: Given the information that was available at the

Figure 18: Measuring Perceptions of Uncertainty



Sherman Kent, the first director of CIA's Office of National Estimates, was one of the first to recognize problems of communication caused by imprecise statements of uncertainty. Unfortunately, several decades after Kent was first jolted by how policymakers interpreted the term "serious possibility" in a national estimate, this miscommunication between analysts and policymakers, and between analysts, is still a common occurrence.[143]

I personally recall an ongoing debate with a colleague over the bona fides of a very important source. I argued he was probably bona fide. My colleague contended that the source was probably under hostile control. After several months of periodic disagreement, I finally asked my colleague to put a number on it. He said there was at least a 51-percent chance of the source being under hostile control. I said there was at least a 51-percent chance of his being bona fide. Obviously, we agreed that there was a great deal of uncertainty. That stopped our disagreement. The problem was not a major difference of opinion, but the ambiguity of the term probable.

The table in Figure 18 shows the results of an experiment with 23 NATO military officers accustomed to reading intelligence reports. They were given a number of sentences such as: "It is highly unlikely that. . . ." All the sentences were the same except that the verbal expressions of probability changed. The officers were asked what percentage probability they would attribute to each statement if they read it in an intelligence report. Each dot in the table represents one officer's probability assignment.[144] While there was broad consensus about the meaning of "better than even," there was a wide disparity in interpretation of other probability expressions. The shaded areas in the table show the ranges proposed by Kent.[145]

The main point is that an intelligence report may have no impact on the reader if it is couched in such ambiguous language that the reader can easily interpret it as consistent with his or her own preconceptions. This

143. Sherman Kent, "Words of Estimated Probability," in Donald P. Steury, ed., *Sherman Kent and the Board of National Estimates: Collected Essays* (CIA, Center for the Study of Intelligence, 1994).
144. Scott Barclay et al, p. 76-68.
145. Probability ranges attributed to Kent in this table are slightly different from those in Sherman Kent, "Words of Estimated Probability," in Donald P. Steury, ed., *Sherman Kent and the Board of National Estimates: Collected Essays* (CIA, Center for the Study of Intelligence, 1994).

time, should analysts 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.[153]

Unfortunately, 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 analysts should have foreseen events that were, in fact, unforeseeable on the basis of the information available at the time.

The experiments reported in the following paragraphs tested the hypotheses that knowledge of an outcome increases the perceived inevitability 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 datum in the event description. Those subjects who were informed of an outcome were placed in the same position as an overseer of intelligence analysis preparing a postmortem analysis of an intelligence failure. This person tries to assess the probability of an outcome

153. Experiments described in this section are reported in Baruch Fischhoff, "Hindsight does not equal Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty," *Journal of Experimental Psychology: Human Perception and Performance*, 1, 3 (1975), pp. 288-299.

Figure 19

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	48.4	9.9
Told Outcome 4	33.0	15.8	24.3	27.0

based only on the information available before the outcome was known. The results are shown in Figure 18.

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 outcome. Similarly, the control group with no outcome knowledge estimated the probability of Outcome 2 as 21.3 percent, while those informed that Outcome 2 was the actual outcome perceived it as having a 38.4 percent probability.

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 one of four possible outcomes has occurred approximately doubles the perceived probability of that outcome as judged with hindsight as compared with foresight.

The relevance that subjects attributed to any datum was also strongly influenced by which outcome, if any, they had been told was true. As Roberta Wohlstetter has written, "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

may refer to anything from a 1-percent to a 99-percent probability. To express themselves clearly, analysts must learn to routinely communicate uncertainty using the language of numerical probability or odds ratios.

As explained in Chapter 2 on "Perception," people tend to see what they expect to see, and new information is typically assimilated to existing beliefs. This is especially true when dealing with verbal expressions of uncertainty. By themselves, these expressions have no clear meaning. They are empty shells. The reader or listener fills them with meaning through the context in which they are used and what is already in the reader's or listener's mind about that context.

When intelligence conclusions are couched in ambiguous terms, a reader's interpretation of the conclusions will be biased in favor of consistency with what the reader already believes. This may be one reason why many intelligence consumers say they do not learn much from intelligence reports.[141]

It is easy to demonstrate this phenomenon in training courses for analysts. Give students a short intelligence report, have them underline all expressions of uncertainty, then have them express their understanding of the report by writing above each expression of uncertainty the numerical probability they believe was intended by the writer of the report. This is an excellent learning experience, as the differences among students in how they understand the report are typically so great as to be quite memorable.

In one experiment, an intelligence analyst was asked to substitute numerical probability estimates for the verbal qualifiers in one of his own earlier articles. The first statement was: "The cease-fire is holding but could be broken within a week." The analyst said he meant there was about a 30-percent chance the cease-fire would be broken within a week. Another analyst who had helped this analyst prepare the article said she thought there was about an 80-percent chance that the cease-fire would be broken. Yet, when working together on the report, both analysts had believed they were in agreement about what could happen.[142] Obviously, the analysts had not even communicated effectively with each other, let alone with the readers of their report.

141. For another interpretation of this phenomenon, see Chapter 13, "Hindsight Biases in Evaluation of Intelligence Reporting."
142. Scott Barclay et al, *Handbook for Decision Analysis*. (McLean, VA: Decisions and Designs, Inc. 1977), p. 66.

There is some evidence that awareness of the anchoring problem is not an adequate antidote.[139] This is a common finding in experiments dealing with cognitive biases. The biases persist even after test subjects are informed of them and instructed to try to avoid them or compensate for them.

One technique for avoiding the anchoring bias, to weigh anchor so to speak, may be to ignore one's own or others' earlier judgments and rethink a problem from scratch. In other words, consciously avoid any prior judgment as a starting point. There is no experimental evidence to show that this is possible or that it will work, but it seems worth trying. Alternatively, it is sometimes possible to avoid human error by employing formal statistical procedures. Bayesian statistical analysis, for example, can be used to revise prior judgments on the basis of new information in a way that avoids anchoring bias.[140]

Expression of Uncertainty

Probabilities may be expressed in two ways. Statistical probabilities are based on empirical evidence concerning relative frequencies. Most intelligence judgments deal with one-of-a-kind situations for which it is impossible to assign a statistical probability. Another approach commonly used in intelligence analysis is to make a "subjective probability" or "personal probability" judgment. Such a judgment is an expression of the analyst's personal belief that a certain explanation or estimate is correct. It is comparable to a judgment that a horse has a three-to-one chance of winning a race.

Verbal expressions of uncertainty—such as "possible," "probable," "unlikely," "may," and "could"—are a form of subjective probability judgment, but they have long been recognized as sources of ambiguity and misunderstanding. To say that something could happen or is possible

139. Alpert and Raiffa, ibid.
140. Nicholas Schweitzer, "Bayesian Analysis: Estimating the Probability of Middle East Conflict," in Richards J. Heuer, Jr., ed., *Quantitative Approaches to Political Intelligence: The CIA Experience* (Boulder, CO: Westview Press, 1979). Jack Zlotnick, "Bayes' Theorem for Intelligence Analysis," *Studies in Intelligence*, Vol. 16, No. 2 (Spring 1972). Charles E. Fisk, "The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning", *Studies in Intelligence*, vol. 16, no. 2 (Spring 1972), originally classified Secret, now declassified. Both the Zlotnick and Fisk articles were republished in H. Bradford Westerfield, *Inside CIA's Private World: Declassified Articles from the Agency's Internal Journal, 1955-1992*, (New Haven: Yale University Press, 1995).

the disaster has occurred, but before the event it is obscure and pregnant with conflicting meanings."[154] The fact that outcome knowledge automatically restructures a person's judgments about the relevance of available data is probably one reason it is so difficult to reconstruct how 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 without knowing the outcome were virtually the same as the test subjects' own retrospective interpretations.

These results indicate that overseers conducting postmortemevaluations of what analysts should have been able to foresee, given the available information, will tend to perceive the outcome of that situation as having been more predictable than was, in fact, the case. 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.

Discussion of Experiments

Experiments that demonstrated 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. Unfortunately, the experimental subjects were students, not members of the Intelligence Community. There is, nonetheless, reason to believe the results can be generalized to apply to the Intelligence Community. The experiments deal with basic human mental processes, and the results do seem consistent with personal experience in the Intelligence Community. In similar kinds of psychological tests, in which experts, including intelligence analysts, were used as test subjects, the experts showed the same pattern of responses as students.

154. Roberta Wohlstetter, *Pearl Harbor: Warning and Decision* (Stanford, CA: Stanford University Press, 1962), p. 387. Cited by Fischhoff.

My own imperfect effort to replicate one of these experiments using intelligence analysts also supports the validity of the previous findings. 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—that is, 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 go back and check the analysts' recollections of their earlier estimates. Because CIA estimates are rarely stated in terms of quantitative probabilities, and because the occurrence of an estimated event within a specified time period often cannot be determined unambiguously, these preconditions are seldom met.

I did, however, identify several analysts who, on two widely differing subjects, had made quantitative estimates of the likelihood of events for which the subsequent outcome was clearly known. I went to these analysts and asked them to recall their earlier estimates. The conditions for this mini-experiment 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 above.

All this leads to the conclusion 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.

Can We Overcome These Biases?

Analysts tend to blame biased evaluations of intelligence performance at best on ignorance and at worst on self-interest and lack of objectivity. Both these factors may also be at work, but the experiments suggest the nature of human mental processes is also 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 analysts are familiar. Moreover, in these experimental situations the biases were highly resis-

those who began with an estimate that was too low. Even the totally arbitrary starting points acted as anchors, causing drag or inertia that inhibited fulladjustment of estimates.

Whenever analysts move into a new analytical area and take over responsibility for updating a series of judgments or estimates made by their predecessors, the previous judgments may have such an anchoring effect. Even when analysts make their own initial judgment, and then attempt to revise this judgment on the basis of new information or further analysis, there is much evidence to suggest that they usually do not change the judgment enough.

Anchoring provides a partial explanation of experiments showing that analysts tend to be overly sure of themselves in setting confidence ranges. A military analyst who estimates future missile or tank production is often unable to give a specific figure as a point estimate. The analyst may, therefore, set a range from high to low, and estimate that there is, say, a 75-percent chance that the actual production figure will fall within this range. If a number of such estimates are made that reflect an appropriate degree of confidence, the true figure should fall within the estimated range 75 percent of the time and outside this range 25 percent of the time. In experimental situations, however, most participants are overconfident. The true figure falls outside the estimated range a much larger percentage of the time.[138]

If the estimated range is based on relatively hard information concerning the upper and lower limits, the estimate is likely to be accurate. If, however, the range is determined by starting with a single best estimate that is simply adjusted up and down to arrive at estimated maximum and minimum values, then anchoring comes into play, and the adjustment is likely to be insufficient.

Reasons for the anchoring phenomenon are not well understood. The initial estimate serves as a hook on which people hang their first impressions or the results of earlier calculations. In recalculating, they take this as a starting point rather than starting over from scratch, but why this should limit the range of subsequent reasoning is not clear.

138. Experiments using a 98-percent confidence range found that the true value fell outside the estimated range 40 to 50 percent of the time. Amos Tversky and Daniel Kahneman, "Anchoring and Calibration in the Assessment of Uncertain Quantities," (Oregon Research Institute Research Bulletin, 1972, Nov. 12, No. 5), and M. Alpert and H. Raiffa, "A Progress Report on The Training of Probability Assessors," Unpublished manuscript, Harvard University, 1968.