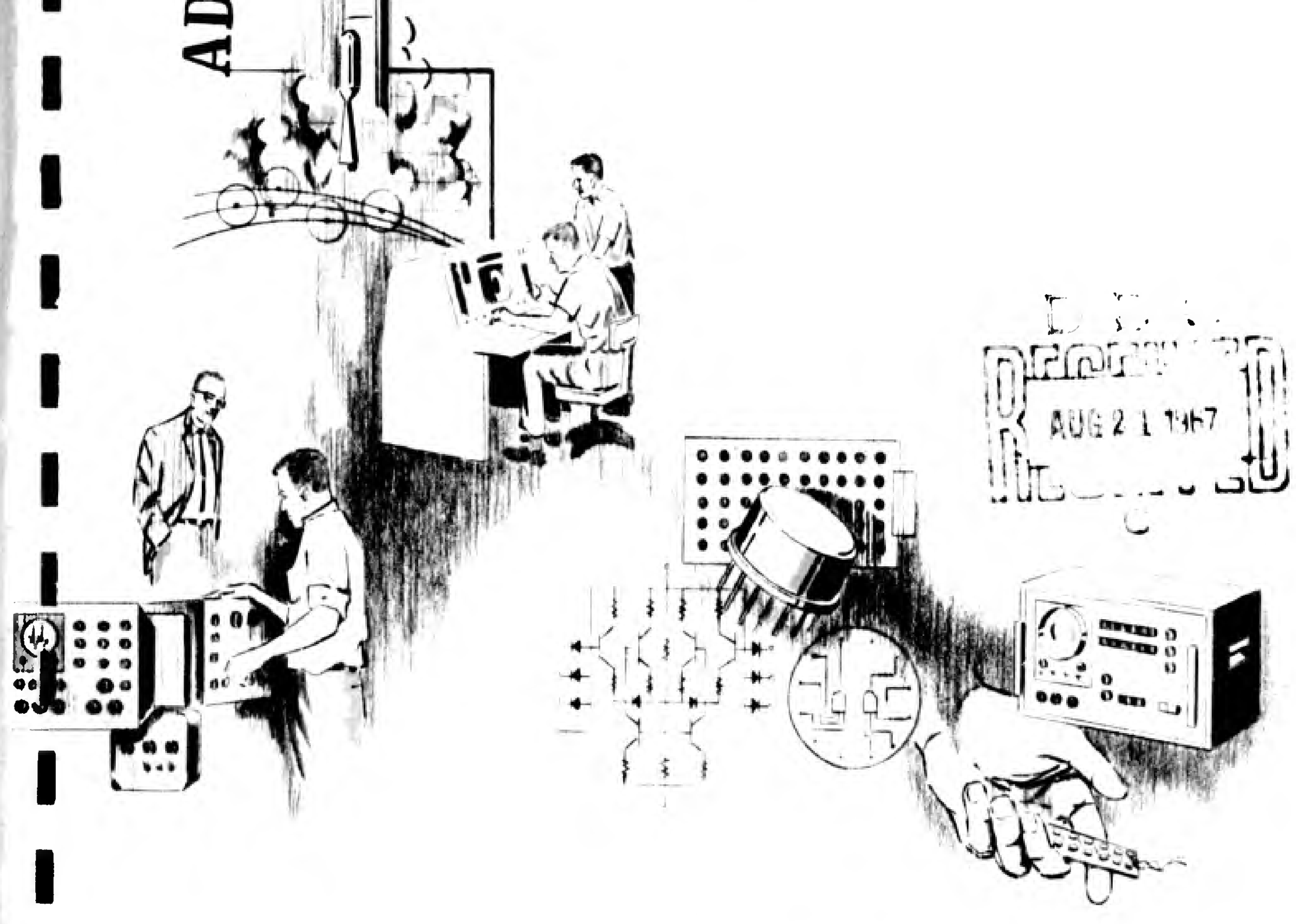
HRB·SINGER, INC.

AD656597

4015.11–R–2

THE ROLE OF PLAUSIBLE REASONING WITHIN MILITARY INTELLIGENCE An Application of Bayes Theorem As A Model For Problem Solving



SINGER
INFORMATION SCIENCES

HRB-SINGER, INC.
SCIENCE PARK, BOX 60 • STATE COLLEGE, PA. 16801
A SUBSIDIARY OF THE SINGER COMPANY

4015.11--R--2

THE ROLE OF PLAUSIBLE REASONING WITHIN MILITARY INTELLIGENCE: An Application of Bayes Theorem As A Model For Problem Solving

Prepared Under Navy Contract N00014-66-C0230

Sponsored By: Information Systems Branch
Office of Naval Research

Prepared By: C.R. Blunt, P.T. Luckie, E.A. Mares, D.E. Smith

MAY 1967

COPY NO. 52 OF 113 COPIES

SINGER
INFORMATION SCIENCES

ABSTRACT

The military intelligence analyst must cope with uncertainty during his problem-solving efforts. This uncertainty maps directly through the analysis and synthesis processes and affects his confidence in the output. A formal methodology for analysis can reduce some of these unfavorable effects while augmenting the analyst's problem solving capabilities. The role of plausible forms of logical reasoning within intelligence analysis is reviewed, followed with an introduction of Bayes Theorem as a model for intelligence analysis. The conjecture is made that Bayes Theorem can also serve as the nucleus of a formal methodology. The application of Bayes Theorem to several types of problems is demonstrated. However, the implementation of such a model as the nucleus of a complete analysis methodology is hindered by several significant problems. Some of the prime hindering aspects are delineated and discussed.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF ILLUSTRATIONS

BLANK PAGE

I. INTRODUCTION

> "...A great part of the information obtained in war is contradictory, a greater part is false, and by far the greatest part, somewhat doubtful."
>
> Karl von Clausewitz

The material presented in this paper highlights a technique that could help the analyst integrate tenuous data with the existing intelligence. This technique is based on an aspect of probability theory (Bayes Theorem) which provides a vehicle for weighting evidence and updating existing opinion in light of new inputs. The basic rationale underlying this research[1] has been essentially that

1. Intelligence analysis and synthesis efforts are hindered by the gaps and noise that are present in the input,
2. Intelligence analysts are fallible and often constrained by time and other limitations--they may not always extract the full measure of useful intelligence from these data--and
3. Improved collection systems may not entirely eliminate these input problems.

Therefore, it is necessary to continue to improve the intelligence capability to exploit data obtained by design or chance from sources of varying reliability. This research program examines methods that tend to alleviate the problem of producing intelligence from data that may be incomplete and not totally reliable.

---

[1]This research program is sponsored by the Information Systems Branch of the Office of Naval Research under contract N00014-66-C0230.

A. UTILITY OF A FORMAL METHODOLOGY IN INTELLIGENCE PRODUCTION

Central to the theme of this report is the very simple concept that although events of the real world can be portrayed in black and white (e. g., a submarine's pendant is 903 or it is not 903), knowledge of these events is often obtained from imperfect sources or is inferred from ancillary data; hence, the data base depicting these events can frequently be best described with shades of grey (e. g., the submarine's pendant is possibly 903). The uncertainty of the information in the data base affects the degree of confidence that can be placed on conclusions derived from these data.

There are two points that are advanced in this research effort. First, in the complex mental processes of data analysis and synthesis, it may be beneficial to make more explicit some of the judgments and inferences used in formulating the intelligence products. This "bookkeeping" aspect increases the likelihood that the analyst will consistently evaluate, weigh and integrate all pertinent evidence. Second, the complexity of the synthesis process suggests the need for utilizing an aid in integrating the reported data into an intelligence assessment.

There is a mathematical method that can aid the analyst in distilling intelligence from imperfect data, discarding irrelevant information, and updating the existing assessment. This method is an adaptation of an existing inference model (Bayesian Decision Model) and incorporates several desirable properties; e. g., it enables one to probe into such things as the inferences, assumptions, judgments and weightings that are used in evaluating input data and synthesizing a solution. Bayes Theorem is a vehicle for turning prior judgments into optimal posterior ones by employing probabilistic statements concerning information. Thus, if degrees of belief or levels of confidence in the reported information can be represented adequately in numerical notation (e. g., only 30% confidence that the submarine's pendant is 903), then it is possible to exploit this formal mathematical method as an aid in intelligence analysis and synthesis. As data are integrated in intelligence processing efforts, the numerical expressions of data fidelity are also combined to yield a new indication of confidence in the resulting product.

Some of the advantages gained in utilizing such a formal methodology in intelligence production efforts include the following:

1. The techniques couple the evaluated input data and the resulting products in an explicit manner. Each integrated item produces a corresponding change on the intelligence output thus reflecting both the utility and fidelity of each input datum. This provides the analyst with

   a. A diagnostic capability to examine and re-examine the rationale underlying his hypotheses.

   b. An indication of the plausibility of each alternative under examination at any point in time. Thus, if time pressures prematurely terminate his efforts, he can report the relative weights of each possible solution.

2. The techniques provide a means for integrating the effects of both conflicting data as well as substantiating information. Since the importance of the conflicts can be coupled to the significance of their impact on the intelligence output, these techniques may be useful in minimizing the efforts necessary to resolve critical differences in the data base.

3. There are indications that the techniques can synthesize the analyst's individual judgments better than man can. Furthermore, the mechanical nature of the synthesis methodology allows the analyst to identify and integrate additional hypotheses during his analysis effort with very few additional assessments.

4. These techniques, when coupled with on-line computer processing technology, offer the analyst a rapid, efficient method of (a) organizing his data base, (b) evaluating new inputs and (c) updating his products.

Thus, a Bayesian-aided analyst can augment his intellectual efforts with an electronic bookkeeping capability that can also organize his quantified judgments in a process of data synthesis.

In addition to the Bayesian aid in intelligence production, it is reasonable to envision a comprehensive system providing other aids to the analyst in many phases of his production efforts. One of the first innovations to any processing system would be to utilize the computer's memory and its logically stimulated recall capability to store information and questions that the analyst should remember when solving recurring problems. Thus, the system begins to take on the flavor of a self-organizing and adaptative system. There are now such branching innovations with feedback common in CAI (Computer-Assisted Instruction). CAI techniques, such as dialogue and interrogation approaches, bring an important training capability to intelligence processing. With appropriate adaptation, the more experienced analysts receive the benefits of being reminded about various aspects of the problem they previously noted in similar efforts. Moreover, new analysts receive the benefits of the stored knowledge of the more experienced analysts, removing some of the mysticism that normally shrouds such processing for the novice. Thus, the novice could be alerted to the existence of various patterns of operation, noted in prior intelligence efforts, that share something in common with the problem at hand.

Incorporating a display mechanism into the system could provide the analyst with a type of "electronic chalkboard" [Newman, 1966]. Research has indicated that the pictorial-verbal display apparently makes numerical data more tolerable. Properly formatted displays allow people to tolerate and absorb much more information than normally would be expected. The stage would then be set for the so-called step-display-look cycle [Newman, 1966], fundamental to intelligence processing. The step is a unit action that the analyst wants taken by his computer, and could include numerical calculations, nonnumerical data manipulation, or simple updating or purging of information. After taking the step, the analyst displays the results and looks at them. The cycle is then complete and ready to be reinitiated. The most important phase is the look phase because the analyst is not only absorbing the information, but is also contemplating his next step. This is the creative phase that exemplifies the extension of man's capability with a computer display system. By providing memory aids and extending the manipulating, recording, and transformating capability of the analyst, his problem solving and decision making ability is greatly expanded.

Thus, a complete system can be formulated to augment the analyst's efforts in intelligence processing, analysis, and synthesis. The inclusion of subassemblies within the processor to facilitate information storage and retrieval, curve fitting and plotting calculations, simulation of real-world situations, graphic presentations of geographic areas, etc., would be a natural modular approach for extending the system in support of a complete analysis methodology.

B. REPORT OVERVIEW

This report starts with a discussion of the intelligence analyst (Chapter II) and his role in the production effort. The analyst is presented as a problem solver who could utilize the logical forms of reasoning if the uncertainty in the data could be quantified to some extent. Chapter III presents an intuitive form of plausible reasoning and illustrates how the uncertainty in the premises of a logical argument map through to the conclusions. This chapter introduces Bayes Theorem as a natural extension of the plausible form of logical reasoning. Following the introduction to Bayesian synthesis, the fourth chapter of this report discusses several types of problems that can be examined with this aid. These examples are hypothetical; their purpose is not to present a solution but, instead, to illustrate the range of problems that admit to this form of synthesis. While studies by HRB-Singer and others[1] indicate that these mathematical techniques show promise, there is considerable research that must be completed before the true potential and limitations of these approaches will be understood. Chapter V discusses some of the problems hindering applications of Bayesian Techniques in intelligence.

The appendices of this report separate several individual topics from the main text. Appendix A gives a derivation of Bayes Theorem and briefly presents what is known as the recursive version. Appendix B discusses one of several pilot studies in which Bayes Theorem was applied to a real-world problem of data synthesis. In this study, an analyst's judgment about the data was synthesized by the analyst and, independently, by a computer program utilizing the recursive version of Bayes Theorem. The results are compared and briefly

---

[1]These studies include the works of Edwards, Peterson, Philips and Hayes of the University of Michigan; Kaplan and Newman of SDC; Briggs, Schum, Goldstein, and Southard of Ohio State University. Some of these studies will be discussed later in this paper.

discussed in this section. Appendix C touches upon the role of probabilistic intelligence outputs and command's decision process. Appendix D examines the particularly difficult problem of "noise" factors in an environment that exhibit characteristics that may be mistakenly identified as belonging to the object of interest. Data about these characteristics, when integrated into the intelligence effort, influence the intelligence assessment (whether it is formulated in a Bayesian manner or not). Appendix E presents some applications of digitized logic as a filter on the inputs to intelligence production efforts. This technique of filtering examines the consistency and redundancy which may exist among the input information. Moreover, the technique of digitized logic can be used to determine the validity of conclusions drawn from propostions of the intelligence assessments.

II. INTELLIGENCE PRODUCTION

The principal mission of the intelligence community is to be responsive to command's requirements for timely, evaluated information about the enemy (present or potential) -- to provide his location, strength, capability, vulnerability and intent. In order to satisfy these broad requirements, sophisticated and expansive systems have been established to provide, on a continuing basis, data from both directed and nondirected collection efforts. The various sources providing this material differ in capability; thus, the data vary in reliability and completeness of coverage.

The production of intelligence from these data requires the combined resources of many agencies and organizations who examine a wide spectrum of intelligence interest areas each day. While there are both multifarious intelligence interests and a wide range of data sources, there are some fundamental aspects invariant among all production efforts. This chapter briefly discusses two of these, i.e.,

1. The intelligence analyst
2. The process of data synthesis

The mathematical techniques discussed in the following chapters of this paper offer promise as an improved capability to aid the analyst in his efforts to synthesize the apparently disconnected bits and pieces of input data.

A. THE INTELLIGENCE ANALYST

One of the most critical of all components in any intelligence system is the man, e.g., the experienced analyst who must screen, evaluate, and transform the input information into useful intelligence products. Figure 1 illustrates one viewpoint of this function. In this effort, the analyst may receive mixtures of raw and partially evaluated data from his prime sources as well as products and evaluated intelligence from other groups. The analyst screens these inputs to determine their pertinence to his interest areas. Because the collection and reporting efforts are not always complete and reliable, the analyst must frequently temper the superficial import of these data with judgments about the

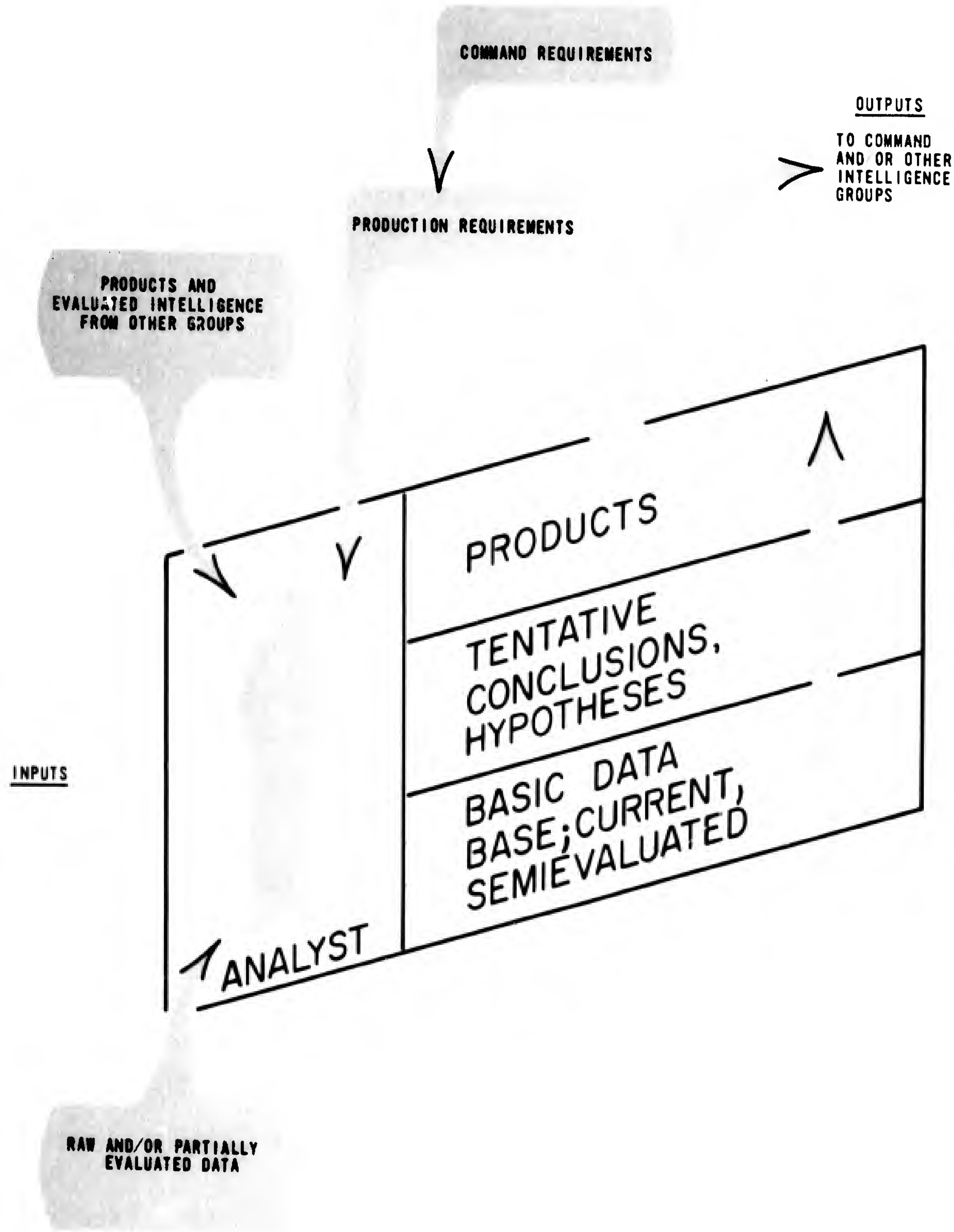


FIG. 1 ONE VIEWPOINT OF THE INTELLIGENCE PROCESS

reliability of the source and the credibility of the information. Data deemed pertinent to an interest area that cannot be rejected as invalid are integrated with the existing intelligence formulation. This integration of new inputs can revise the prior formulation in two ways, i.e.,

1. It can change the scope of a postulated estimate. For example, the possible appearance of a new pendant may tend to revise the estimate on the number of submarines operating in a fleet area. The observation of a series of tests may revise the estimate on the completion date of a new weapon system.

2. It can affect the degree of confidence one has in a postulated hypothesis. For example, the analysis of a seismogram could reinforce the belief that a nation has commenced underground nuclear testing. The possible appearance of a new pendant may tend to lower the belief that a ship building program is being reduced.

As more data are examined and integrated, the analyst's confidence in a tentative assessment may be raised to the level where he is satisfied that the gathered evidence uniquely supports one conclusion and he reports this as a finished item of intelligence. [1]

In general, the production of intelligence from collected data requires an intellectual effort encompassing both analysis (the breaking down of complex structures so as to discern their fundamental elements and relationships) and synthesis (the formulation of a whole from disconnected bits and pieces). Traditionally the intelligence community groups both of these processes under the general heading of intelligence analysis; thus the title "analyst." Both processes, however, are used in most production efforts. Sherman Kent [1965, p. 157], for example, cites the application of analysis in the early stages of strategic intelligence research"...to discover which facets of (the problem) are of

[1]Unfortunately, however, the pressures of time sometimes dictate that the analyst furnish his best estimate of the situation before he has put all the pieces together and has validated their fit in the intelligence picture.

actual importance to the U.S. and which of several lines of approach are most likely to be useful to its governmental consumers." In the last stage of this formulation is the "establishment of one or more hypotheses as truer than others and statement of these hypotheses as the best present approximations of truth." This last stage is predominately a process of data synthesis.

B. THE PROCESS OF DATA SYNTHESIS

Intelligence production can be viewed as an effort to provide "solutions" to problems that are pertinent to our national interests.[1] These problems may be extremely broad, e.g., "what weapon systems will nation X have in the 1970-1980 time frame?" or they may be quite specific, e.g., "what is the clearance of the railroad tunnel at location Y?" In many instances, a general intelligence problem may be decomposed into a series of smaller and more specific sub-problems. For example, the estimate of a nation's future weapon systems may be based, in part, on a knowledge of its present weapon systems, its research and development efforts, its evolving military tactics, etc. One role of the process of data synthesis is to suggest intelligence problems that should be explored (this may, in turn, aid in establishing collection requirements); another role of this intellectual effort is to determine which of several hypothesized solutions is most plausible.

1. The Role of Synthesis in Problem Formulating

The intelligence community continually collects and examines data about events occurring in the real world. Some objectives of this surveillance are to (1) provide early warning of impending offensive actions, (2) provide command with timely information about the locations of mobile units of interest, (3) detect the development of new weapon capabilities, (4) identify emerging political, military or economic factors among groups of interest, etc. In this examination of these data, the analyst usually has a problem area identified and is seeking to

[1]It should be noted that one of the more challenging aspects of intelligence is the recognition and formulation of meaningful problems to be explored. The posing of an unimportant problem can consume the energies of people already hard-pressed with commitments. On the other hand, the failure to identify a substantive problem can leave command unprepared for developing situations.

update his products. During these efforts, however, it is quite possible that an assembly of information may suggest a specific problem to be explored. For example, an analyst might assemble a series of disconnected items such as

a. Country X is believed to be expanding its air defense systems.

b. Country X has obtained a site in country Y.

From these two items, the analyst might attach more significance to the site than if it had come to his attention alone. If the interest in the potential significance of these items is high enough, these data may form the basis of a problem to be explored, e.g., "what is country X doing with the site in country Y?"

Once sufficient data have been gathered that are pertinent to the problem, the analyst will usually formulate several possible hypotheses about the solution, e.g., country X is building (a) an early warning station, (b) an airfield, (c) a missile base, etc. This formulation is based on the analyst's examination of the data, his prior experiences and knowledge of the subject area. The hypotheses "fit" the data that have been assembled; their function is mainly that of establishing more specific subproblems to solve: e.g., instead of asking "what is country X doing in country Y," the analyst can ask "is country X building an airfield at this site in country Y?"

> The hypothesis serves somewhat as a psychological pattern or stencil in permitting the needed facts to come to our attention and in excluding the irrelevant ones. The operation of the hypothesis may be likened to the 'stencil' we are carrying in mind as we look for a silver quarter that has been dropped in the yard. We walk about and glance over the ground. We actually 'see' or note little except objects that in some way resemble a quarter. Our drifting glance is fixed specifically by any small, round, bright objects; other things are usually ignored. In much the same way the hypothesis guides us to the relevant facts. We must be careful not to limit ourselves to one hypothesis, for if we do, the probability is high that we will exclude the very facts that would provide the key to the problem. Also, if we begin with only one hypothesis, we are likely to develop a liking for it and unconsciously to search only for those facts that will support it. We must avoid trying to establish any particular hypothesis.
>
> [Little, et al., 1952, p. 184]

In this process of data synthesis, the analyst attempts to fit various disconnected items into a proper relationship and to discern the significance of the construction. The hypotheses that he poses represent possible solutions to problems of importance to intelligence. The integration of new, pertinent, discriminating data will affect the degree of confidence that the analyst holds for each possible solution; increasing his belief in one (or more) of these and decreasing his belief in others.

2. The Role of Syntheses in Problem Solving

Once hypotheses have been established that exhaust the possible solutions of a problem, the processes of problem solving can continue by testing these hypotheses. One strategy in this approach is to earnestly attempt to discredit all of the hypotheses. The hypotheses surviving this test become stronger candidates as possible solutions to the problem. Another strategy is to examine all supporting data to determine which of the possibilities are most credible. Both strategies are useful methods in problem-solving and incorporate the processes of logical deduction and plausible induction.

a. Deduction. Logical reasoning includes the derivation of conclusions by inference. The deductive process starts with premises and moves by inference to a conclusion. The premises may be facts, convictions, hypotheses, assumptions, etc. "...the propositional rules transmit truth from premises to conclusions in any sound inference. If there is a deduction -- i. e., a sound inference -- then it is impossible for all its premises to be true when its conclusion is false." [Anderson et al., 1962, P. 79]

One form of deductive arguments can be represented in the following structures:

If A, then B		
A	premises	(1)
B	conclusion	

If A, then B		
not B	premises	
not A	conclusion	(2)

where the symbols "A" and "B" represent sentences like "the army is attacking" and "its supply lines are difficult to maintain," etc. Thus if it is known that all attacking armies have difficulty maintaining their supply lines, then the observation of an army in an attack permits one to conclude that the army is having difficulty maintaining its supply lines. Conversely, the knowledge that an army is not having difficulty maintaining its supply lines permits the conclusion that the army is not attacking. The reliability of conclusions reached by deduction depends upon (1) the truth of the premises and (2) the validity of the inference.

False conclusions can be drawn from true premises by fallacious reasoning, e.g.,

> if a submarine is "G" class, then it has a ballistic missile capability
> Submarine 958 has ballistic missiles
> Conclusion: Sub 958 is "G" class

In this situation, there are several classes that have a ballistic missile capability; thus, while all "G" class units are ballistic missile types, not all ballistic missile types are "G" class.

False conclusions can also be derived by valid argument if the premises are not true, e.g.,

> if country X is constructing an airfield, then Smith will direct the project.
> Smith is not directing the project
> Conclusion: country X is not constructing an airfield.

In this situation, it may be a fact that Smith is not directing the project. However, if the premise

> "if country X is constructing an airfield, then Smith will direct the project"

is an unverified hypothesis; it may be possible that someone else will direct the project; hence, it could be possible that country X is building an airfield under the direction of some other engineer.

One method of rejecting an hypothesis (H) is to pose it as the antecedent of a conditional statement and demonstrate that the consequent (C) is false, i.e.,

$$\frac{\begin{array}{c}\text{If H, then C}\\ \text{not C}\end{array}}{\text{Conclusion: not H}}$$

Thus, an analyst might form the hypothesis

H = Country X is constructing an airfield.

Now if he can construct a nontrivial true premise based on this hypothesis, he may be able to test the hypothesis. For example, assume the following premise to be true:

If H (country X is constructing an airfield),
then C (heavy construction equipment will be used).

If the surveillance effort shows that heavy construction equipment will definitely not be used, then the analyst can reject this hypothesis.

Unfortunately, the data available to intelligence are frequently not sufficiently reliable to determine if the premises are true. In the above illustration, for example, it may not be certain that <u>all</u> airfield construction utilizes heavy equipment. Also, in the matter of surveillance, the failure to observe the equipment may not mean that it was not used, nor need it indicate that it won't be used at some future point in time. In general, there are at least three simple structures that denote the varying points of uncertainty in the premises typical of intelligence, analysis, and synthesis efforts. These are:

1. $\dfrac{\begin{array}{c}\text{Possibly if A, then B}\\ \text{A}\end{array}}{\text{Conclusion: ?}}$	<u>Exemplifying</u> uncertain pattern or relationships

2. It is true that A implies B
   possibly A
   ______
   Conclusion: ?

   uncertain input data

3. Possibly if A, then B
   possibly A
   ______
   Conclusion: ?

   both problems

Intuitively, "B" exists as a plausible conclusion in each situation; however, the level of confidence in the truth of "B" certainly varies among the different cases. These situations represent examples of plausible deductive inferences. In the next chapter it will be shown that if it is possible to quantify the degree of one's belief in the premises, then it may be possible to derive a level of confidence for the conclusions derived by deductive reasoning.

b. Induction. Another method of plausible reasoning is the inductive form [Polya, 1945] and is typified by the structure

If A, then B
B is true
______
Conclusion: A is more credible

where again, "A" and "B" represent sentences. Consider a previous example illustrating how one might test the hypothesis (H) that country X is constructing an airfield. In this test, the analyst postulated the premise

> If country X is constructing an airfield, then heavy construction equipment will be used.

Assume this premise is true and suppose that the collection effort does in fact uncover the presence of heavy construction equipment at work on the site. This evidence does not in itself permit a reliable conclusion that H is true, i. e., that country X is constructing an airfield. It does, however, tend to support the conclusion. The degree of the support is dependent upon how many other plausible hypotheses have the same consequent. If, for example, heavy construction equipment is connected with all of the hypothesized solutions, then this item does not discriminate among the possibilities. On the other hand, if H is the only possible hypothesis explaining the use of heavy equipment, then the hypothesis not only implies the consequent, but the consequent implies the hypothesis, i.e.,

(If H, then C) and (If C, then H)

Verification of C (heavy construction equipment is being used at the site) enables the establishment of both inductive and deductive structures:

1. If H, then C
   C is true
   ―――――――
   H is more credible

2. If C, then H
   C is true
   ―――――――
   H is true

Thus, in this special case, the hypothesis becomes quite credible; in fact, it is confirmed. In general, the verification of the consequent of a conditional statement whose antecedent is an hypothesis does not establish the truth of the hypothesis. This verification does, however, usually tend to support its plausibility. The extent of this support may range from equal support to <u>all</u> hypotheses to confirmation of the tested case. Polya [1945, p. 189] identifies the structure[1]

If A, then B
B is true
―――――――
A is more credible

as the "heuristic syllogism" and offers that

> The conclusion of the heuristic syllogism differs from the premises in its logical nature; it is more vague, not so sharp, less fully expressed. This conclusion is comparable to a force, has direction and magnitude. It pushes us in a certain direction: 'A' becomes <u>more</u> credible. The conclusion also has a certain strength. 'A' may become <u>much more</u>

[1]It should be noted that the uncertainty of the premises of intelligence efforts also give rise to at least three inductive forms, i.e.,

1. Possibly if A, then B
   B is true
   ―――――――
   ?

2. If A, then B
   B is possibly true
   ―――――――
   ?

3. Possibly if A, then B
   B is possibly true
   ―――――――
   ?

> credible, or just a little more credible. The conclusion is not fully expressed and is not fully supported by the premises. The direction is expressed and is implied by the premises, the magnitude is not.

His later work [1954] incorporates applications of probability theory in examining the magnitude or degree of credibility added by the integration of new evidence. It is this line of attack that has captured the interests of many researchers. In many areas (e.g., Command and Control, Intelligence, Business Management, Medicine) decisions are frequently made under pressures of time and are often based on inconclusive data. An ability to identify which kinds of data offer the most discrimination in a problem enables one to examine the more fruitful evidence in the available time. An ability to examine the credibility of each alternative enables one to weigh the likelihood of the choices against the consequences of the results. Thus, an ability to quantify the uncertainty of information can aid the analyst who must reach some conclusion(s) as to the significance of these data; it can aid the decision maker who incorporates these conclusions into his decision processes. [1]

---

[1] Appendix C presents a discussion of the role of probabilistic intelligence assessments and the decision-making process.

BLANK PAGE

III. DEGREES OF BELIEF, PROBABILITY AND BAYES THEOREM

There are frequently several transformations that take place between the occurrence of some event and the reporting of some aspect of the event. The first transformation takes place as some sensor records stimuli associated with the occurrence of the event. The second transformation takes place as these are interpreted; thus, an observer sighting an object may perceive size and shape, and report that he has sighted a tank at coordinate x, y. An intelligence analyst, receiving such a report, must determine how much of the material reflects the true state of the event; e.g., if the enemy is practicing deception, the characteristics observed may not typify the true state of affairs.[1] On the other hand, the characteristics may be true, but the observation and/or interpretation may be faulty; e.g., the object may have been a personnel carrier.

In the context of intelligence, data are unreliable if the analyst cannot be confident that they adequately reflect the true state of the real world. Data reliability is often a judgment that is based on parameters such as basic capability of the collection system, prior reliability of the source, credibility of the report, etc. It is, of course, possible that an analyst may believe an erroneous or false report and also that he may reject valid input. The uncertainty of intelligence collection and processing often does not immediately admit to extreme judgments (i.e., the data <u>are</u> or <u>are not</u> reliable); but, instead, admit to different degrees of belief (i.e., the data <u>may</u> be reliable or they <u>probably</u> are reliable, etc.).

In the previous chapter, a brief discussion of logical reasoning was presented to demonstrate the power of deduction and the utility of plausible induction. It was noted, however, that both forms of reasoning encounter difficulty when the premises may not be true. The fact that the premises of intelligence inferences may not always be true does not preclude the useful application of logical reasoning in intelligence. This fact does, however, suggest the need for additional techniques of reasoning that enable one to extract useful information from the assembled data.

---

[1]For example, tank mock-ups may be strategically placed to give the illusion that tanks are present.

> Ordinary logic seems to be inadequate by itself to cope with problems involving beliefs. In addition a theory of probability is required. Such a theory is defined here as a fixed method which, when combined with ordinary logic, enables one to draw deductions from a set of comparisons between beliefs and thereby to form new comparisons. [Good, 1950, p. 3]

In his preface, Good defines probability as "... the logic (rather than the psychology) of degrees of belief and of their possible modification in the light of experience."

There are many schools of thought concerning the nature of a probability; two of these will be briefly discussed here. As noted above, one suggests that probabilities are subjective; that they are a measure of one's degree of rational belief about a particular action or state of nature. These probabilities are often referred to as personal probabilities. The other school views probabilities as being objective or measuring a "relative frequency." Here probability is defined as the relative frequency with which members of a class exhibit a specified property. This relative frequency theory is especially suited to take account of probability judgments arising out of statistical investigations. Assignment of such a probability can only be on the basis of the evidence available to the person making the assignment and are frequently referred to as "counts." Both schools define that a probability ranges between 0 and 1, where "0" denotes impossibility and "1" denotes certainty. This paper will not discuss the merits of either school, but does suggest that both views offer utility in the production of intelligence. Under some circumstances, expert judgment may provide the most reasonable assessment of the situation. In other instances relative frequencies, derived from the data base, may serve as reasonable expressions of probability. Moreover, there may occur situations where a relative frequency offers a value as a guide for human judgment to accept, increase or lower for the situation at hand.

For the moment, this paper will only distinguish between a "degree of belief" and a "probability" by notation. In the discussion which follows, "A," "B," etc., denote statements which may be either true or false. The degree of belief in A is represented by b(A) and is a number between 0 and 1

($0 \leq b(A) \leq 1$). Thus, if one believes that A is true, $b(A) = 1$; if one believes that A is false, $b(A) = 0$. Similarly, the probability of an event A is denoted as $p\{A\}$ ($0 \leq p\{A\} \leq 1$). If the likelihood of event A is certain, $p\{A\} = 1$; if A cannot occur, $p\{A\} = 0$.

One contention of this paper is that a degree of belief may be usefully substituted for an unknown probability value at least as an aid in analysis. Consider, for example, an old-fashioned riverboat poker game in which a traveler has been dealt a straight flush to the king. If the game is honest, the probability of his having the winning hand is exceedingly high (the odds against a Royal Flush being dealt is 649,739 to 1). On the other hand, if his belief that the game may not be honest is also quite high, this belief should temper his willingness to stay in the game.

A. QUANTIFYING THE DEGREE OF BELIEF IN CONCLUSIONS DRAWN FROM PLAUSIBLE IMPLICATIONS

It has been frequently noted that the intelligence analyst encounters data in which he does not have 100% confidence. He may draw conclusions from these data on the basis of implications which also may not always be true; thus, his conclusions are uncertain. This section defines rules by which one can quantify one's degree of belief in the truth of conclusions which have been derived by plausible implication. The assumption is made that one can quantify one's belief in the premises of the argument, i.e., that one can ascribe values to one's belief in the truth of statement "A" and in the implication "if A, then B" (which will be symbolically written "$A \rightarrow B$").

While the rules derived can actually be viewed as definitions and theorems in the context of probability theory, an attempt has been made to define them only by intuitive feelings about degrees of belief. They are rules, therefore, that have been derived in a pseudomathematical way, but which may, nevertheless, have appeal and practical value.

1. Derivation of the Theorem

Let A, B, C, ... denote statements which are either true of false; and let $\overline{A}$, $\overline{B}$, $\overline{C}$, ... represent their negations. The following definitions are made:

a. $b(A)$ is a number between 0 and 1 ($0 \leq b(A) \leq 1$), representing the degree of belief in the truth of A.

b. If one believes that A is true, then $b(A) = 1$.

c. If A_1, A_2, ...A_n are mutually exclusive, in the sense that no two of these can be true simultaneously, then $b(A_1 \text{ or } A_2 \ldots \text{or } A_n) = b(A_1) + b(A_2) + \ldots + b(A_n)$

d. $b(A \rightarrow B)$ is a number between 0 and 1 representing the degree of belief in the truth of B when A is true.

By b. and c. it follows that $b(A) + b(\bar{A}) = 1$ since A and $\bar{A}$ are mutually exclusive (they cannot be true simultaneously) and one of them must be true (A or $\bar{A}$ is true). Thus, $b(\bar{A}) = 1 - b(A)$.

The question is, given $b(A)$ and $b(A \rightarrow B)$, how should one define $b(B)$? A suitable definition of $b(B)$ will be postulated after specifying certain conditions that the definition should satisfy. These conditions are:

a. First, if $b(A) = 1$, then $b(B) = b(A \rightarrow B)$. This reasonably stems from definition d, for $b(A \rightarrow B)$ is the degree of belief in the truth of B when A is true and the assumption that $b(A) = 1$ means that it is believed that A is true.

b. Second, if $b(A) = 0$, then $b(B) = b(\bar{A} \rightarrow B)$. This is also reasonable because $b(A) = 0$ means $b(\bar{A}) = 1$ (i.e., it is believed that A is false) and $b(\bar{A} \rightarrow B)$ is the degree of belief in the truth of B when A is false.

c. Third, if $0 < b(A) < 1$ (i.e., the belief is that A is neither definitely true nor definitely false), then $b(B)$ should reflect the possibility of B being true even if A is not true. That is, it is reasonable to expect the degrees of belief in both "$A \rightarrow B$" and "$\bar{A} \rightarrow B$" to affect the degree of belief in B. In addition, if $b(A \rightarrow B) = 1$ (B is always true if A is true), then $b(B)$ should be at least as large as $b(A)$.

The following definition satisfies the conditions a.-c. above:

Def. 1: $b(B) = b(A) \cdot b(A \rightarrow B) + b(\bar{A}) \cdot b(\bar{A} \rightarrow B)$

Note that when $b(A)$ and $b(A \rightarrow B)$ are given, also $b(\bar{A})$ is known, but a value for $b(\bar{A} \rightarrow B)$ must be determined from the knowledge of the frequency with which B is true when A is false. If it is known that B can never be true without A being true, then $b(\bar{A} \rightarrow B) = 0$ and the above equation reduces to $b(B) = b(A) \cdot b(A \rightarrow B)$.

In the above definition it should also be noted that A and $\bar{A}$ can be viewed as two events which are mutually exclusive and either A or $\bar{A}$ must be true. The question now is: Suppose there are n different, mutually exclusive events $A_1, A_2, \ldots\ldots A_n$, in conjunction with one of which the event B will occur, the frequency of its occurrence depending on which of the A_i is true. Assuming one can express at a given moment a degree of belief in each A_i and a degree of belief in B given any one of the A_i, what is, at that moment, a reasonable degree of belief in B?

By reasoning similar to that made in deriving Def. 1, the following definition is a natural generalization of Def. 1, and the latter becomes then a special case of it:

Def. 2: Given $b(A_i)$ and $b(A_i \rightarrow B)$, where A_i is any one of n mutually exclusive events, one (and only one) of which must be true, the degree of belief in B is defined as

$$b(B) = b(A_1) \cdot b(A_1 \rightarrow B) + b(A_2) \cdot b(A_2 \rightarrow B) + \ldots + b(A_n) \cdot b(A_n \rightarrow B)$$

or $b(B) = \sum_{i=1}^{n} b(A_i) b(A_i \rightarrow B)$.

2. Discussion and Examples

In chapter II, three simple structures were presented that illustrated some points of possible uncertainty in deductive inferences of intelligence efforts. Using the notation introduced in this chapter, these structures collapse into one form, i.e.,

$$\frac{\begin{array}{c} A \rightarrow B \\ A \end{array}}{\text{Conclusion: } B}$$

where the confidence of the conclusion, "B," is derivable from the belief in the premises "A→B" and "A," and the belief in the implication "$\bar{A} \rightarrow B$." If <u>either</u> premise is true, the confidence that rests in the conclusion is directly connected with the degree of belief in the uncertain premise, i.e.,

$$(1) \quad \frac{\begin{array}{c} b(A \rightarrow B) = 1 \\ b(A) = U \end{array}}{b(B) \geq b(A) = U} \qquad (2) \quad \frac{\begin{array}{c} b(A \rightarrow B) = U \\ b(A) = 1 \end{array}}{b(B) = b(A \rightarrow B) = U}$$

Where U is the value given to the belief in the uncertain premise. An example of the first situation is as follows:

> It is known that a foreign submarine is within 200 nm. of Oahu, Hawaii.[1]
> It is believed that this unit is possibly of the Golf class.
> What is the likelihood that a ballistic missile system is within 200 nm. of Oahu?

Assume that all G class submarines carry ballistic missiles. Let G denote "submarine of the G class"; and let B denote "ballistic missiles." Hence, G→B is true, i.e., $b(G \rightarrow B) = 1$. Now, according to expression (1) above,

$$b(B) \geq b(G).$$

The belief in ballistic missiles is at least as great as the belief that the observed unit is a submarine of the G class since this class carries missiles. The belief in B may be greater than the belief in G; however, because other possible class assignments for the unit may also carry ballistic missiles.

There are two aspects of the second situation that should be noted. The expression $b(A \rightarrow B)$ may be uncertain for two different reasons. First,

[1]Examples used in this subsection have been fabricated using data currently reported in Jane's Fighting Ships.

"A" may often (but not always) imply "B." In this sense, there exists some probability for the occurrence of B given the occurrence of A. Second, it may be hypothesized that A implies B. Moreover, it may be that all observations to date have shown that B may be inferred from the presence of A. However, the testing of this hypothesis may not have been conclusive, and there exists some uncertainty that the hypothesis is true.

An example of the second situation is as follows:

> It has been estimated that 10 "Z" class submarines have been fitted with a new weapon system; the remaining 25 have not been modified. Given a positive identification of a unit as a Zulu, what is the likelihood that it carries the new weapon system?

Let Z denote "unit is a submarine of the 'Z' class"; let W denote "unit has the new weapon system," then

$$\begin{array}{c} b(Z \rightarrow W) = U \\ \underline{\quad b(Z) = 1 \quad} \\ b(W) = U \end{array}$$

If the contacted unit is considered equally likely to be any one of the 35 Zulus, then $b(W) = \frac{10}{35}$ or about 28%.

It is possible, of course, that both premises of the simple, plausible deductive argument may be uncertain. Consider the following example:

> It is reported that 10 "H" class submarines are currently operational. It is also believed that five of these boats are armed with the newer long-range ballistic missile found on both units of the "J" class. A submarine contact has been made and the unit is identified as being "H," "J" or possibly "N" class. What is the likelihood that the contacted unit carries the new missile (designated M)?

In this example, the belief that the contacted unit carries the new missile can be expressed as

$$b(M) = b(H)b(H \rightarrow M) + b(J)b(J \rightarrow M) + b(N)b(N \rightarrow M)$$

where

$b(H) + b(J) + b(N) = 1$

$b(H \rightarrow M) = U$; an unknown value between "0" and "1."

$b(J \rightarrow M) = 1$; all units believed to carry M.

$b(N \rightarrow M) = 0$; units do not carry missiles.

Thus

$$b(M) = b(H)b(H \rightarrow M) + b(J)$$

Derivation of a value for b(M) necessitates quantifying the degree of belief in "H," "H→M" and "J."[1] This may be accomplished by directly estimating values for these factors, or they may be derived by weighing other data pertinent to the problem. Previous operational characteristics of the classes plus a knowledge of the locations of some of the units, could, for example, influence the analyst's expression of b(H) and b(J).

B. BAYES THEOREM AND PROBABILISTIC INFORMATION PROCESSING

Bayes Theorem is a relatively old aspect of probability theory that has received renewed attention over the past six years as a method of revising

---

[1]It should be noted that while the "N" class units do not carry missiles, the factor b(N) does influence b(M). Since b(H) + b(J) = 1-b(N), a high degree of belief that the contact is "N" class lowers the belief that it is "H" or "J," hence, that it is armed with missiles.

opinion in the light of new evidence.[1] A fundamental aspect of this rule stems from the concept of conditional probabilities. A conditional probability simply refers to the probability that an event of a certain class will have a given outcome under the condition that it belongs to a specified subclass of the whole class. A conditional probability is an expression concerning one action or state of nature, assuming another action or state of nature exists. Thus, it is meaningful to talk about the probability of event B occurring, assuming event A has occurred; the conventional notation for this expression is $p\{B/A\}$; moreover, the conditional probability $p\{B/A\}$ is analogous to the plausible implication $b(A \rightarrow B)$. It will be shown that Bayes Theorem and the techniques of plausible implications are analagous; in fact, their main differences are found in the methods of application.

1. Expression of the Theorem

Placed in the context of intelligence analysis and synthesis, Bayes Theorem can be viewed as follows:

> Consider an intelligence problem being examined at some time T. Suppose that an analyst can postulate several mutually exclusive hypotheses that exhaust the possible solutions to this problem. It is quite possible that these alternative "solutions" may not be considered to be equally likely; but, instead are given some preference weighting according to the background knowledge prior to T.
>
> Pertinent data examined after time T will have some impact on these prior weightings. In general, the extent of the impact will be dependent upon the utility of the data in discriminating among the alternative possible solutions.

Let S_1, S_2, S_3, . . ., S_n denote mutually exclusive hypotheses that exhaust the possible solutions to an intelligence problem. The following definitions are made:

---

[1]This Theorem was stated by the Reverend Thomas Bayes and published posthumously in 1763. Appendix A of this paper presents a derivation of this Theorem.

a. $P\left\{S_i\right\}$ is a number between 0 and 1 representing the prior probability that the ith solution is true in light of the intelligence prior to time T.

b. D represents a new pertinent input datum examined after time T.

c. $P\left\{D/S_i\right\}$ is the likelihood of obtaining "D" if the analyst assumes solution S_i to be correct.

d. $P\left\{S_i/D\right\}$ represents the revised opinion about solution S_i after datum "D" has been integrated into the intelligence problem.

Bayes Theorem can be expressed as:

$$P\left\{S_i/D\right\} = \frac{P\left\{D/S_i\right\} \; P\left\{S_i\right\}}{P\left\{D\right\}} \, . \qquad (1)$$

The probability of obtaining the datum D is the total probability of obtaining D from all hypothesized solutions; i. e.,

$$P\left\{D\right\} = P\left\{D/S_1\right\} \; P\left\{S_1\right\} + \ldots + P\left\{D/S_n\right\} \; P\left\{S_n\right\} \, .$$

Thus, Bayes Theorem can be written

$$P\left\{S_i/D\right\} = \frac{P\left\{D/S_i\right\} \; P\left\{S_i\right\}}{P\left\{D/S_1\right\} \; P\left\{S_1\right\} + \ldots + P\left\{D/S_n\right\} \; P\left\{S_n\right\}} \cdot \qquad (2)$$

2. Discussion and Examples

The utility of D in discriminating among the hypotheses is reflected in the variations among the likelihoods of obtaining D assumming each solution to be true. If the value $P\left\{D/S_i\right\}$ is the same for all solutions, then D does not discriminate among the hypotheses; it is cancelled from the expression and

$P\{S_i/D\} = P\{S_i\}$ signifying that D does not influence the prior probability for S_i. On the other hand, if this value is zero for every case but $P\{D/S_i\}$, then

$$P\{S_i/D\} = \frac{P\{D/S_i\}\ P\{S_i\}}{P\{D/S_i\}\ P\{S_i\}} = 1$$

signifying that D designates a solution.

It should also be noted that the value of $P\{D/S_i\}$ does not reflect the relationship between an element and its characteristics; it reflects, instead, the likelihood of obtaining the data assumming the hypothesized solution to be true. This value, therefore, encompasses considerations of errors that may occur in collecting, processing and reporting information. Consider, for example, the following accuracy table for an hypothetical shape sensor:

	Sensor Output	
Real World	Square	Circle
Square	80%	20%
Circle	30%	70%

The sensor, observing a square in the real world, correctly identifies the shape as a square 80% of the time and erroneously identifies it as a circle 20% of the time. Observing a circle in the real world, the sensor is correct 70% of the time and in error 30% of the time. Let "circle" and "square" represent the sensor's output. The following conditional probabilities reflect the data presented in the above table:

$P\{D/S_i\}$

$$P\{\text{"circle"} \mid \text{CIRCLE}\} = .70$$

$$P\{\text{"circle"} \mid \text{SQUARE}\} = .20$$

$$P\{\text{"square"} \mid \text{CIRCLE}\} = .30$$

$$P\{\text{"square"} \mid \text{SQUARE}\} = .80$$

Another aspect of Bayes Theorem which is imporant to intelligence is that the application can be dynamic, facilitating the integration of a number of items into an assessment of a problem. Consider several items of information $D_1, D_2, \ldots, D_m$. The new probability for the ith solution after D_1 has been integrated is $P\{S_i/D_1\}$. This value is now the best estimate for S_i prior to the integration of other data (such as D_2) and can be substituted for $P\{S_i\}$ in Bayes Theorem. The impact of D_2 on this new assessment of S_i is:

$$P\{S_i/D_2\} = \frac{P\{D_2/S_i\} \quad P\{S_i/D_1\}}{P\{D_2/S_1\}\ P\{S_1/D_1\} + \ldots + P\{D_2/S_n\}\ P\{S_n/D_1\}}$$

This expression is known as the recursive version of Bayes Theorem and assumes that the data $D_1, D_2, \ldots, D_m$ are independent. Examples of this form will be presented in Chapter IV.

3. Related Works

Although Bayes Theorem as a model for intelligence processing has been advanced and utilized previously on other contracts at HRB-Singer, the stimulus which prompted its incorporation into an intelligence analysis methodology was research reported by Dr. Ward Edwards of the University of Michigan. Dr. Edwards, writing in 1963 on Command and Control Systems [Edwards, 1962; 1963] coined the acronym, PIP (for probabilistic information processor), a system which would combine human probability estimators with a Bayesian Processor. An example of one such system is shown in Figure 2. Prior to this, there had essentially been no elementary discussion highlighting the application of Bayesian statistics to military information processing. Dr. Edwards gives the following reasons why a PIP concept has merit:

a. The procedure is optimal for extracting as much certainty as possible from the available information.

b. The system can accept and use with profit information so seriously fallible or degraded that it would be excluded or ignored in deterministic systems.

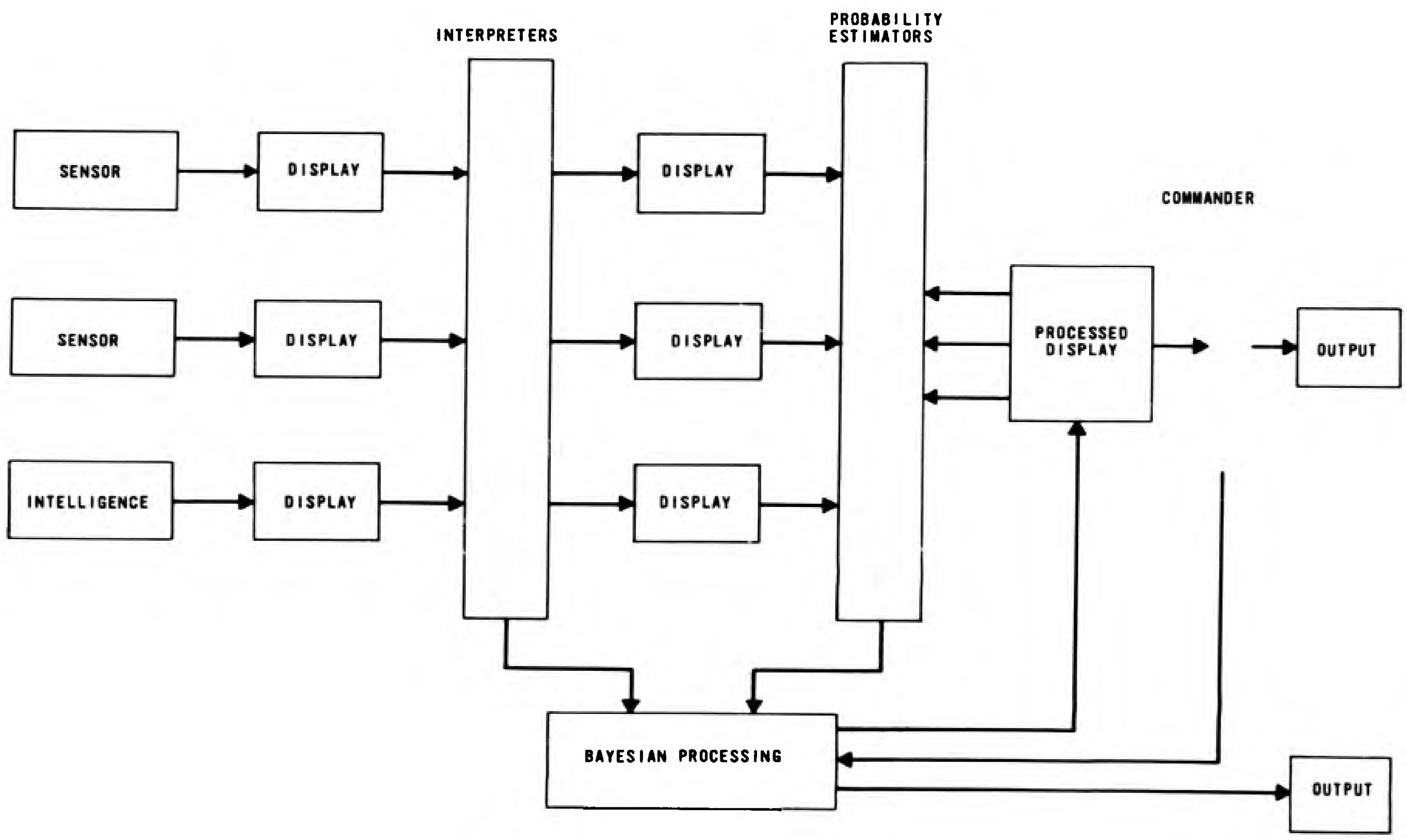


FIG. 2 PIP SYSTEM

c. The technique permits novel allocation of the elements of probabilistic inference between men and machines which may be better than that possible in deterministic systems.

However, two psychological propositions fundamental to the whole idea of PIP had to be established:

a. Men are effective transducers for probabilities.
b. Men are sufficiently inferior to machines at translating their probability estimates into conclusions about the truth of hypotheses.

Again writing in 1966 about his own research and that of others [Edward et al., 1966], including Peterson, Philips, and Hayes of the University of Michigan; Kaplan and Newman of SDC; Briggs, Schum, Goldstein and Southard of Ohio State University, Dr. Edwards states that two crucial findings of the Man-and-Bayes Theorem experimentation to date are:

a. Men are conservative estimators of posterior probabilities.
b. A PIP technique processes information more sufficiently than its competitors.

An expansion of these findings shows:

a. Conservatism in estimating posterior probabilities diminishes with experience, but never is superior to the Bayesian Processor. The suboptimal behavior may be the result of intellectual, not motivational deficiencies. Conservatism increases as the amount of revision required increases.
b. The likelihood probability estimates improved with experience. Some persons were consistently optimistic while others were consistently pessimistic, but all that is necessary is that the behavior in making the judgments be reasonably consistent. For many situations a man can estimate closely related numbers called likelihood ratios, equally acceptable to Bayes Theorem.
c. It may be possible to improve performance in multihypothesis situations by reducing the number of hypotheses under active consideration as rapidly as the data permit. In addition,

providing the probability estimators with the output of Bayes Theorem after each datum was a hindrance rather than an aid.

d. The people who make the probability judgments should not also be the final decision makers. The task of making likelihood judgments or some comparable response mode is enough for them to do, and they can be trained to be quite good at it.

Edwards and his colleagues at the University of Michigan, Schum and his colleagues at Ohio State University, and Newman and his colleagues at SDC are continuing to experiment.

Schum is looking at repeatable situations in which the set of possible observations is quite limited so that subjects can reasonably expect to accumulate relevant relative frequencies linking data with hypotheses. [Edwards et al., 1966].

Edwards is primarily concerned with vague verbal data and vague verbal hypotheses for which no hope of frequentistic information linking data with hypotheses exists. [Edwards et al., 1966].

Newman's research is related to those investigations of human problem solving or "thinking" in which the emphasis has been on the process rather than on the output or end product of a problem solving activity. [Newman et al., 1966].

In summary for this section, the current research into problem solving methodologies has been predicated upon two assumptions. The first is that man can filter input uncertainty excellently, and, in fact, can translate this uncertainty into appropriate indices. But man is not very good at integrating his filtered information into a cohesive output. The second is that machines cannot filter input uncertainty very well but are excellent at combining numerical indices into some formal output. Therefore, the marriage of man and machine should produce a powerful problem solving system. However, one question remaining unanswered is whether an intelligence report should be only the best estimate of a solution (a decision) presented as a "conclusion," or the actual list of potential solutions and their associated weights. The eminent statistician, Dr. J. W. Tukey, presents a very enlightening presentation of the differences between conclusions and decisions [Tukey, 1960]. Further discussions on this topic and a method for incorporating intelligence outputs into a decision-making process are presented in Appendix C.

IV. EXAMPLE APPLICATIONS OF BAYES THEOREM TO INTELLIGENCE PROBLEMS

If Bayes Theorem is to be used as the basis of a model for emulating the synthesis phase in an intelligence producing process, it is important to examine its application to different types of intelligence problems. Examples will be presented in the following sections for three types of intelligence inputs--qualitative characteristics, quantitative characteristics, and subjective indicators.

Qualitative characteristics may be represented by discrete random variables. Examples would include color, the presence or absence of physical items, monitored parameters, etc. Errors of observation or transformations are incorporated into the appropriate probability density by a weighting scheme.

Quantitative characteristics may be represented by continuous or approximately continuous random variables. Examples would include length, width, height, weight, velocity, location, etc. All associated errors are incorporated within the appropriate probability density.

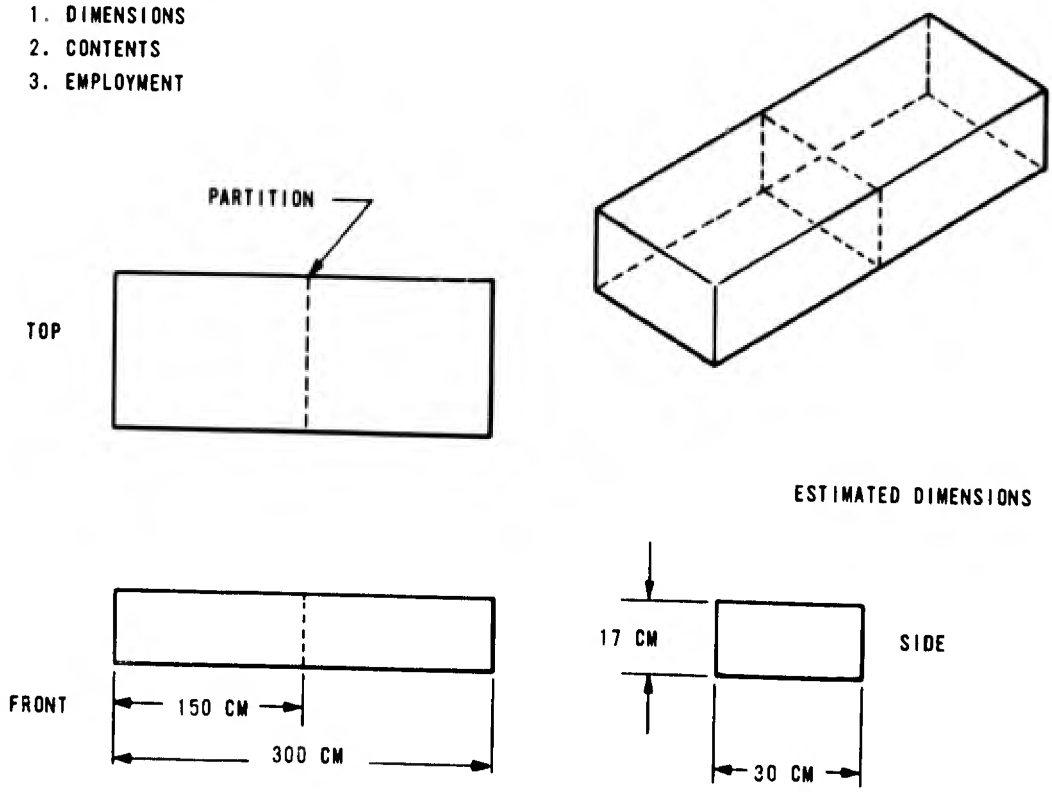
Subjective indicators have associations with members of the solution set which are very loosely defined as well as being common to innumerous other states of nature outside the solution space. All associated confidence is incorporated into an appropriate weighting. Data of this type, such as observation of key personnel at particular locations or allocation of combat units to geographic areas, are usually employed to sharpen the solution selection.

A. EXAMPLE OF QUALITATIVE CHARACTERISTICS SYNTHESIS

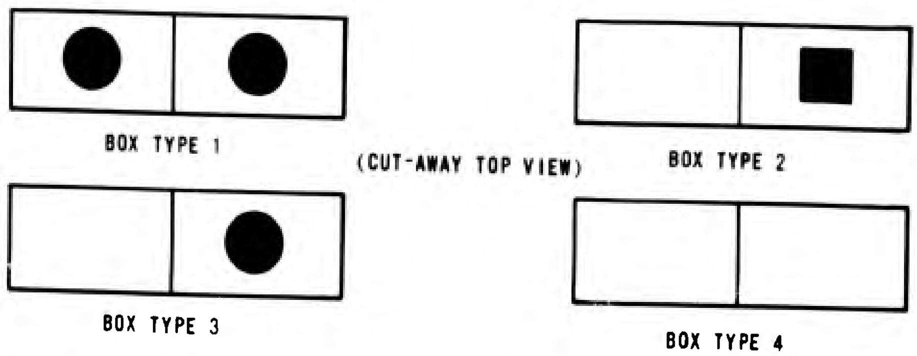
Suppose it has been noted that the enemy is employing a new "Black Box" component in the fire control subsystem of some of their tactical weapon systems. The scientific and technical intelligence efforts have concluded that, in fact, there are four distinctive types of boxes being utilized. The objective in this intelligence problem example is to identify which type of "Black Box" has been employed in the observed situation. In order to simplify the example, the term "Black Box" will be taken literally; i.e., the intelligence collection efforts will be directed towards the identification of a box having length, width, and height. Figure 3 illustrates the initial S & T intelligence description of the four box types.

DESCRIPTION
BLACK BOXES

1. DIMENSIONS
2. CONTENTS
3. EMPLOYMENT



FOUR TYPES OF BLACK BOXES HAVE BEEN EMPLOYED; EACH BOX CONTAINS TWO ELEMENTS SEPARATED BY A THIN PARTITION (POSSIBLY A HEAT SHIELD). THE DIFFERENT BOX TYPES CAN BE IDENTIFIED BY THEIR CONFIGURATION AS FOLLOWS.



THE BOXES ARE OPEN AT EACH END (POSSIBLY FOR VENTILATION) EXPOSING THE INTERNAL COMPONENTS. THERE ARE NO DISTINGUISHING EXTERIOR CONFIGURATIONS OR MARKS TO FACILITATE BOX TYPE ORIENTATION.

WHILE EACH BOX TYPE HAS BEEN OBSERVED IN PRODUCTION, NO RELATIONSHIP HAS BEEN ESTABLISHED BETWEEN THE TYPE OF BOX USED AND THE TYPE OF WEAPON SYSTEM EMPLOYED.

FIG. 3 DESCRIPTION OF BOX TYPES

Assume aerial reconnaissance detected a new construction site (designated V-3) and photographed a Black Box component with a visible element, but the color and the shape of the element could not be determined. Assume also that a second aerial reconnaissance, employing a combined sensor platform incorporating a highly reliable color sensor and a fairly reliable shape sensor overflew the V-3 site. The Black Box was again detected, and the visible element reported to be a "DARK SQUARE." Question; which Box Type was observed?

The initial assumptions and inputs to the problem are these:

1. The observed box is equally likely to be any of the four types.
2. The observed element is one of two elements in the box.
3. The color and shape sensors are detecting the same elements.
4. The report is <u>not</u> totally reliable.

The reliability of the color sensor is:

REAL WORLD	SENSOR OUTPUT: DARK	SENSOR OUTPUT: LIGHT
DARK	95%	5%
LIGHT	10%	90%

The table shows the discriminating capability of the sensor between the two colors. The data indicate that when the color sensor is actually viewing a DARK object, it will usually (95 times out of 100 total times) correctly report the object color as DARK. Sometimes (5 times out of 100 total times), however, the sensor will erroneously report that the object color is LIGHT. Similarly, when the color sensor is actually viewing a LIGHT object, it will accurately report the color as LIGHT 90% of the time, but erroneously report the color as DARK 10% of the time. The reliability of the shape sensor is:

REAL WORLD	SENSOR OUTPUT: SQUARE	SENSOR OUTPUT: CIRCLE
SQUARE	80%	20%
CIRCLE	30%	70%

The combined platform reliability is therefore

REAL WORLD	COMBINED SENSOR OUTPUT: DARK SQUARE	DARK CIRCLE	LIGHT SQUARE	LIGHT CIRCLE
DARK SQUARE	76%	19%	4%	1%
DARK CIRCLE	28.5%	66.5%	1.5%	3.5%
LIGHT SQUARE	8%	2%	72%	18%
LIGHT CIRCLE	3%	7%	27%	63%

The Bayesian formulation begins by assessing the prior probability of each solution, i. e., $P\left\{S_i\right\}$. Since one of the initial assumptions was that the observed box is equally likely to be any of the four types, the prior probabilities are:

$$P\left\{S_1\right\} = P\left\{S_2\right\} = P\left\{S_3\right\} = P\left\{S_4\right\} = \frac{1}{4}.$$

The solution subscripts refer to the corresponding Box Type.

The probability of reporting "DARK SQUARE" conditional on each Box Type (i. e., $P\left\{D/S_i\right\}$) considers both the array of elements among the four Box Types and the reliability of the collection system as depicted in the COMBINED SENSOR OUTPUT table. In general, the probability that the combined sensor platform will report "DARK SQUARE," (given one of the Box Types) is equal to the following summation:

$$P\left\{D/S_i\right\} = \sum_{i=1}^{4} P\left\{\text{Reporting "DARK SQUARE"/Object "j"}\right\}^1 \cdot P\left\{\text{Object "j"}/S_i\right\}$$

where the Object "j" can be DARK SQUARE, DARK CIRCLE, LIGHT SQUARE, or LIGHT CIRCLE.

---

[1]Actually this should be the conditional probability of Reporting "DARK SQUARE," given the Object "j" and Box Type "i"; i. e., $P\left\{\text{Reporting "DARK SQUARE"/Object "j"} \cap S_i\right\}$, but since the addition of Box Type "i" does not influence the value of the probability, the expression may be written and interperted as shown.

Let "RDS" represent "Reporting DARK Square," then the probability of obtaining a report of "DARK SQUARE" if the combined sensor platform is viewing a component from Box Type 1 is:

$$P\{D/S_1\} = P\{\text{RDS/DARK SQUARE}\} \cdot P\{\text{DARK SQUARE/BOX TYPE 1}\} + P\{\text{RDS/DARK CIRCLE}\} \cdot P\{\text{DARK CIRCLE/BOX TYPE 1}\} + P\{\text{RDS/LIGHT SQUARE}\} \cdot P\{\text{LIGHT SQUARE/BOX TYPE 1}\} + P\{\text{RDS/LIGHT CIRCLE}\} \cdot P\{\text{LIGHT CIRCLE/BOX TYPE 1}\},$$

or $P\{D/S_1\} = (76\%)\ (0) + (28.5\%)\ (1) + (8\%)\ (0) + (3\%)\ (0) = 28.5\%$.

Similarly,

$$P\{D/S_2\} = (76\%)\ (\tfrac{1}{2}) + (28.5\%)\ (0) + (8\%)\ (0) + (3\%)\ (\tfrac{1}{2}) = 39.5\%$$
$$P\{D/S_3\} = (76\%)\ (0) + (28.5\%)\ (\tfrac{1}{2}) + (8\%)\ (\tfrac{1}{2}) + (3\%)\ (0) = 18.25\%$$
$$P\{D/S_4\} = (76\%)\ (0) + (28.5\%)\ (0) + (8\%)\ (1) + (3\%)\ (0) = 8.0\%$$

Applying Bayes Theorem to determine the probability that the unidentified box is Box Type 1 yields:

$$P\{S_1/D\} = \frac{P\{D/S_1\} \cdot P\{S_1\}}{\sum_1^4 P\{D/S_i\} \cdot \{P\ S_i\}}$$

$$P\{S_1/D\} = (28.5\%)\ (\tfrac{1}{4}) / \left[(28.5\%)\ (\tfrac{1}{4}) + (39.5\%)\ (\tfrac{1}{4}) + (18.25\%)\ (\tfrac{1}{4}) + (8.0\%)\ (\tfrac{1}{4})\right]$$

$$P\{S_1/D\} = 7.125/23.5625 = 30.24\%.$$

Similarly,

$$P\{S_2/D\} = 9.875/23.5625 = 41.91\%$$
$$P\{S_3/D\} = 4.5625/23.5625 = 19.36\%$$
$$P\{S_4/D\} = 2.00/23.5625 = 8.49\%$$

Because of the uncertainty attached to the input from the combined sensor platform, the data do not discriminate greatly among the four hypotheses, e.g.,

	Prior Probability	$P(S_i/D)$
S_1	25%	30.24%
S_2	25%	41.91%
S_3	25%	19.36%
S_4	25%	8.49%

Had the data been 100% reliable, the report of "DARK SQUARE" would have yielded S_2 as the solution.

B. EXAMPLES OF QUANTITATIVE CHARACTERISTICS SYNTHESIS

1. Discrete Prior and Posterior Distribution

This example uses Bayes Theorem to discriminate among states of nature on the basis of quantitative observable characteristics whose errors are assumed to be normally distributed. The technique employed here involves tempering the a priori odds with an expression of likelihood to obtain a posteriori odds.

Suppose it is known that there is an enemy submarine in a given area, and that there are four types of enemy subs. The problem, then, is to classify the sub as either type #1, #2, #3, or #4.

Assume that available intelligence information is summarized in the following table.

Type of Sub	No. of Subs of this Type	Length	Width	Height
#1	20	300	30	17
#2	40	275	28	22
#3	60	305	32	19
#4	80	295	27	19

Now, in terms of the discussion above, the possible solutions are S_1, S_2, S_3, and S_4, where S_i = Submarine of Type i. The a priori probabilities are given as

$$P\left\{S_1\right\} = \frac{20}{200} = .1$$

$$P\left\{S_2\right\} = \frac{40}{200} = .2$$

$$P\left\{S_3\right\} = \frac{60}{200} = .3$$

$$P\left\{S_4\right\} = \frac{80}{200} = .4$$

Also, there is a set of characteristics $c_{1i} \equiv$ length of type #i sub, $c_{2i} \equiv$ width of type #i sub, and $c_{3i} \equiv$ height of type #i sub.

Assume now that there are two observers (sensors) O_1 and O_2 who report observations $D_1 = \left\{d_{11}, d_{21}, d_{31}\right\} = \left\{295, 32, 18\right\}$ and $D_2 = \left\{d_{12}, d_{22}, d_{32}\right\} = \left\{300, 29, 20\right\}$, respectively. Further assume that the accuracy of these observations is reflected by the following assignment[1] of σ_{kt}^2 : $\sigma_{11}^2 = 25$, $\sigma_{21}^2 = 4$, $\sigma_{31}^2 = 1$, $\sigma_{12}^2 = 100$, $\sigma_{22}^2 = 16$, and $\sigma_{32}^2 = 4$.

Thus,

$$f(\underline{D}/S_1) = (2\pi)^{-3}(800)^{-1} e^{-\frac{1}{2}\left[\frac{(295-300)^2}{25} + \frac{(32-30)^2}{4} + \frac{(18-17)^2}{1} + \frac{(300-300)^2}{100} + \frac{(29-30)^2}{16} + \frac{(20-17)^2}{4}\right]}$$

$$= (2\pi)^{-3}(800)^{-1} e^{-2:6562}$$

$$= (2\pi)^{-3}(800)^{-1} (.07021)$$

$$f(\underline{D}/S_2) = (2\pi)^{-3}(800)^{-1} (.00000)$$

---

[1]The assignment of values to the σ_{kt}^2's is not a set procedure. One way of assigning a value to σ_{kt}^2 is to construct (subjectively) the point $x_\mu \ni \text{prob}(x_{kt} > x_\mu / c_{ki}) = .05$. Thus, if $c_{ki} = 300$ and $x_\mu = 310$, $\sigma_{kt} = \frac{10}{1.645} = 6.08$, and hence $\sigma_{kt}^2 = 36.97$.

$f(\underline{D}/S_3) = (2\pi)^{-3}(800)^{-1}(.03317)$

$f(\underline{D}/S_4) = (2\pi)^{-3}(800)^{-1}(.00865)$

Combining these values with the prior probabilities yields

$$
\begin{aligned}
P\left\{S_1\right\} f(\underline{D}/S_1) &= (2\pi)^{-3}(800)^{-1}(.007021) \\
P\left\{S_2\right\} f(\underline{D}/S_2) &= (2\pi)^{-3}(800)^{-1}(.000000) \\
P\left\{S_3\right\} f(\underline{D}/S_3) &= (2\pi)^{-3}(800)^{-1}(.009951) \\
P\left\{S_4\right\} f(\underline{D}/S_4) &= (2\pi)^{-3}(800)^{-1}(.003460)
\end{aligned}
$$

and hence,

$$\sum_{1}^{4} P\left\{S_i\right\} f(\underline{D}/S_i) = (2\pi)^{-3}(800)^{-1}(.020432) .$$

Therefore, the posterior distribution is

$$
\begin{aligned}
P^*\left\{S_1\right\} &= \frac{.007021}{.020432} = .344 \\
P^*\left\{S_2\right\} &= \frac{.000000}{.020432} = .000 \\
P^*\left\{S_3\right\} &= \frac{.009951}{.020432} = .487 \\
P^*\left\{S_4\right\} &= \frac{.003460}{.020432} = .169 .
\end{aligned}
$$

For comparison purposes, the following table of prior and posterior probabilities is given.

	Probability	
Type of Sub	Prior	Posterior
#1	.1	.344
#2	.2	.000
#3	.3	.487
#4	.4	.169

This example used a discrete prior, and, of course, the final distribution was also discrete. Although this type of problem is fairly common, many cases arise where the prior and posterior distributions are continuous.

2. Continuous Prior and Posterior Distribution

This example uses Bayes Theorem for updating a continuous probability distribution in the light of new data. How this differs from problems where these distributions are discrete will be evident in the example.

Suppose an object is contacted (e.g., a submarine) and it is desired to determine its length, l_o. It can be postulated that the length was somewhat around 100; that there was only a 1% chance that this length might be less than 50, and likewise, that there was only a 1% chance that this length might be greater than 150.

Assuming normality (which is often a good approximation), the prior feeling about l_o can be represented by a normal distribution with mean l_o^* and variance σ^2, where l_o^* and σ^2 may be determined by solving the following equations:

$$P\left\{l_o < 50\right\} = P\left\{l_o > 150\right\} = .01 .$$

This is equivalent to

$$P\left\{Z < \frac{50-l_o^*}{\sigma}\right\} = P\left\{Z > \frac{150-l_o^*}{\sigma}\right\} = .01$$

where Z is a random variable having a standard normal distribution which is well-tabulated and can be found in almost any statistics textbook. Resorting to the tables,

$$\frac{l_o^* - 50}{\sigma} = \frac{150 - l_o^*}{\sigma} = 2.33,$$

and solving for l_o^* and σ, or $l_o^* = 100$ and $\sigma = 21.5$. Thus, the prior distribution of the length, l_o, is

$$P\left\{l_o\right\} = \frac{1}{(21.5)\sqrt{2\pi}} e^{-\frac{1}{2(21.5)^2}(l_o - 100)^2} .$$

Suppose that a sensor is employed to measure the object. As is usually the case in practical applications, the sensor will not be perfect. In fact, assume that the length reported by the sensor is accompanied with error which is normally distributed with mean 0, and variance 9. (These numbers are chosen arbitrarily for illustrative purposes; however, for an actual application, the distribution of the sensor error can be determined by running the sensor under test conditions.)

If an observation $\underline{x}$ is reported, an analyst can determine the posterior distribution of l_o by applying Bayes Theorem:

$$P\left\{l_o/x\right\} = \frac{P\left\{x/l_o\right\} \quad P\left\{l_o\right\}}{\int P\left\{x/l_o\right\} \quad P\left\{l_o\right\} dl_o} .$$

For this example, the prior distribution $P(l_o)$ is given above, while

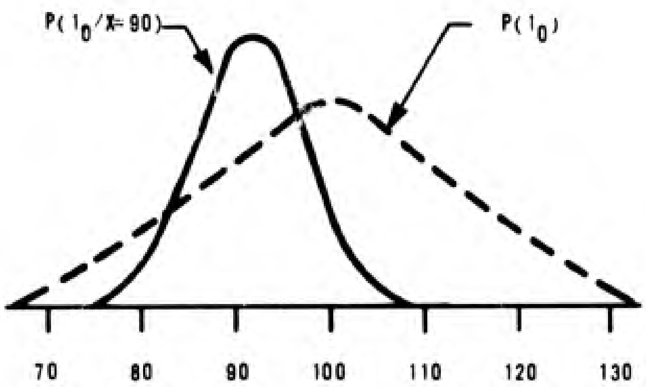
$$P\left\{x/l_o\right\} = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{18}(x - l_o)^2} .$$

Now, suppose the sensor reports a length of $x = 90$. The resulting posterior distribution is, then,

$$P\left\{l_o/x = 90\right\} = \frac{P\left\{x=90/l_o\right\}P\left\{l_o\right\}}{\int P\left\{x=90/l_o\right\}P\left\{l_o\right\}dl_o}$$

$$= \frac{\frac{1}{3\sqrt{2\pi}}e^{-\frac{1}{18}(90-l_o)^2}\frac{1}{(21.5)\sqrt{2\pi}}e^{-\frac{1}{2(21.5)^2}(l_o-100)^2}}{\int_{-\infty}^{\infty}\frac{1}{3\sqrt{2\pi}}e^{-\frac{1}{18}(90-l_o)^2}\frac{1}{(21.5)\sqrt{2\pi}}e^{-\frac{1}{2(21.5)^2}(l_o-100)^2}dl_o}$$

$$= \frac{1}{\sqrt{2\pi\,(8.8)}}e^{-\frac{1}{2(8.8)}(l_o-90.2)^2}$$



That is, based on the observation of $x = 90$, the resulting posterior distribution over l_o is normal with mean 90.2 and variance 8.8. The diagram indicates the prior and posterior distributions over the length l_o.

Of course, if more observations are made on the object, the posterior may be calculated by means of the recursive version of Bayes Theorem.

C. EXAMPLE OF SUBJECTIVE INDICATORS SYNTHESIS

This last example represents an application of discrete probabilities for discriminating among the states of nature on the basis of subjective indicators. The problem inputs are weighted, using average values from the Sherman Kent Chart (Figure 4) as a technique for expressing "degrees of belief."

AVERAGE VALUE	CHANGES FOR - AGAINST		
100	100 - 0	CERTAINTY	NO ESTIMATE
90	99 - 1	ALMOST CERTAIN. HIGHLY LIKELY.	
	85 - 15		
70	84 - 16	PROBABLE. LIKELY. PROBABLY. WE BELIEVE.......	
	55 - 45		
50	51 - 49	CHANCES JUST BETTER THAN EVEN. ON BALANCE......	
	50 - 50	CHANCES ARE EVEN.	
	49 - 51	CHANCES JUST LESS THAN EVEN	
30	45 - 55	IT IS DOUBTFUL. WE DOUBT. IMPROBABLE. UNLIKELY. PROBABLY NOT.	
	16 - 84		
10	15 - 85	ALMOST CERTAINLY NOT. HIGHLY UNLIKELY. CHANCES ARE SLIGHT.	
	1 - 99		
0	0 - 100	IMPOSSIBILITY	NO ESTIMATE

FIG. 4 SHERMAN KENT CHART

Suppose that on the basis of information received to date, the analyst knows that some type of construction is going on in enemy territory at location L. Using this information, the analyst may establish hypotheses as to the possible types of construction which might be going on at location L. Suppose that the analyst uses average values of the Sherman Kent Chart to express his views about the construction, e.g.,

1. Likely Airfield Construction
2. Likely Submarine Pen Construction
3. Slight Chance of Radar Site Construction
4. Slight Chance of Missile Site Construction
5. Other types of construction impossible.

Thus, the analyst is left with four alternative hypotheses, namely

S_1: Airfield Construction (weighted as .70)

S_2: Submarine Pen Construction (.70)

S_3: Radar Site Construction (.10)

S_4: Missile Site Construction (.10)

The weight (W_i) for each hypotheses can be converted to values expressing a degree of belief, b (S_i), by normalizing so that

$$\sum_{i=1}^{4} b\,(S_i) = 1.$$

This can be accomplished by defining the degree of belief in each hypothesis as

$$b\,(S_i) = \frac{W_i}{\sum_{i=1}^{4} W_i} .$$

Thus, the degree of belief in Airfield Construction is expressed as

$$b\,(S_1) = \frac{.70}{.70 + .70 + .10 + .10} = .438.$$

Similarly,

$$b\,(S_2) = .438$$

$$b\,(S_3) = .062$$

$$b\,(S_4) = .062.$$

The analyst requests more data on location L, and in response to his request receives the following data:

Within the past month at location L:

D_1: Prof. X has made at least two trips to location L. (Prof. X has some background in many areas of civil enginneering; however, he has extensive experience in the construction of submarine pens).

D_2: Naval Supply Company arrived at L.

D_3: 400 Laborers arrived at L.

D_4: 125th Heavy Equipment Batt. arrived at L.

D_5: Large Steel Shipment arrived at L.

D_6: 16th Finance Office sent a 25,000 Ruble Bank Draft to location L.

At this point, the analyst will not be able to use D_3, D_4, D_5, or D_6 to discriminate among S_1, S_2, S_3, and S_4 and he may start a search to see if additional information on D_3, D_4, D_5, or D_6 may be uncovered.

With respect to D_1, however, the analyst (using the Sherman Kent Chart) may evaluate

D_1/S_1	chances about even	$b(D_1/S_1)$	$= .5$
D_1/S_2	highly likely	$b(D_1/S_2)$	$= .9$
D_1/S_3	chances about even	$b(D_1/S_3)$	$= .5$
D_1/S_4	chances about even	$b(D_1/S_4)$	$= .5.$

Also, for D_2, the analyst may evaluate

D_2/S_1	likely	$b(D_2/S_1)$	$= .7$
D_2/S_2	certain	$b(D_2/S_2)$	$= .1$
D_2/S_3	likely	$b(D_2/S_3)$	$= .7$
D_2/S_4	likely	$b(D_2/S_4)$	$= .7.$

At this point in his analysis, the weightings of the four alternative hypotheses would be:

$$\left.\begin{array}{ll} b(S_1/D_1D_2) \propto b(D_1/S_1)\, b(D_2/S_1)\, b(S_1) \propto = .35(.438) & = .1526 \\ b(S_2/D_1D_2) \propto = .90(.438) & = .3924 \\ b(S_3/D_1D_2) \propto = .35(.062) & = .0217 \\ b(S_4/D_1D_2) \propto = .35(.062) & = \underline{.0217} \\ & \;\; .5884 \end{array}\right\} \begin{array}{ll} b(S_1/D_1D_2) & = .259 \\ b(S_2/D_1D_2) & = .667 \\ b(S_3/D_1D_2) & = .037 \\ b(S_4/D_1D_2) & = .037. \end{array}$$

Now suppose that the search for further information on D_3, D_4, D_5, or D_6 turns up the information that:

> The 16th Finance Office has to date (as far as is known) financed no missile sites, no radar sites, but has financed construction of some airfields and all sub pens.

Thus, the data D_3, D_4, D_5 still do not discriminate, but the data D_6 now may be evaluated by the analyst as follows:

D_6/S_1	chances even	$b(D_6/S_1) = .5$
D_6/S_2	almost certain	$b(D_6/S_2) = .9$
D_6/S_3	unlikely	$b(D_6/S_3) = .3$
D_6/S_4	unlikely	$b(D_6/S_4) = .3.$

Hence, finally

$$\left.\begin{array}{ll} b(S_1/D_1, D_2, D_3, D_4, D_5, D_6) \propto b(D_6/S_1)\, b(S_1/D_1D_2) = .5(.259) & = .1295 \\ b(S_2/\underline{D}) \propto \quad .9(.667) & = .6003 \\ b(S_3/\underline{D}) \propto \quad .3(.037) & = .0111 \\ b(S_4/\underline{D}) \propto \quad .3(.037) & = \underline{.0111} \\ & \;\; .7520 \end{array}\right\} \begin{array}{ll} b(S_1/\underline{D}) & = .172 \\ b(S_2/\underline{D}) & = .798 \\ b(S_3/\underline{D}) & = .015 \\ b(S_4/\underline{D}) & = .015 \end{array}$$

If this is all the data the analyst can obtain, his best report would be his probability distribution over S_1, S_2, S_3 and S_4, or at least a report using the Sherman-Kent Chart. That is;

$$\begin{array}{l}\text{Certain}\\\text{Construction}\end{array}\begin{cases}\text{Probable Sub Pen Construction}\\\text{Doubtful Airfield Construction}\\\text{Almost certainly not Radar Site Construction}\\\text{Almost certainly not Missile Site Construction}\end{cases}$$

V. PROBLEMS HINDERING APPLICATION OF BAYES THEOREM IN INTELLIGENCE

There are at least three significant problem areas that hinder application of Bayesian synthesis techniques in intelligence production efforts. These problem areas center upon the following:

1. Establishment of meaningful hypotheses.
2. Derivation of values for the expressions of belief or probability.
3. Dependency relationships among the data.

The first problem listed is not unique to a Bayesian approach to problem solving; it exists as a difficulty in present-day production efforts. The second and third problems, although they explicitly confront mathematical approaches to analyses and syntheses, are frequently implicit in the difficulty compounding non-numerical approaches.

A. ESTABLISHMENT OF HYPOTHESES

In problem solving efforts, the role of postulated hypotheses is to relate and explain data that have been assembled and are pertinent to the problem at hand. The assessment of subsequent data may indicate that some of these hypotheses (a) should be revised, (b) can be discarded; or it may occur that (c) new hypotheses should be formed. A Bayesian approach to problem solving imposes some restrictions on the establishment of the hypothesized solutions. These restrictions are:

1. The hypotheses must be mutually exclusive; i. e., only one of these can be correct.
2. The hypotheses must be exhaustive; i. e., at least one of these must be correct.

Additionally (as in all problem-solving efforts), the hypotheses should be non-trivial, and the analyst should be able to evaluate his data with respect to each possible solution.[1]

Theoretically, it is very simple to satisfy the two restrictions; e.g., the analyst can postulate that

$$S_1 = \text{the hypothesis is correct}$$

$$S_2 = \text{the hypothesis is not correct.}$$

Unfortunately, however, this dichotomous postulation complicates the process of data evaluation to the point where it is frequently the case that these assessments cannot be made. A good example of this problem is the difficulty in classifying a submerged contact as a "sub" or as a "non-sub." The hypothesis "non-sub" encompasses objects such as whales, schools of fish, wrecks, etc. This diversity of the population makes it difficult to assess the conditional probability

$$P\left\{D_i \mid \text{NON-SUB}\right\}$$

where D_i can be the interpreted output of various systems such as MAD; Jezebel, Julie, etc. The likelihood of a sub-like return from MAD given a metal wreck is not the same as the likelihood given a school of fish. Deriving a single, meaningful value for this conditional probability maybe impossible.

One method of alleviating the problem of weighing data with respect to a complex hypothesis is to partition the hypothesis into a complete set of mutually exclusive sub-hypotheses, e.g.,

$$(\text{NON-SUB}) = (\text{whales}) \text{ or } (\text{schools of fish}) \text{ or } \ldots$$

---

[1]It should be noted that the failure to include a possible solution in a set of hypotheses can bias the impact of the data in the synthesis process. In a Bayesian approach, however, should a new plausible hypothesis come to light <u>after</u> a quantity of data have been integrated, the mechanical nature of the process enables the analyst to efficiently reassess the problem without re-weighting the collected data with respect to the original hypotheses.

There are two problems that exist in this approach. First, there is the practical problem of having too many hypotheses. Since all data must be evaluated with respect to each possible solution, a problem having 25 items and 50 hypothesized solutions would require the derivation of 1,250 conditional probabilities and 50 prior probabilities. Second, there still exists the problem of completeness; i.e., there may exist a possible solution that has escaped attention. One method of completing the set is to define a catch-all hypotheses such as "the solution is something other than what has been explicitly postulated." Although this does exhaust the possible solutions, it is difficult to derive meaningful values for the probability of obtaining the data given this hypothesis to be true.

These problems exist in all approaches to problem solving; they are not unique to the Bayesian approach. The extent that these difficulties hinder useful application of this mathematical synthesis technique may well be dependent upon the type of intelligence problems that are to be solved. Basic research is continuing to be directed at methodological problems that arise from improper establishment of hypotheses; however, test applications to real or realistic problems of intelligence will most likely provide the greatest insight to the limitations imposed by this difficulty.

B. DERIVATION OF NUMERICAL VALUES

The application of Bayes Theorem (or the techniques of plausible inference) requires the numerical quantification of the probabilities of both the prior situations and the relationships between input data and the potential solutions. In some circumstances, observational error may be derived by appropriate testing of the sensor and can be used to determine the likelihood of observing a characteristic under different circumstances. Combining such results with the real-world characteristic associations should produce a reasonable "physical probability."[1] In other instances, source reliability and data accuracy will be totally dependent upon human judgment or a "psychological probability." A key question, then, is can the intelligence analyst express his degree of belief as a measurable probability?

---

[1]Good (1965, p. 6) discusses several types of probabilities, e.g., (1) physical--an intrinsic property of the material world, (2) psychological--a degree of belief and (3) subjective--consistent psychological probabilities.

In the preface to his monograph [1965], Good notes that

> The estimation of probabilities is of practical and philosophical interest, . . . The difficulties become clear when it is realized that we estimate probabilities every minute of the day, at least implicitly, and that how we do this is unknown. When this problem is solved, a potential pathway to artificial intelligence will be cleared, apart from easier applications, such as to character recognition and medical diagnosis.

Additionally many of the studies presented in the bibliography of this paper indicate that man can effectively produce reasonably consistent quantifications of his judgments as "subjective probabilities." There are still, however, some basic questions concerning the meaning and usage of these numerical expressions that must be explored. "Although controversy can be avoided in the mathematics of probability, it is unavoidable when the mathematics is applied to the outside world [Good, 1965, p. 9]. Is the value ".99," for example, sufficiently close to "1.00" to accept such a weighted output as certain? Does one analyst's expression of belief as ".85" have an equivalent meaning to another analyst? If an analyst cannot derive a meaningful value for an item how should the item be treated?

> My own view, following Keynes and Koopman, is that judgments of probability inequalities are possible but not judgments of exact probabilities; therefore, a Bayesian should have upper and lower betting probabilities [Good, 1965, p. 5].

The derivation of prior probabilities for the hypothesized solutions of a problem offer special challenge to the analyst. Past history, which often provides patterns useful in estimating prior probabilities, may not always serve as a model of today's events. In estimate intelligence, for example, the hypotheses may be predictions. Can techniques be developed to establish prior probabilities, for events that essentially have no history (e.g., determining the capabilities of a proposed weapon system)?

> Circumstances are always changing, but humans have the facility of estimating the probabilities of many events that have never previously occurred. Their sample is their past experience, and they must often make predictions without the benefit of an increased sample ...
>
> Nevertheless, for the purpose of making decisions, we do manage to make approximate estimates of probabilities. How this is done is an interesting problem in psychology and in neurophysiology. It might, for example, be conjectured that neural circuits automatically use a maximum entropy estimation (see Chapter 9). The problem of estimating probabilities of events that have never occurred is philosophically interesting, and in my opinion, likely to be important for the design of ultraintelligence machines [Good, 1965, p. 4].

It is reasonable to expect that a Bayesian approach to data synthesis may not enjoy equal utility for all areas and types of intelligence. In fact, one possible indication of utility might rest in some measure of difficulty one has in deriving meaningful probability values for the hypotheses and the data.

Although there are some problems in expressing opinion as a precise or consistent numerical value, there are possible benefits that may overshadow these difficulties in utilizing Bayes Theorem as a method of data synthesis. One major benefit is that the method explicitly connects the evaluations of the inputs with a degree of confidence that can be placed upon the output solutions. This connection enables an analyst to re-examine his rationale underlying a conclusion drawn from the data. Moreover, if at a later date, other evidences are uncovered; these can be readily integrated with the existing material to provide a new assessment of the problem.

C. DATA DEPENDENCIES

In the theory and examples discussed up to this point, it has been explicitly assumed that the data are independent; that is, the observation of one item of data does not alter the likelihood of observing another item of data. If this is not true, the data are said to be dependent. For example, if a card is drawn from an ordinary deck of playing cards and is observed to be a red face card, the likelihood that it is the queen of spades decreases to zero. That is, the data "red" and "spade" are dependent.

In mathematical notation, two pieces of data D_1 and D_2 are said to be independent in the light of a hypothesis (or solution) S if

$$P\left\{D_1, D_2/S\right\} = P\left\{D_1/S\right\} P\left\{D_2/S\right\}.$$

Otherwise, D_1 and D_2 are said to be dependent.

In general, if there are n possible solutions $S_1, \ldots, S_n$, Bayes Theorem states that

$$P\left\{S_j/D_1, D_2\right\} = \frac{P\left\{D_1, D_2/S_j\right\} P\left\{S_j\right\}}{\sum_1^n P\left\{D_1, D_2/S_i\right\} P\left\{S_i\right\}}$$

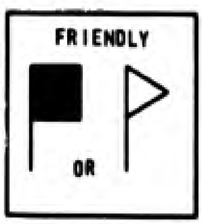
If D_1 and D_2 are independent, Bayes Theorem may be written in recursive form as

$$P\left\{S_j/D_1, D_2\right\} = \frac{P\left\{D_2/S_j\right\} P\left\{S_j/D_1\right\}}{\sum_1^n P\left\{D_2/S_i\right\} P\left\{S_i/D_1\right\}}$$

where

$$P\left\{S_j/D_1\right\} = \frac{P\left\{D_1/S_j\right\} P\left\{S_j\right\}}{\sum_1^n P\left\{D_1/S_i\right\} P\left\{S_i\right\}}$$

However, if the data are not independent, the use of the recursive version of Bayes Theorem may lead to erroneous conclusions.



For example, suppose in an exercise "enemy" ships fly a black triangular flag half of the time and a white square flag the rest of the time, while "friendly" ships fly a black square flag half of the time and a white triangular flag the remainder of the time.

Assume that an unidentified ship is sighted, and it is desired that the ship be classified as friend or foe.

Let D_1 denote the presence of a white flag

D_2 denote the presence of a triangular flag

S_1 denote that the unidentified ship is an enemy ship

S_2 denote that the unidentified ship is a friendly ship.

Thus, the appropriate probabilities are

$$P\left\{D_1/S_1\right\} = P\left\{D_1/S_2\right\} = .5$$
$$P\left\{D_2/S_1\right\} = P\left\{D_2/S_2\right\} = .5.$$

However,

$$P\left\{D_1, D_2/S_1\right\} = 0 \neq P\left\{D_1/S_1\right\} P\left\{D_2/S_1\right\}$$
$$P\left\{D_1, D_2/S_2\right\} = 1 \neq P\left\{D_1/S_2\right\} P\left\{D_2/S_2.\right\}$$

Thus, D_1 and D_2 are dependent.

Assume that it is thought equally likely, a priori, that the unidentified ship is friend or foe. That is, $P\left\{S_1\right\} = P\left\{S_2\right\} = .5$. If the recursive version of Bayes Theorem is used, the resulting posterior probabilities would be calculated as follows:

$$P\left\{S_1/D_1\right\} = \frac{P\left\{D_1/S_1\right\} P\left\{S_1\right\}}{P\left\{D_1/S_1\right\} P\left\{S_1\right\} + P\left\{D_1/S_2\right\} P\left\{S_2\right\}} = \frac{(.5)\ (.5)}{(.5)(.5) + (.5)(.5)} = .5$$

$$P\left\{S_2/D_1\right\} = \frac{P\left\{D_1/S_2\right\} P\left\{S_2\right\}}{P\left\{D_1/S_1\right\} P\left\{S_1\right\} + P\left\{D_1/S_2\right\} P\left\{S_2\right\}} = \frac{(.5)\ (.5)}{(.5)(.5) + (.5)(.5)} = .5$$

and hence,

$$P\left\{S_1/D_1, D_2\right\} = \frac{P\left\{D_2/S_1\right\} P\left\{S_1/D_1\right\}}{P\left\{D_2/S_1\right\} P\left\{S_1/D_1\right\} + P\left\{D_2/S_2\right\} P\left\{S_2/D_1\right\}} = \frac{(.5)\ (.5)}{(.5)(.5) + (.5)(.5)} = .5$$

$$P\{S_2/D_1, D_2\} = \frac{P\{D_2/S_2\}\, P\{S_2/D_1\}}{P\{D_2/S_1\}\, P\{S_1/D_1\} + P\{D_2/S_2\}\, P\{S_2/D_1\}} = \frac{(.5)\ (.5)}{(.5)\ (.5) + (.5)\ (.5)} = .5.$$

That is, based on the data D_1 and D_2, it appears that nothing was gained, since the posterior probabilities are the same as the prior probabilities.

This is an incorrect conclusion, of course, since it is obvious that the data D_1 and D_2 may arise only if the unidentified ship is a friendly one. Thus, S_2 must be the true solution. The nonrecursive form of Bayes Theorem would have provided the correct answer since

$$P\{S_1/D_1, D_2\} = \frac{P\{D_1, D_2/S_1\}\, P\{S_1\}}{P\{D_1, D_2/S_1\}\, P\{S_1\} + P\{D_1, D_2/S_2\}\, P\{S_2\}} = \frac{(0)\ (.5)}{(0)\ (.5) + (1)\ (.5)} = 0$$

$$P\{S_2/D_1, D_2\} = \frac{P\{D_1, D_2/S_2\}\, P\{S_2\}}{P\{D_1, D_2/S_1\}\, P\{S_1\} + P\{D_1, D_2/S_2\}\, P\{S_2\}} = \frac{(1)\ (.5)}{(0)\ (.5) + (1)\ (.5)} = 1.$$

Of course, this example is very elementary, and the dependencies may be handled by tabulation of the joint conditional probabilities. However, for larger and more realistic problems, determining the existence of dependencies and effectively dealing with them may be a difficult task, and will usually be resolved by using human judgments. In general, a mathematical model which provides exhaustive consideration of all data dependencies may imply an extremely large investment of computer time and of the analyst's time and labor (especially if subjective estimates are used) with perhaps only a marginal increase in accuracy. Instead, it might be advisable to balance any investment in overcoming dependencies with the corresponding expected gain in accuracy. It is quite possible that erroneously assuming independence may provide a workable model which provides an approximation good enough for most practical applications.

VI. SUMMARY

A basic contention expressed in this report is that a formal synthesis methodology could improve intelligence production efforts. Formal logic, the methodology of sound reasoning, can serve as a guide for intelligence processing, but is limited as an aid to the analyst because of the uncertainty of the data and the relationships that are used in intelligence problem solving. Plausible reasoning, on the other hand, can be used as a vehicle for mapping the effects of this uncertainty onto the conclusions derived from the assembled data. The techniques of plausible implication are introduced as a method of quantifying conclusions derived by deductive inference. Bayes Theorem is presented as a practical dynamic method of synthesizing probabilistic information pertinent to the hypotheses of intelligence.

The adaptation of Bayes Theorem as an intelligence information processing model provides a systematic analysis methodology that generates an ordered, weighted list of postulated solutions. A well structured data base should be achieved as a by-product of the efficient assessment of relationships between problem inputs and solution. In addition, a shorthand mathematical notation is developed which, if properly catalogued, provides a method for keeping track of any current problem-solving effort. This same tally can also be viewed as a permanent historical record which, when appropriately stored and retrieved, generates a review of any analysis effort. Such records can also be employed to identify the discriminating contribution of specific inputs to any particular problem, thus providing a basis for assessing the input requirements of an intelligence area, or evaluating the various collection systems.

This entire capability is mainly predicated upon the ability to quantify the analyst's knowledge. Extensive research and testing, both in the areas of measurement theory and probabilistic processing of information, have indicated that a numerical quantification approximating a subjective probability (i.e., a consistent probability that obeys the axioms of probability theory) can be teased from an individual, or individuals. However, it is not now possible to identify which methods can be best employed by different groups working on different problems.

The objective of the continuing research effort is to identify the contribution that this analysis methodology brings to a particular intelligence area compared with existing operating procedures. The research will investigate the utility of different aids that enable the analyst to express his knowledge and judgments in forms compatible with the mathematical techniques discussed in this report. Moreover, the continuing research effort will explore aspects of data handling, problem solving, patterning, etc., that may lend themselves to the establishment of a complete methodology for intelligence problem solving.

APPENDIX A

BAYES THEOREM

APPENDIX A

BAYES THEOREM

There are many presentations of Bayes Theorem in literature today. The following one is not presented under the guise of being unique, but rather to provide a convenient reference for readers of this report. This particular form is what some authors designate as NEW BAYES.

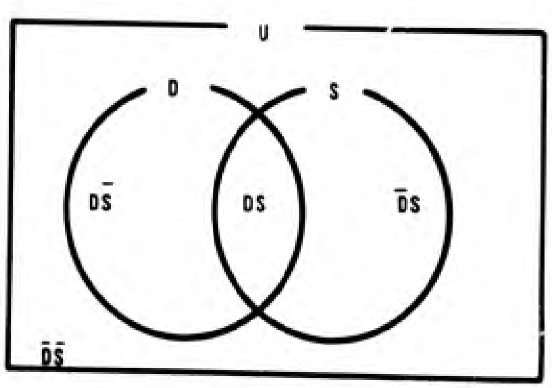
For the sake of clarity and ease of presentation, consider the situation of two events where the first event is the known existence of a particular submarine SS-12, and the second event is the observation of a forward deck gun on an unidentified submarine. The problem is to identify the unknown submarine by the inference that the observation may imply that the unknown submarine is in truth the particular submarine SS-12. However, the fact is also recognized that the observation of a forward deck gun would be common to other submarines, and also, although not germane to our problem, that the submarine SS-12 may be sighted without the observation of the forward deck gun. Let

S = The particular submarine, SS-12

$\bar{S}$ = All other operational submarines,

D = The observation of a forward deck gun, and

$\bar{D}$ = The observation of no forward deck gun.



The diagram consolidates these assumptions. The letter S with the associated circular area represent the event S or the submarine SS-12; the letter D with the associated circular area represent the event D or the observation of the forward deck gun on the unidentified submarine; and the letter U with the associated rectangular area represent the extreme boundaries for the problem or its universe. In this situation, U is the set of all operational submarines.

Two letters written together such as AB are read as "A and B" and in the diagram represent specific areas or the assumptions of our problem. These areas are identified as

$\bar{D}\bar{S}$ -- The area representing some other particular submarine(s) existing with the observation of no forward deck gun.

$D\bar{S}$ -- The area representing some other particular submarine(s) existing with the observation of a forward deck gun.

$\bar{D}S$ -- The area representing the particular submarine SS-12 existing with the observation of no forward deck gun.

DS -- The area representing the particular submarine SS-12 existing with the observation of a forward deck gun.

It is this last area, known as the intersection of the events D and S, that is of interest.

An acceptable assumption is that the probability of the intersection of two events D and S is the same as the probability of the intersection of the same two events designated as S and D; or

$$P\left\{DS\right\} = P\left\{SD\right\}.$$

The multiplication rule for probabilities states that

$$P\left\{DS\right\} = P\left\{D/S\right\} \cdot P\left\{S\right\} \text{ for } P\left\{S\right\} \neq 0 \text{ and likewise}$$
$$P\left\{SD\right\} = P\left\{S/D\right\} \cdot P\left\{D\right\} \text{ for } P\left\{D\right\} \neq 0.$$

The terms $P\left\{D/S\right\}$ and $P\left\{S/D\right\}$ are conditional probabilities, which simply mean that assuming the one event to the right of the vertical line has occurred, this is the probability of the other event occurring.

The term $P\left\{S/D\right\}$ represents the desired output of the problem; that is, given the observation of a forward deck gun on the unidentified submarine, what is the probability of the unidentified submarine being the SS-12. Therefore, equating the relationships yields

$$P\{S/D\} \cdot P\{D\} = P\{D/S\} \cdot P\{S\}.$$

Providing now that $P\{D\} \neq 0$ and $P\{S\} \neq 0$,

$$P\{S/D\} = \frac{P\{D/S\} \cdot P\{S\}}{P\{D\}} \quad \text{and} \quad P\{\bar{S}/D\} = 1.-P\{S/D\}.^{*}$$

The universe of this problem was bounded so that S and $\bar{S}$ are mutually exlusive and exhaustive; i. e., $P\{S\bar{S}\} = 0$ and the $P\{S \cup \bar{S}\} = 1$. This means that the probability of S and $\bar{S}$ existing together is zero or cannot occur since the unidentified submarine is either the SS-12 or some other submarine, but not both; and the probability of either S or $\bar{S}$ occurring is one, or that likewise the unidentified submarine is either SS-12 or some other particular submarine. Consequently, it follows that D can only occur in connection with S alone or $\bar{S}$ alone. Therefore, the probability of D is

$$P\{D\} = P\{DS\} + P\{D\bar{S}\}.$$

Again applying the multiplicative rule of probability,

$$P\{D\} = P\{D/S\} \cdot P\{S\} + P\{D/\bar{S}\} \cdot P\{\bar{S}\}.$$

Now the desired output can be shown to be

$$P\{S/D\} = P\{D/S\} \cdot P\{S\} / [P\{D/S\} \cdot P\{S\} + P\{D/\bar{S}\} \cdot P\{\bar{S}\}]$$

which is Bayes Theorem applied to the simple case of a single event S and its negation $\bar{S}$.

Consider now identifying N mutually exclusive and exhaustive submarines in place of $\bar{S}$. That is, rather than simply state that it is some other submarine,

*It should be noted that all probabilities are really conditional although not written as such. The $P\{S\}$ values are conditional on all information about S before D is learned.

assemble a list of candidates, even if it must be all submarines ever built. Again the phrase mutually exclusive and exhaustive simply means that only one of the submarines in the list can be the unidentified one and that the unidentified submarine is truly one of the submarines contained in the list. Then the formula becomes

$$P\{S_i/D\} = \frac{P\{D/S_i\} \cdot P\{S_i\}}{\Sigma_1^n P\{D/S_i\} \cdot P\{S_i\}} \quad \text{with} \quad \Sigma_1^n P\{S_i/D\} = 1. \text{ and } \Sigma_1^n P\{S_i\} = 1.$$

The subscript "i" indicates which submarine from the list of candidates is being considered. The summation sign, Σ, merely indicates that the denominator which before contained the sum of only two products -- $P\{D/S\} \cdot P\{S\} + P\{D/\bar{S}\} \cdot P\{\bar{S}\}$ --now is the sum of all n products -- $P\{D/S_1\} \cdot P\{S_1\} + P\{D/S_2\} \cdot P\{S_2\} + \ldots + P\{D/S_n\} \cdot P\{S_n\}$.

Consider now the arrival of many pieces of input. In such cases, the posterior estimate $P\{S_i/D\}$ for the first piece of information becomes the prior estimate $P\{S_i\}$ for the second piece of information. The posterior for the second piece of information is then :

$$P\{S_i/D_2\} = \frac{P\{D_2/S_i\} \cdot P\{S_i/D_1\}}{\sum_1^n P\{D_2/S_i\} \cdot P\{Si/D_1\}}$$

Generally the equation is written as

$$P\{S_i/D_j\} = \frac{P\{D_j/S_i\} \cdot P\{S_i\}}{\sum_1^n P\{D_j/S_i\} \cdot P\{S_i\}}$$

This equation is known as the recursive version of Bayes Theorem and assumes independence among data.

The probabilities incorporated into Bayes Theorem may be any combination of discrete and continuous, objective or subjective probabilities. Thus, the prior might be discrete (the probability of submarine SS-12 being one of say ten submarines under consideration maybe 0.1) and the input continuous (the length of the unidentified submarine is between 150 and 165 feet with a best estimate at 155) or vice versa. For a more detailed presentation of the various facets of Bayesian Statistics, see Weir, [1966].

APPENDIX B

A BASIC STUDY OF THE APPLICATION OF BAYESIAN ANALYSIS TO AN INTELLIGENCE PROBLEM AREA

BLANK PAGE

APPENDIX B

A BASIC STUDY OF THE APPLICATION OF BAYESIAN ANALYSIS TO AN INTELLIGENCE PROBLEM AREA

In addition to theoretical calculations, some basic pilot studies were conducted in order to explore the application of one approach of Bayesian Analysis to the intelligence analysis-synthesis process. This appendix presents (1) the rationale, (2) the methodology and (3) observations on some results of these studies.

If an intelligence problem area is well defined and the data normally associated with the area are well structured, it is possible to assemble this information into an analysis table such that the column headings represent solutions under consideration and the row headings represent the known associated problem inputs. A portion of such a "characteristics matrix" could be represented as.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
C_1	✓				✓			✓	
C_2			✓			✓			
C_3									
C_4	✓			✓			✓		
C_5									

where the nature of the association between the characteristics and the solutions are given in the intersecting cells. The solutions could be very distinct items such as the specific submarines SSN-578 Seawolf, SSN-578 Skate, SSN-586 Triton, SSN-649 Sunfish, etc. Alternatively, the solutions could represent classes of items where individual members are not identified. For example, the solutions could represent specific classes of the Douglas Skyhawk such as A4A, A4B, A4C, A4E, etc., without regard for individual members (in this case 165, 542, 638, 500, respectively).

The problem inputs represent individual items whose historical files have identified an association with the various solutions. Examples would include the presence or absence of colors, characteristic shapes, physical items, monitored parameters, etc. These examples would be representative of qualitative characteristics or inputs analogous to discrete random variables.

Inputs such as length, width, height, velocity, location, etc., would be representative of quantitative characteristics or inputs analogous to continuous or approximately continuous random variables. These types of inputs would best be acknowledged by entering an expected value in each of the cells.

A characteristic matrix that simply acknowledges the existence of an association between a particular characteristic and a particular solution may not, however, provide the analyst with the type of information he desires. In fact, if the analyst interprets such acknowledgments as 0-1 relationships (i.e., the solution either has or does not have the characteristic), then the utility of such a matrix may be deceptive. There are, in general, three problems hindering the utility of a 1-0 characteristic matrix. These are

1. Some characteristics are temporal; they may exist only under certain circumstances or they may periodically change.
2. If the solutions represent a set or class of units, some characteristics may not exist for all members of the class.
3. The fact that a solution exhibits a characteristic doesn't mean that it will be reported. Similarly, the absence of a characteristic does not mean that one may not be mistakenly reported.[1]

	S_1	S_2	S_3	S_4	S_5
C_1	.90	--	.75	.30	.60

A much more useful matrix can be created if the contents of the cells denote the history of reported data associated with each solution. For example, this diagram indicates that the characteristic has been reported more often with S_1, followed by S_3, S_5, S_4, and never with S_2.[2] Incorporating appropriate considerations, such as the reliability of the collecting system, etc., a matrix containing

[1] The failure to observe an existing characteristic is an example of a TYPE I or α error. The reporting of an observation of a nonexisting characteristic is an example of TYPE II or β error.

[2] It should be noted that the contents of the cell indicate the frequency with which the data are reported given the solution to be true. It does not indicate the frequency with which those solutions are true given the reported data.

probabilistic statements concerning the expected relationships between inputs and solutions are created, not just acknowledgements between characteristics and solutions.

With these observations in mind, some basic studies were performed at HRB-Singer to examine the application of a Bayesian analysis technique to a quasi-invariant intelligence problem area. This effort represented the application of discrete probability distributions for discriminating among the states of nature on the basis of qualitative characteristics. A senior analyst with extensive experience in the selected area participated in the study. The procedure utilized was selected because the problem was well defined and the data were well structured. This procedure was as follows:[1]

1. First, a characteristics matrix was constructed for the problem area. The column headings represented each member of the solution set while the row headings represented each piece of data that has been an input to this problem area. A portion of the matrix is shown as follows:

	S_1	S_2	S_3	S_4	S_5	S_6
D_1	.5	.5	.65	.8	.8	.7
D_2	.2	.2	.6	.75	.75	.5
D_3	.4	.4	.75	.8	.9	.7
D_4	.2	.2	.6	.75	.75	.01
D_5	.05	---	.05	.1	.04	.8

The values in each cell represent the conditional probability $p(D/S)$ and were obtained from the analyst either as frequency counts from his historical files, or as his own expression of judgment; i.e., "If I assume that S_6 is the true solution, what are the chances of observing D_4 as an input?" Thus, the relationship is not only acknowledged, but is actually weighted to account for those times when the relationship does not always exist.

2. Next, selected inputs were presented to the analyst while a computer program mathematically combined the selected inputs into output estimates using the recursive version of Bayes Theorem and the values from the

---

[1]As discovered later, this procedure is very similar to the scenario development technique employed by W. Edwards, et al., when testing the PIP concept.

characteristics matrix. It should be emphasized that the analyst and the computer received the same inputs. Also, theoretically both the analyst and the computers weighed the inputs identically since the computer incorporated the values developed by the analyst for the characteristic matrix. Thus, any differences in their outputs should reflect only differences in the integrating methodologies.

3. Finally, the outputs from the analyst and the computer were compared. For example, the analyst was given twelve pieces of information and asked to give his solution. Sometimes he was also asked to give a ranked intermediate solution listing after receiving only part of the information. Sometimes the analyst was asked to estimate his confidence by numerical quantification in the ranked solution listing by "betting" up to $100 on each member of the solution set. These numbers were then normalized. Although these weightings may not be totally interpretable, it was felt that they would provide some insight into the rankings. At the same time, the computer, starting with the input that all solutions were equally likely, calculated a solution set using the values from the characteristics matrix for the same twelve pieces of information. The intermediate probabilities distributed over the solution set after considering each input were calculated in addition to the final output; only a discussion of some of the observations will be given here. A typical result is given in Figure 5. The upper part of the figure shows a comparison of some intermediate and the final outputs of the analyst with his confidence and the computer output. The lower portion represents the confidence plot for solution S_4 only.

The first series of studies incorporated several assumptions based on the analyst's statements that proved to be incorrect. For example, the analyst stated that he assumed any solution of the 14-member set was equally likely of being the true state of nature prior to his receiving any data. However, upon additional questioning this proved to be false and the prior probabilities were changed accordingly. The impact of the correction on some solutions is shown in the following example.

D_4 ANALYST	D_4 COMPUTER	D_7 A	D_7 C	D_8 A	D_8 C	D_9 A	D_9 C	D_{10} A	D_{10} C	D_{12} A	D_{12} C
4(59%)	4(73%)	4(77%)	4(82.14%)	6(50%)	4(88.03%)	4(53%)	4(99.86%)	4(83%)	4(99.99%)	4(98%)	4(99.99%)
5(11%)	5(22.8%)	5(15%)	6(17.85%)	4(25%)	6(11.96%)	6(28%)	6(.13%)	6(11%)		6(1%)	
6(6%)	3(2%)	6(8%)		5(25%)		5(19%)		5(6%)		5(1%)	
14(6%)	14(2%)										
1(6%)	1(.2%)										
3(6%)	6(.05%)										
2(6%)	2(.01%)										

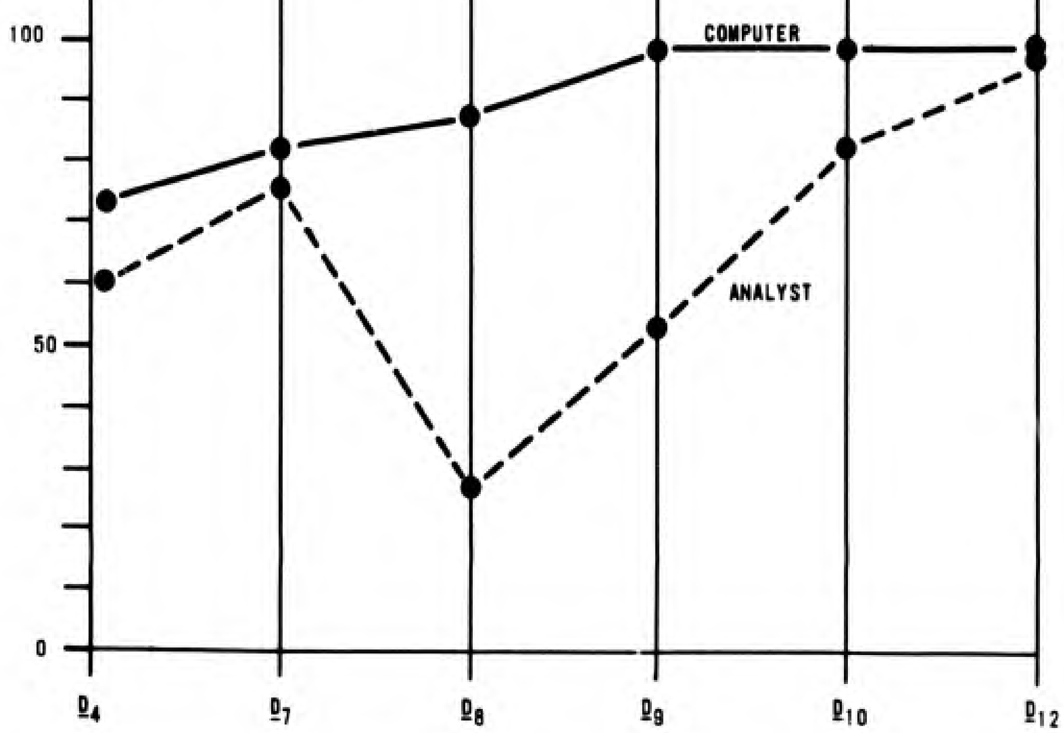


FIG. 5 SOLUTION #4 OF PILOT TEST

Computer Output For "Equally Likely"		Computer Output For "Weighted Prior"	
S_{11}	62.24%	S_8	78.45%
S_8	37.74%	S_{11}	21.53%

In this example the solution set ranking flip-flopped. In others the net result was a change in weights but not in ranking. However, the combination of this updated prior and other changes was sufficient to generate other flip-flops. This updated prior weighting was checked several times and found to be consistent.

Another assumption was that the analyst only considered positive inputs, that is, information he received, not the lack of certain inputs. Again this proved false as shown in the following case.

Computer Output For "Positive Inputs"		Computer Output For "Lack of D_8"	
S_6	97.52%	S_6	83.37%
S_{14}	2.06%	S_{14}	15.73%
S_5	0.41%	S_5	0.89%

At first, the change might not seem too drastic. But again this change in association with other changes, caused a complete re-ranking of the solution set. Perhaps more dramatic was the situation for a certain subset of the solution set.

Computer Output For "Positive Inputs"		Computer Output For Addition of "Seven Lacking Inputs"		Computer Output For Addition Of "Three More Lacking Inputs"	
S_5	73.45%	S_5	44.85%	S_1	99.99%
S_6	16.13%	S_{11}	38.59%		
S_1	6.17%	S_6	6.67%		
S_{14}	3.00%	S_{14}	4.70%		
S_2	0.69%	S_2	4.37%		
S_3	0.50%	S_3	0.78%		
S_4	0.02%				

As the reader can see, the increased number of lacking inputs strongly changes the output. The problem here, of course, is whether or not these inputs were truly not there or just not observed or reported. It does demonstrate once again, however, that "negative" input also has utility.

One important difference noted between computer generated solution weights and those directly estimated by the analyst was the effect of individual items on the assessment. In the following example, the analyst reconsidered some of the solutions that he had previously eliminated after examining datum D_{22}. The Bayesian processer integrated this item with all previous items, producing quite different results.

AFTER RECEIVING 4 INPUTS		AFTER CONSIDERING THE NEXT INPUT, D_{22}	
Analyst	Computer	Analyst	Computer
S_4	S_4	S_4	S_4
S_5	S_5	S_5	S_5
S_1	S_3	S_6	S_6
S_2	S_{14}	S_{10}	
S_3	S_1	S_{13}	
S_6	S_6	S_7	
S_{14}	S_2	S_8	
		S_9	
		S_{14}	

Another initial assumption of this particular approach that proved slightly erroneous was the belief in an invariant situation. For example, although the analyst utilized frequency counts from his historical files, the specific time period was very important. The relationship between say, D_{14}/S_7 may be 10% over a five year period, but only 1% over the last year. Therefore, a forecasting method emphasizing the more recent information or some other stepwise heuristic search pattern to facilitate such decisions would seem to be in order. The amount of noise that this condition added to these studies was not immediately attainable.

Two types of dependency problems existed. The first involved certain solutions that could be the true state of nature with D_1 or D_2 or D_3 but not with D_1 and D_2 and D_3. For example:

	S_1	S_8
D_{23}	5%	1%
D_{27}	10%	40%
D_{32}	75%	90%

The discriminating capability of these inputs between S_1 and S_8 is essentially nil for these three pieces of data so that the only output difference would be the prior weighting ratio. However, S_1 would actually exist only with D_{23} or D_{27} or D_{32}, but never with all three. So S_8 would be the unique solution if all three inputs were received. Therefore, this must be handled by some different technique. The other type was what might be described as quasi-twins or a concatenate relationship, i.e., if solution 8, then also solution 10 is in contention. This type of situation may only be unique to this problem area. A third type of dependency problem--the combinatorial type--was not specifically pinpointed.

The implicit assumptions made explicit were particularly impressive. For example, a disagreement in solution was quickly resolved after the analyst noted that he questioned one of the inputs.

ANALYST		COMPUTER		COMPUTER AFTER CORRECTION OR INPUT	
S_8	46.23%	S_8	78.45%	S_8	75.59%
S_{10}	37.63%	S_{11}	21.53%	S_{10}	20.67%
S_{11}	16.13%			S_{11}	3.73%

The numerical values perhaps should reflect to some extent a weighted amount of disbelief the analyst has created. This type of problem can create havoc in analysis as soon as the validity of an input is questioned. Therefore, some scheme must be provided to allow for reassessment of the $p(D_1/S_j)$ as required.

One thing that was not resolved is the effect of a "new" solution on the data input and the ability to identify a new solution on the basis of its input. For example, would a seemingly random selection of inputs generate sufficient contrast to stimulate the thought of a new solution previously not considered? In reality, the situation of two or more unidentified solutions occurring simultaneously was not considered either. Some strategy of pattern checking would be necessary to patrol these situations.

There were several runs that could not be resolved even after interrogation. Whether these constituted simply combinatorial dependency problems--unique patterns to this particular analyst--or something more complex was not fully recoverable. Because the total amount of information for the problem area was purposely not incorporated, and because the analyst seemed to demonstrate several inconsistent results when the same problems were represented from time to time, these could not be fully resolved.

There is a need to monitor the computer operation. If two solutions are under consideration and the discrimination between them generated by a piece of input is .1% versus .01%, the former may very well go to 99.99%. However, the analyst should be warned of the reason for such a drastic change if the characteristic matrix is fixed in a machine since it would appear that the particular piece of information is rare for both.

Two aspects of synthesis were quite obvious. A solution with unique inputs was quickly isolated by both the analyst and the machine.

	S_{10}	S_{12}	ALL OTHER SOLUTIONS
D_{40}	20%	30%	---
D_{42}	---	95%	---
D_{47}	---	20%	---

Obviously, any input combination with D_{42} and D_{47} pinpointed S_{12}. In addition, a very common solution was consistently identified, although the uniqueness was not as obvious until all inputs were set forth within a reduced matrix. The computer consistently identified these solutions with less information than the analyst. A simple example of such a problem would be

	S_1	S_2	S_3	S_4	S_5	S_6	S_{14}
PRIOR WEIGHTS:	2.4%	0.24%	1.2%	21.64%	6.01%	2.4%	1.2%
INPUTS: D_1	50%	50%	65%	80%	80%	70%	65%
D_2	20%	20%	60%	75%	75%	50%	60%
D_3	40%	40%	75%	80%	90%	70%	75%
D_{22}	---	---	---	80%	40%	50%	---
OUTPUTS: CUMULATIVE							
AFTER D_1	4.48%	0.44%	2.91%	64.72%	17.97%	6.28%	2.91%
D_2	1.28%	0.12%	2.51%	69.69%	19.35%	4.50%	2.51%
D_3	0.63%	0.06%	2.33%	69.11%	21.59%	3.91%	2.33%
D_{22}	---	---	---	83.92%	13.11%	2.96%	---

Generally, increasing the number of inputs still will not lessen the obvious answer because of the systematic presentation within the matrix. Atypical inputs naturally generated the greatest confusion, particularly when a solution represented a hybrid.

The analyst's solutions seemed to be consistently conservative, as predicted by other reported research and shown in the following examples.

AFTER 17 INPUTS

Analyst	Computer
S_{10}-43%	S_{10}-99.02%
S_8 - 30%	S_8 - 0.97%
S_{13}-27%	

AFTER 11 INPUTS

Analyst	Computer
S_4-80%	S_4-99.19%
S_6-20%	S_6- 0.80%

AFTER 4 INPUTS

Analyst	Computer
S_4-60%	S_4-83.92%
S_5-27%	S_5-13.11%
S_6-13%	S_6- 2.96%

Again the absolute values of the analyst's weights are not fully understood but should provide an indication of his confidence in the selected solution set.

The fact that many of the relationships are invariant and that the computer quickly moved in on them should be an indication that a technique of searching for inputs that would generate contradictions may be more in order. Thus, the analyst function would be properly shifted towards analysis to insure that the maximum amount of information had been made available. Of course, the computer role could be expanded in many other ways to further facilitate this analysis function.

APPENDIX C

THE ROLE OF PROBABILISTIC INTELLIGENCE OUTPUTS AND LOSS FUNCTIONS

At the end of Chapter III, the question was raised whether intelligence outputs should be decisions in the disguise of "conclusions," or simply presented as inputs to a decision process. Since intelligence outputs can be categorized as decisions or conclusions, the question is raised as to the differences between these concepts. In developing the definitions, the writer will quote liberally from Dr. J. W. Tukey's paper "Conclusions vs. Decisions," (Tukey, 1960).

A conclusion is a statement which is to be accepted as applicable to the conditions of an observation unless and until unusually strong evidence to the contrary arises. Conclusions are established with careful regard to evidence, but without regard to consequences of specific actions in specific circumstances. Thus, conclusions are withheld until adequate evidence has accumulated.

The definition of a conclusion has three crucial parts; two explicit and the third implicit. First, it emphasizes "acceptance" in the original, strong sense of the word. The conclusion is accepted, and taken into the body of knowledge, not just into a guidebook of advice for immediate action. Secondly, the definition speaks of "unusually strong evidence." This implies that only a small percentage of all conclusions will, in due course, be upset. Finally, it does imply the possibility of later rejection. A conclusion is something of lasting value extracted from the data; but it is to be of lasting value, not necessarily of everlasting value.

Hence, conclusions must be reached cautiously, firmly, not too soon, but not too late. They must be judged by their long-run effects, by their "truth," not by specific consequences of specific actions.

Decisions, on the other hand, are more nearly of the form "let us decide to act for the present as if," rather than "we accept." The distinction is important and too often neglected. The restrictions "act...as if" and "for the present" convey two separate and important ideas, ideas which serve to distinguish conclusions from decisions.

When it is stated to "act as if A is greater than B," no judgments as to the "truth" or "certainty beyond a reasonable doubt" is made of the statement "A is greater than B." When it is stated "for the present," the decision maker is referring only to the particular situation under consideration at present. During his decision-making processing, he is weighing both the evidence concerning the relative merits of A and B and also the probable consequences in the present situation of various actions. Finally he decides that the particular course of action which would be appropriate if A were truly greater than B is the most reasonable one to adopt in the specific situation that faces him. The consequences in other situations of acting as if A is truly greater than B have not been considered. It is most important that the decision maker be totally aware of these assumptions.

Decisions to "act for the present as if" are attempts to do as well as possible in specific situations, to choose wisely among the available gambles. But this raises the second question -- who is the decision-maker? From the presentation thus far, it is evident that in most cases the intelligence output from a Bayesian Processor will not be a single solution with a posterior probability equal to one, i. e, a solution that can be regarded as a certainty; but rather, there will be a positive probability associated with several of the solutions. The intelligence analyst may be tempted to report a solution as true if the probability of that solution is greater than all the other solutions. However, this approach may lead to serious error as the following example will demonstrate.

Assume that a strike mission is being planned against a target V-3 and that intelligence has been requested to identify the defensive weapon system at V-3. Suppose the intelligence output at this point, on the basis of previously monitored construction is AAA-.76, and SAM-.24.

Intelligence might be tempted to respond with the report:

> "Evidence to date indicates that the target V-3 is probably defended by at least one AAA unit."

This statement does not convey as much useful information as it could. The associated probability estimates could be combined with Command's assessment of the tactical situation.

Suppose Command has the following information at its disposal and seeks to maximize the chances of a successful mission.

1. It is possible to draft a flight plan and ECM usage plan that could effectively counter the use of AAA or SAM, but NOT both.
2. If AAA is countered and V-3 is defended by AAA, the estimated probability of a successful mission is .80.
3. If SAM is countered and V-3 is defended by SAM, the estimated probability of a successful mission is .90.
4. If, however, AAA is countered when SAM is the defensive system, the probability of success is only .10.
5. If, on the other hand, SAM is countered when AAA is the defensive system, the probability of success is .60.

This information can be conveniently assembled in matrix form as the decision maker utility table, loss table, or loss function. For example,

	ACTION	
REAL-WORLD SITUATION	Counter AAA	Counter SAM
AAA defense	.8	.6
SAM defense	.1	.9

The cell values represent the probable consequences in the present situation for two actions -- counter AAA or counter SAM. The decision maker can now incorporate his knowledge (the posterior probabilities) concerning the relative merits of AAA and SAM to derive the chances of a successful mission.

If the action to counter AAA is made, the probability of success is

$$(.8)(.76) + (.1)(.24)$$

or .632. Likewise, if the action to counter SAM is made, the probability of success is

$$(.6)(.76) + (.9)(.24)$$

or .672. Therefore, to maximize the chances of a successful mission, the planners should assume the target is defended by SAM and act accordingly; i. e., "act as if SAM defense is greater than AAA defense for the present situation."

The decision is really the opposite of the one that would have been indicated from the first output. This fact indicates that in order to make a decision, not only the probability of occurrence of each solution is needed, but also the expected loss or gain for each of the possible actions that might be taken. Since an analyst could not have possession of all the required utility tables, loss functions and their applications will not be discussed further since this is not within the scope of an intelligence analyst's job. Instead, the analyst should supply the best estimate of the probabilities of each solution.

APPENDIX D

MODIFICATIONS TO BAYES THEOREM TO COPE WITH INPUT UNCERTAINTY

The intelligence effort is directed toward the determination of an unknown state of nature when one assumes that it is one of n possible solutions $S_1 \ldots S_n$. In order to attack the problem by a Bayesian formulation, it is necessary that in addition to a prior probability distribution existing over the S_i's, there are k characteristics $C_1 \ldots C_k$ and conditional probability distributions connecting the C_j's and the S_i's. These characteristics may be one of two types, depending on whether or not observation of the characteristics gives rise to a discrete or continuous random variable.

The former type, which can be labeled as a "discrete characteristic," is such that it may be observed in one of a possible number of states. For example, the characteristic "color" may have as states "red", "blue", "white", and "black", assuming these are the only colors that the true solution may possibly admit. Another example of a discrete characteristic is, if considering submarines, "snorkel", where there may be only two possible states, namely "presence of snorkel" and "absence of snorkel".

The other type of characteristic, labeled as a "continuous characteristic", is such that it may be observed over a continuous spectrum. Examples of such characteristics are lengths, widths, and heights of objects. In some cases, a continuous characteristic may be used as an approximation when the characteristic is really discrete. Thus, if the characteristic is the number of men in a particular regiment of enemy forces, no great error is introduced by assuming that this characteristic is a continuous one.

Despite the type of characteristic C_j, if its value or state is known, the new or posterior probability distribution may be calculated. However, in many cases the exact value or state is not known, and although an observation D_j of C_j has been made, a degree of uncertainty still exists as to the exact value or state of C_j. This uncertainty may be caused by an error of observation or transformation. Such errors may occur at the source by the sensor or during the transmission of the data to the analyst. Another source of uncertainty may be created when an input is generated within the environment, but not by the unknown state of nature. The uncertainty generated by these types of inputs may be accounted for by considering the probability distribution of the errors, which may be derived subjectively or with the aid of empirical results.

The Continuous Case

Since, in many cases, when dealing with continuous characteristics, one can expect the distribution of measurement errors to be normally distributed (at least approximately), the normal distribution will serve as a model of error in the continuous case. An example of how such an error model may be used was presented in Chapter IV. As another example, consider only two possible solutions, S_1 and S_2. Furthermore, assume that both solutions are equally likely, a priori, and that a posterior probability assignment will be calculated with the aid of an observation of length. The following table summarizes the problem:

			Length
$p\{S_1\}$	$= .5$	S_1	50 ft.
$p\{S_2\}$	$= .5$	S_2	60 ft.

Thus, an observation without error would of necessity be either 50 ft. or 60 ft., and no other values. However, when error is present in the observation, any value becomes possible, although it would be reasonable to assume that the observed value D would be closer to 50 or 60 ft. than to, say, 1000 ft. In general, it is reasonable to assume that D will be distributed according to a normal distribution, with population mean equal to the true length of the object being observed. What would be unknown, of course, would be the variance. Assuming that this may be determined, either from test situations or from subjective estimates, the posterior distribution over S_1 and S_2 may be calculated taking into consideration the inherent uncertainty in the observation D.

If it were known, for instance, that the error in the measuring device being used had a variance of 5, then the corresponding probability distribution for the observation D would be

$$P\left\{D\right\} = \frac{1}{\sqrt{10\pi}} e^{-\frac{1}{10}(D-C)^2}$$

where C denotes the true length of the object being observed. In the example under consideration, C may be one of two possible values, either 50 or 60.

Suppose now, that an observation D = 53 were reported. Using Bayes Theorem, the desired posterior probabilities may be calculated as follows:

$$P\left\{S_1/D=53\right\} = \frac{P\left\{D=53/S_1\right\} P\left\{S_1\right\}}{P\left\{D=53/S_1\right\} P\left\{S_1\right\} + P\left\{D=53/S_2\right\} P\left\{S_2\right\}}$$

and

$$P\left\{S_2/D=53\right\} = \frac{P\left\{D=53/S_2\right\} P\left\{S_2\right\}}{P\left\{D=53/S_1\right\} P\left\{S_1\right\} + P\left\{D=53/S_2\right\} P\left\{S_2\right\}} .$$

Since

$$P\left\{D=53/S_1\right\} = \frac{1}{\sqrt{10\pi}} e^{-\frac{1}{10}(53-50)^2} = \frac{1}{\sqrt{10\pi}} e^{-.9} = \frac{.406570}{\sqrt{10\pi}}$$

and

$$P\left\{D=53/S_2\right\} = \frac{1}{\sqrt{10\pi}} e^{-\frac{1}{10}(53-60)^2} = \frac{1}{\sqrt{10\pi}} e^{-4.9} = \frac{.007447}{\sqrt{10\pi}} ,$$

it follows that the posterior distribution over S_1 and S_2 is

$$P\left\{S_1/D=53\right\} = \frac{.406570(1/2)}{.406570(1/2) + .007447\ (1/2)} = .98$$

and

$$P\left\{S_2/D=53\right\} = \frac{.007447(1/2)}{.406570(1/2) + .007447(1/2)} = .02.$$

It should again be noted that this posterior distribution takes into account the uncertainty of the observation D=53.

<u>The Discrete Case</u>

In general, for a problem dealing with discrete characteristics, it is assumed that the following are given:

1. n possible solutions $S_1, \ldots, S_n$
2. An *a priori* probability distribution over $S_1, \ldots, S_n$
3. k characteristics $C_1, \ldots, C_k$
4. Each characteristic C_j may be observed in one of k_j possible states c_{jm}
5. The probabilities $P\left\{c_{jm}/S_i\right\}$ for $i=1, \ldots, n$; $j=1, \ldots, k$, and $m=1, \ldots, k_j$ are known.

Thus, if it is definitely known that the unknown solution has a characteristic in a given state, the posterior distribution over $S_1, \ldots, S_n$ may be calculated.

However, many times the exact state c_{jm} of characteristic C_j is not known, although some collateral information about it exists. For example, consider the characteristic C_j of "color", having possible states of $c_{j1}, \ldots, c_{j6}$ which might be, say, red, white, blue, green, yellow, and orange. If, of ten observers looking at an object, eight report the color red, that is, the state c_{j1}, and two report the color orange, that is, the state c_{j6}, it is obvious that observational error exists (assuming the observed object possesses only one color). Based on the report D only, it would be reasonable to assign

$$P(c_{j1}/D) = .8$$

$$P(c_{jm}/D) = 0 \text{ for } m = 2, 3, 4, 5$$

$$P(c_{j6}/D) = .2.$$

In order to arrive at a posterior distribution $P_1(S_i)$ over $S_1, \ldots, S_n$, Dodson [1961] suggests the formula

$$P_1(S_i) = \sum_{m=1}^{6} P(S_i/c_{jm})\, P(c_{jm}/D).$$

Although this may seem to be an intuitively satisfying procedure, it may lead to incorrect results. This is because the knowledge and information on which the prior distribution over the S_i's was based may or may not influence the report D. What Dodson tacitly assumes is that this prior knowledge is taken into account when the report is made. This becomes evident if one considers the knowledge, K, on which the prior probability assignments are made. In symbols, the posterior probability which needs to be computed is $P(S_i/DK)$. This may be written as

$$P(S_i/DK) = \sum_{m=1}^{6} P(S_i c_{jm}/DK) = \sum_{m-1}^{6} P(S_i/c_{jm}DK)\, P(c_{jm}/DK).$$

Since the report does not alter the probability of S_i if c_{jm} is known, $P(S_i/c_{jm}DK) = P(S_i/c_{jm}K)$. Thus,

$$P(S_i/DK) = \sum_{m=1}^{6} P(S_i/c_{jm}K)\, P(c_{jm}/DK).$$

This explicit inclusion of the information K into the probability formulas stresses the fact that Dodson's $P(c_{jm}/D)$ is, in fact, $P(c_{jm}/DK)$ and does indeed depend upon this prior knowledge. As is often the case, however, the observation is made without the use of this prior knowledge, and thus, the probability which results is $P(c_{jm}/D)$, *per se*, and *not* $P(c_{jm}/DK)$.

To illustrate this point, assume that there exists an unknown S which is either S_1 or S_2, and for which the existing prior distribution is $P(S_1) = .9$ and $P(S_2) = .1$. Further assume that based on characteristic C_1, having the two states c_{11} and c_{12}, the conditional probabilities $P(c_{11}/S_1)$, $P(c_{11}/S_2)$, $P(c_{12}/S_1)$, and $P(c_{12}/S_2)$, are 1, 0, 0, and 1, respectively. Note that the prior knowledge is implicit in these probability assignments. Summarizing in tabular form yields

	.9	.1
	S_1	S_2
c_{11}	1	0
c_{12}	0	1

Suppose now that an observer files a report in which he states that, based on his observation, he couldn't tell whether C_1 were in state c_{11} or state c_{12}, and thus felt that there was about a 50-50 chance that characteristic C_1 was in state c_{11}. That is, based on this report D, $P(c_{11}/D) = .5$ and $P(c_{12}/D) = .5$. Hence, since $P(S_1/c_{11}) = P(S_2/c_{12}) = 1$ and $P(S_1/c_{12}) = P(S_2/c_{11}) = 0$, Dodson's method would give

$$P_1(S_1) = \sum_1^2 P(S_1/c_{1m})\, P(c_{1m}/D) = 1(.5) + 0(.5) = .5$$

$$P_1(S_2) = \sum_1^2 P(S_2/c_{1m})\, P(c_{1m}/D) = 0(.5) + 1(.5) = .5$$

which indicates a large shift in the probability distribution. However, it seems realistic to assume that in light of the report which adds nothing to existing knowledge, the distribution over S_1 and S_2 should remain the same, i.e., should stay at $P(S_1) = .9$ and $P(S_2) = .1$.

It is evident in this case that the probabilities $P(c_{11}/D)$ and $P(c_{12}/D)$ were not calculated with respect to the prior knowledge K. Since the probabilities involving the S's and c's were, however, the marginal probability $P(c_{1m})$ is actually the probability $P(c_{1m}/K)$. What is needed is to combine $P(c_{1m}/R)$ and $P(c_{1m}/K)$ in order to arrive at the probability $P(c_{1m}/DK)$. In general, there is no way to do this. However, if it may be assumed that the probabilities $P(c_{1m}/D)$ arise as a result of an implicit uniform prior (sometimes referred to as a prior of ignorance), this prior is $P(c_{1m}) = \frac{1}{n_1}$ for $m = 1, \ldots, n_1$. Hence, based on this uniform prior, the probabilities $P(D/c_{1m})$ may be calculated by the formula

$$P(D/c_{1m}) = \frac{P(c_{1m}/D)P(D)}{P(c_{1m})} = n_1 P(c_{1m}/D)P(D).$$

Now, these probabilities represent the chance of obtaining the given report if characteristic 1 is truly in state c_{1m}. Thus, the additional information obtained from the prior knowledge K is negligible, because the overwhelming impact arises from the knowledge that C_1 is in state c_{1m}. An excellent approximation should then be given by $P(D/c_{1m}K) = P(D/c_{1m}) = n_1 P(c_{1m}/D)P(D)$.

The probability $P(S_1/DK)$ may be written as

$$P(S_1/DK) = \sum_{m=1}^{n_1} P(S_1/c_{1m}DK)P(c_{1m}/DK).$$

Now, it may be assumed that

$$P(S_1/c_{1m}DK) = P(S_1/c_{1m}K),$$

since if c_{1m} is known to be the true state of C_1, then any report about C_1 will not alter the associated probability distribution. Also, one may write

$$P(c_{1m}/DK) = \frac{P(D/c_{1m}K)P(c_{1m}/K)}{P(D/K)},$$

where
$$P(D/K) = \sum_{m=1}^{n_1} P(D/c_{1m}K)P(c_{1m}/K).$$

Thus,
$$P(S_1/DK) = \sum_{m=1}^{n_1} P(S_1/c_{1m}K)\frac{P(D/c_{1m}K)P(c_{1m}/K)}{P(D/K)}.$$

Applying this procedure to the problem discussed above yields

$$P(D/c_{11}K) \doteq 2(.5)P(D) = P(D)$$

and

$$P(D/c_{12}L) \doteq 2(.5)P(D) = P(D).$$

Thus,

$$P(S_1/DK) = \frac{1 \cdot P(D) \cdot (.9)}{1 \cdot P(D) \cdot (.9) + 1 \cdot P(D) \cdot (.1)} = .9$$

and

$$P(S_2/DK) = \frac{1 \cdot P(D) \cdot (.1)}{1 \cdot P(D) \cdot (.9) + 1 \cdot P(D) \cdot (.1)} = .1.$$

That is, the resulting posterior agrees with the one which appears most satisfying intuitively.

In general, it would seem wise to examine the procedure used to arrive at the reported probabilities $P(c_{1m}/D)$. If it may be assumed that the report D takes into consideration the available prior knowledge, Dodson's approach will provide the best probability assignment, while if it may be assumed that the report D is based only on the observation with an explicit or implicit uniform prior over the c_{1m}'s, then the procedure suggested above appears to provide the best probability assignment.

Noise Factors in the Environment

In the discussion up to this point, it has been assumed that observations were made only on the true unknown. This is sometimes not the case, because one may observe something he thinks may be associated with the unknown but which actually is not. For example, an observer may think he observes a submarine when he is actually looking at a whale.

Thus, in order to consider a case such as this, assume that it is known that either $S_1, S_2, \ldots,$ or S_n is in a given environment, and it is desired to know which S_i is present. In order to do this, of course, observations are taken on the unknown state of nature. Suppose, however, that in addition to the unknown state of nature there is something else which may be observed instead. This something else will be designated as a "noise factor" and will be denoted by N. In general, N will exhibit, or be capable of exhibiting, some or all of the characteristics of the S_i's. If the proportion of times N will be present can be determined, and also the frequency with which certain characteristics will be exhibited, then the effect of this noise factor may be taken into account.

Assume that the noise factor is independent of any of the S_i's, and consider the hypotheses H_i, $i=1, \ldots, n$, where H_i denotes the event that solution S_i is present, although it may be accompanied by a noise factor. Thus, the problem becomes one of obtaining a posterior probability distribution over the H_i's. Now, based on these assumptions, the prior distribution over the H_i's may be computed from the prior distribution of N and the S_i's. That is,

$$P(H_i) = P(S_i N \text{ or } S_i \bar{N}) = P(S_i)\,.$$

Hence, the distribution over the H_i's is equivalent to the distribution over the S_i's. Now, the probability of an observed characteristic C_j in state c_{jm} given hypothesis H_i is

$$P(c_{jm}/H_i) = P(c_{jm}/NS_i)P(N) + P(c_{jm}/\bar{N}S_i)P(\bar{N})\,.$$

Hence, given that c_{jm} is observed, the posterior probability of any H_i may be computed as:

$$P(H_i/c_{jm}) = \frac{P(c_{jm}/H_i)P(H_i)}{\sum_{i=1}^{n} P(c_{jm}/H_i)P(H_i)}\,.$$

As an example, consider a box containing six bags of type 1 and four of type 2, where type 1 bags contain nine red balls and one blue ball, and type 2 bags contain two red balls and eight blue balls. Assume that a bag is chosen at random from the box and its contents are dumped into another box. To complicate matters, a coin is tossed and if heads results, an additional five red and five blue balls are placed in this box, while if tails results, no extra balls are added. Thus, the possible introduction of these ten balls is essentially the noise factor N.

The corresponding characteristics matrix is

	S_1	S_2	N
c_{11}	.9	.2	.5
c_{12}	.1	.8	.5

where $P(S_1) = .6$, $P(S_2) = .4$, and $P(N) = .5$. Of course, S_1 denotes a bag of type 1, S_2 denotes a bag of type 2, N denotes the presence of five additional balls of each color, c_{11} denotes the color red, c_{12} denotes the color blue, and each entry in the table is a probability of the type $P(c_{11}/S_1)$, etc.

Suppose the procedure outlined above is followed. That is, one of the ten bags in the original box is chosen at random and its contents dumped into a second box. Further, a coin is flipped and the additional ten balls are placed in the second box if the toss results in heads (of course, assuming one does not know the outcome of this toss). Suppose that a person is presented with the second box, and draws one ball from it. If this ball should happen to be blue, what can one say about the posterior probability distribution over the hypotheses H_1 and H_2? That is, what are the chances, having observed the blue ball, that it was bag type 1, as opposed to bag type 2, that had its contents dumped into the box?

The probabilities obtained from the characteristics matrix are

$$P(c_{11}/S_1) = .9, \qquad P(c_{12}/S_1) = .1$$

$$P(c_{11}/S_2) = .2, \qquad P(c_{12}/S_2) = .8$$

$$P(c_{11}/N) = .5, \qquad P(c_{12}/N) = .5$$

$$P(S_1) = .6, \quad P(S_2) = .4, \quad P(N) = .5 .$$

Since the observation was a blue ball, i. e., characteristic C_1 was observed in state c_{12}, the probabilities $P(c_{12}/NS_1)$ and $P(c_{12}/NS_2)$ are required. When N is present with S_1, there are 14 red balls and 6 blue balls.

Thus,

$$P(c_{12}/NS_1) = \frac{6}{14+6} = .3$$

Likewise,

$$P(c_{12}/NS_2) = \frac{13}{7+13} = .65$$

Therefore,

$$P(c_{12}/H_1) = .3(.5) + .1(.5) = .200$$

$$P(c_{12}/H_2) = .65(.5) + .8(.5) = .725$$

and the resulting posterior distribution over H_1 and H_2 is

$$P(H_1/c_{12}) = \frac{.200(.6)}{.200(.6) + .725(.4)} = .293$$

$$P(H_2/c_{12}) = \frac{.725(.4)}{.200(.6) + .725(.4)} = .707$$

It should be noted that if the presence of the noise factor had not been taken into account, the resulting posterior distribution would have been

$$P(H_1/c_{12}) = \frac{P(c_{12}/S_1)P(S_1)}{\sum_1^2 P(c_{12}/S_i)P(S_i)} = \frac{.1(.6)}{.1(.6) + .8(.4)} = .158$$

$$P(H_2/c_{12}) = .842.$$

In any event, ignoring noise factors will result in an incorrect posterior distribution. Thus, if it is possible that noise factors may occur, this should be taken into account when computing the posterior distribution.

APPENDIX E

THE APPLICATION OF DIGITIZED LOGIC TO NONNUMERICAL PROBLEMS IN INTELLIGENCE ANALYSIS

A computational technique, referred to in this paper as "digitized logic," makes possible an attack on certain nonnumerical problems which may arise in intelligence analysis and which would be difficult, if not impossible, to solve by intuitive "logical" thinking.

This technique can be applied in conjunction with a Bayesian formulation and could be useful in examining the date for logical consistency, redundancy, and to determine the validity of conclusion drawn from the information. The foundations of the method lie in that part of mathematical logic known as "the Boolean algebra of propositions."

In a sense, the application of this mathematical theory to the problems that will be discussed is not new. What is new about this approach is that the technique uses sentences in a digitized form and can be structured for computer processing. The computerizable application of Boolean algebra to nonnumerical problems such as might arise in intelligence analysis was first described by Robert S. Ledley[1] in a number of research papers which have appeared in the last eight years. [Ledley, 1954; Ledley, et al., 1960]. While the same methods have been applied with success in some fields, it seems that up to now they have not been tried, on any appreciable scale, as aids in solving problems in intelligence analysis.

Foundations of the Computational Method

In the following discussion, A, B, C, . . . denote sentences about each of which it is meaningful to say that the content is either true or false. One can refer to A, B, C, . . . as "elementary elements" which shall be combined to form "combined elements" according to the following rules:

[1] President of the National Biomedical Research Foundation. Formerly consultant mathematician to the National Bureau of Standards and on the staff of George Washington University.

A + B, (read "A or B"), is true if and only if A is true, or B is true, or both are true.

A · B, (read "A and B"), is true if and only if A is true and B is true.

$\overline{A}$, (read "not A"), is true if and only if A is false.

A → B, (read "if A then B"), is false if and only if A is true and B is false. Otherwise A → B is true.

Thus, the sentences A, B, C . . . can be combined by the operations + , · , ‾ , → and the truth value of the "combined element" is determined by the above rules according to the truth values of the elementary elements. Both the elementary and combined elements shall be referred to as "propositions."

By definition, the proposition X is equivalent to the proposition Y, written, X=Y, if and only if X and Y have the same truth value for any assignment of truth values to the elementary elements. Note that some propositions are always true, regardless of the truth value of the elementary elements. For example, $X+\overline{X}$ is always true, for any proposition X. Similarly, the proposition $X\cdot\overline{X}$ is always false. Whether or not two propositions are equivalent can be verified by constructing a truth table.

For example, letting 0 stand for "false" and 1 for "true," the following truth table (left) shows the equivalence of the propositions "X → Y" and "$\overline{X}$ + Y:"

X	0	1	0	1
Y	0	0	1	1
X→Y	1	0	1	1
$\overline{X}$ + Y	1	0	1	1

ALL POSSIBLE ASSIGNMENTS OF TRUTH VALUES FOR THE PAIR X,Y.

To see the motivation for the use of the implication arrow → in the proposition X → Y, note the colloquial use of implications in English: when one says "X implies Y" or, meaning the same thing, "Y can be inferred from X" or "Y is implied true by X," one means that if X is true, then Y is true; and the case "X true, Y false" cannot occur. However, one should note that if Y is true, X may or not be true and still the assertion "X implies Y" is true, as long as the truth of Y can be inferred from the truth of X.

In applications, one is interested in the set of propositions generated by a finite number of elementary elements. For the sake of simplicity, consider the set of all propositions that can be formed from two elements A and B. Two such propositions are called "distinct" if and only if they are not equivalent. Consider the proposition A + B. Letting 0 stand for "false" and 1 for "true," the truth values for A + B can be displayed by the following truth table:

A	0	1	0	1
B	0	0	1	1
A + B	0	1	1	1

The columns of the first two rows of the table correspond to the four possible cases, or assignments of truth values to A and to B. The four digits in the third row represent the truth value of A + B corresponding to each case. The two sequences of zeros and ones corresponding to A and to B are called a "basis," and the sequence of zeros and ones in the third row is called the "designation number" of A + B with respect to this basis. There are as many bases as there are ways of ordering the columns in the table. In this paper, however, designation numbers will always be discussed with respect to the basis shown in the table because of its visual simplicity. Thus, for 4 elementary elements A, B, C, D, the basis would be

A	0101	0101	0101	0101
B	0011	0011	0011	0011
C	0000	1111	0000	1111
D	0000	0000	1111	1111.

The designation number of any proposition can be written down, using the following rules of logical addition, multiplication and inversion: $1+1 = 1$, $1+0 = 1$, $0+1 = 1$, $0+0 = 0$; $1 \cdot 1 = 1$, $1 \cdot 0 = 0$, $0 \cdot 1 = 0$, $0 \cdot 0 = 0$; $\bar{1} = 0$, $\bar{0} = 1$. Thus, assuming only 2 elementary elements, the designation number of $A \cdot B$ is 0001, that of $A+\bar{A}$ is 1111 and of $A \cdot \bar{A}$ is 0000. Furthermore, it can be shown that to any such sequence of zeros and ones, there corresponds a proposition formed from A and B with this sequence as its designation number. There are systematic procedures for searching combinations of the elementary elements and their negations to produce propositions in various canonical forms whose designation number is the given sequence. Recalling the meaning of equivalence of two propositions, one sees that there are $2^4 = 16$ distinct propositions that can be formed from two elementary elements.

Consider now what it means, in terms of designation numbers, to say that "Y is implied true by X." For each case that X is true Y must be true and, therefore, one can assert, that Y is implied true by X if and only if the designation number of Y has a 1 in at least the same positions as that of X.

Before turning to an example, it is necessary to introduce the notion of redundancy. If $F_1, F_2, \ldots\ldots\ldots F_n$ is a set of propositions, then any one F_i is called redundant if it is implied true by the truth of any one or more of the other propositions in the set. It should be clear how one can interpret this condition in terms of the designation numbers.

An Example Concerning Submarines

It is assumed that an intelligence agency has compiled the following information about Soviet submarines of the oceangoing type: (The "S-type missiles" referred to in the report are hypothetical and can be considered as referring to a new type of missile which is known to have been recently developed.)

1. A conventionally powered medium range sub does not carry S-type missiles.
2. An S-type missile-carrying sub is either of the H class or not of the H class but has long range.
3. A conventionally powered sub is not of the H class.
4. If a sub is not of the H class and either has medium range or is conventionally powered, then it does not carry S-type missiles.
5. A long range sub carrying S-type missiles is of the H class.

It is assumed, furthermore, that the information comprising the report has come from fairly, though not completely, reliable sources. The method of digitized logic described in the preceding sections will be used to investigate (1) logical consistency of the five statements, (2) redundancy, and (3) validity of conclusions. By this method the report can be analyzed in such a way as to obtain the maximum amount of information from it.

It is known that an oceangoing submarine can be considered as either having medium range or long range, but not both. Similarly, it is either conventionally powered or nuclear powered, but not both. Furthermore, a submarine is either of the H class or not of the H class and it either carries S-type missiles or does not carry S-type missiles. With this in mind, the following elementary elements are introduced:

M = the sub has medium range

H = the sub is of the H class

S = the sub is carrying S-type missiles

C = the sub is conventionally powered.

The negations of the elementary elements are

$\overline{M}$ = the sub has long range

$\overline{H}$ = the sub is not of the H class

$\overline{S}$ = the sub is not carrying S-type missiles

$\overline{C}$ = the sub is nuclear powered.

The following table, showing the basis as well as the designation numbers of the negations of the elementary elements, is constructed:

M	0101	0101	0101	0101
H	0011	0011	0011	0011
S	0000	1111	0000	1111
C	0000	0000	1111	1111
$\overline{M}$	1010	1010	1010	1010
$\overline{H}$	1100	1100	1100	1100
$\overline{S}$	1111	0000	1111	0000
$\overline{C}$	1111	1111	0000	0000 .

The five statements comprising the report are now written in symbolic form and their designation numbers are found.

1.	$C \cdot M \rightarrow \bar{S}$	: $\overline{C \cdot M} + \bar{S}$	: 1111 1111 1111 1010
2.	$S \rightarrow H + \bar{H} \cdot \bar{M}$	: $\bar{S} + (H + \bar{H} \cdot \bar{M})$	: 1111 1011 1111 1011
3.	$C \rightarrow \bar{H}$	: $\bar{C} + \bar{H}$	: 1111 1111 1100 1100
4.	$\bar{H} \cdot (M+C) \rightarrow \bar{S}$	: $\overline{\bar{H} \cdot (M+C)} + \bar{S} = H + (\overline{M+C}) + \bar{S}$	: 1111 1011 1111 0011
5.	$\bar{M} \cdot S \rightarrow H$	: $\overline{\bar{M} \cdot S} + H = M + \bar{S} + H$	: 1111 0111 1111 0111

If F_i = the ith proposition above, then

$$\pi F_i = F_1 \cdot F_2 \cdot F_3 \cdot F_4 \cdot F_5 \; : \; 1111 \quad 0011 \quad 1100 \quad 0000.$$

By examining the designation number of πF_i, which has 1's in eight positions, one can conclude that the five statements are logically consistent, since there is at least one assignment of truth values to the elementary elements (in fact, there are eight assignments) for which πF_i is true. If this were not the case, that is, if the designation number of πF_i consisted only of zeros, then it would not be possible to accept all the statements as true, and one would conclude that they are logically inconsistent.

By examining the designation numbers one sees readily that F_2 is implied true by F_4 (since the designation number of F_2 has 1's in at least the same positions as that of F_4), and F_1 is implied true by $F_4 \cdot F_3$. Thus, F_1 and F_2 are redundant if F_3, F_4 and F_5 are true. The report, therefore, reduces to the statements 3., 4., and 5. Accepting these statements as true, and denoting them by F_3, F_4, F_5 respectively, the report is equivalent to the assertion that $F_3 \cdot F_4 \cdot F_5$ is true. The designation number of $F_3 \cdot F_4 \cdot F_5$ is 1111 0011 1100 0000 which, of course, is the designation number of πF_i.

The most general conclusion that can be drawn, in the sense that every other valid conclusion is a statement that is implied true by it, is obtained by finding a proposition in a relatively simple form which is equivalent to $F_3 \cdot F_4 \cdot F_5$, i.e., which has the same designation number as $F_3 \cdot F_4 \cdot F_5$. As was previously mentioned, this can be done in a systematic way, and the details of the method can be found in Ledley [1954; 1960]. It can be checked that the proposition $\overline{C}\cdot(\overline{S}+H\cdot S) + C\cdot\overline{H}\cdot\overline{S}$ is equivalent to $F_3\cdot F_4\cdot F_5$, and thus it can be concluded that a sub is either (nuclear and either carries S-type missiles and is of the H class or does not carry S-type missiles) or (it is conventionally powered and not of the H class and does not carry S-type missiles).

Any asserted conclusion in terms of the elementary elements and their negations can be checked automatically for validity. It is understood that a conclusion is valid, if it is a statement that is implied true by $F_3\cdot F_4\cdot F_5$. For example, the assertion "It is not possible for a sub to be conventionally powered and carry S-type missiles" is a valid conclusion. The validity follows from the fact that the proposition $\overline{C\cdot S}$ is implied true by $F_3\cdot F_4\cdot F_5$, as can be seen by examining the designation numbers:

$$\overline{C\cdot S}: \quad 1111 \quad 1111 \quad 1111 \quad 0000$$

$$F_3\cdot F_4\cdot F_5 : \quad 1111 \quad 0011 \quad 1100 \quad 0000 .$$

On the other hand, the assertion "Every nuclear powered sub is of the H class" is not a valid conclusion, since H is not implied true by $(F_3\cdot F_4\cdot F_5)\cdot\overline{C}$, as can be seen from the designation numbers:

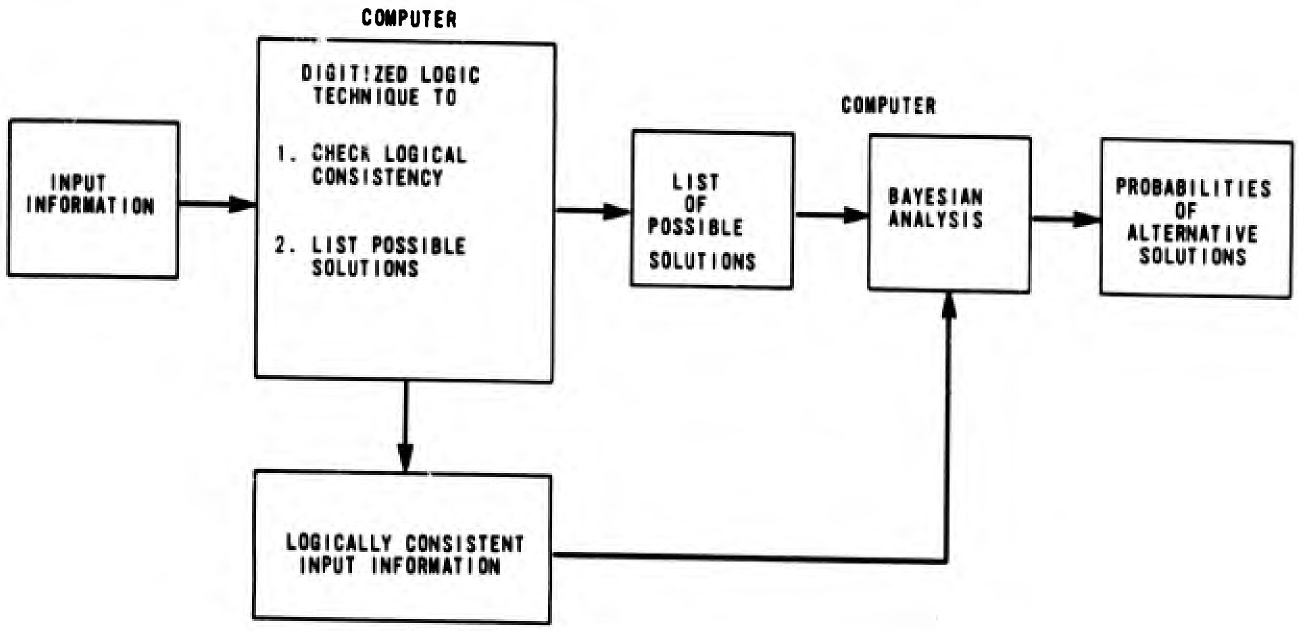
$$H : \quad 0011 \quad 0011 \quad 0011 \quad 0011$$

$$(F_3\cdot F_4\cdot F_5)\cdot\overline{C} : \quad 1111 \quad 0011 \quad 0000 \quad 0000 .$$

Conclusions

A computational technique based on the Boolean algebra of propositions and utilizing sentences in a digitized form has been described in the preceding sections. This technique makes possible a rigorous analysis of certain "logical" problems which might arise in the analysis and evaluation of intelligence and military reports. It can also be useful in the initial stages of Bayesian analysis, by providing a means for mechanically checking input information for logical consistency and for determining the set of possible solutions to the given problem. When applied in this way, one can consider the technique of digitized logic as being incorporated within intelligence analysis in a similar way as it is employed in computer aided medical diagnosis [Ledley, et al., 1960].

The diagram below shows how the technique of digitized logic might be applied to the first of the two principal steps involved in analyzing a problem via Bayes Theorem.



That is, the application of Bayesian analysis presupposes that (a) information (data) has been collected that is logically consistent and (b) on the basis of the collected data, a set of possible solutions has been determined with the property that one and only one of the possible solutions is the true solution. It is these two aspects of the analysis procedure to which the method of digitized logic might profitably be applied.

It should be emphasized, however, that certain conditions must be satisfied in order to computerize this initial aspect of the Bayesian analysis procedure. The problem must be such, to put it broadly, that all possible relevant outcomes must be defined and stored in the computer memory. This means, specifically, that all data that might be entered as input information to the problem must be categorized and stored, as well as all possible solutions that would conceivably be considered. The elementary elements for the problem would be defined during this initial phase. Next, the pertinent relationships between combinations of data items and possible solutions must be stored. From these would be determined a set of propositions in terms of the elementary elements which would be accepted as true for the analysis. When the input data are entered, the logical consistency of it with the stored true propositions would be checked, and a list of valid conclusions in the form of possible solutions would be generated by an algorithm based on the theoretical considerations described in the preceding sections.

BIBLIOGRAPHY

Anderson, J. M., et al., Natural Deduction, The Logical Basis of Axiom Systems, Wadsworth Publishing, Belmont, Calif., 1962.

Dodson, J. S., "Simulation Systems Design for a TEAS Simulation Research Facility," Planning Research Corporation, Los Angeles, PRCR-194, Nov. 1961(AD 272 844).

Edwards, W., "Dynamic Decision Theory and Probabilistic Information Processing," Human Factors 4, (2), Apr. 1962.

"Probabilistic Information Processing in Command and Control Systems," University of Michigan, Institute of Science and Technology, Ann Arbor, ESD-TDR-62-345, Feb. 1963, (AD 299 291).

"Nonconservative Probabilistic Information Processing Systems," University of Michigan, Institute of Science and Technology, Ann Arbor, ESD-TR-66-404, Dec. 1966 (AD 647 092).

Edwards, W., et al., "Special Issue on Revision of Opinions by Man and Man-Machine Systems," IEEE Trans. on Human Factors in Electronics HFE-7, (1), Mar. 1966.

Good, I. J., Probability and the Weighing of Evidence, Hafner Publishing Company, New York, 1950.

The Estimation of Probabilities: An Essay on Modern Bayesian Methods, MIT Press, 1965

Kent, S., Strategic Intelligence for American World Policy, Anchor Books, Hamden, Conn., 1965.

Ledley, R. S., "Mathematical Foundations and Computational Methods for a Digital Logic Machine," J. Ops Res. Soc. Amer. 2, (3), Aug. 1954.

Digital Computer and Control Engineering, McGraw Hill, New York, 1960.

Ledley, R. S., et al., "The Use of Electronic Computers in Medical Data Processing: Aids in Diagnosis, Current Information Retrieval, and Medical Record Keeping," IRE Trans. on Medical Electronics, Jan. 1960.

Little, W. W., et al., Applied Logic, The Riverside Press, Cambridge, Massachusetts, 1952.

Newman, J. R., "Extension of Human Capability Through Information Processing and Display Systems," System Development Corporation, Santa Monica, SP-2560/000/00 Dec. 1966 (AD 645 435).

Newman, J. R., et al., "Experiments in Computer-Aided Inductive Reasoning," System Development Corporation, Santa Monica, TR-3227, Dec. 1966 (AD 645 422).

Polya, G., How to Solve It, A New Aspect of Mathematical Method, Doubleday Anchor Books, Garden City, 1945.

Patterns of Plausible Inference, Princeton University Press, Princeton, New Jersey, 1954.

Schum, D. A., "Inferences on the Basis of Conditionally Non-independent Data," Ohio State University Research Foundation, Columbus, AMRL-TR-65-161, Dec. 1965 (AD 630 662).

"Inference on the Basis of Conditionally Nonindependent Data," Ohio State University, Columbus, AMRL-TR-66.215, (reprint from J. of Exp. Psy. 72, (3), 1966), Sept. 1966 (AD 645 588).

Schum, D. A., et al., "The Influence of Experience and Input Information Fidelity Upon Posterior Probability Estimation in a Simulated Threat-Diagnosis System," Ohio State University, Laboratory of Aviation Psychology, Columbus, AMRL-TR-65.25, Apr. 1965 (AD 615 758). (a)

"Further Investigation of the Effects of Reduced Input Data Fidelity Upon the Determination of Posterior Probabilities in a Simulated Threat-Diagnosis System," Ohio State University Reasearch Foundation, Columbus, AMRL-TR-65.233, Dec. 1965 (AD 631 781). (b)

"Research on a Simulated Bayesian Information Processing System," Ohio State University Research Foundation, Columbus, AMRL-TR-66-78 (reprint from IEEE Trans. on Human Factors in Electronics HFE-7, (1), 1966), Mar. 1966. (AD 633 904).

Southard, J. F., et al., "An Application of Bayes Theorem As a Hypothesis-Selection Aid in a Complex Information-Processing System," Ohio State University, Laboratory on Aviation Psychology, Columbus, AMRL-TDR-64-51, Aug. 1964 (AD 607 256).

Tukey, J. W., "Conclusion vs. Decisions," Technometrics 2, (4), Nov. 1960.

Weir, T. D., et al., "Proceedings of the Second Missile and Space Division Seminar on Bayes Theorem and Its Application to Reliability Measurements," General Electric Company, Philadelphia, 65SD995, Jan. 1966, (AD 481 645).

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

HRB-Singer, Inc.
State College, Pennsylvania

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

The Role of Plausible Reasoning Within Military Intelligence: An Application of Bayes Theorem as a Model for Problem Solving

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Charles R. Blunt, Peter T. Luckie, Erika A. Mares, Dennis E. Smith

6. REPORT DATE

May 1967

7a. TOTAL NO. OF PAGES

116

7b. NO. OF REFS

25

8a. CONTRACT OR GRANT NO.

N00014-66-C0230

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

401511-R-2

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Distribution of this Document is Unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Information Systems Branch
Mathematical Sciences Division
Office of Naval Research
Washington, D.C. 20360

13. ABSTRACT

The military intelligence analyst must cope with uncertainty during his problem-solving efforts. This uncertainty maps directly through the analysis and synthesis processes and affects his confidence in the output. A formal methodology for analysis can reduce some of these unfavorable effects while augmenting the analysis problem solving capabilities. The role of plausible forms of logical reasoning within intelligence analysis is reviewed, followed with the introduction of Bayes Theorem as a model for intelligence analysis. The conjecture is made that Bayes Theorem can also serve as the nucleus of a formal methodology. The application of Bayes Theorem to several types of problems is demonstrated. However, the implementation of such a model as the nucleus of a complete analysis methodology is hindered by several significant problems. Some of the prime hindering aspects are delineated and discussed.

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Probability Theory						
Statistics						
Bayes Theorem						
Military Intelligence						
Information Processing						
Mathematical Prediction						

Unclassified

Security Classification