For Official Use Only

Handbook of Bayesian Analysis for Intelligence

For Official Use Only

OPR-506
June 1975

Approved For Release 2000/09/14 : CIA-RDP84B00506R000100070010-7

CENTRAL INTELLIGENCE AGENCY

DIRECTORATE OF INTELLIGENCE
OFFICE OF POLITICAL RESEARCH

June 1975

HANDBOOK OF BAYESIAN ANALYSIS FOR INTELLIGENCE

In the preparation of this study, the Office of Political Research consulted other offices of the Central Intelligence Agency. Their comments and suggestions were appreciated and used, but no formal attempt at coordination was undertaken. Comments would be welcomed by the author, [redacted] (Code 143, x4091).

25X1A9a

CONTENT

SUMMARY

Within the Intelligence Community, the term "Bayes" has been bandied about from time to time over the past decade, as one of a number of techniques to supplement traditional analysis. There have been frequent assessments of the Bayesian technique and infrequent applications of it to the analysis of an intelligence question, usually on an experimental basis. During 1974 and 1975, the Office of Political Research used Bayesian analysis in three separate studies—dropping the qualifier "experimental" after the first—and the office intends to continue its application in the future, when such an approach is relevant. The technique produced a style of report which conveyed much information in a very concise format. It is our view and the frequently-voiced opinion of others that this represented an advance in communications over traditional methods of reporting.

This handbook will attempt to explain the use of the Bayesian approach as a forecasting tool for intelligence. It is written for potential users of the technique, both to make OPR's experience available to others, and to encourage and facilitate its proper use. It is meant as a self-help guide on how to conduct such an exercise. Our experiences are intended to be instructive in a general way; the techniques described represent a particular solution to problems of political intelligence, but could be adapted with minor or major changes to other intelligence and bureaucratic situations. The presentation explains the rationale and format of the three OPR analyses, the procedures developed to administer them, and the benefits and problems of the Bayesian method. Neither the analyses nor this handbook wallows in methodological soul-searching; in both cases, the purpose is utility rather than perfection.

There will be, of necessity, some examination of the mathematical formulae involved. The bulk of this handbook, however, is a catalog of the practical administrative details of such a project. To avoid "perfection of means and confusion of goals [which] seem to characterize our age" (Einstein) the need for careful definition of the problem and careful project design is repeatedly stressed.

BAYES

1. What are its Capabilities and Benefits?

The task of intelligence is to answer questions about the capabilities and intentions of various international actors; information may be abundant or meager, but it is usually incomplete or incoherent. The main feature of Bayesian analysis is that it provides a rational, consistent, and objective process whereby many apparently unrelated facts can be combined to produce an overall assessment. If an intelligence problem can be formulated into a set of questions which can be answered in probabilistic terms—how likely one of a set of outcomes is—it may lend itself to Bayesian analysis.

There is no magic and no inherent wisdom in Bayes. In simplest terms, the Bayesian technique consists of a statistical formula and a procedure for its use. (Statistics is a discipline which allows one to deal with uncertainty in an organized fashion without being vague or imprecise; Bayesian statistics differ from classical statistics in that they allow an analyst to use his own expert understanding of a situation along with probabilistic judgments based on evidence.) The Bayesian technique is an organizing device which allows an analyst to assign probabilities to the likelihood of various carefully drawn scenarios about an intelligence problem, and thereafter to evaluate fragments of evidence in terms of those hypothetical scenarios. The Bayesian formula then aggregates these numbers mathematically, rather than by the inductive logic of an analyst, into an overall set of probabilities. This has several advantages:

—More information can be extracted from the available data because the technique allows each piece of evidence, central or marginal, to add its weight to the final assessment in a systematic way; thus, a number of small items can outweigh a large one, and the probabilities are not at the mercy of the most recent or most visible item.

—The procedure provides a reproducible sequence of steps for arriving at the final figures; a disagreement among analysts can thus often be seen to be a disagreement over the meaning of certain items of evidence rather than an unresolvable difference of opinion.

—The formulation of the questions forces the analyst to consider alternative explanations of the facts he sees, thus loosening the bonds of established opinions. In other words, he is asked to look at how well the evidence explains possible scenarios other than the one he has already decided is most likely.

—The use of quantified judgments allows the results of the analysis to be displayed on a numerical scale, rather than through the use of terms like "probable," "likely," "unlikely," or "possible." Also, the work of a number of analysts can be arrayed in graphic form, with ranges and averages.

—The formal procedure has been shown to be less conservative than analysts' informal opinions, and to drive probabilities away from 50–50 faster and farther than the analysts' overall subjective judgments do. This is often initially unsettling for the analysts, but most have admitted that they later agreed with the assessment.

—The mere fact that a team of experts is asked to assess periodically the evidence on an important intelligence question provides managers of intelligence production with a degree of assurance that the question is indeed being monitored.

Before embarking on their initial Bayesian project, the OPR analysts assigned to coordinate the exercise reviewed the Agency's previous experiments with Bayesian analysis, read the available literature on the theory and application of Bayes, and consulted with individuals familiar with the technique both in and outside of the Agency. The earlier work of [redacted] [redacted], and the advice of [redacted] in various offices of the Agency, provided especially useful guidelines for the design of these projects. 25X1A9a 25X1A9a 25X1A9a

The rule of Bayes is an established statistical technique with a variety of applications in the social sciences and industry, but its application to intelligence questions is both more complex and less precise. The political, economic, strategic, and social events of the world are imperfectly understood and difficult to measure. Hence, to use a technique like Bayes it is necessary to turn to expert judgments expressed quantitatively. The values assigned by expert analysts are, of course, approximate, but they provide a rough basis for comparison and analysis. OPR's experience suggests that it is relatively easy to induce analysts accustomed to qualitative expressions of probability to shift to numerical assessments. An ever-present danger, however, is the tendency to attribute more precision to the numbers than is warranted, and it should be stressed that the numbers are always only approximations.

2. When is it Useful?

The starting point for any investigation must always be the careful formulation of a relevant question. Before discussing the rule of Bayes further or the procedures developed to work with it, it is important to remember several overriding substantive considerations.

What is the nature of the intelligence problem; what is the question and what data are available? In the intelligence process the nature of the problem must dictate which methods and approaches will be most successful. There are many techniques for enhancing traditional analysis, and it is possible that something other than Bayes would be more appropriate to the level and complexity of the issue involved; trend analysis, an econometric model, or a Delphi exercise might, for example, be more relevant. Many of the more rigorous techniques require good data, and to the extent that a problem becomes more data-rich, it may actually cease to be an intelligence problem. There is also a certain delay involved in the genesis of a project, which might preclude its use in fast-developing situations. To be the type of question which is susceptible to Bayesian analysis:

—It must lend itself to formulation in categories which overlap very little, such as war versus no war, or development of a complete nuclear capability versus development of a peaceful nuclear capability versus no nuclear development. Bayes is useful only when the question is expressed as a specific set of outcomes; thus, the Bayesian approach would be useless as a predictor of something as amorphous as future Middle East relations. The question would have to be re-cast in terms of specific alternatives, that is, a set of more or less mutually exclusive scenarios of Middle East developments. In this process, however, there is a danger that the question will be so simplified that the answer would be neither relevant nor interesting.

—There must be a fairly rich flow of data at least peripherally related to the question. For example, in the nuclear example above, data on all related materials and processes would be relevant. If data are sparse, the technique is very sensitive to each piece and is less reliable although it may still be useful.

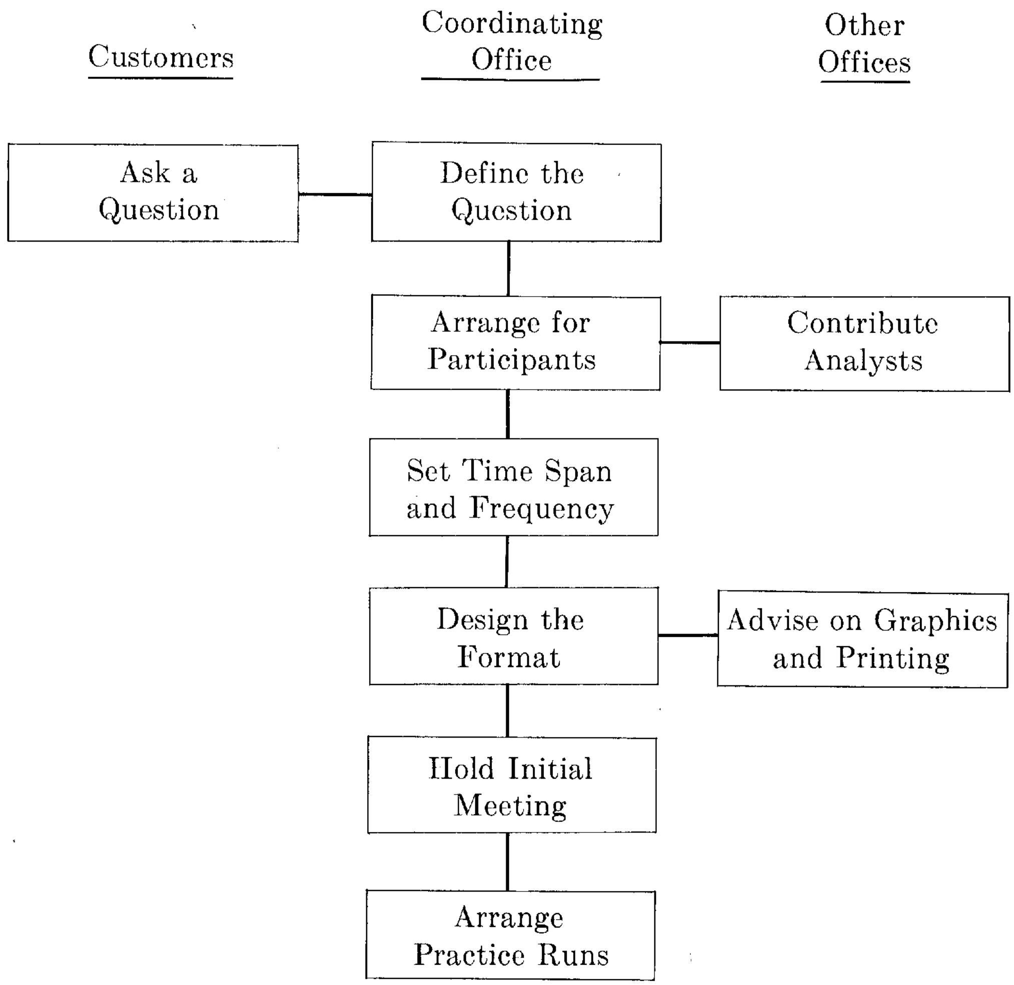
—It must be the type of activity which produces preliminary signs and is not a chance or random event. For example, it would be fruitless to attempt to predict whether Giscard would be killed in an automobile accident. Bayesian analysis can measure only preparations for and indications of the hypothesized outcomes.

For whom and at what level is the question important? This will have a direct bearing on the level of resources which can be commanded for an analytical effort. The consumer receives a product which displays results concisely and graphically, and unless he or she reacts negatively to such displays, this can be an important consideration.

Each production office would have to weigh the benefits and costs of the method—on one side the rigor, reproducibility, and sex appeal of the method, and on the other the need for education, monitoring, and coordination of a number of analysts who may not find the method immediately congenial. Having considered the above questions, one should ask whether Bayes is applicable.

HOW TO INITIATE A BAYESIAN EXERCISE

1. Flowchart of Setup Procedure



2. Defining the Scenarios

Once a question has been selected or assigned as a candidate for a Bayesian analysis, the first step is to formulate it into categories which are more or less mutually exclusive, that is, which cover the relevant possibilities and overlap as little as possible. It is imperative not to let the question run away with the formulation; that is, it must not be chosen simply because it can be answered using the Bayesian technique. If so, it may not necessarily be the question that the customer really wants answered.

This formulation is easiest when there is a single event which one is interested in predicting. For example, the actual formulation of the first OPR study, in what are termed "scenarios" or "hypotheses," was:

A. The North Vietnamese will launch a major military offensive (defined as a countrywide offensive on the scale of the Tet 1968 or March

1972 fighting, or attacks on a similar level generally confined to one or two military regions) before the end of June, 1974.

B. The North Vietnamese will not launch such an offensive.

When this formulation as a dichotomy can be done, it allows a simplified assignment and computation technique, as explained later.

Often, however, the problem is not so simple, and to formulate an intelligence problem as an either-or question would be to strip it of its complexity and interest. For example, the second project was designed to monitor the Sino-Soviet conflict, and the following possible scenarios were identified:

A. The Soviets will undertake a nuclear strike against Chinese strategic or nuclear targets within six months.

B. The Soviets will launch a large-scale conventional attack against China within six months.

C. The Soviets will launch a localized cross-border attack, with limited objectives, on a scale larger than the 1969 incidents within six months.

D. The Chinese will launch a localized cross-border attack, with limited objectives, on a scale larger than the 1969 incidents within six months.

E. One or more minority groups on either side of the border will revolt, following instigation by the opposite side within six months.

F. Neither side will undertake any of the above types of hostilities within six months.

3. Participants

The second step is to decide the number and variety of participants that are needed to carry out the Bayesian analysis. OPR has been able to enlist participants in its various Bayesian projects from a wide variety of offices within CIA, analysts who concentrate on economic, strategic, political, and geographic affairs, or who analyze propaganda or imagery, or who follow the broad spectrum of current events. For Intelligence Community projects, participants have been enlisted from the Army, Navy, Air Force, DIA, NSA, INR, the Intelligence Community Staff, and the National Intelligence Offices. As stated before, the importance of the question will affect the number of persons willing or even available to give time to this type of effort. The number of participants in our reports has varied from six to thirteen, and although a larger number would clog the logistical mechanisms and clutter certain types of graphs, there is no theoretical limit. Nor is there any reason the analysis could not be done by one person.

Depending on the purpose of the exercise, it is possible for an individual, an office, an agency, or an interagency group to carry out a Bayesian analysis. OPR, in its projects thus far, has had the benefit of working on topics of high-level interest, with a specific consumer demand that this technique be applied. On a topic of lesser moment, or where there is little visible encouragement from above, it would still be practical and quite useful for a small group or even an individual to undertake a Bayesian analysis. The scale on which OPR has been able to carry out its projects is not essential to a useful application of the technique.

Monitoring an intelligence situation is most frequently done by an office or an individual within an office, and the Bayesian technique can be applied just as profitably on that level, without the overlay of logistics and cartography. In a scaled-down procedure, the logistics of the exercise would diminish to the point where everything but the selection and evaluation of evidence could be recorded, stored, processed, and printed via a computer terminal. A member of OPR's Analytical Techniques Group could provide assistance in setting up the exercise and training the participants, after which another individual could administer the project. A small in-office Bayes could serve well to support or supplement conventional analysis, and would be better suited to following a fast-moving situation than the elaborate productions OPR has designed. It might even serve as the basis for a later full-scale project.

After the exercise has been initiated, a single analyst or a small number of analysts in a single office could draw up a list of individual items of evidence; probabilities could then be assigned to each item and consolidated by computer or by hand. This could be done on any schedule: weekly, daily, or whenever new evidence appears. Readouts of revised probabilities would be immediately available at any time, and the results could be circulated via narratives, hand-drawn graphs, or computer-drawn graphs.

4. Time Span and Frequency

The third step is to determine how long the exercise should run and how often reports should be issued. This depends on the volume of data and the expected speed of development. For a data-rich question of great urgency and some probability, it may be necessary to run the exercise once or twice a week over a period of many months. Lacking any of these attributes, however, the problem may lend itself to a more leisurely schedule. OPR's reports have been issued at intervals of one week, two weeks, four weeks, and six weeks, and have lasted for periods of six months to one year. The reporting period can be changed in the course of an exercise, as the question becomes more or less urgent, without upsetting continuity.

Another consideration which has great practical impact upon frequency is the cost, in man-hours of analysis and coordination, of running a large project. As developed by OPR, these studies have kept to a minimum the time required of analysts, though this was achieved by a large expenditure of time on the part of the coordinators. Whereas each coordinator spent approximately two full days for each periodic assessment, each analyst spent only one-quarter to one hour. Even this large investment of coordination time might be decreased by a greater use of computer terminals and an interactive data-gathering program, already developed but untested.

5. Designing the Format

Finally, the format for presenting the results should be designed. If the exercise is of sufficient import to be published, that is, if there is a known or perceived consumer demand for it, then the design of the graphic presentation is a very important step in communicating the analysis. Any cartographic aid, either for initial design or for continuing support, should be enlisted early in this process. If there is no perceived high-level demand for such an analysis, meaning that it would be done for the benefit of working-level analysts, a less formal and expensive presentation would probably be more appropriate and just as effective.

6. An Initial Meeting

It will be necessary to have an initial meeting of participants at which the rationale, format, and procedures are explained. It is also convenient at this time to collect the following information, some of which will be used for the list of participants in the publication:

—Name as it is to appear in print;
—Title and office name;
—Number of years in office;
—Number of years on subject;
—Office telephone number(s);
—Tube station or LDX address;
—An alternate analyst to stand in during vacations;
—Starting probabilities for each hypothesis.

The latter numbers should represent the participants' current overall feeling for the probability that each event will occur as stated. (See the section on calculations for a further description of the starting probabilities.)

7. Practice Runs

It will be useful to arrange one and perhaps two practice runs, to familiarize the analysts with the routine, and also to work out the logistics, the cartography, and the printing of the report. The wording of the hypotheses should be reviewed at this point, to assure that the proper question is being addressed. Although the basic design of the exercise and the formulation of the questions will remain fairly constant once the formal publication begins, the work of familiarizing analysts with the routine will never completely end, as occasionally someone will forget to cite sources or to return contributions on time.

All of the foregoing constitutes the design and definition of the project. Careful attention at this stage is extremely important to the subsequent conduct of the exercise. It is also the point at which basic questions about goals and intentions must be resolved. Much of it can be handled by one person—preferably the one who will later coordinate the exercise—in consultation with both consumers and participants.

THE OPR BAYESIAN ANALYSES

The three Bayesian analyses run by OPR in 1974–75 relied heavily for their success on supplementary techniques relating to data-gathering, presentational format, and administration. The Bayesian aspects of the reports formed the core of the analysis, around which a number of other carefully-tailored aspects were draped.

1. More than One Analyst

One of the central features of the studies was the use of a group of analysts rather than a single expert. This more than anything else influenced the data-gathering process, the format of the publication, and the actual production procedure. The reasons for this decision were:

—To bring to the exercise a range of expertise beyond the skills and experience of any single analyst;

—To supply a richer mix of evidence on the questions by asking each analyst to contribute anything he or she considered important. Most political, strategic, or military intelligence problems involve such varied inputs as propaganda analysis, photographic interpretation, and logistic matters. (The evidence was then shown to all participants without identifying the contributors); and

—To provide a balance of expertise in which the effects of organizational and individual biases are minimized.

It is well-known that different analysts will tend to place greater reliance on different types and sources of intelligence. The consolidated list provided an opportunity for each analyst to call an item of evidence to the attention of his colleagues, and assured that the participants would consider each item explicitly before judging its relevance.

To avoid the unpredictable and oft-decried effects of group dynamics, however, each analyst worked on the probabilistic assessments individually, and relayed them to the OPR coordinator. This also avoided the time wasted in a group meeting and the problem of assembling the group on a convenient schedule.

2. Printing the Evidence

Each periodic report contained the pieces of evidence identified and used by the participants, along with only a paragraph or two on the principal trends during the period. No attempt was made to formulate or coordinate a lengthy textual analysis of the situation. This allowed the reader

—To see the basic evidence rather than just a summary and hence to better understand the analysts' assessments;

—To make his own direct assessments if he so desires, or just to keep up with the topic by viewing the evidence regularly; and

—To keep a concise record of the situation.

3. Graphs and Visibility

The ability to portray the results of the analysis graphically was one of the strongest arguments for using a quantitative method like Bayes, and the graphs in the publications have been well-received. In each study, the probability of the hypothesized event, usually "how likely are major hostilities within a certain time period?" was immediately visible on a broken line or bar chart. This conveyed much information at a glance, especially with the broken-line charts, which illustrate trends far more concisely and vividly than do words. On both the broken-line and bar charts the range of estimates around the central measure showed clearly and concisely how much disagreement there was.

Different formats were tried in the various reports, each demonstrating certain advantages and disadvantages. Arrayed on the foldout page are examples of the graphs and combinations of graphs (with representative values) used in the various reports. As a rule, the aim was to make the graphs as prominent as possible, while trading off the amount of information against the degree of clutter, the use of color against the cartographic and printing constraints, and the size and placement of the graphs against the requirements of the overall publication.

In figure 1, from the Vietnam study, each analyst was represented by a separate line showing his or her position in successive weeks. This extensive detail was possible because there was only one positive question to be answered and there were only six analysts participating. (The negative scenario could of course be omitted because it was the mathematical complement of the positive one.) The assessments were represented by six lines on a single graph, plus one for the average (arithmetic mean).

For the Middle East study with four scenarios and thirteen analysts, this format was reduced to four charts—one for each scenario—each of which displayed the median, the interquartile range, and the range, because the full representation would have been confusing, time-consuming, and expensive in printing. One of those four charts is represented in figure 2. This format does have the disadvantage, however, that individuals are not identified; consequently a full-page supplementary graph was supplied for that purpose, similar to figure 3b, on which each analyst gave a hunch judgment of hostilities in the next ninety days. This supplementary graphic was totally non-Bayesian. The provision of individual lines allowed each analyst to be visibly identified, as in the Vietnam reports, and thus to advocate a certain position if desired. This visibility was pleasing to the analysts and interesting for the readers, with the result that there was little if any pressure on the analysts to try to manipulate their positions on the other graphs (see the section on manipulation).

In figure 3a, the assessments of thirteen analysts for six scenarios of Sino-Soviet hostilities were condensed into a bar graph which showed, for each scenario, the average of the thirteen probabilities along with the lowest and the highest values assigned. To provide continuity from period to period and to illustrate trends, a single broken-line chart was added which showed the progression of the average figures over time. In addition, individual lines were provided on a single supplementary chart, figure 3b, which was an attempt to quantify and measure a related problem, the "level of tension" between the two countries; this was also a non-Bayesian assessment. Each analyst was asked to gauge his overall feeling of the level of tension between the two countries on a pre-arranged historical scale, independently of the chances for hostilities. Gossamer though this concept is, the graph performed well as an indicator of changes in the atmosphere between the two countries; once each individual chose his or her position on the chart—as a hawk, a dove, or a middle-of-the-roader—the ups and downs of the graph showed clear trends from month to month, and there was no need for the analysts to agree on a single number.

4. Identifying the Participants

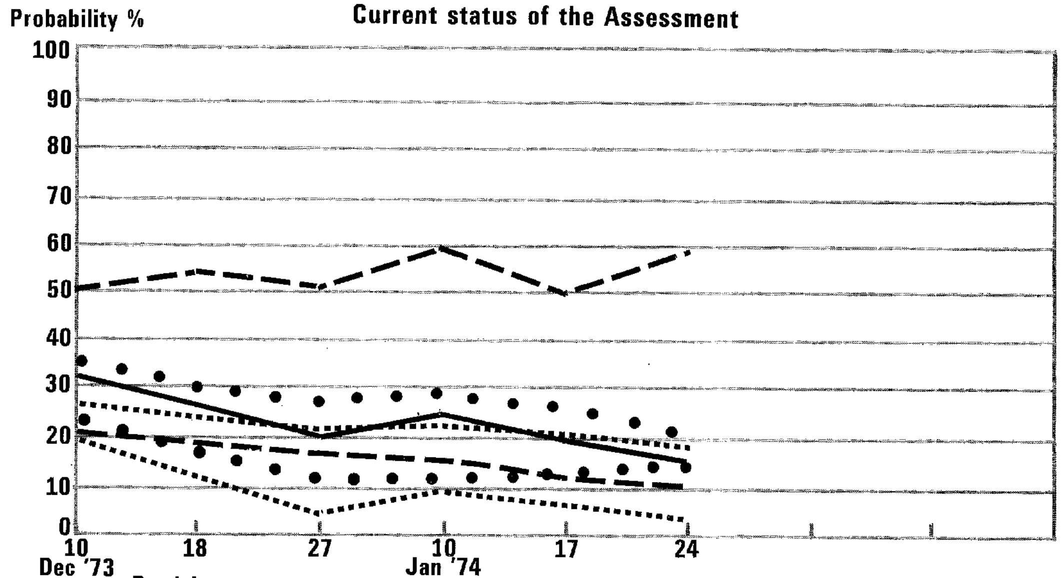
The final presentational technique was the listing of all participants by name, and their identification on at least one graph with a trend line of assessment. This visibility, as mentioned, helped to motivate participants, and in the latter two studies provided them a direct medium for expressing themselves in accord with their opinions on the issue.

Figure 1

BAYESIAN ANALYSIS ON THE LIKELIHOOD OF A MAJOR NORTH VIETNAMESE MILITARY OFFENSIVE

10 December 1973-30 June 1974

Current status of the Assessment



Participants:

25X1A9a [redacted]

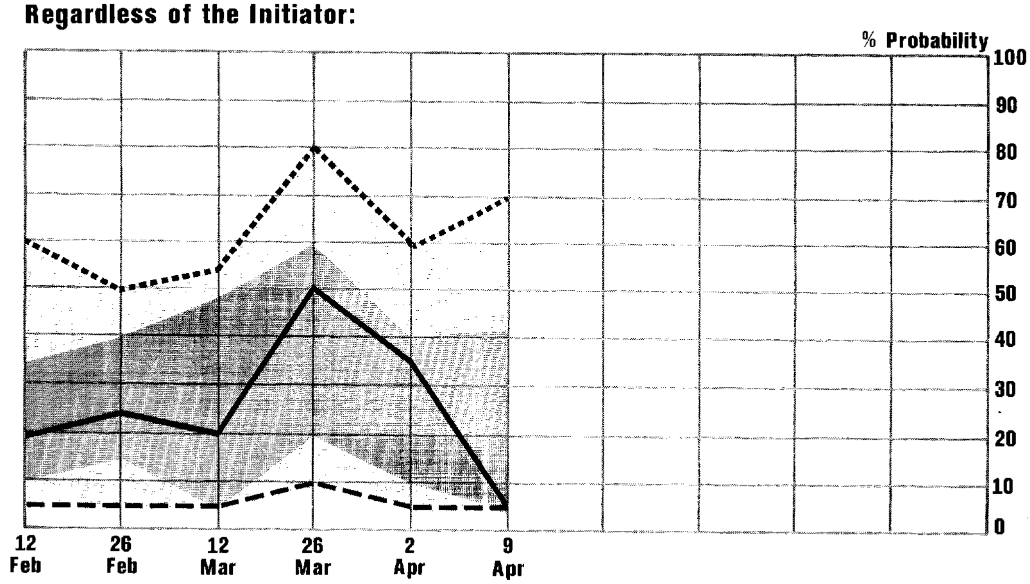
— Central Tendency of the Group's Estimates

Figure 2

Bayesian Analysis on the Likelihood of Major Hostilities in the Middle East in the Next 30 Days

Assessment as of 9 April

Likelihood of any Major Hostilities, Regardless of the Initiator:



- - - Lowest estimate by any of the analysts
— Average of all analysts (median)
······ Highest estimate by any of the analysts

Most deeply shaded area shows inter-quartile range of opinion

566533

Figure 3a

A Bayesian Analysis of the Likelihood of Sino-Soviet Hostilities

Before 1 September 1974

Current Status of the Assessment -- as of 1 March

A The Soviets will undertake a nuclear strike against Chinese strategic or nuclear targets.

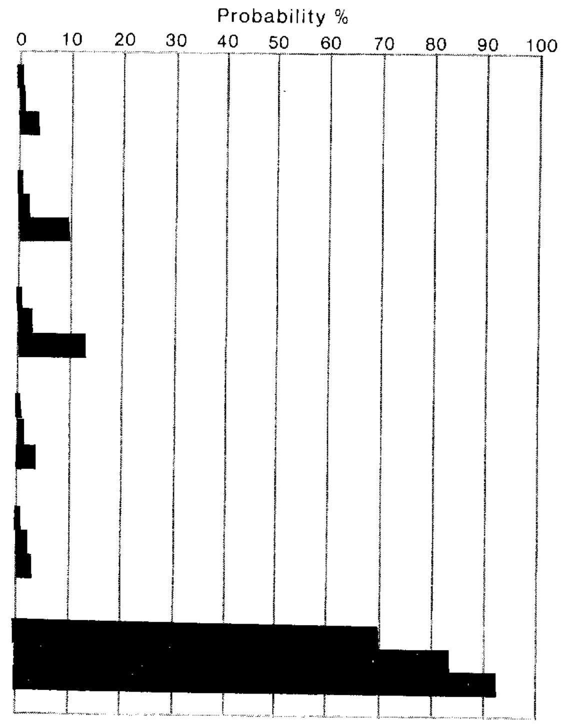
B The Soviets will launch a large-scale conventional attack against China.

C The Soviets will launch a localized cross-border attack, with limited objectives, on a scale larger than the 1969 incidents.

D The Chinese will launch a localized cross-border attack, with limited objectives, on a scale larger than the 1969 incidents.

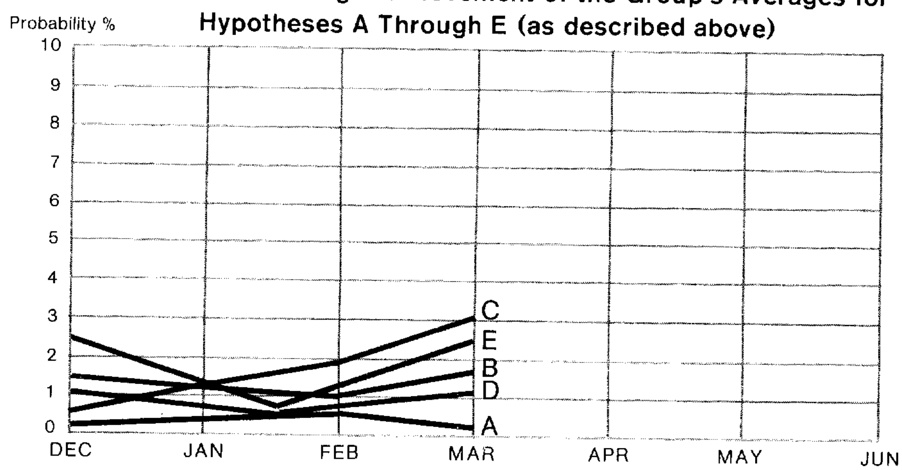
E One or more minority groups on either side of the border will revolt, following instigation by the opposite side.

F Neither side will undertake any of the above types of major hostilities.



Lowest estimate by any of the participants | Average of all estimates | Highest estimate by any of the participants

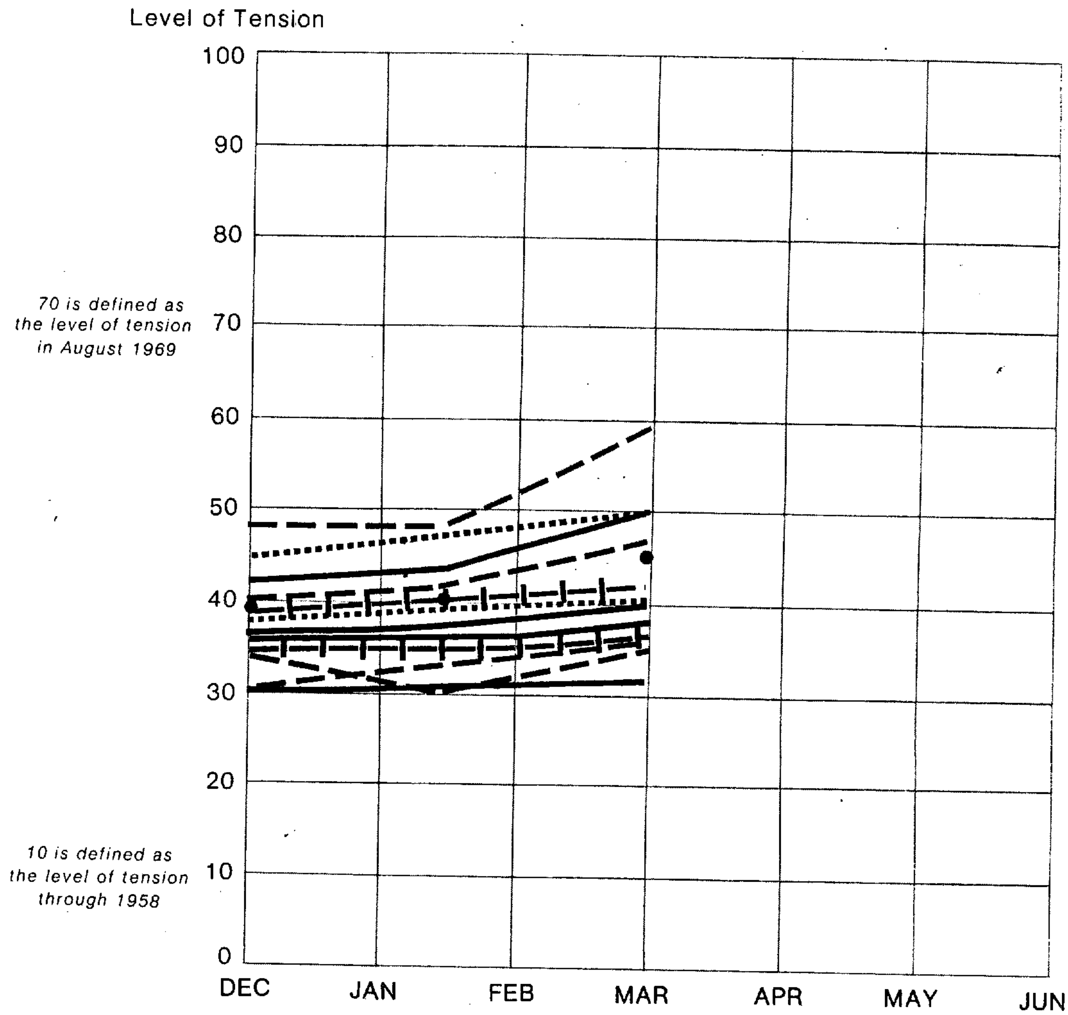
A Time Chart Showing the Movement of the Group's Averages for Hypotheses A Through E (as described above)



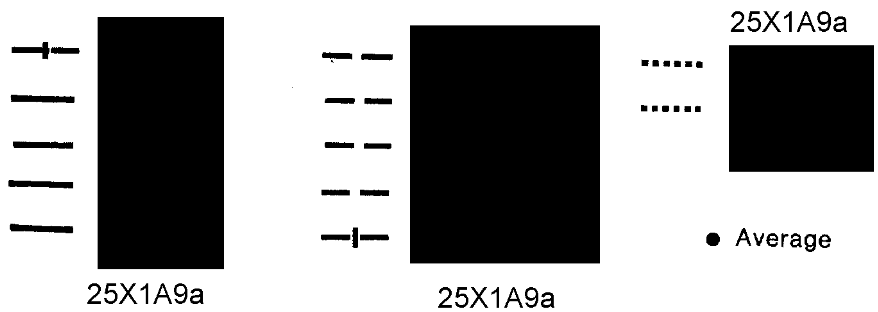
566534

Figure 3b

Individual Assessments Of Sino-Soviet Tensions



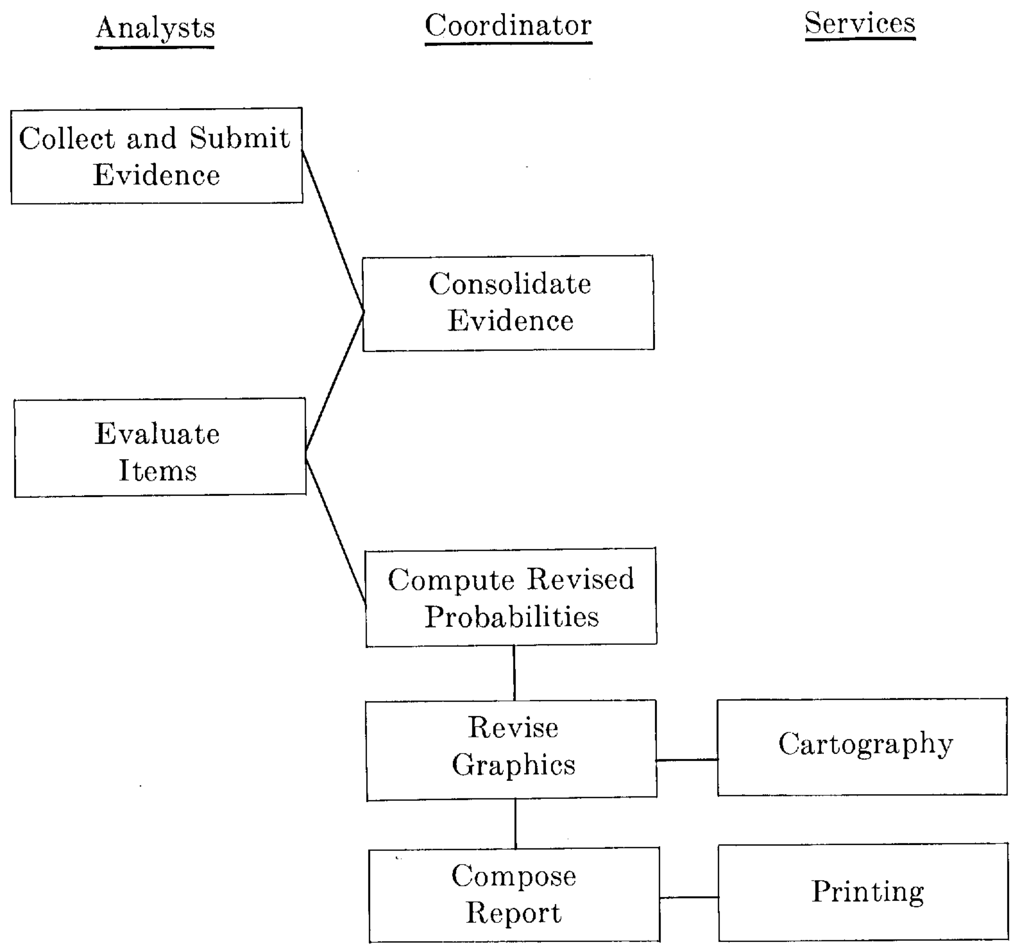
The points on this chart were chosen by each participant on an intuitive basis, using the rough guidelines shown on the chart.



566535

HOW TO MANAGE AN ON-GOING BAYES

1. Flowchart of Periodic Procedure:



2. Routine Procedure

Once an exercise is started, the procedure is periodic and recurring. On the first day of each period, each analyst submits the items of evidence he has seen since last round relating to the question. The submission is in the form of one or two sentences summarizing the item, along with the date, source, and classification, e.g.,

> Statement in the Soviet journal, Problems of the Far East by Yumjaagiyn Tsedenbal, First Secretary of the Central Committee of the Mongolian People's Revolutionary Party, alleging that groups of Chinese soldiers violate the Mongolian border, fell trees, start forest fires, and herd diseased cattle into his country. (Foreign Broadcast Information Service Daily Report, 10 January, UNCLASSIFIED)

This can best be done on preprinted forms. The choice of data is left entirely to the analyst, who should include anything relevant, and exclude what can be judged to be irrelevant. There will be some overlap and considerable diversity in what is submitted.

Later the same day, the coordinator to whom the data is sent consolidates the items, resolving differences of wording, emphasis, and meaning, and distributes the complete list of items to all participants. At the coordinator's discretion, one standard item may be added which asks the participants to evaluate the volume of data itself. This is explained later in the section on weaknesses of the method.

By the following day, the analysts working individually should evaluate the items, as explained below, and return their numerical assessments to the coordinator.

3. The Calculations

There are two ways in which the Bayesian assignments and subsequent calculations can be made, a basic and a simplified way. They are both based on the same mathematical relationships, but one works with ratios of probabilities and the other with probabilities directly. The simplified is restricted to use with questions formulated in an either-or fashion, such as war or no war. The basic method, however, can be used in all cases, for either-or questions and for multiple scenario questions. Both methods require a set of starting probabilities which represent each analyst's best judgment of the situation at the start of the exercise, and the subsequent evaluations build on these.

A. The Basic Method

For example, using the basic method and the following scenarios,

A: Large-scale aggression by country X within six months
B: Small-scale provocations by country X within six months
C: No planned hostilities within six months

an analyst might assign a set of starting probabilities to A, B, and C of 25%, 25%, and 50% likely at the start of the exercise. (The assignment of these probabilities is done entirely subjectively, although techniques exist to refine what a person really means by these numbers.) Note that the sum of these probabilities equals 100, because the scenarios supposedly cover all possible events. This is the only time the analyst will have to assign numbers which total 100.

In the course of the exercise, the analysts will evaluate each item of evidence as it relates to the scenarios, formulating the question as "how likely (0–100%) is it that this item would occur if A is planned?", "how likely is it that this item would occur if B is planned?", and "how likely is it that this item would occur if C is planned?" The peculiar formulation of the questions is dictated by the nature of Bayes' theorem, which depends upon conditional probabilities as explained in the first appendix. The set of probabilities assigned for each item need not add to 100. In this case, the wording of a very bellicose speech by the leader of country X might be 90% likely if that country were contemplating outright aggression against its neighbor, 90% likely if small-scale border provocations were planned, and 80% likely even if nothing were planned.

This procedure is followed for each item independently, without reference to the other items. The only exception is the discretionary final item, which refers explicitly to the other items: "How likely is it during this period that this volume of information (no more and no less) would occur if A is planned?", et cetera.

The mathematics of revising the estimates, once the coordinator receives them, is straightforward though tedious. It is recommended that an interactive computer program be used to save time and reduce arithmetic errors. The calculations are performed for each analyst separately; in other words, analyst A's set of probabilities from the previous round are updated by the conditional probabilities for each item in turn, resulting in a final set of revised probabilities for this round, which are used for the publication and as starting probabilities for the next round. The same procedure is then followed for each other analyst.

The general formula for determining the revised probability for each new item is

$$P(S_i/I) = \frac{P(S_i) \times P(I/S_i)}{\sum_{i=1}^{m}(P(S_i) \times P(I/S_i))}$$

where

$P(S_i)$—is the starting probability for scenario i
$P(I/S_i)$—is the conditional probability of an item
$P(S_i/I)$—is the revised probability of scenario i
m—is the number of scenarios

For a full elaboration of the mathematics, see Appendix 1; Appendix 2 contains the code for interactive programs in APL and BASIC which will perform these calculations. The following is an example of the interaction using the starting probabilities assigned above, the probabilities assigned to the hypothetical speech, and two other sets of probabilities chosen for illustrative purposes. (The user's entries start in column 5, the computer's responses in column 1):

```
    BAYES
THIS PROGRAM CALCULATES EVENT PROBABILITIES BY THE
BAYESIAN FORMULA. YOU WILL BE ASKED TO ENTER A SET
OF PRIOR PROBABILITIES, THE NUMBER OF ITEMS OF EVI-
DENCE, AND THE ITEM PROBABILITIES.
```

PRIOR PROBABILITIES FOR THIS ANALYST, IN ORDER.

25 25 50

HOW MANY PIECES OF EVIDENCE?

3

PROBABILITIES FOR ITEM 1, IN ORDER.

90 90 80

INTERMEDIATE RESULTS:

26.47 26.47 47.06 (a very small change)

PROBABILITIES FOR ITEM 2, IN ORDER.

90 90 90

INTERMEDIATE RESULTS:

26.47 26.47 47.06 (no change)

PROBABILITIES FOR ITEM 3, IN ORDER.

60 20 10

INTERMEDIATE RESULTS:

61.36 20.45 18.18 (a significant change)

$P(S_i/I) = 61.36 \quad 20.45 \quad 18.18$

It should be noted from perusing the above, as one of the lessons to be learned early about the diagnosticity of evidence, that the items which have the greatest impact upon the final values are the items with the greatest spread of numbers. Item 2 had no effect on the probability of the scenario because, although it had large probabilities associated with it, the item gave no indication of what scenario was more likely; it was not diagnostic. The assignment of three nineties implied that the item was very likely to occur no matter what scenario was true. Item 3, however, was a sensitive indicator, because it was very unlikely if the second and third scenarios were true; the effect of item 3 was to depress significantly the revised probability for those two scenarios. This concept of diagnosticity can usually be conveyed to the participants by showing them the results of the calculations for a round or two.

B. The Simplified Method

Using the simplified method when the question is a yes-no either-or formulation reduces the effort for both the analyst and the coordinator. It is even feasible to perform all the calculations by hand. In this procedure, each item is evaluated as before and assigned two numbers—one for each hypothesis—but they are reported as a ratio, of one over the other. Identical results are obtained whichever hypothesis is in the numerator, but the practice must of course be consistent. For example, if an item is only 10% likely if A is true, and 90% likely if B is true, and the ratio is to be reported as A over B, the evaluation for the item would be 10/90.

The assignment procedure is followed for each item independently, and the mathematics of revision is then considerably simplified. For each analyst the starting probabilities are formed into a ratio (A—25% and B—75% would become 25/75) and simply multiplied against all the item ratios consecutively. Since multiplication is commutative, this can be done in any order, and much cancellation can usually be done if the whole set of ratios is considered. For example,

$$\frac{75}{25} \times \frac{\cancel{90}}{\cancel{80}} \times \frac{\cancel{80}}{\cancel{40}} \times \frac{\cancel{20}}{\cancel{90}} \times \frac{\cancel{40}}{10} \times \frac{50}{\cancel{20}} = \frac{3750}{250} = \frac{15}{1}$$

The resulting figure is the revised ratio, and should be kept as the starting ratio for the next round.

The ratio must be converted to probabilities (percentages) by the following simple method for reporting purposes: add the numerator and denominator together and divide the total into 100. In the case above this produces 100/16, or 6.25. The result of this division can be used to multiply both the numerator and the denominator, producing the percentages for A and B, respectively. Again in this case, $A = 15 \times 6.25 = 93.75$, and $B = 1 \times 6.25 = 6.25$. Remember that for the next round the starting number is the ratio, not the percentages.

This procedure can be further simplified by reducing the numerical scale. If probabilities are only allowed to be multiples of 10, such as 10, 20, 50, and 90, the calculations become much easier. This level of precision is thought by some to be as much as an untrained mind can estimate, although many analysts have demonstrated their ability to use the full 100-point scale. Also, because 10/90 is the same as 1/9, the ratios can be assigned using just the numbers from 1 through 9.

Whichever method is used, the basic or the simplified, each analyst will end up with a set of probabilities, which can be reported in a number of ways, as described in the section on graphics.

4. Problems in Mid-course

The most frequent problem in conducting a Bayesian exercise is one common to any group project: the management of a group of individuals, some of whom will make unusual demands and require special attention. There will also be occasional logistics problems, and delays in cartography and printing, for which there is no ready solution.

Holidays, vacations, and dropouts are another problem which should be anticipated. Ideally each participant should designate an alternate who would be familiar with the exercise and could take over when necessary. When that is impractical, the exercise can be run with a diminished number of participants; in that case, the individual's line can either be left out or extended without change, and any averages can of course be calculated on the basis of a smaller number of analysts.

Finally, there is always the possibility that the basic design of the exercise or the formulation of the question will be found lacking, but that is the reason a practice run or two is recommended.

5. Terminating and Evaluating a Bayes

In some exercises the scenarios specify a fixed deadline at which time the event either has or has not taken place, such as the OPR report on the likelihood of a North Vietnamese offensive before July, 1974 (the end of the dry season). It is clear that such a project would be terminated at that time, if not before. Certain other questions will be formulated to look at the probabilities of various scenarios within a continuously-moving time frame, such as the report on the likelihood of Sino-Soviet hostilities within six months. A project such as this may be terminated at any time, or extended indefinitely.

The Bayesian method upon completion results in an archive of evidence, evaluations, and predictions which lend themselves to various forms of evaluation; OPR has undertaken only the grossest measures of performance, however. A wealth of material lies untouched for rating collection systems, analyst bias, the use of intelligence in the community, or—that most dubious measurement of all—analyst performance.

The main criterion for evaluation is the accuracy of prediction, by individuals and by group, although this may not be as straightforward as it seems. Because of the myriad variables in the prediction equation, an event may occur which was only 10% probable the day before, or an event which was scheduled to take place may fail to materialize. Thus there have been times of great uncertainty during our reporting periods when the probability of a scenario rose, only to fall back again. Did this mean that the high probability of the event occurring was somehow in error? In retrospect, it seems to mean that the event could very well have occurred at that time if other factors had coincided; the evaluation really cannot be considered "wrong."

Generally, OPR's studies in 1974 successfully predicted non-events, that is, they showed that the evidence did not support any of the positive scenarios, and none of them occurred. In this case, the only point to be noted is how early in the exercise the evaluations moved away from an indeterminate figure toward a strong probability of no change. It has been our experience that the Bayesian calculations show this movement earlier than the analyst's own judgment would.

Until one of the positive scenarios actually occurs during the course of a Bayesian exercise, there is no way of knowing the predictive value of the technique. When the positive event does take place, it will be possible to conduct a much more searching evaluation of the Bayes procedure. What were the earliest indicators? What evidence was missing, overlooked, or misperceived? When did the trend lines signal a significant alteration in the situation? How did the Bayesian assessment compare with other intelligence assessments?

INHERENT WEAKNESSES AND PROBLEMS OF THE BAYESIAN TECHNIQUE

1. Limited Applicability

The first and foremost reservation in the use of the technique, as noted earlier, is that it is applicable only to certain types of questions. They must be capable of definition as a set of fairly distinct outcomes or hypotheses. Also, the procedure as developed is cumbersome enough to discourage its use for questions on a crisis schedule, although it is conceivable that the technique could be so adapted, using computer terminals and greatly simplified presentational techniques.

2. Data Problems

There is the problem of identifying which evidence is relevant—whether certain peripheral items should be included. And if included, whether they should carry less weight than other items. The OPR exercises have delegated that decision to the analysts. After all, they are experts, and their frequent disagreement over items shows that objective measures of relevance would be virtually impossible to devise. Little editorial judgment is imposed by the coordinator in the process of consolidating evidence, and any item which appears to be even peripherally relevant is included for evaluation. Nevertheless, each analyst is then allowed to ignore any item he or she considers irrelevant. This gives the participants great leeway over what they rate, but insures that they at least see the evidence and make an explicit decision whether it is relevant. Furthermore, if two or more items are seen as overlapping, the participant is asked to rate only one of them.

Related to this is the problem of source reliability. Although some methodologists have suggested that each analyst assign a numerical rating of source reliability along with each item, to be incorporated into the calculations as a weight, the OPR studies have avoided placing this extra burden on the analyst by requesting that the probabilities reflect how much faith the analyst places in each item. As mentioned earlier, the effect of an item will depend on the spread of probabilities assigned, and items of greater salience and reliability will automatically be assigned a greater range of probabilities, if the analyst understands the process and rates items thoughtfully. The problem of discerning items of disinformation or fabrication is left to the judgment of the participants; indeed, one of the benefits of the Bayesian process as a group exercise may be the opportunity to identify and sift out such items.

The problem of negative evidence is another point which bothers strict methodologists. This refers to the fact that the absence of any positive evidence may in itself be highly indicative, and the journalistic bias toward reporting events rather than non-events compounds the situation. That is, we tend to get news only of events or changes, whereas the fact that the status quo is being maintained can be quite significant. This problem has been recognized—as discussed in previous sections—and is partially satisfied by including an item such as "how likely is it that exactly this volume of evidence would occur (and be seen) if hypothesis A is true?", et cetera.

3. Problems Over Time

There are indisputable difficulties in the use of the method in a project continuing over many months. First of all, the questions probably require some reference to a time period (explicit or implicit), that is, an event to occur within the proximate month or year. And as the project continues, the timeframe must either contract or move forward. Contraction is illustrated by OPR's first reports, which investigated the probability of an event before a certain fixed date. In this case, the passage of time and the reduction of the period remaining may itself be of significance. The administrator may wish to include an item to that effect for evaluation.

Moving the timeframe forward is illustrated by the second and third OPR studies, which looked at the probability of events "within the next X months." This "sliding window" approach brings up the problem of retaining or discarding data which was evaluated months earlier with regard to an earlier frame of possibilities. One solution, adopted by the second study, which covered a year, is to drop earlier evaluations, maintaining only the most current three months of evidence multiplied against the original probabilities. The starting probabilities were also updated at intervals.

4. Problems with Numbers

There are also two numerical problems. The first is that multiplication by zero must not occur; once a probability becomes zero it can never recover. Thus any evaluation of zero probability must be replaced by a very small number. The second is more profound, and is the problem of how individual participants handle numbers, especially probabilities. It has been OPR's experience that some analysts think easily in probabilities, others have to work at it each time, and a few need constant attention and retraining. It is essential that the process of assigning probabilities or ratios be reviewed on an individual basis when necessary. We have found that the administrator can recognize when a participant is uncertain of the procedure, and one of the most effective teaching devices for sharpening the ability to assign probabilities is to show him or her the effects of each of the evaluations on the revised probabilities.

5. Manipulation

Finally, there is the problem of conscious manipulation. An analyst may assign his probabilities in a manner which reflects a pre-determined goal rather than unbiased judgment. This is especially true where the Bayesian probabilities affect the individual trend lines, as in the first study. Solutions to the problem could include disciplinary action or expulsion from the exercise, but the later OPR reports to some extent circumvented this problem by providing another outlet—the supplementary graphs—for the expression of individual opinions.

APPENDIX 1

THE STATISTICAL/MATHEMATICAL BASIS FOR THE CALCULATIONS

The statistical formula which underlies this method first appeared in an essay by the Reverend Thomas Bayes in 1763. It is a tool of statistical inference, used to find the probability that an observed event was caused by one source rather than another.

1. Rationale

The purpose of Bayesian analysis is to determine the probability that each scenario (hypothesis) is true, given that certain items of evidence have been seen. The Bayesian formula allows this to be calculated in an indirect fashion, from the probabilities that the items would be seen if each scenario were true. In our application of the method, analysts assign subjective values to the latter probabilities (that items would be seen), and the OPR coordinators calculate the former (that the scenarios are true) by applying the Bayesian formula. Only one other thing is necessary: a beginning set of probabilities for the scenarios, which the analysts supply for the initial round only, based solely on intuition. (This initial assignment, and the subsequent assessments, could be refined by various academic techniques, but we have not risen to that level of perfection or analyst rapport.)

2. Derivation

The formula itself is derived from the following basic statistical identities and transformations:

S—represents a scenario, or hypothesis, such as "The USSR is planning to launch a nuclear attack on China."
I—represents an item of evidence, such as "TASS reports that the Chinese are deploying nuclear missiles, thereby threatening world peace."
P(S)—is the probability of a scenario being true.
P(I)—is the probability of an item occurring and being seen.

$S \cap I$—is the intersection of S and I, that is, both S and I.
P(I/S)—is the conditional probability of I given S, that is, that an item I would occur if a particular scenario S were true.
P(S/I)—is the conditional probability of S given I, that is, that a scenario would be true if an item were seen.

1. $P(S \cap I) = P(I \cap S)$ commutativity of intersection

2. $P(I/S) = \dfrac{P(I \cap S)}{P(S)}$ definition of conditional probability

3. $P(I/S) = \dfrac{P(S \cap I)}{P(S)} = \dfrac{P(I \cap S)}{P(S)}$ from 1 and 2

4. $P(I \cap S) = P(S) \times P(I/S)$ transposition of 2

At this point, we define m mutually exclusive scenarios, S_i, of which one must occur. That is,

$$\sum_{i=1}^{m} P(S_i) = 1$$

5. $P(I) = P(I \cap S_1) + P(I \cap S_2) + \ldots + P(I \cap Sm)$ by definitions

6. $P(I) = (P(S_1) \times P(I/S_1)) + \ldots + (P(Sm) \times P(I/Sm))$ from 4 and 5

7. $P(I) = \sum_{i=1}^{m}(P(S_i) \times P(I/S_i))$ restatement of 6

The desired probability is the probability that each S_i is true given that a piece of evidence has been seen, or, using this notation, $P(S_i/I)$.

8. $P(S_i/I) = \dfrac{P(I \cap S_i)}{P(I)}$ from 3

9. $P(S_i/I) = \dfrac{P(S_i) \times P(I/S_i)}{P(I)}$ substitution of 4 in 8

10. $P(S_i/I) = \dfrac{P(S_i) \times P(I/S_i)}{\sum_{i=1}^{m}(P(S_i) \times P(I/S_i))}$ substitution of 7 in 9

The last formula, which is the rule of Bayes, says that, given an analyst's starting probabilities $P(S_i)$ for all scenarios, and his assessments of how likely an item would be if S were true, $P(I/S_i)$, then the new probability of S_i, now that I has occurred, can be calculated using formula 10.

3. Example

Assume an analyst has assigned the following probabilities, intuitively, to scenarios one and two, that the USSR is planning to launch a nuclear attack on China within six months, and that she is not:

$$P(S_1) = 10\% \text{ or } .1$$
$$P(S_2) = 90\% \text{ or } .9$$

Also assume that the following item comes in and that the analyst assigns probabilities that it would occur, first assuming that Russia is planning a nuclear attack, and second, assuming that she is not planning one:

"TASS reports that the Chinese are deploying nuclear missiles, thereby threatening world peace."

$$P(I/S_1) = 99\% \text{ or } .99$$
$$P(I/S_2) = 80\% \text{ or } .8$$

This information can be used to revise the probabilities that each scenario is true by using the Bayesian formula:

$$P(S_1/I) = \frac{P(S_1) \times P(I/S_1)}{\sum_{i=1} (P(S_i) \times P(I/S_i))} = \frac{.1 \times .99}{.099 + .72} = \frac{.099}{.819} = .12$$

Similarly,

$$P(S_2/I) = \frac{P(S_2) \times P(I/S_2)}{\sum \cdots} = \frac{.9 + .8}{.099 + .72} = \frac{.72}{.819} = .88$$

Notice that the two new probabilities add to 1 (100%) even though the item probabilities did not. As this is a recursive process, in which a succession of items of evidence are assessed, the new probabilities are used as $P(S_i)$ in calculating the effect of the next item.

APPENDIX 2

INTERACTIVE PROGRAMS IN APL AND BASIC

1. Interactive APL Program

```
     ∇ B;INSTR;INSTR1;INTERMEDIATE;COUNT;PRIOR;REVISED;TIMES;PROB;TOTAL
[1]    ' '
[2]    ' '
[3]    'INSTRUCTIONS? TYPE YES OR NO.'
[4]    INSTR1←3-(ρ(INSTR←⍞))
[5]    →16×⍳INSTR1
[6]    'THIS PROGRAM CALCULATES EVENT PROBABILITIES'
[7]    '         BY THE BAYESIAN FORMULA'
[8]    ' '
[9]    'YOU WILL BE ASKED TO ENTER A SET OF PRIOR PROBABILITIES,'
[10]   'THE NUMBER OF ITEMS OF EVIDENCE, AND THE ITEM PROBABILITIES.'
[11]   ' '
[12]   'TO CORRECT A MISTAKE:  ON A DELTA, MOVE THE CURSOR BACK AND HIT
[13]   '                       ON A 2741, BACKSPACE AND HIT ATTN.'
[14]   'IF YOU WANT TO STOP, TYPE THE  →  KEY.'
[15]   ' '
[16]   'PRINT INTERMEDIATE RESULTS? TYPE YES OR NO.'
[17]   INTERMEDIATE←3-(ρ(INTERMEDIATE←⍞))
[18]   ' '
[19]   COUNT←0
[20]   'PRIOR PROBABILITIES FOR THIS ANALYST, IN ORDER'
[21]   PRIOR←⎕
[22]   REVISED←PRIOR
[23]   'HOW MANY PIECES OF EVIDENCE'
[24]   TIMES←⎕
[25]   COUNT←COUNT+1
[26]   'PROBABILITIES FOR ITEM ';COUNT;', IN ORDER'
[27]   PROB←⎕
[28]   →31×⍳((ρPROB)=(ρPRIOR))
[29]   'WRONG NUMBER OF ENTRIES'
[30]   →26
[31]   TOTAL←+/(REVISED×PROB)
[32]   REVISED←(REVISED×PROB)÷TOTAL
[33]   →35×⍳(INTERMEDIATE)
[34]   'INTERMEDIATE RESULTS: ';REVISED×100
[35]   →25×⍳(COUNT<TIMES)
[36]   ' '
[37]   'P(Sι|I)  = ';REVISED×100
[38]   ' '
[39]   ' '
[40]   ' '
[41]   ' '
[42]   ' '
     ∇
```

2. Interactive BASIC Program

```
100 DIM B(10)
105 DIM I(10)
110 PRINT "INSTRUCTIØNS? TYPE YES ØR NØ."
120 INPUT A$
130 IF A$ = "NØ" THEN 200
140 PRINT "THIS PRØGRAM CALCULATES EVENT PRØBABILITIES"
150 PRINT "BY THE BAYESIAN FØRMULA."
160 PRINT "YØU WILL BE ASKED TØ ENTER THE NUMBER ØF"
170 PRINT "SCENARIØS, A SET ØF PRIØR PRØBABILITIES,"
180 PRINT "THE NUMBER ØF ITEMS ØF EVIDENCE, AND THE"
190 PRINT "ITEM PRØBABILITIES."
192 PRINT
194 PRINT "TØ CØRRECT A MISTAKE, USE THE @ KEY."
196 PRINT "IF YØU WANT TØ STØP, HIT THE BREAK KEY."
198 PRINT
200 PRINT "PRINT INTERMEDIATE RESULTS? TYPE YES ØR NØ."
210 INPUT A$
220 PRINT
300 PRINT "NUMBER ØF SCENARIØS, INCLUDING NULL:"
310 INPUT S
320 IF S<2 THEN 990
330 IF S>10 THEN 990
350 PRINT "PRIØR PRØBABILITIES FØR THIS ANALYST, IN ØRDER:"
360 FØR C1 = 1 TØ S
370 PRINT C1;":";
380 INPUT B(C1)
390 NEXT C1
400 PRINT "HØW MANY PIECES ØF EVIDENCE?"
410 INPUT E
430 FØR C2 = 1 TØ E
440 PRINT "PRØBABILITIES FØR ITEM";C2;"IN ØRDER:"
460 FØR C1 = 1 TØ S
470 PRINT C1;":";
480 INPUT I(C1)
490 NEXT C1
500 D = 0
510 FØR C3 = 1 TØ S
520 D = D + ( B(C3) * I(C3) ) / D
530 NEXT C3
550 FØR C3 = 1 TØ S
560 B(C3) = ( B(C3) * I(C3) ) / D
570 NEXT C3
630 IF A$ = "NØ" THEN 690
650 PRINT "INTERMEDIATE RESULTS:"
660 FØR C1 = 1 TØ S
670 PRINT C1;":";B(C1)
680 NEXT C1
682 PRINT
690 NEXT C2
710 PRINT "P(S/I) = "
720 FØR C1 = 1 TØ S
730 PRINT C1;":";( B(C1) * 100 )
740 NEXT C1
790 GØ TØ 998
990 PRINT "TØØ FEW ØR TØØ MANY SCENARIØS."
998 STØP
999 END
```

Approved For Release 2000/09/14 : CIA-RDP84B00506R000100070010-7