(12)

AD A115067

METHODS OF ESTIMATING

STRATEGIC INTENTIONS

Frank J. Stech
Kenneth C. Hoffman

Research Report
May 1982

Prepared for:

Office of Naval Research
Department of the Navy
Contract No. N00014-81-C-0508

Prepared by:

MATHTECH, Inc.
4630 Montgomery Avenue
Bethesda, Maryland 20014
Telephone: (301) 657-1610

CONTENTS

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Block 20. Abstract (continued)

1. A description of the estimation process and of related problems including the outline of a conceptual representation of the process against which specific methods may be evaluated.

2. A description of analytical aids useful in the intention estimation process including judgmental methods as well as analytical aids using techniques of extrapolation, structural, and causal analysis.

3. A comparison of analytical aids in the context of the conceptual estimation process.

4. The evaluation of specific analytical aids.

5. Development of a catalogue of 18 analytical aids matched by methodology to steps in the conceptual estimation process.

Unclassified

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A115 067
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) METHODS OF ESTIMATING STRATEGIC INTENTIONS	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Frank J. Stech, Kenneth C. Hoffman	8. CONTRACT OR GRANT NUMBER(s) N00014-81-C-0508
9. PERFORMING ORGANIZATION NAME AND ADDRESS MATHTECH, Inc. 4630 Montgomery Avenue Bethesda, Maryland 20014	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, code 430B 800 North Quincy Street Arlington, Virginia 22217	12. REPORT DATE May 1982
	13. NUMBER OF PAGES 143
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Unlimited

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Capabilities & Intentions	Methods of Analysis
Cognitive Processes	Military Intelligence
Intelligence Estimates	Military Psychology
Intention Estimation	Naval Intelligence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The estimation of strategic intentions is an important and evolving part of the intelligence process. The research program described in this report was undertaken to catalogue a number of quantitative and behavioral methods of analysis used in support of various phases of the estimation process. Thus, this research builds on earlier work performed by MATHTECH in which the intention estimation process was defined and various supporting methods identified.

Specific tasks undertaken in this research project are:

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

I. SUMMARY

The estimation of strategic intentions is an important and evolving part of the intelligence process. The motivation for focusing on estimates of peacetime Soviet naval intentions in previous research studies performed by MATHTECH arises from the increased contacts between U.S. or Western fleets and the Soviet Navy. These more frequent and extensive contacts in off-shore defensive zones, sea lanes, and in proximity to Third World countries may occur in crisis situations or in situations with the potential for escalation into a major crisis because of the behavior and responses among the parties involved. The way in which the navies of the West and the Soviet Union conduct themselves in these peacetime and crisis encounters has a considerable impact on subsequent escalation and responses. The previous research performed by MATHTECH dealt with an assessment of methods for estimating peacetime Soviet naval intentions. In the course of that research a number of quantitative and behavioral methods were identified (Stech, 1981). A follow-on research program, the results of which are described in this report, was then undertaken to catalogue those methods in relation to estimation problems and issues. In addition criteria were established to guide the selection of appropriate methods to deal with specific estimation issues. As background for this research, a quite detailed description of the intention estimation process was developed.

The individual tasks that were performed in this research project are:

TASK 1. <u>Description of the Estimation Process and of Problems that Arise in the Process</u>:

The conceputal model developed in this task to provide the context for intention estimates includes 8 sequential steps;

- o preceptions (reactions to the environment and situation based on capabilities and behavior),
- o estimates and assessments of current and future capabilities, risks, actions and reactions,
- o integration and judgements related to evaluation of alternate hypotheses,
- o definition of intentions based on logic embodied in previous steps,
- o formulation of a strategic concept outlining courses of action,
- o strategy statement to guide future action,
- o development of plan to guide actions, and
- o execution of the plan.

TASK 2: Description of Analytical Aids:

Analytical aids useful in the intention estimation process vary considerably in scope, methodology, and logic. Consistent with the complexity and range of possibilities of intentions, these aids are similarly complex and wide ranging; there are no individual methods that are to be preferred even in a given class of analytical problems. The methods vary considerably in their dependence on data, judgements, and relational or behavioral descriptions, as well as in their dependence on quantitative techniques. They range from purely subjective or judgemental methods to highly quantitative probabilistic methods. A general categorization of analytical aids is as follows:

- o Judgemental Methods; supported by attributes of sagacity, control, and acumen,
- o Analytical Aids; quantitative methods that support specific elements of the estimation process,
- o Extrapolation Models; based on past events and data, and
- o Structural models; taking into account structural realities and changes in leadership, physical capabilities, and other influencing factors.

The third and fourth categories make extensive use of quantitative methods, in one case involving trend extrapolation, in the other attempting to represent the physical and behavioral relationships that affect intentions in a given time and place. Clearly, these categories are not mutually exclusive and intention estimates will be based on some combination of methods drawn from these categories. The art of estimation is to select and blend appropriate methods and the objective of this research project is to assist in that process through the cataloguing and organization of analytical aids used to support and extend human judgement and interpretation.

The specific analytical aids covered in this review are listed as follows in relation to the elements of intention estimating:

1. Perceiving Data: statistical sampling, record of events, coding categories.

2. Weighting Data: policy capturing, Bayesian methods, correlation and variance analysis.

3. Characterizing Data: memory aids, fuzzy sets, factor analysis.

4. Assessing Covariations: actuarial models, backcasting, bootstrapping.

5. Cause and Effect Assessment: causal search, causal analysis, search trees, stepping analysis, hypothesis, regression analysis.

6. Predictions: Backcasting, boot strapping, decision analysis.

7. Theories: scenarios, judgemant heuristics, etc.

TASK 3: Comparison of Aids and Problems:

This task compares the analytical aids in the context of the intention estimation process. General criteria that apply to the comparison include:

- o scope and content of method in relation to estimation problem. (can method incorporate information on capabilities, force structure, deployment, logistics, manpower, risk propensity, strategic patterns, responsive patterns, etc),
- o descriptive power
- o inferential power
- o treatment of constraints (operational, tactical, and strategic),
- o logic structure (deterministic, probabilistic, inference, etc),
- o transparency of relationships and assumptions,
- o analytical method (state of development and prior experience),
- o application and documentation of method,
- o validation, verification, and peer review of method,
- o cost of initial system and upkeep,
- o treatment of risk and uncertainty, and
- o sensitivity (robustness of result and effect of changes in assumptions).

TASK 4: Evaluation of Aids:

Problems with particular analytical aids and an assessment of strength and weaknesses are discussed in the context of the intention estimation process with reference to historical experiences.

TASK 5: Catalogue of Aids and Process Summary:

A summary table relating specific analytical aids, analytical problems, and criteria was developed and is included in this report.

Chapter II of this report covers Tasks 1 and 2, the description of the intention estimation process and the description of analytical aids pertinent to that process. Chapter III covers Tasks 3 and portions of 4, the comparison of analytical aids and their evaluations. The summary catalogue of aids is presented in Chapter IV.

Major conclusions that are drawn from this project are:

1. Estimation of intentions may usefully draw upon a variety of analytical aids.

2. The selection of specific analytical aids and their application to deal with estimation problems depends upon such factors as scope of problem, data availability, and timing. This process cannot be described in cook-book fashion with any absolute guarantees of success. While the process depends a great deal on creativity, criteria have been developed and can prove quite useful as a checklist for the selection and use of aids.

3. The criteria and evaluation methods employed in this project to catalogue and evaluate aids in the estimation of peacetime Soviet naval intentions are sufficiently general to apply to a wide range of intelligence, logistic, and strategic analytical aids of interest to the office of Naval Research.

II. DESCRIPTION OF THE INTENTION ESTIMATION PROCESS AND ANALYTIC AIDS

The estimation of intentions by intelligence analysts can be characterized in a variety of ways and in terms of: the estimation steps taken by the analysts, the analysts' logic and reasoning processes, and the types of evidence and information analysts might consider to form an intelligence estimate of intentions. This chapter uses these three outlooks on intention estimation to describe the estimation process and to identify problems that might affect various steps of the process.

A. MODEL OF INTENTIONS

This model is intended to provide the context or background for the comparison and evaluation of analytical aids. Intentions can be given very complex definitions, or simple ones; e.g., Lawrence (1972: 83) writes that intentions are "desirous foreknowledge or expectant desire." Basically intentions and psychological states that involve

> Mental images of future events in which the intender pictures himself as a participant and makes choices as to which image he will try to bring to reality. (Griffin 1976: 5).

An intelligence professional (Gazit, 1980) has divided intentions estimation into identifying a decision already taken, analyzing responses taken in reaction to actions by others, and analyzing the outcome of a developing, ongoing situation. Behaviorally, intentionality manifests several features: expectations of the outcome of an act, selection among alternatives for the fulfillment of a goal, sustained effort in a given direction in the sequence of actions taken to make the resulting state resemble the expected state, and a flexible plan for using alternative means and actions to compensate and correct for discrepancies between expectations and results (Foster and Brandes, 1980: 326). Longer, philosophical dissertations on the meaning of "intentions" (Anscombe, 1969; Lawrence, 1972; Stech, 1979, ch. 2) are interesting but not essential to outlining a descriptive model of the elements of intentions which an intelligence analyst may need to analyze.

In the following model we are assuming that the adversary, or enemy, or ally, that is, the party whose intentions the intelligence analyst hopes to ascertain, is not psychopathological. Model of intentions can be created for analyzing the psychopathologic adversary (e.g., Langer's, 1972, model of the "mind of Adolf Hitler" created during World War II to aid Allied intelligence estimations), but each such model is a special case.

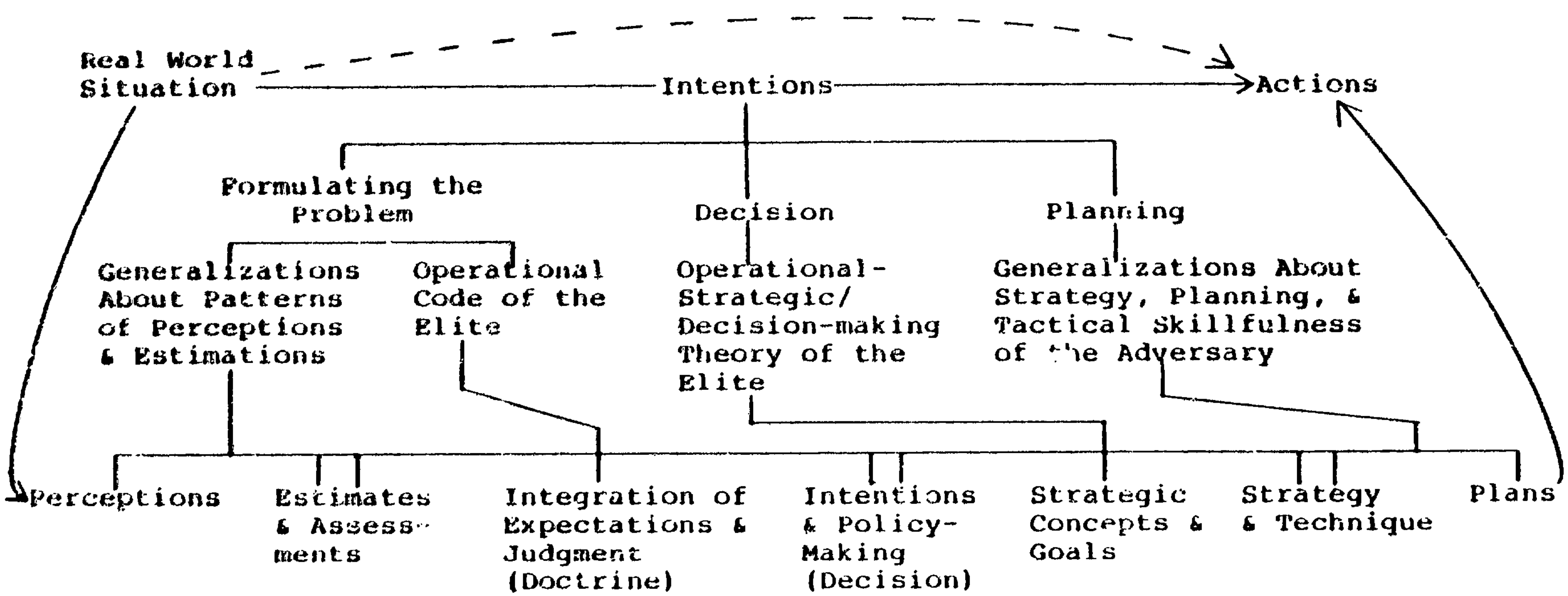
Our objective is to outline a model that can hope to deal with the estimation of the intentions of most normal actors, functioning without obvious psychotic disturbances. We do not assume that the adversary is wholly rational, in the logical, mathematical, or economic meaning of "rationality." In fact, one limitation of the "rational actor model" (as used, e.g., by Allison, 1971, to assess the Cuban Missile Crisis) is that no human actors are strictly rational except when they do mathematics, symbolic logic, or economic utility calculus, and even then biases, errors, and paradoxes can lead to behavior that is not completely rational. The only real requirement for the model is that the actor, whose behavior the analyst is attempting to predict through the analysis of intentions, not be widely perceived as totally irrational, i.e., the model does not work if the adversary is crazy.

The qualification in the preceding paragraph is fairly important from the analysts' standpoint. The customers of intelligence estimates, particularly those which hope to estimate intentions, will often justify their ignoring such estimates with the assertion that Prime Minister So-and-so is "unpredictable" or, in the extreme case, President Such-and-such is simply "crazy". Neither argument is tenable; the first asserts that we are as wise as we will ever be, and that the fact (if it is a fact) that no accurate predictions were made in the past means none will ever be made in the future. The second ignores the fact that truly crazy people are quite easy to predict, use very circumscribed routines of behavior, and suffer from the lack of behavioral flexibility and choice, i.e., their behavior is compelled, compulsive, and hence predictable. The arguments may justify scepticism in the predictions, but not complete ignorance of them.

This model (Figure 1) has two main sources. Its general characteristics come from a model used in World War II to predict Axis intentions through the analysis of propaganda, and it has been subsequently systematized by George (1959). Grafted onto this basic skeleton are concepts taken from the general areas of decision-making and problem-solving behavior. That is, the adversary is perceived to be solving a problem: what actions are needed to reach some goal; and making a decision: what action should be implemented to reach the goal. The analyst's task is to try to penetrate the adversary's problem-solving and decision-making processes and procedures, i.e., to recreate the steps the adversary has taken to solve the problems and make a decision.

The top line of the figure shows intentions, which can only be inferred, as intervening between the real world situation and the adversary's actions, both of which can be observed directly, as well as inferred from other evidence. The dotted arrow reflects the fact that some of the adversary's actions are not intentional but instead are accidental or miscarriages of intended actions. Unintended action is difficult to separate analytically from intended action; there is a tendency to attribute all action to some motive or intention of the

Figure 1. Conceptual Model of Intentions



actor (see Stech, 1980: 104-120 for a discussion of this tendency in naval analysis).

The second line of the figure reflects the approach taken here to intentions, that intentional behavior reflects problem formulation, decision-making, and then planning for the execution of the decision taken. These steps are conditioned by the adversary's perceptions of the real world situation (i.e., the adversary considers reality in formulating the problem) and are presumed to influence the actions taken.

The sequential steps of problem formulation (logic of estimation, estimation, and decision/planning have been the subjects of a variety of theories and concepts used by political scientists to assess governmental actions. Some of these political science formulations are reflected on the third line of the figure. For example, generalizations about the typical perceptions and estimations made by a nation faced with various situations contribute to understanding how the nation formulates problems, and offers information on the important dimensions and value criteria the nation will tend to use to structure its decisions. Similarly the operational code of a national elite provides information on the instrumentalities and basic strategic values of the top decision-makers. Again, such information contributes to understanding how a real world problem might be solved, or more basically, what will be perceived as a problem in the first place.

At the bottom of the figure are seven steps which reflect the process that intervenes between the stimulus of the real world situation and the action taken in response. These are represented in Figure 2 as sequential steps indicating activities preparatory to and subsequent to the definition of intentions. An eighth step, action, has been added to Figure 2 for completeness. First the real world produces an immediate perceptual response. The adversary reacts to

Figure 2. Sequential Decision Process
(Context for Intention Estimation)

1. Perceptions →	2. Estimates + Assessments →	3. Integration + Judgments →	4. Definition of Intentions →	5. Strategic Concept →	6. Strategy →	7. Plans →	8. Action
Reactions to environment, own actions, opponents' actions, capabilities, etc.	Estimates of current + future resources, risks, actions, and likely outcomes. Assessment of current and future motives, utilities, and values. Done for own side + opponent's side	Integration of resources, risks and values. Evaluation of alternative actions and impact of various contingencies. Raking of goals.	Selection of a subset of desired and possible goals and outcomes (objectives) given available resources, feasible actions, and anticipated contingencies.	Formulation of general objectives, and general courses of action necessary.	Statement of specific actions and subobjectives in order of priority for action. Outline of permissible contingent actions.	Program of Actions in sequence, assignment of responsibilities, specification of subobjectives, allocation of resources. Definition of contingent actions.	Execution of Program. Reactions to contingenci.
Formulating the Program-Estimation Logic			Intentions	Decision and Plan of Action			Action

the perceived environment specifically by noting needs, reacting to its own previous actions (e.g., a previous failure to satisfy a need might provide the stimulus for subsequent action), the actions of opponents to block or facilitate need satisfaction, the capabilities of the adversary and of opponents to satisfy needs through actions. In this first step the adversary is selecting aspects of the environment to attend (e.g., what is the need, what is the problem), coding those aspects along important dimensions, and generally characterizing the nature of the environment into those parts that require action or attention (i.e., identifying needs), and those that do not.

The second step of the intention process involves the organization of perceptions, estimates, and expectations into orderly categories of information. Among these categories are the following: estimates of current and future resources, both for one's own side and for the opposite side, estimates of possible actions on each side and the risks attached to them and also the utilities attached to the outcome of the various actions. The latter information requires the analyst to form some impression of the adversary's value system. The analyst needs to estimate how the adversary evaluated outcomes for each side. It is helpful to separate the elements of this step into those that deal with issues of fact (which can also be thought of as predictions) and those that deal with issues of value and evaluation. For example, the enemy may be contemplating five different modes of attack. Several issues of fact must be estimated: what resources are needed for each mode, what is the likelihood of each mode succeeding, what is the opponent capable of doing in reaction to each mode and what are the probabilities of each of those reactions. Additionally, various issues of value must be assessed: how willing is the adversary to expend resources at various levels (i.e., what is the cost of the resources), how good are the various possible outcomes of the different attack modes (i.e., what are the benefits) and what are the values to the parties on both sides of the enemy's reactions to the various possible models. To make these assessments the analyst must form some impression of how the adversary makes estimates of facts and probabilities and how he assesses values, both for the adversary's own actions and for his estimates of the analysts' country's reactions and initiatives.

The third step in the intention process integrates the information on facts and values to determine what actions would be better than other actions and the likelihood of various outcomes and benefits. Possible goals are then ranked according to value, risk, probability, cost, or any other criteria of importance to the adversary decision-making system. In this step, the adversary is formulating expectations and clarifying those "mental images of future events" which might be brought to reality.

To assess this step in the intention process the analyst may have to consider the adversary's decision-making rules. For example, suppose one mode of attack offers the adversary low risks of loss but little in the way of tactical advantage, i.e., low gains. Another mode of attack offers spectacular gains, but at the possible risk of huge losses. If the adversary allows these dimensions of risk and gain to compensate for each other (i.e., a lot of gain makes up for a lot of risk) the riskier, more advantageous course might receive the higher ranking. However, if the adversary cannot affort such compensating trade-offs (e.g., because resources are so limited that any major risk of loss is intolerable) then an unacceptable state on a single dimension, such as risk, may be sufficient to eliminate that option from further consideration. Obviously, a similar conclusion might be reached by the analyst who attends to whether the adversary is generally a gambler or conservative. The decision analysis approach has the added advantage that it attempts to identify <u>why</u> the adversary might be conservative or intrepid.

One aspect of the adversary which may play an important role at this step is doctrine, those prescriptions as to what types of action outcomes should be highly valued (e.g., sieze the high ground) or avoided (e.g., never concentrate in range of enemy guns). Doctrine can be viewed as codified decisions, or stereotyped integrations of estimates and assessments from past experience. Such codes may offer important clues as to how the adversary will integrate the estimates and assessments bearing on the current problem It is always possible, however, that the adversary will flout doctrine, either intentionally (e.g., as an explicit means to achieve surprise), or unintentionally (e.g., out of ignorance or confusion). It is also possible for doctrine, when applied to a specific detailed case, to yield incompatible recommendations (e.g., the principles of war include both "concentration" and "economy of force," two recommendations that are rarely simultaneously compatible), and thus offering little guidance to either the adversary decision-maker or the intention estimating analyst.

In the fourth step the adversary selects a subset of goals from the rankings created in the third step. This constitutes the intention decision, or the adversary's policy. The adversary selects a desired and possible outcome (or objective) based on the benefits that would seemingly result from it and the apparent costs of the path to that outcome, given the resources that are available for accomplishing that outcome, the feasible actions that are required, and the anticipated contingencies of enemy response. The steps preceding this one involve estimating these aspects of the decision. The steps subsequent to this one entail refinement of the policy selection made in this step.

If the challenge to the analyst in the previous step was to narrow down the range of possible courses of action to those the adversary would consider feasible and worthwhile, the challenge in this step is to determine which objective or policy the adversary will value most highly. An analysis of successful intention estimation by intelligence analysts in the past (Stech, 1979) reflects two characteristics that might aid other analysts.

Successful intention estimators used rather explicit analytical models of how the adversary made decisions. These models related the general war objectives of the adversary (the two historical cases took place in World War II) to the adversary's resources and constraints. These explicit models were refined by a process of inference). That is, the analysts determined how evidence of different kinds would lead them to change the models. The analysts then sought the necessary evidence to determine which direction the models should evolve. In effect, the analyst anticipated the possible behaviors of the adversary and considered the implication of each possible move for the evolving model of the adversary's intentions. As the evidence became available, the analyst was able to refine the model, making it more explicit. Eventually, the enemy's actions and reactions could be estimated with considerable accuracy. This process by the analysts was one of "sagacity," i.e., understanding how the adversary had behaved in the past and the nature of his motives and "acumen," i.e., the analyst's ability to understand and anticipate the enemy's response and to duplicate the mental decision-making logic of the adversary (see Stech, 1981, ch. 2, on sagacity and acumen). To a lesser degree the analysts were able to use "control" to assist in the development of their models. That is, the analysts were at times able to control the information or the situation the adversary faced (e.g., by sending false information through double agents or deception operations, or by knowing in advance what one's own forces were planning to do) and could thus use this control to settle open questions about the enemy's reactions and decisions.

The fifth step in the intention process consists of translating the intention decision, or policy, into a strategic concept with a general objective and general courses of action outlined. This outline forms the basis for subsequent detailed planning. It identifies <u>what</u> the adversary intends to achieve and <u>how</u> this will be accomplished to a degree sufficiently detailed so that planners can turn this strategic concept into concrete actions.

The sixth and seventh steps involve detailed planning so that the intention can be carried out in action. First (step six) planners must determine from the strategic concept what specific actions and subobjectives are required to attain the overall strategic objective. The necessary steps must be ordered into a sequence of operations so that each sub-step is completed before subsequent steps that build on it are undertaken. Enemy actions have to be anticipated and losses calculated. Orders for required resources are issued. Warnings are given to troop and naval untis. Training operations focus on anticipated missions. Studies are made of uncertain aspects of the plan. Reconnaissance of the target area is increased. Deception operations may be undertaken to conceal these detailed preparations.

In the last step of the intention process, the adversary writes and issues the detailed plans for the intended operation. This may have the traditional form of the military or naval operational order, or it may be far more informal and instantaneous, for example, it may simply consist of the codeword that signals the execution of an operation that was planned long before. In this step, those actors that must carry out the adversary's intentions must carry out the final steps before actual action itself: assigning responsibilities, specifying subobjectives, allocating resources, issuing maps and plans, arming forces, fueling vehicles, sending out scouts, etc. The adversary has not committed himself to the plan, but is ready to do so and has taken all but the final step of action itself.

The description of these seven steps in the intention process makes it clear that the adversary's behavior becomes easier to monitor as the intention process approaches the action step. More physical activities and preparations are needed to translate the strategic goal into strategic technique and then into tactical plans then are needed at earlier stages. Thus the closer the adversary comes to actually carrying out his intentions, the greater the amount of physical intelligence the adversary will be forced to generate and, if he wishes to conceal his intentions, to hide or disguise. (But hiding and disguising are also actions which generate physical intelligence.)

David Kahn (1978: 39-41) makes a useful distinction between "physical intelligence" and "verbal intelligence." Physical intelligence is derived from the natural resources, physical installations, numbers of weapons and troops available, volume of commercial trade, and so forth. Verbal intelligence is derived from words through plans, orders, morale, perceptions, intentions, estimates, promises, or motives.

Plans and intentions take time to translate into physical reality. Knowing an adversary's intentions gives time to react, whereas knowing what actions the adversary has already begun to take may leave little or no time for reaction. While the latter steps of the intention process generate more physical intelligence than the earlier steps, the earlier steps may tend to generate more verbal intelligence, i.e., information about how the adversary perceives, estimates, makes judgments and decisons, and so on. In modeling the intention process of the adversary, the analyst will have to use verbal intelligence and physical intelligence to infer the events between the real world situation and the adversary's actions.

B. CONTENT OF ANALYTICAL METHODS

Intentions with certain characteristics are easily predicted. To the degree that the analyst can determine that the intentions are of the more predictable type he can put greater confidence in his estimates.

Stable intentions are predictable. Two types of stability apply to intentions. Some behaviors are habitual, i.e., prompted by recurring needs and drives. The need of a nation's people for food and shelter during a crisis such as war may lead to an intention by the nation's leadership to support minimal living standards. Other behaviors are customary or respond to strong cultural norms and are thus stable so long as these norms remain effective. Habitual and customary behaviors are highly predictable, if the analyst develops the ability to recognize the repetitive pattern and the stimuli that summon such behaviors. In the presence of such conformity inducing situations the adversary will tend to demonstrate the same habits or customs. In these eliciting situations the adversary's behavior is highly autocorrelated, and shows low variance over these situations. Habits are reliable reactions to aroused needs and customs are reliable reactions in response to a stable intention to conform. The main task of the analyst is to recognize the autocorrelated, reliable behaviors and then determine what needs, situations or events trigger these stable responses.

A second form of stable intention occurs when a form of behavior is constantly repeated over time regardless of particular situations. These behaviors have very high base rates and will appear to be normal, routine behavior. These behaviors may serve some functional purpose (e.g., regular patrolling and reconnaissance of sensitive terrain) or may be merely a tradition serving only ceremonial purposes (May Day parades). Some behavior may lack any recognized tradition and may be repeated merely because no alternative occurs to those who perform the behavior. An analyst may create a plausible explanation for such behavior, but its actual "cause" may be simply that "it was always done that way," i.e., mere behavioral inertia. It takes effort to change things, and there is no point to fixing something that is not broken. Similarly, behaviors that cause no harm may be tolerated and even reinforced although they produce no benefits except to keep people busy. Behavior that has been repeated under a wide variety of situations and circumstances with little variation in character is likely to continue. The intention for such behavior may relate to identifiable goals or it may be simply the intention that what was done be done again. High base rate behavior is highly predictable. Behavior that conflicts with or prevents high base rate behavior will require strong motives and intentions to overcome the inertia of the repeated behavior.

Some situations bring a multitude of pressures to bear on a nation, all acting to motivate behavior in a certain direction. For example, a massive sudden surprise attack will mobilize a nation's defenses. To the degree that the forces and pressures on the attacked nation tend to work in the same direction, the nation's behavior and intentions are more predictable. For example, imagine a small nation, unsure of its survival, with many enemies and few allies, vulnerable borders and trade routes, a militaristic tradition, capable standing armed forces, a unified and resourceful population, a tradition of antipathy toward the attacker, an economic, political, or territorial interest in the outcome of war, and a history of racial or religious disputes with the attacker. Such a nation is more predictably going to respond militarily to a surprise attack than a nation with opposite characteristics and pressures. Many predisposing forces and pressures acting in a given direction allow a more confident prediction or estimate of intentions than cross-cutting and conflicting pressures. In other words, when a nation faces intense demands, pressures, and forces to act in a certain way, and little or no opposing pressures to such action, it is likely to respond to the pressures. The more pressures that act in the same direction, the easier the task of forecasting the nation's intentions.

There is a nontrivial danger in using a causal analysis of the converging pressures on a nation to estimate its intentions, however. What may appear to the analyst to be an irresistable pressure might in fact pose only a temporary problem to the adversary nation. On the other hand, what appears to be a trivial factor to the analyst might be an overwhelming demand that the leaders of the adversary state cannot ignore. An example of the former case might be Egypt prior to the 1973 attack on Israel. Egypt seemingly faced severe limitations of military capability which would prevent any successful attack. In fact, Egypt had evolved alternative solutions to its military limitations (e.g., ingenious means for rapidly cutting through the sand dunes along the Suez Canal with high pressure fire hoses) and of adjusting its military intentions to its military capabilities so as to accomplish overriding political and diplomatic goals. An example of the latter case might be the underestimation by Western analysts prior to 1941 of the criticality of resources to Japan and the central role of oil supplies to Japanese intentions. The Western embargo on oil sales, an attempt to use economic force to pressure Japan to moderate its China policy, had the effect of increasing Japan's militaristic ambitions (to secure reliable oil supplies in Southeast Asia) rather than curtailing them.

Just as there are behaviors that are compelled by converging pressures, there are other behaviors that require converging preparations, i.e., complex and multifaceted multiple actions that lead up to and prepare for some behavior the state intends to act out. Such preparations are like a pitcher's wind-up, they signal an oncoming event, even though they may not reveal the exact nature of the pitch that is thrown or the base the pitcher may throw to. The entire complex of preparations may suggest one single integrated intention, or may be associated with disparate, disconnected events. The chronology of observations (by the intention estimating analysts) may or may not reflect the logical order of acts dictated by a single plan; instead the observational chronology may result from the mxing of actions called for by several unconnected plans. On the other hand, certain plans and operations require that various essential preparations be made in predictable order. If a watch is set for these preparations and none are detected, this is a negative indicator; it is unlikely the adversary is preparing that particular operation. Of couse, the adversary may anticipate such a watch and disguise such preparations or cover them with a plausible deception. If both predisposing pressures and necessary preparations for a particular behavior are absent, there is little likelihood that the adversary intends to perform that behavior.

Certain planned actions by an adversary require various acts be taken or processes be set in motion to carry out the intention. The attack on Pearl Harbor for example, required that the Japanese fleet sail to Hawaii. These acts may partially determine the courses of action open to the adversary. While at some point on its route to Hawaii the Japanese fleet could have aborted the attack (e.g., if it were discovered by U.S. reconnaissance), there was a particular moment at which there could be no turning back — the air raid had to be launched regardless of whatever else was taking place.

Some plans by an adversary may be sufficient determined by their early events that the conclusions can be predicted. The analyst observing the unfolding of the plan may be able to anticipate the ultimate target by noting that certain options and directions have been closed off as the operation has progressed, indicating the intended direction of the adversary's actions.

To the extent that the adversary's behavior is under the control of some agency that the analyst is able to monitor, the behavior is more predictable. For example, in combat the analyst's own nation will conduct operations which may determine the adversary's reactions. If the analyst knows of the impending operations against the adversary, he or she can anticipate the impact on the adversary's response. The analyst's nation may be able to control the adversary's access to information, as for example in Britain in World War II when all Nazi espionage agents were captured and controlled by British security forces. This control of information can allow the analyst to better model the adversary's decision process by providing direct access to some of the information that goes into that process.

An example of how information control assisted intelligence analysts to estimate intentions occurred in the Pacific in World War II prior to the battle of Midway. Through code-breaking, U.S. naval intelligence was able to determine that the Japanese intended to launch a major attack to destroy the few remaining U.S. aircraft carriers. The Japanese intended to draw on the U.S. carriers by first attacking a kev U.S. land base. The question unanswered by the code-breaking was where this land attack was to take place: Midway, Oahu, the Aleutians, the U.S. west coast were all possibilities with Midway being the prime candidate. To confirm that the Japanese intended to attack Midway, U.S. Navay intelligence conducted an "intelligence experiment," leaking the information that Midway Island was short of water. This was picked up by Japanese listening posts which reported to Tokyo that the target for the forthcoming attack was having water problems. These signals were intercepted and read by the U.S. Navy confirming the suspicion that the Japanese were aiming at Midway. The successful trap set by the U.S. Navy for the Japanese depended on this confirmation by information control.

Analyst Logic and Reasoning Processes

It is possible to separate intention estimation into primary steps of estimation logic, and then determine where parts of the process might be strong or weak. For this we divide estimation into seven different steps needed to reach an analytic conclusion (see Figure 3). These steps are somewhat arbitrary and could be combined into fewer or expanded into more steps. Nonetheless, they seem characteristic of inference and deductive reasoning as studies by cognitive psychologists and seem to capture the various capabilities of individuals to process information. All intention estimation involves at some point the thinking of individual analysts so an appraisal of the analytic logic needed to make estimates should serve to reveal some of the nonobvious perennial difficulties of intention estimation.

The mental methods people use to process information have been labeled "heuristics". These are usually adequate for the reasoning we typically need to do. Certain problems go beyond the capabilities of these heuristics, however, and require more elaborate, less speedy, and less efficient methods than everyday heuristics. Often it is difficult to recognize that the everyday heuristics of thinking are inadequate and people continue to use them when they should use stronger reasoning methods.

Perceiving Data

One such heuristic, "availability", is used when people perceive and encode data (see Figure 4). An analyst often decides whether an event has often occurred, and is thus typical, or whether it is rare and unusual behavior. Various characteristics of events influence how memorable they are; the frequency of their occurrence, the salience to the analyst of the event, its vividness. Frequent, salient, vivid events are more memorable. Memorability in

Figure 3. Steps in Estimation Logic

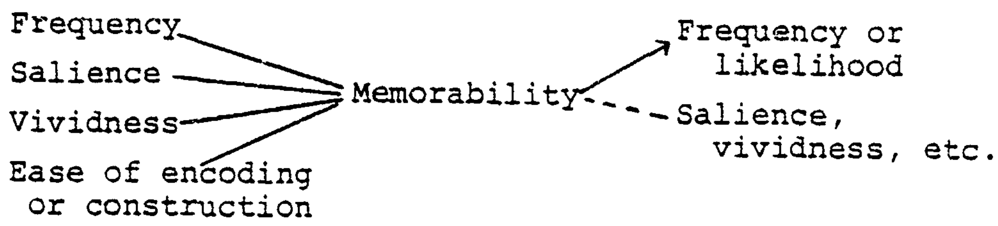
- o Perceiving Data
- o Weighting Data
- o Characterizing Data
- o Assessing Covariations
- o Assessing Causes and Effects
- o Prediction
- o Theorizing

Figure 4. Perceiving Data: Availability

ENCODING DATA

INFERENCE

Data Characteristics



turn influences judgements of frequency and likelihood. The events we remember best we tend to believe are more likely or more frequent. Note, however, that we can believe an event is frequent even if frequency were not the characteristic of the event that led to it being memorable. We may perceive a vivid or salient event as frequent, even when it is not, because we remember it well. Or to put it the other way, we might remember something very well because it was salient or vivid, and then, because it was memorable, estimate that it was frequent or likely. The more events are available to memory, for whatever reason, the more we tend to overestimate their frequency and likelihood. Uncommon events may be perceived as common.

In intention estimation, characteristics of events _other than_ their frequency or typicality seem to enhance their memorability and, in turn, inflate analysts' estimates that these are typical behaviors. International crises are highly salient and vivid to the intelligence analysis community and very memorable. It is likely that analysts tend somewhat to overestimate the likelihood that an adversary will shift from noncrisis to crisis behavior in periods of tension. The "false alarm" rate for predictions of adversary crisis action will thus tend to be high.

The proximity of one event in time or space to another event allows the two to be more easily remembered as connected. Such memorability of a connection may lead analysts to perceive coincidental events as causally related. One example seems to have occurred in the case of the 1969 Libyan coup and the 1973 Cold War. The mere presence of Soviet Navy ships plus these two events seemingly led analysts to suspect causal connections between the events and the Soviets which better evidence tends not to support.

Causal links between events are among the most memorable data characteristics. To the degree that an analyst perceives a causal link between an adversary's behavior and U.S. behavior, not only is the causal link likely to be remembered, but there is the risk that this memorability will lead the analyst to overestimate the frequency with which such causal influence takes place or is likely in the future. Causal links are difficult or impossible to prove, however, but are readily perceived, even when they don't exist, as is discussed further below. So the combination of a vivid, memorable, but spurious, causal link, and the availability heuristic, could lead an analyst to overestimate the likelihood of a causal relationship that, in fact, does not really exist, or is extremely rare. Again, false alarms about what the adversary is able to cause, or intends to cause, are likely.

A second heuristic which affects the way analysts perceive data is "representativeness" (see Figure 5). When the analyst assesses an event, certain characteristics and features of the event lead him to infer the event is of one type rather than another, and to estimate what population of events this particular one came from (e.g., is this particular behavior "hostile" or just "unfriendly"?). There is nothing wrong with this logic _if_ the analyst gives at least some thought to the population pool itself; asking if it is likely that any event, regardless of its characteristics, came from that particular pool of behavior (e.g., in general, is the adversary's behavior more often "hostile" or "unfriendly"?). Some behavior populations are highly improbable (e.g., the launching of surprise attacks) making it unlikely that a given event under study came from that population, no matter how strong the resemblance might be between the features of the particular event and the character of that population of events. (That is, many more behaviors may _look like_ preparations for surprise attacks than actually are preparations for surprise attacks.) Unless the analyst considers the base rate frequency of the population of behaviors as well as considering the features of the specific event, there is the possibility he

Figure 5. Perceiving Data: Representativeness

DATA SAMPLE FEATURE 1 → MEMBERSHIP INFERENCE → ESTIMATED POPULATION

DATA SAMPLE FEATURE 2 → MEMBERSHIP INFERENCE

DATA SAMPLE FEATURE 3 → MEMBERSHIP INFERENCE or she will overestimate atypical and unusual events and perceive them as more common than they really are.

Analysts who attend to the doctrinal statements of an adversary may find a statement or set of pronouncements that suggest a particular future intention. The specific characteristics of these statements may be very suggestive of a future action, e.g., open hostility with other countries. These features of these particular statements have to be balanced against the base rate of analogous statements and the base rate of behavior. If the entire body of analogous statements reflects a peaceful outlook, or a cautious attitude, the analyst should reduce the weight given to the hostile remarks. Similarly, if the nation being monitored has had a history of peaceful relations the analyst should consider the hostile statements of representing an unusual departure and a possible anomaly rather than a clear-cut indicator of the future.

Similarly, analysts often make major deductions about an adversary's tactics or strategy using military equipment construction as evidence. That is, the construction of a particular weapon is taken as evidence of a particular intention. Such arguments may lose sight of the other naval or military construction programs which may represent even greater emphasis or effort. The use of naval events, for example, as evidence of naval policy or doctrine may fall victim to the same problems, that is, using rare and unusual events to draw conclusions about typical or general naval behavior.

A fascinating example of the tenuousness of estimates of enemy intentions extrapolated from estimates of enemy military capabilities is given in McLachlan's (1968) account of British Naval Intelligence before and during World War II. In 1936, when the Germans were constructing their great battleship Bismarck, the consensus of British diplomatic opinion was that Germany would adhere to the Anglo-German Naval Treaty of 1936, which limited German battleship size to 35,000 tons. This consensus provided the basis for the design of the British battleship King George V. The dimensions of the Bismark and Tirpitz as released by the Germans indicated that if these ships did in fact displace only 35,000 tons as the Germans claimed, they were of much shallower draft than the British ship. Although the lower tonnage was doubted by Naval Intelligence nontechnical officers, intelligence opinion was divided. From the inferred characteristics (shallow draft and 35,000 tons) the British Naval Plans Division concluded "the present delsgn of German capital ships appears to show that Germany is looking towards the Baltic with its shallow approaches more than in the past" (quoted by McLachlan, 1968: 136), that is, the German ships were aimed more at Russia than at Britain. In fact, Bismark and Tirpitz were designed to be roughly 45,000 tons, and the Germans released false figures to the British, depending on British belief in the readiness of the Germans to honor the 1936 agreement. The Germans read their opponent well; the British Director of Plans at the time wrote, "our principal safeguard against such an infraction of treaty obligations lies in the good faith of the signatories" (p. 137 in McLachlan). Not only had Germany deceived Britain as to her capabilities, she had the additional, unintended benefit of an erroneous British estimate of Germany's naval intentions. Similar underestimates based on similar German deceptions regarding submarines, cruisers, and battlecruisers also occurred prior to World War II and may also have misled British estimates of Germany's willingness to engage British naval power with what the British took to be a far less capable navy than Germany in fact possessed.

Similarly, U.S. underestimates of the range and performance of the Japanese Zero and the estimate that shallow water torpedo attacks were infeasible probably Contributed to the Japanese surprise in attacking Pearl Harbor (Wohlsetter, 1962: 394).

Fear of the Luftwaffe's strategic bombing capability, overestimated in 1938 by the British Air Staff, inflated British Cabinet estimates of Germany's willingness to go to war over the Czechoslovakia crisis and contributed to Chamberlain's motivation to capitulate to Hitler, according to some historians (see Bracken, 1977).

Finally, it is quite easy to perceive plausible causes for almost any event, and to detect causes that have no actual connections to events, and which may be quite rare and atypical. Features of a data sample may lead the analyst to perceive a causal relationship, but it is important that analysts stop to ask if such relations are generally likely or typical. If not, the analyst should lower his estimate that such unusual causal behavior is occuring in the case under study.

Weighting Data

Not only must analysts perceive data, they must weight it, deciding which pieces of information are important and which are less critical to their judgments (see Figure 6). However, psychologists have found that, in general, people are not at all accurate in identifying what information they actually use to make judgments and decisions. People also often use highly redundant evidence as if it were completely independent information, in effect, counting as two sources what is really just one. Cues which vary are given more weight than stable cues, which is logical, but people sometimes do not notice when a variable cue becomes stable, and may go on giving it light weight on the mistaken assumption that it continues to vary. The more cues that are available, the more confidence people feel about their estimates, even though they often fail to integrate all the additional information into their judgments, or to make more accurate judgments with the additional information. People seek out those data cues which are most like the answer that is being sought. To estimate a

Figure 6. Problems in Assigning Weight to Data

- o Lack of introspective accuracy
- o Reliance on redundant or highly correlated cues
- o Emphasis on variable cues
- o More cues, more confidence
- o Cue-response compatibility
- o Salience and vividness
  - -- Case studies versus base rates
  - -- Neglect of nonevents

future attack submarine threat, for example, an analyst might seek data on attack submarines, although data on naval aircraft or space systems might be equally important to answer the question. Highly salient and vivid data are given more weight than statistical data, or information on nonevents. For example, although analysts sometimes use nonevents in their estimates, important nonevents are often not attended or analyzed.

Categorizing Data

As analysts perceive and weigh data they also categorize them — perception involves the act of categorizing. That is, people attach a description to a piece of information to code it for memory and later inferential work. Mental theories and conceptions aid in this process of classifying and organizing perceptions and facts.

Perhaps the major difficulty people have in categorizing complex phenomena is in trying to impose either/or distinctions on objects or events which have too many dimensions and too few cut and dried boundaries to be so narrowly pigeon-holed. Psychologists have found that people typically do not rely on either/or categorizations of complex objects or events in making judgments but instead seem to make judgments as to whether the object is more or less like a prototype or schematic model of the category in question. In other words, category boundaries are fuzzy and probablistic rather than hard and fast.

The fuzzy nature of complex natural categories and the difficulty in specifying all the features which give an object or event a family resemblance to a category may contribute importantly to conflicts between intelligence analysts. For example, there are characteristics of Soviet ships that suggest they would be effective deterents to interventions during crises by Western aircraft carrier task groups. These and other characteristics of these ships give them capabilities to serve in a "blue belt defense plan" against nuclear strikes from the sea on the Soviet homeland. Similarly, deployments of Soviet ships share features that relate to both the diplomatic and strategic missions. Deciding whether a particular event was diplomatic or strategic in character can be very difficult because of the fuzzy boundaries of these categories. Often analysts can do no more than to conclude that a particular event was more like one category and less like the other but also a little like both. Nor does it seem to be any easier to categorize doctrinal writings, military equipment construction, operational or deployment patterns, or diplomatic events; all these complex events can be categorized as having a family resemblance to strategic or diplomatic missions, offensive or defensive objectives, cautious or expansive tendencies. At best the analyst will be able to make only probabilistic judgments as to the categories in which an event belongs. Debating pigeonholes is far less useful than attempting to measure the strength of family resemblances.

It is important, when generalizing from a sample of data, to be conscious that certain data sampling methods bias the information in predicatble ways. One such bias involves roles. For example, the role of deterrer in a crisis, i.e., the party that acts to deter another party, has a built-in role conferred bias. Namely, if the adversary does nothing, the deterrer can credibly claim success in deterring, even though the deterrer actually may have done little and the lack of action by the deterred was due to other factors (e.g., the adversary lacked any incentive to act). The deterrer is a little like the man who scatters corn flakes around himself to keep the Bengal tigers away. When you point out that there are no tigers in the neighborhood, he beams at you and takes credit for his amazing deterrent powers.

A second type of sampling bias occurs when an analyst attends a part but not all of a data configuration (see Figure 7). In panel A of the hypothetical example, the entire data configuration is shown. Various biases can be seen in panels B, C, and D. When the whole pattern is considered all one would want to conclude is that Soviet naval events of all types are twice as frequent in the latter time period. Biased sampling, however, might lead to the conclusions that Soviet diplomatic naval activity has doubled, or that strategic forward deployment has doubled, or that the Soviets are now twice as likely to engage in strategic as opposed to diplomatic forward deployments.

Real examples of such biased sampling can be found in naval analysis. For example, McGwire (1976: 166) argued that the Soviets surveyed the Indian Ocean from 1967 to 1969, making many diplomatic port visits to scout out facilities, and then visited very few ports thereafter, having found the facilities they needed. Peterson (in Dismukes and McConnell, 1979: 91-92) disagreed, writing:

> While it is true that a drop in diplomatic vistis to Indian Ocean ports was registered in 1970, it is <u>not</u> true that "very few" visits have been made elsewhere in the region since then. Between 1970 and 1971, for example, no fewer than 30 diplomatic visits were made to Indian Ocean countries other than Somalia and South Yemen. In comparison, only 28 were made to Mediterranean ports during the same period.

The total data configuration is shown in the figure (see Figure 8). There were nearly twice as many Indian Ocean visits as Mediterranean ones, but in the latter period the ratio of visits is not 2 to 1 but 1.4 to 1. While visits per year increased in the Mediterranean over the two periods from 2.3 to 4, in the Indian Ocean they decreased from 10 per year to 5.6. Petersen's comparison of 30 to 28 visits is biased and misrepresents the far higher level of overall visiting in the Indian Ocean and the sharp decline of visiting in that ocean in the latter time period to a level comparable to that in the Mediterranean.

Figure 7. Hypothetical Example of Biased Sampling

A. Soviet Naval Events

	Diplomatic Forward	Strategic Forward	Total
1954-1967	15	30	45
1964-1974	30	60	90
Total 1954-1974	45	90	135

B. Column Bias: (a) Diplomatic Forward Perspective
(b) Strategic Forward Perspective

	Diplomatic Forward	Strategic Forward
1954-1963	15	30
1964-1974	30	60

C. Row Bias: New Soviet Navy Perspective

	Diplomatic Forward	Strategic Forward
1964-1974	30	60

D. Total Bias: Overall Perspective

	Diplomatic Forward	Strategic Forward
Total 1954-1974	45	90

Figure 8. Number (and Percentage of Diplomatic Port Visits in the Indian Ocean and Mediterranean by the Soviet Navy 1967-1969, 1970-1976*

	Indian Ocean	Mediterranean	Total
1967-1969	30 (43%)	7 (20%)	37
1970-1976	39 (57%)	28 (80%)	67
Total 1967-1976	69 (100%)	35 (100%)	104

* Data from Petersen, Table 3.2 (p. 92 in Dismukes and McConnell, 1979.)

Data Covariations

Analysts must be able to discern relationships between sets of data to infer that one set of events covaries with another set. For example, military analysts must assess the covariation between an adversary's strategic needs and military activities, between political commitments and crisis behavior, between capabilities and actions, etc., to understand and predict military intentions.

People have difficulty perceiving unexpected covariations and tend to perceive expected covariations even when there are none. That is, people are far too ready to detect theory-driven covariations in the data they process and unable to see unexpected data-driven covariations unless they are overwhelmingly strong. Psychologists have found that people underestimate covariations, i.e., strong correlations (e.g., r = .6 to .8) are perceived as weak relations, and near perfect covariations (r = .8 to 1.0) are perceived as merely strong relations. Unexpected relationships, even when near perfect, tend to be missed.

Unless an analyst expects to find a covariation relationship, he or she will tend not to notice one unless it is very strong. Consequently, strong relations may go unnoticed, or underestimated. Potential indicators of future activity may be overlooked and important predictable patterns may be neglected.

On the other hand, psychologists find that people will perceive an expected, theory-driven covariation even when none exists in the data set. What seems to happen is that positive occurrences are noted, i.e., those cases when the two expected events do, in fact, occur together. Those cases where one, but not the other, event occur are neglected. The analyst thus amasses a convincing list of confirmations of the relationship but has ignored all the disconfirming cases.

Such difficulties in assessing covariations can be found in naval analysis (see Figure 9). For example, one analyst argued that Soviet diplomatic naval

Figure 9. Problems in Assessing Covariations

- Theory-driven correlations are perceived, even if absent in the data.

  Theory: The expansion of Soviet diplomatic visits in the Indian Ocea, 1967-1976.

  Data: Correlation between year and number of visits: -.38.

- Data-driven correlations are not perceived if absent from theory.

  Theory: The intensity of diplomatic visits reflects the prominance of political concerns in the Indian Ocean squadron's mission structure.

  Data: Correlation between ship days and diplomatic visits: -.65.

  Correlation between ship days and lengths of visit: -.83.

- Illusory Correlations

  e.g., naval presence and coup occurrences.

visits in the Indian Ocean expanded in the years 1967 to 1976, reflecting Soviet interest in that area. This implies a positive relation; more visits per year as the years go by. In fact the relation is negative: there were less visits as the years went by in this period. The data suggest there was a contraction, not an expansion, of Soviet diplomatic visits in the Indian Ocean. On the other hand, unexpected correlations can go unnoted. One analyst expected that as the Soviet Naval Squadron spent more ship days in the Indian Ocean, the prominance of Soviet political concerns would intensify diplomatic visits. The data reflected just the opposite; the more ship days the Soviets spent in the Indian Ocean, the fewer the number and the shorter the duration of diplomatic visits. These unexpected (and unnoted) negative relations were quite strong. Finally, when a theory suggests a relationship should exist, coincidental co-occurrences may be taken as confirming evidence. On several occasions the presence in the area of Soviet ships has been interpreted as related to the occurrence of coups, or crises, although there was little evidence to substantiate any connection between the events and the Soviet's presence.

Cause and Effect

Covariations are important in their own right, but also because they are crucial data for inferring cause and effect relationships. The analysis of causes is one of the intelligence analyst's major objectives; having causal understanding of an adversary enables you to predict his future actions. Determining causes is one of the most difficult of intellectual and information processing operations, however, and one for which a myriad of analytic aids are not just useful but often necessary. Our intuitive, unaided reasoning and judgments about cause and effect are often in error. Even scientists sometimes find nonexistent causes for events.

Psychologists find that there is a strong tendency to perceive as having a causal role anything on which we focus our attention when considering cause and effect (see Figure 10). Futher, if an actor is perceived to have a certain intention, and an event occurs bringing about the desired outcome, there is a strong tendency to view the actor as having caused the event, although the event may have occurred for other reasons. We attribute the behaviors of actors we are observing to their intentions, dispositions, and motives and underestimate the degree to which their behavior is controlled by the environment, the situation, or the context they are in.

To the degree an actor's behavior has important consequences and the actor can foresee the consequences of his actions, the more we attribute the actor's behavior to a profound motive. That is, the cause is perceived to match the effect: if the effect is profound, the motive, or cause, must also be equally profound. This is a manifestation of the representativeness heuristic, the tendency to assume that features of the cause should be representative of the features of the effects. The problem is that rather trivial causes and motives can lead to very profound consequences.

If an actor is perceived to foresee his actions' consequences, the observer tends to attribute the actions to intention. For example, the Soviet Union's war fighting outlook implies foresight; the anticipation of general nuclear war. There is a tendency to assume that, if the Soviets are able to plan and prepare for such a war, they may intend to wage such a war to win their own ends. The Soviets may thus be seen as building a nuclear blackmail capability. Of course, it is also possible that the Soviets have no confidence in their own deterrent capability, and prepare for nuclear war, not optimistically for blackmail, but pessimistically out of a fear that they cannot forestall it.

In "magical thinking", the features of effects are used to guide the search for causes -- causes are sought which resemble effects. Perhaps the worst

Figure 10. Problems in Assessing Causes and Effects

- Dispositional vs. situational attributions
  e.g., explanations of forward deployment
- Profound motive fallacy
- Foreseeability implies causality
- Magical thinking -- causes resemble effects
  e.g., capabilities cause intentions, can do -- will do
- Minimal causation
- Causal hydraulics
  e.g., desire for military superiority precludes desire for detente
- Parsimony and "Indiscriminant Pluralism"

example of magical thinking is when the analyst lets himself believe the adversary's intentions (causes) follow from or resemble their military capabilities (effects). In particular, there is the danger of calculating the worst the adversary could do and then reasoning that that is what the adversary will do. Alternatively, the analyst may be tempted to calculate the limits of the adversary's capability and then estimate that the adversary would not undertake missions that exceed those apparent limits. "Minimal causation" refers to the tendency to accept the first reasonable cause that fits the data as the only cause. After people find a plausible cause to explain an effect, they seem to stop searching and the hypothesis that multiple causes contribute to an effect is rarely tested.

"Causal hydraulics" are a colorful way of describing the tendency to believe that causes compete and compensate for each other in producing effects. That is, an analyst might believe that if strong political forces are producing some behavior, then military or naval factors cannot also be contributing at the same time. The notion that bureaucratic politics shape an adversary's behavior and that competing agencies and interests are contending for influence is especially subject to causal hydraulic reasoning. The danger is that necessary causes may be mislabeled as sufficient causes.

A final problem in causal reasoning is the tendency to seek a plausible cause for every behavior. This can easily lead to what one historian termed "indiscriminant pluralism" -- an attempt to propose a cause for everything. The problem with efforts to explain everything is that one ends up with too many explanations and no ability to predict. There is no way of knowing which explanation to use for the future situation. This is why scientists seek parsimony, the fewest possible causes sufficient to explain and predict. The smaller the number of causal explanations that can account for events, the better the chance these causes will predict the future.

Prediction

Predicting and forecasting are central tasks in intelligence analysis. Analysts, however, have been charged with being unable to accurately predict either the capabilities or intentions of adversaries. Some recent prediction failures are shown in the figure (see Figure 11).

Many explanations of failures of intelligence predictions seek the causes in unique characteristics of intelligence estimation. These explanations are made less compelling by the fact tha forecasting and predicting are not done with much accuracy in other fields. Recent reviews by Ascher (1978) and by Hogarth and Makridakis (1979) of population, economics, energy, transportation, technology, and business forecasts found that prediction in these areas is rarely accurate. This suggests that problems of forecasting and prediction are general ones, not limited to intelligence analysis.

People need to perceive their world as orderly and predictable. The more people believe they control events or that events are controllable (even if such control is illusory), the more predictible the world seems to them. This illusion of control can lead to predictions made with great confidence of events that are in fact controlled by chance. Forecasters may perceive the future as more predictable than it is simply because of their efforts to predict it.

People see pattern and order where none exists because of their need for an orderly world. There are extremely strong perceptual tendencies to structure the environment and make sense out of it in order to organize perceptions. People rarely consider the possibility that the environment may have random or probabilistic elements, instead they see events as determined and fixed or ordered by regular mechanistic (not probabilistic) processes.

Figure 11. Prediction Problems

- o Prediction Failures
  - Bombers
  - ICBMs
  - IRBMs
  - Launchers
  - MIRVs
  - Defense Expenditures
  - Surprise Attacks
  - Crises
  - Peace Offensives

- o General Causes
  - Illusion of Control
  - Illusion of Order
  - Faulty Assumptions
  - Unreliability of Judgment Vice Statistical Relations
  - Illusions of validity and overconfidence
  - Inappropriate Techniques
  - Theory-driven overprediction

Forecasters often reach different conclusions which all reflect a similar biased inaccuracy. Ascher (1978) concluded that forecasters making estimates in the same time period all tend to share the same key assumptions, and to the degree these are wrong, tend to make the same mistakes. The analysts' key or core assumptions may be much less carefully scrutinized than data or methods, but bad core assumptions may produce greater error than flawed data or inappropriate methods.

For many medium- and short-term forecasting tasks, simple statistical trend extrapolations are often more accurate than are analytic judgments, or theoretically elaborate or methodologically elegant forecasting methods. People have very high confidence in the superior accuracy of their judgments, a level of confidence unwarranted by the demonstrated inaccuracy of their predictions relative to simple statistical methods. People are sometimes most confident when they are least accurate.

Often forecasters employ highly inappropriate techniques which virtually assure an inaccurate prediction. One very overworked and abused technique is Delphi. The famous economist, Paul Samuelson, noted that:

> the greatest error in forecasting is not realizing how important are the probabilities of events other than those everyone is agreeing upon.

To the extent that Samuelson is right (and there are several reasons why he probably is) then a technique like Delphi is doomed to inaccurate predictions because it focuses the experts' attention upon the probabilities on which everyone has agreed. Delphi systematically seeks out the extreme predictions and the unexpected, low probability conjectures, leaving only the middle range, concensus probabilities most of the experts had already thought about. The extremes that Delphi eliminates, however, are the most likely to predict surprises and unexpected events. Delphi is appropriate only if surprises or unexpected events are not going to occur. But if we can be assured of that degree of orderliness, we do not need Delphi; simple statistical extrapolations will do better and cost less.

Theory

The use of theory, the last analytic step, is much to be commended in any prediction or forecasting effort. There is the chance, however, that the theory will suggest more order and predictability in the data than really exist, leading analysts to over-predict. That is, the analyst with a theory may make more extreme predictions with greater confidence than the data and the analyst's prediction record would warrant (see Figure 12). One of the main reasons for using theories is to make accurate predictions. The analyst must take care not to use theory wishfully, and to keep predicting distinct from theory testing.

Creating and modifying theories of the adversary should be one of the major goals of the intelligence analyst. Whether they acknowledge them or not, analysts are always using something like a theory as they proceed through the previous six estimation steps. The danger of using theory is that it can so easily and pervasively color the other data processing steps, leading to perceptions of data, covariations, and cause and effect that are not accurate or valid.

Philosophers of science argue that the best procedure for refining and testing theory is to try to disprove it in whole or part. In general, however, people (including scientists) tend to test theories by the far less productive and less valid method of accumulating confirming evidence. A theory that survives repeated attacks is likely to be more valid than one that has masses of confirming data to support it but has never been subjected to a disconfirmation

Figure 12. Theorizing Problems

- Theory-biased consideration of evidence
- Confirmation bias
- Difficulty of using disconfirming evidence
- Infrequency of explicit multiple hypothesis testing
- Perseverance of discredited theories
- Overconfidence -- inference versus perception
- Failure to learn from experience

effort. It is no simple matter to arrange valid disconfirmation tests for complex, real-world phenomena; it is difficult to establish the necessary controls and to sufficiently isolate the crucial elements of the theory to determine the meaning of the tests. Similar problems occur with confirmation tests but they are psychologically more easily overlooked. Analysts tend much more to amass confirming data for a theory than to explicitly attempt to disconfirm a theory.

Like most people, analysts involved in hypothesis testing rarely set out to pit opposing theories against each other and explicitly test them against the same data bases. Some efforts are very worthwhile from a methodological viewpoint, but require diligent and rigorous methods to keep orderly all the logical implications of the data. It is also difficult to evaluate cases of partial data support, i.e., there is no good metric for measuring partial support for a theory.

Psychologists and historians of science have found that people in general and scientists in particular are often unwilling to give up their theories even in the face of massive discrediting evidence. Once a theory is accepted it becomes quite easy to explain away any conflicting data as artifactual, biased, improperly analyzed or interpreted, irrelevant, etc. An accepted theory quickly becomes entwined with other theories and beliefs which tend to support and bolster the theory when conflicting or anomalous data are received. A far higher standard of evidence is demanded for disconfirming evidence than is required for confirming evidence. Such an attitude is approriate only to the degree that the theory is making proven, accurate predictions. Analysts should be more skeptical if the theory is unproven in application, or has a checkered past in making predictions and forecasts.

People have remarkably high levels of confidence in their own analytic and inferential powers -- actually overconfidence, since they are often more confident in their judgments than they are accurate. On the other hand, people also seem reluctant to credit as accurate their perceptual senses and memory for psychophysical judgments. People can make very accurate judgments with these capabilities but seem to have less confidence in them than in their deductive and inferential reasoning powers. This suggests that analysts might try to make greater use of methods that capitalize on their perceptual sense (e.g., using data reduction and display techniques that employ graphic comparisons) and rely on mental inference and deduction only to the degree that some external criteria of validity can be summoned for assistance.

The characteristics reviewed above of how people use and maintain theories make it very hard for people to learn from experience that their theories are wrong or inaccurate. Bad theories can survive a great amount of painful experience. People tend to assume a deterministic, well-ordered world with neat conceptual boundaries and they have difficulty using the stochastic data the world usually generates.

Fortunately, many of these same information processing problems have been encountered before in different fields, and methods have evolved to cope with them. Some of these aids and their features are described in the next section.

C. ANALYTICAL METHODS

For the purpose of this research project we have considered four general categories of analytical aids to the problem of estimation of intentions.

1. Judgmental methods employing sagacity (the understanding of an analyst or estimator of relationships between subtle cues and behavior), control (intelligence experiments that provide predictive information by controlling the information available to an adversary), and acumen (the result of pure reasoning and

the appreciation of the others' point of view in a particular place and time).

2. Analytical aids involving such techniques as event coding and content analysis that support specific elements of the estimation process.

3. Extrapolation models based on past events and data, and

4. Structural representations or models which take into account structural realities or changes with respect to leadership, capabilities, and other state-of-the-world factors.

These categories are not mutually exclusive in that specific analytical aids may fall into several of the categories. Further, any estimate of intention must generally draw upon all four categories of aids. Clearly, the ultimate estimation must be based on judgments of the reality of critical assumptions, situations, and responses that might be indicated by an analytical aid. Quantitative analysis in the extrapolation or structural representation categories can comprise between 10 and, say 80% of an intention estimation depending on the situation. In all cases subjective judgments must be applied at some stage. Also, some portion of most estimation problems will be amenable to quantitative analysis. The art of estimation is to select and blend appropriate methods and the objective of this research is to assist in that process.

Specific analytical methods that may be used in supprot of judgmental analyses include:

- scenario writing to explore a range of circumstances and responses in an internally consistent framework that can be reviewed and evaluated by a third party
- availability and representative heuristics.

The heuristics can be aids or mental traps depending on the situation and their application, but must be listed since they represent basic psychological processes in the perception and integration of data. The availability heuristic deals with the ease with which memories and stereotyped constructions of events can be retrieved by an individual. The representative heuristic in a judgment strategy that involves the use of salient features of a sample to infer its membership in a specific category or class representing behavior or response.

Analytical aids to the extrapolation category of estimation include the statistical methods of covariant analysis and regression analysis. These techniques may be used to extrapolate both intentions and capabilities based on past observations. Both probabilistic and determininistic methods may be employed.

A wide range of analytical aids may be drawn for the structural representations category of estimation. This category of analysis breaks free from the problems inherent in the extrapolation processes, but requires a more detailed understanding of cause-effect relationships and characterization of changes in capabilities; leadership, and other pertinent factors. Techniques employed here include Bayesian analysis, factor analysis, decision analysis, and other structural modeling techniques. Intention estimates using these methods must draw upon related studies of elite structures, technical capabilities, logistics, and deployment.

III. COMPARISON AND EVALUATION OF ANALYTICAL AIDS

This chapter describes a variety of quantitative aids that have promise in dealing with the analytic problems noted for the various steps in the estimation process in the previous chapter. It also outlines how the important analytic features of quantitative methods can be assessed by the analyst considering their use.

In describing these techniques we attempt to touch on several aspects. The technique is briefly described, if possible using an example from intelligence analysis (several cases are taken from an earlier study of Soviet naval analysis, Stech, 1981). The major theoretical and methodological assumptions of the method are described and the nature of the input data required is outlined. The output of the method is discussed and related to estimation problems that might be alleviated. Some other aspects that are touched upon include the descriptive, inferential, and deductive power that the technique might provide, the startup and maintenance costs, the flexibility of the technique to deal with new or different research questions, the need for special equipment training or expertise, any any interesting characteristics of the methods (see Brownell, Stoll, and Thomann, 1980 for a similar evaluative format for estimative and analytic techniques).

There are many promising analytic methods and experimental techniques that apply to the estimative problem areas. Some of these techniques are used now by analysts (but apparently infrequently) and have been applied to various types of intelligence estimation. Some have not yet been used by intelligence analysts, so far as we know. All of these methods are aimed at compensating for or preventing the information processing biases that can occur in unaided analysis. They are all analytic aids, not substitutes for analysts or analytic reasoning.

The editor of a recent compendium on quantitative approaches to political intelligence in the CIA observed (Heuer, 1978: 1):

> The behavioral revolution in academic political science has been virtually ignored by the (CIA) and the intelligence community as a whole.

Although Heuer saw the narrative essay as continuing to be the dominant form for intelligence estimates, he recommended greater use of quantitative techniques that:

> ...help to trace the logical consequences of subjective judgments, extend the mental capacity of the individual analyst, force the analyst to make his assumptions explicit, or help organize complexity (p. 8).

The aids described below are designed to enhance analysis of intentions by moving toward the goals Heuer listed.

A. PERCEIVING CATA

The three main problems regarding the perception of data are (1) the nonperception of nonevents and negative evidence, (2) the use of availability information, and (3) the use of representative features of samples to estimate population characteristics, to the neglect of base rate data.

Events and Nonevents

Analysts must render a continuous, undifferentiated stream of information on the adversary into discrete, discriminable, describable entities. Perception is discrete rather than continuous. The adversary is perceived as performing a series of discrete actions. These actions divide the stream of information into segments or units.

Psychologists have begun to investigate how people segment this stream of information into series of discrete action events. One main finding is that the density of events becomes greater the more unexpected the action. That is, when people observe highly organized, predictable, step-by-step actions, with a clear hierarchy of subordinate and superordinate goals, they tend to segment the action perceptually into grosser units. People resort to shorter units when perceiving unexpected action. Organization of the perceived action becomes extremely fine-grained immediately after an unpredicted, significant event occurs.

Most intelligence analysts are familiar with the impact of crises or "surprises" on normal operations -- requirements on intelligence production increase greatly. Much more fine-grained analysis is called for than normally.

One consequence of this difference in event perception is that different coding categories are applied to the cases of expected and unexpected events. Actions during a crisis may appear different from everyday actions because the former are subjected to fine-grained coding, while the latter fit grosser, more familiar schemas. This suggests that analysts make strong efforts to keep their fine-grained segmentations comparable (through aggregation) to their day-to-day coding of actions. This will allow for meaningful comparisons of crisis and noncrisis episodes. Such comparisons may be precluded if the analysts' fine-grain categories of coding events in crisis are not comparable with the grosser normal coding.

It is also important that psychologists have found that as perceived behavior becomes more motivationally important to the perceiver, grosser coding units tend to be used. That is, during an important crisis, the consumers of estimates may tend to use grosser codings of events at the same time that analysts are using more fine-grained codings. (This assumes that decision-makers and operators are under greater motivational pressures than are analysts.) That is, arousal leads to a tendency to focus on a few relevant cues and decreases the range of attention, while amplifying its intensity.

Several studies have shown that people segmenting action into larger units make neutral attributions as to causes: i.e., the actions are attributed to both situational factors and dispositions of the actor. When using smaller units of analysis, people tend to attribute action to the actors alone, not to the situation. This suggests that analysts during crisis should be alert to (1) a tendency toward "hypervigilance", or the close monitoring of only a few indicators, (2) the possibility that estimate consumers may be looking for very gross organizations of action while analysts are generating extremely fine analyses, and (3) analytic categories for crisis action may be incompatible with the categories used for normal actions, possibly producing a false analytic dichotomy, especially regarding the causes of action.

A method which might help analysts to code a stream of information into useful categories was demonstrated by O'Leary and Coplin (1975; 182 ff.) for Statement Department Intelligence analysts using data on conflict acts between Egypt and Israel. Rules for coding each type of event on a scale of violence were used to score the entire event series. Analysts were then able to graphically follow events in terms of either the frequency of events of different levels of violence, or in terms of the level of violence itself. Such graphs and codings enable the analyst to capture, respectively, the grosser relationships between events and the finer-grained details of actions. By allowing the analyst to demonstrate patterns and trends in events, coding rules and methods aid in the observation of nonevents and negative evidence. Interruptions, omissions, and nonoccurrences are easier to detect against an orderly background of action trends.

Data on Frequencies

The availability heuristic is fundamentally the tendency to substitute memorability for an estimate of frequency. To the degree that more objective data on frequency are readily accessible, the analyst should feel less compelled to use availability as a substitute. Two techniques for systematic frequency recording in intelligence analysis are event coding and content analysis.

Event coding has become a staple method in efforts at crisis forecasting and prediction (see, e.g., the March 1977 issue of International Studies Quarterly on "International Crisis: Progress and Prospects for Applied Forecasting and Management" or Kaplan, 1981 for a coding of the political use of Soviet military power). In general, elaborate rules for coding events or actions are specified which are then applied uniformly to the stream of events. This allows the analyst to make quantitative comparisons on any of the dimensions created by the coding scheme. Given the widespread use of event analysis in early warning intelligence and crisis forecasting, its absence in other areas of analysis is striking. Some naval analysts make use of event tabulations (see Petersen, in Dismukes and McConnell, 1979: ch.2, Tables 2.12-2.15, for a detailed listing of Soviet naval operations), but the events are not elaborately coded and frequency analyses are uncommon. Naval analysts now use some event categories (e.g., frequencies of diplomatic port visits) in their studies, so they do not seem adverse to the concept. It may be that more complex coding of Soviet naval events waits on a taxonomy of naval actions to provide the coding framework. McConnell's analysis of the "rules of the game" (ch. 7 in Dismukes and McConnell, 1979) might provide an initial step for developing such an event coding system, and his classification of cases (Tables 7.1 and 7.2) provides a rough coding scheme (see Cohen, 1980, for an assessment of "rules of the game" analysis in estimation).

Event coding techniques make several methodological assumptions the analyst should consider. In order to code events the coder must receive some report of the event, a cable from the embassy, an intelligence report, a wire service story, a news article. Those individuals and agencies doing the reporting are also affected by the same event perception phenomena we outlined above. As events become more unexpected, the density of reporting will increase. This changes the character of the reporting from crisis to non-crisis periods, making comparisons across periods problematic unless specific efforts are taken to control for the finer grained perceptions of the crisis periods. There is likely to be a significant feedback effect from the analytic community to the policy community to the reporting sources. If an embassy, intelligence collection organization, or news gathering team learns that the analytic and policy communities are monitoring a particular problem or area they are likely to intensify their reporting of events and will report events that might otherwise have been ignored. The consequence is that the basic rate and character of the reports changed. This is critical because volume of reporting is one indicator used in event analysis. Positive feedback effects may significantly distort such an indicator.

Event coding requires varying levels of input data, depending on the purpose of the analyst. If used to monitor all world activity in the hope of anticipating crises and "hot spots" a continuous stream of events must be coded, i.e., from the public press or monitored news broadcasts. On the other hand, individual analysts might develop a coding scheme to track various events over time for a particular country, e.g., incidents of civil disorder, requiring considerably less coding volume and effort. Like any pre-determined coding scheme, event analysis is vulnerable to the problem that the initially established coding categories may be inappropriate for later analytic tasks. Coding events in terms of cooperation and conflict, for example, might be inappropriate if the analyst later must determine variation in types of conflict. No coding scheme is sufficiently flexible to anticipate all the future research questions the analysts may need to answer.

Content analysis has been applied to intelligence problems since World War II (e.g., George, 1959) to determine objective frequencies for actions, events, or statements. Content analysis is generally applied to verbal or written statements (e.g., propaganda, speeches, memoirs) to determine such things as authenticity, trends in semantics or rhetoric, shifts in interests.

Soviet naval analysts Friedheim and Hehn (1977) made imaginative use of content analytic methods to determine Soviet positions in Law of the Sea (LOS) negotiations. They described (p. 345) their technique as follows:

> Soviet positions on five of the major issues that the USSR had entered into the UN record were measured by thematic content analysis of statements by official speakers who expressed for their governments a preferred position...This provides a systematic record of all major points made by all states in these negotiations since they began in 1967.

This technique allowed Friedheim and Hehn to conclude, for example, that the Soviets have been somewhat flexible on the issue of free transits of straits, but inflexible on fishing rights. They were also able to "score" national positions on LOS issues so that Soviet views can be compared with U.S., Japanese, or other national positions. They also compared the USSR positions with those of important individuals, e.g., Admiral Gorshkov.

Content analysis has been used to address traditional Kremlinological issues, e.g., what the attitudes of Soviet elites were toward Leonid Brezhnev (Heuer, 1978). Heuer analyzed how sixteen Soviet elites referred to Brezhnev. He found an index of personal reference rank-ordered the sixteen elites in terms of their political support for Brezhnev much like a panel of three CIA experts. A recent study by Kirk (1980) of all public speeches by members of the Soviet Politburo between 1972 and 1979 offers some interesting content analysis results. During World War II analysts became very adept at inferring Nazi intentions from the analysis of the content of Axis propaganda (See George, 1959).

Content analysis has the advantage of being systematic and having clear rules of coding and inference, aspects lacking in more impressionistic modes of inquiry. It has a variety of drawbacks however. It is extremely labor intensive. The reading and coding of every document or speech possibly relevant to a particular issue can require large numbers of man-hours and multiple coders, introducing the problems of intra- and inter-coder reliability. Texts in foreign language may not code equivalently in translated form, requiring original language coding, or validation experiments with parallel coding in the original language and in English to determine if coding of the translations is feasible.

Often the coding in content analysis consists of highly objective categories such as frequencies: how often has a given expression occurred in various media. Other codings are more subjective, as when pronouncements are coded as cooperative or conflictual. The rich symbolism in the use of language makes clear-cut coding of all linguistic meaning impossible. For example, irony is always subject to possible misinterpretation, and may not be appropriately coded. Euphemism, tone, analogy, simile, and Aesopian language are various forms of sophistocated expression that may confuse content coding and lead to miscategorization of expressions. The context of the communication may or may not affect its content. Speakers may or may not express what is on their minds. These problems influence the interpretation of the results of content analysis.

Each content analysis project requires the establishment of coding categories and rules for classification so that the coders know how to assign content to classes. While it is these rules that lend objectivity to the practice of content analysis, it also leaves the technique inflexible if new research questions arise that were not considered when the original categories were determined. New questions may require the re-coding of the same material to check for the new categories. The most modern methods of computer text reading and text storage will alleviate this problem of inflexibility to some extent, but such techniques have not been applied widely to content analysis and it will be some time before a fully flexible macnine-assisted content analytic procedure exists.

Features of Samples and Populations

The representativeness heuristic is the tendency (1) to assume that a sample is entirely representative of the population from which the sample was selected, and (2) to neglect features of the population that are not in the sample. The methods that reduce representativeness are those that improve the chances that the analyst's appraisal of a sample will accurately reflect the features of the population.

A common manifestation of the representativeness heuristic is the tendency to overemphasize case-specific information and underemphasize base-rate data. This suggests that anaysts should give more attention to statistics on central tendencies (means, medians, and modes) and dispersion (variance) in samples. Analysts rarely report such information and typically seem not to use simple descriptive statistics. Even the moving average (a smoothing statistic), commonplace in trend analysis and forecasting, is rare in intelligence analysis estimates, with the exception of those that deal with economic statistics.

One technique used by analysts to develop the features of a sample is expert opinion. A panel of experts is requested to specify features or aspects of a problem that are important for analysis. The analyst then collects data on these features for a sample and uses these for projection to a population. For example, Ivanoff and Murphy (NPSP, ch. 7) developed nine composite technical parameters that could be used to assess and project Soviet technical progress in anti-ship cruise missiles.

The trends in these composite parameters are estimated and then Ivanoff and Murphy wrote (p. 153):

> ...conclusions are drawn on future adversary systems that will be developed...This...requires a synthesizer rather than an analyst...all considerations of Soviet practices are merged with the factual evidence of the quantitative analysis. Future systems are synthesized and described...

Thorpe (NPSP, ch. 8) used a panel of experts to determine the mission priorities for each Soviet ship, aircraft, and submarine. Since many Soviet naval platforms are deemed multi-purpose, Thorpe's objective was to quantify their multiple features.

Dawes (1974) proposes that the role of the expert in predictive systems should be to determine which variables seem to be important and how they should relate to the prediction:

> There is no way of knowing apart from (the expert) what variables should be looked at. And the man knows what variables to look at only because he knows something about how they predict (p. 524).

Once the possible predictive features are identified, Dawes recommends that they be systematically tested in a model to verify that the expert-selected variables do predict. That is, in contrast to the traditional use of experts, as in the Ivanoff and Murphy study, where experts selected variables, and then synthesized judgements from them, Dawes recommends the experts select the variables and the synthesis and data integration be done mathematically or mechanically. This recommendation is based on the fact that humans are consistently inferior to mechanical information integration systems when synthesizing complex data patterns.

The use of experts to single out predictive variables and the combination of the variables in a mathematical model has been termed "paramorphic representations" (Hoffman, 1960) of the experts, and the use of the mathematical combination of variables to improve the judgements of the experts has been termed "bootstrapping" (Dawes, 1971; Goldberg, 1970). Models of the expert judges outperform the judges themselves because the mathematical integration of information is more reliable and consistant than human integration. Human assessments suffer from inconsistancies and distortions that are irrelevant to the problem being analyzed, whereas the mathematical analysis is more reliable. The prediction model is immune from boredom, fatigue, distractions, or variable application of the integration rules. To the degree that the human expert has less than perfect reliability, error is added to his or her predictions and can only reduce the accuracy of the estimate.

Paramorphic and bootstrapping models usually are simple linear combinations of predictive variables. Experts often reject such simple models as being inappropriate for what they perceive as complex, nonlinear, configural tasks (e.g., Slovic, 1969). The evidence from repeated studies suggests however that the experts perceptions of their tasks do not conform to their own predictions, which tend to be linear combinations (see Chan, 1981 for a review of studies). Simple linear extrapolations of known trends are usually more accurate forecasting methods than complex prediction systems, at least for the short and medium term (Hogarth and Makridakis, 1979).

A psychological technique that has had very little application in intelligence could aid greatly in systematizing expert opinions on the features of events or any other samples of stimuli. This is the technology of metric and nonmetric multidimensional scaling, tree-fitting and clustering. These methods will be discussed more fully below, under Characterizing Data, but are introduced here to demonstrate how they may aid in constructing meaningful perceptual categories for analysts from expert or analyst generated data.

Shepard (1980) recently summarized research on these methods, which are "computer based methods for constructing representations of the psychological structure of a set of stimuli on the basis of pairwise measures of similarity or confusability". These techniques yield three complementary representations of psychological structures: dimensional scales, taxonometric tree-structures, and clusterings.

Generally, this technique uses judges to assess the similarity between pairs of stimuli or to sort stimuli into categories. Alternatively, stimuli pairs can be presented to people for judgments of "same" or "different". The similarity ratings, in the first case, or confusion scores, in the second, can then be mathematically fitted into a dimensional space or into sets which preserve the psychological similarities and differences between the stimuli items. Either individual experts, or groups of experts, can have their judgments thus scaled or clustered. The output of such techniques is a set of dimensions or categories on which stimuli objects can be measured or compared.

In other words, one means by which Thorpe (NPSP, ch. 8) might have approached the problem of mission priorities of Soviet navy platforms would be to present each pair (of all possible pairings of platforms) and ask for a rating of the similarity of missions. These expert generated data would yield a set of dimensions or a taxonometric tree structure on which platforms with missions perceived to be similar would be closely located, while platforms which shared no missions (in the experts' opinions) would be found far apart. These dimensions or the shape of the taxonomy tree would be the fundamental mission aspects of the platforms in the perceptions of these experts. However, it is likely that the clustering or a hierarchical tree-figure would better represent the mission variables perceived by the experts than a set of dimensional scales.

Why should one use scaling/clustering procedures to abstract dimensions/structures of mission priority for Soviet Naval platforms? Why not merely ask experts about these missions, as Thorpe did? First, experts do not all perceive stimuli in exactly the same manner. While Thorpe's method may allow an estimate to be made of the differences between experts (e.g., range of opinion), how it does this is not clear since the experts' opinions are shifting due to the Delphi technique. In contrast, scaling/clustering gives a precise measure of unaccounted variance. Second, experts may have highly complex multidimensional preceptions which they cannot readily dissect without aid. Introspection may be inadequate to abstract these conceptual categories or dimensions. Similarly, an analyst could use such techniques on himself to learn what categories or dimensions seemed to be important in a complex, multidimensional problem. Third, it is feasible, at least in theory, to obtain these scaling/clustering results unobtrusively, at a distance, e.g., from the writings of Soviet naval officers or authorities.

The latter application of scaling requires content analysis of the co-occurrence of descriptors with objects. For example, suppose Soviet Admiral X, a naval expert, always decribes the *Kresta I* and *Kresta II* cruisers with identical modifiers. Furthermore, assume some of these modifiers are used to describe *Karas*, but none are used to describe *Krivaks*. A measure of similarity can be obtained by means of the degree of overlap in use of modifiers for these and other ships. These similarity measures, in turn, can be used to determine the degree to which (and the categories or dimensions on which) Admiral X perceives these ship types as similar or different (Rosenberg and Jones, 1972; Rosenberg and Sedlak, 1972). Similarly an analyst might investigate the perceptual dimensions of Soviet statements on missions (e.g., sea denial, anti-sea lines of communication, etc.), doctrine (e.g., protection of state interests), events, capabilities, etc. That is, merely by describing stimuli, Soviet spokesmen are revealing considerable information on the perceptual categories and dimensions they apply to complex objects and events. This information can be abstracted from their statements by analysts and evaluated with scaling or clustering techniques.

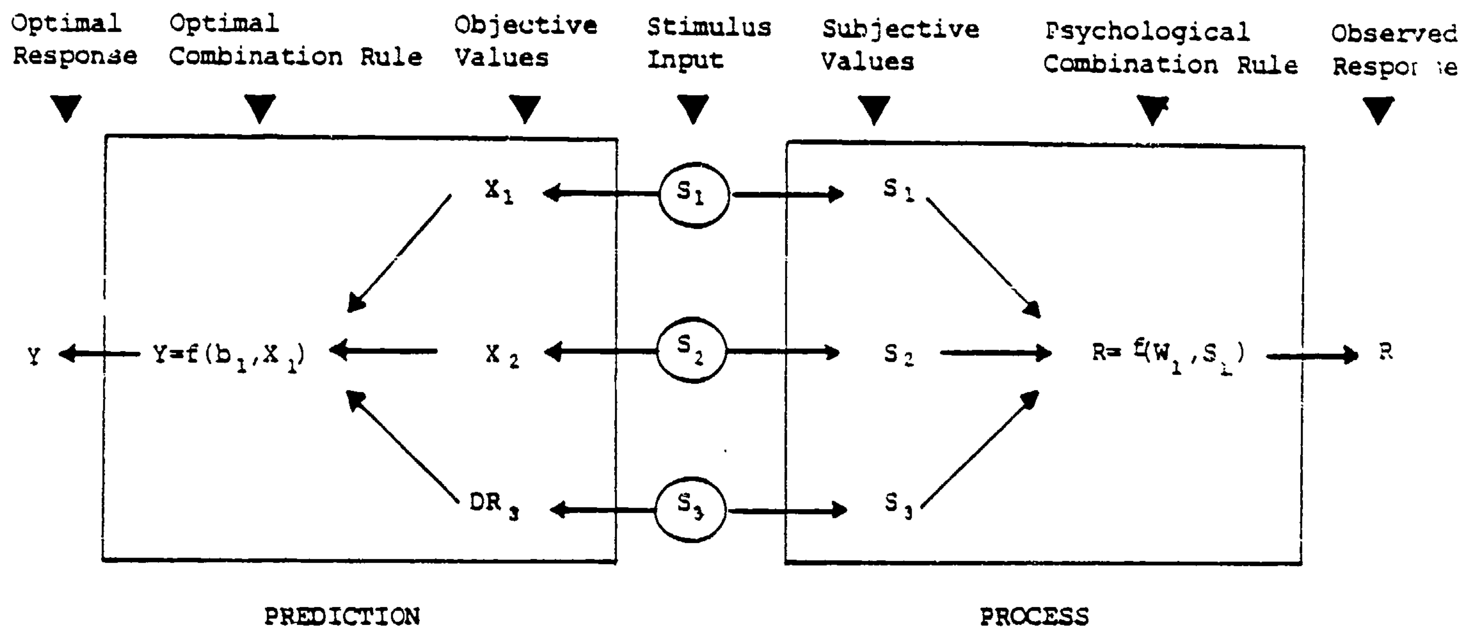
B. WEIGHTING DATA

The fact that people often cannot accurately report the weights they attach to data in making estimates suggests that explicit "policy capturing" assessments of analysts may assist them to understand and improve their estimation processes. That is, an explicit effort can be made to model or capture the quantitative elements of the estimation process of the analyst including data weighting.

Policy Capturing

Figure 13 shows schematically how this can be done. A controlled set of data stimuli (S) are presented to the analyst and the analyst's estimative response is observed (R). The right side of the diagram suggests that analyst's estimation process. The input stimuli and the output response for this process must be quantified; i.e., they may be qualitative in nature originally but they must be scaled, coded, or rated by the analyst to yield at least a more-less,

Figure 13. General Estimation Diagram Illustrating Prediction Analysis (on the left) and Process Analysis (on the right). Stimuli (S_i) a common to both.*



*From Shanteau and Phelps, 1977: 258.

plus-minus quantitative estimate. If the inputs can be quantified directly, analyst coding of them is not needed. This process is repeated with varied stimuli sets, i.e., the analyst sees a new set of data stimuli and makes a new estimate.

The analyst's weighting policy is captured by using the same coded data stimuli sets to mathematically predict the analyst's estimative responses, R. Shown on the left in the diagram. That is, using the same inputs, S, we solve for the combination rule which makes the optimal responses, Y, as close as possible to the analyst's responses, R. The combination rule found will reflect a set of objective weight values that indicate the weights used by the analyst in his or her estimates. The objective weights provide indices on the degree to which the analyst used each of the input data stimuli dimensions.

Two mathematical policy capturing methods are commonly used, linear regression and Bayesian analysis (Slovic and Lichtenstein 1971). In general, the linear regression method has been used more often for assessing data weights, but the Bayesian assessment of the "diagnosticity" of data can also provide information on the degree to which analysts weight data (see Edwards, 1978).

If the important dimensions are known on which analytic judgments are made (e.g., from a multidimensional scaling of analysts' similarity judgments, see above) analysts might simply be asked to rate or rank the dimensions in importance. These weights can then be compared to the policy capturing objective weights to determine the self-insights of the analysts into their judgment and information processes.

Bayesian Techniques

Bayesian estimation methods have been used extensively in intelligence analysis (see Slovic and Lichtenstein, 1971: 717-721, and Slovic, Fischhoff, and

Lichtenstein, 1977: 25-28, for reviews and references). In general, these methods help analysts integrate probabilistic data into their judgments, avoiding the tendency to underweight such data and not adjust posterior probabilities sufficiently. A recent application of Bayesian methods to the problem of estiamting probabilities of a Middle East conflict was described by Schweitzer (in Heuer, 1978, ch. 2). As Schweitzer noted (p. 19) these techniques have been applied to a variety of intelligence estimation issues: the likelihood of a North Vietnamese offensive in 1974, the probability of a Sino-Soviet conflict, the chances of an Arab-Israeli war, and the analysis of order of battle data (p. 13).

Bayesian methods have also been recommended as a means by which the estimates of different experts can be effectively combined. Hennessey (1977) recommended a Bayesian paradigm for the systematic cumulation of evidence from related studies. He suggests this would overcome three widespread difficulties which make research data difficult to interpret: (1) the circuitous and nonintuitive logic of traditional statistics, (2) lack of agreement (often latent and implicit) among analysts on the substantive and technical premises adopted in research arguments, and (3) the low diagnosticity (weakness) of research data for distinguishing among alternative hypotheses. Morris (1974, 1976) recommended Bayesian methods for integrating the judgments of experts into a single estimate and outlined a possible combination of mechanism for this (1976).

Sample Size and Base-Rates

Bayesian methods can also help the analyst with a common weighting problem, the tendency to overweight case data and to underweight base-rate data. By successively updating prior probabilities, the Bayesian techniques "build" base-rate data into the estimation process. They also tend to scale down overweighted case data by adjusting its impact downward via the prior probability base rate.

Simpler methods can help analysts give more weight to groups of cases then to the most recent, vivid, or salient case. When case data are coded, use of a moving average tends to "smooth out" extreme data points (i.e., highly salient or vivid cases) and adjusts recent points to reflect the recent base-rates (i.e., the span of the moving average). Comparing case statistics to the central tendency for all cases (e.g., averages) allows the analyst to put the case into perspective.

The use of averaging, moving averages, and Bayesian probability updating also reduces the tendency to overweight data from extremely small samples, and helps reinforce the law of large numbers, that estimates based on large samples of cases are more representative than estimates based on small samples. The two main objectives in these methods are to help the analyst to avoid under- or over-reacting to a single piece of information, and to use the cumulative information contained in base-rate data.

<u>Redundancy and Variance</u>

Analysts can control the tendency to overweight redundant cues by computing the correlations between them and reducing the weight attached to a highly redundant (correlated) variable. Similarly, the analyst should note the variability in his data and reduce his weighting of a cue which stops varying and becomes static.

C. CHARACTERIZING DATA

A variety of memory biases were discussed which affect the characterization of data and its organization into categories, factors, classes, and generalizations. Problems were noted that result from attempts to impose "either/or" categories on events or objects which share similarities as well as differences. The tendencies to place events and objects into taxonomies was discussed and the possible biases that may result were described. A variety of analytic methods and paradigms can help offset these difficulties.

Aiding Memory

The remarkable ability of naval analysts to recall information is due largely to their complex conceptual schemas for perceiving relationships between aspects of the Soviet Navy. These schemas allow analysts to efficiently code, store, and retrieve information. In general, however, the schematic rules which govern these processes are tacit and implicit in the analysts' narrative estimates. One implication of this is that analysts may disagree becasue their different schemas lead to different perceptions and memories of the same or similar data sets. Since the schematic rules for processing these data sets are not explicit, such sources of differences cannot be explicitly determined.

This suggests that if analysts make their schematic coding and storage of information more explicit, the job of determining the sources of differences among analysts in categorizing data would become easier. Analysts would be able to compare categorization systems and contents explicitly, as well as comparing their conclusions.

The tremendous growth of data base management systems and management information systems in business, adminstration, and government reflects the appreciation of the need for extensive and flexible means for accurately coding, storing, retrieving, and organizing information. While these systems tend to be most often applied to quantitative data, they can also be applied to coded event data or content analysis data. Kirk's (1980) content analytic data on Soviet political elite speeches, for example, is stored in a data base management system, greatly facilitating extensive data retrieval or manipulation and empirical investigations.

In earlier work by Stech (1981, ch. 4) an example was used to demonstrate that a naval analyst's conclusion was based on biased sampling; Stech used that analyst's explicit coding of events (diplomatic port visits). Had these explicitly coded data not been available, we could not have determined that the analyst's conclusion was based on a biased sample.

Naval analysts themselves complain of being unable to deduce the implicit coding schemes of other analysts. For example, McConnel (SNI 612) wrote:

> McGwire and Erickson...count on their subjective impressions of "tone" and "thrust" to tell them that Groshkov is an advocate. I respect the conviction behind this approach, and perhpas I avoid it only because I'm not good at it...It has been my experience that others have a so-so record in this, too, as often as not.

Making subjective impressions into explicit coding rules is often a fairly straightforward (if not always simple) process that is in keeping with the scientific requirement that subjective impressions be replaced by quantitative measurements. Such coding greatly helps analysts determine the validity of their own and others impressions, as well as facilitating accurate recall.

Event coding and content analysis also aid the analyst in efforts to avoid "selective retrieval", by facilitating the recall of the actual original data rather than a retrospective reconstruction of it. The tendency to "constructively remember" events, a process affected by hindsight and a variety of memory biases, is minimized if the analyst can quickly ascertain all the other cases that fall into a given category, or that compare favorably on salient dimensions to a case in point. Mechanical coding systems provide the analyst this ability to organize and manage stored information and help minimize the analyst's need to rely on limited and fallible memory.

Another reason analysts should make greater efforts to code their impressions is the tendency for memory to distort psychophysical impressions of extreme cases. Extremes (biggest, worst, smallest, etc.) are well remembered as being extreme, but their recalled dimensions tend to be less extreme than the original dimensions. Psychologists have made considerable progress in developing methods for scaling, scoring, rating, or otherwise measuring psychological sensations (e.g., see Anderson, 1979). Analysts might employ such functional measurement techniques to record the magnitudes of their original impressions so that current impressions could be maintained for later, accurate comparisons with future impressions. Anderson (1979) reported results from a study by Leon that are relevant to the present study. Adults' and childrens' impressions of "naughtiness" were recorded for various incidents that varied in terms of the degree of severity of damage done, and in the intent of the person doing the damage. Naughtiness was found for both adults and children to be a linear function of both severity and of intent, but children weighted intent less and damage more in determining naughtiness than did the adults. It is highly likely that many naval analysts conceive of threats as being a function of capabilities and intentions. It would be very useful to try these psychological techniques to determine how analysts emphasize these two components of threat for various specific issues and questions, and to be able to record their impressions over time in some form of comparable metric.

Assessing Prototypes, Categories and Dimensions

Stech (1981) argued that people conceptualize complex events and objects in "fuzzy" categories with loose, overlapping boundaries, rather than in "either/or" pigeonholes. This implies that category membership is a matter of "family resemblance" and that stimuli are coded in terms of many dimensions relative to a central prototype. Analysts would be best served if their impressions could be coded on the natural dimensions they themselves use to perceive events or objects, and if their subjectively meaningful concepts of family resemblance and prototypicality were the bases for the coding and content analyzing of data. Rather than imposing arbitrary coding dimensions, or content analysis categories on the analyst, recent psychological research suggests that it is possible to determine explicitly the natural categories or dimensions the analyst uses. These natural categories, or dimensions once made explicit, could then provide the analyst the ability to systematically characterize his or her impressions without grossly distorting the analyst's cognitive process. This psychological research also suggests that the natural categories or dimensions are rarely completely explicit in the analyst's mind before such assessment. These techniques may thus help make the analyst's methods and assumptions more accessible, in keeping with Heuer's (1978) recommendations.

Objects or other stimuli can be considered to have a set of features or attributes. A person's total data base concerning any given object is rich in content and complex in organization and form. It includes features of appearance, meanings, functions, relationships, history, and all other properties that are known or can be deduced. When faced with a particular analytic task (e.g., to identify the object, or determine its similarity or dissimilarity from other objects) people extract and compile from their data base a limited list of relevant features to perform the task (Tversky, 1977).

There are two approaches that can be taken to relate objects to one another. One measures the distances between the features of objects in a geometric sense. The other considers the overlap of common features relative to uncommon features in a set-theoretic sense. Which of these two approaches are used to identify, quantify, and organize prototypes, categories, or dimensions will depend on the objects, the task, or both (Stech and Tversky, 1977; Shepard, 1980). Some problems facing naval analysts have a natural dimensional structure, e.g., estimating the severity and probability of threats. Others have hierarchical structure that may reflect an evolutionary process in which the objects all have an initial common structure and later develop additional distinctive features. An example of the latter might be the classification of Soviet Navy platforms.

Soviet naval analysts have no widely accepted method for classifying the Soviet general purpose navy into its component missions. Thorpe (NPSP, ch. 8) attempted to develop such a method using Delphi techniques. A variety of problems occur when Delphi techniques are used (see, e.g., Morgenstern, Knorr, and Heiss, 1973: 23-26). For example, the range of expert opinion converges sharply, although there is no normative reason why such convergence should lead to greater estimative accuracy. The central tendency of expert judgments often shifts, but in a manner that has no discernible relation to the new information available to experts. That is, it is unclear whether Delphi is an appropriate mechanism for information integration. Delphi is also costly and time-consuming.

Clustering and scaling techniques offer a far more promising method of categorizing judgments by experts and analysts than the questionable and costly Delphi method. For example, to determine how analysts classify Soviet general purpose platforms in terms of wartime missions, Thorpe's objective, one could follow the procedure used by Rosch and Mervis (1975) to classify vehicles (see also Tversky, 1977: 338). Analysts would be asked to list all the possible wartime missions, or alternatively, one could give all analysts a list of missions, as Thorpe did. For each wartime mission and each platform, analysts would be asked to list those features of the platform that were relevant (positively or negatively) for that mission. These lists provide the wartime mission-relevant features of each Soviet platform for each analyst or expert. A master catalog (something like Jane's Ships) of features and attributes for each platform could be made available to all analysts, listing weapons, electronics, beam, draft, propulsion, etc., to refresh the analyst's impressions and to serve as a codebook for the listings.

It is then possible, using the lists of features, to determine for each pair of platforms the number of common and distinctive features. From these data it is possible to predict with high accuracy the analysts' ratings of similarity between platforms given any wartime mission. using the data on shared and nonshared features and (derived or obtained) ratings of similarity, clustering programs can be used to determine a hierarchical clustering diagram. This diagram provides a detailed classification of each platform's perceived capabilities, relative other platforms, to perform each wartime mission. The diagrams created by these programs reveal the main cognitive categories used by the analysts to make these judgments. That is, not only are Soviet Navy platforms categorized by wartime mission, as Thorpe attempted, but the clustering algorithms allow us to determine the main dimensions the analysts used to make these judgments. The latter information cannot be derived from Thorpe's Delphi method. The net result from the clustering approach would be a classification diagram for each wartime mission showing how each Soviet platform compares with all other platforms in accomplishing that mission, how the analysts grouped platforms of similar capabilities , and how (and why) individual platforms and groups of platforms differ from each other in performing that mission.

These feature analysis techniques also allow us to determine for each wartime mission the prototypic features for that mission. A measure of family resemblance (distance from the prototype) for each platform can be derived from these data which allows direct scaling of the rankings of individual, multipurpose naval ships and aircraft among a variety of missions. That is, a ship with an extremely close resemblance to the prototype for a particular mission would be scaled to have a high ranking for that mission.

Thorpe's analysis requires that if a platform has a high weight (percent) for one mission, it must have a low weight for other missions (i.e., percents can only add up to 100). This forces "either/or" distinctions into what are actually "and/both" judgments. For example, a modern Soviet ship such as the Kresta II may be a far better antiship platform that the obsolete Krupny, but in Thorpe's method, the former gets 20 percent for antiship (because of its heavy weight for ASW) while the latter gets a 70 percent. While Thorpe's method may be necessary for the economic analyses he performs, it is probably highly misleading as a reflection of analysts' categorization of platforms and missions. In Thorpe's method a Soviet choice of an obsolete Krupny over a Kresta II would be scored as an increased Soviet Navy emphasis on the antiship mission simply because the Krupny, ineffective as it may be at antiship warfare, is even less capable of ASW. This seems to be an absurd conclusion no naval analyst would make. A more meaningful measure of mission priority would be to measure the family resemblances of each year's new Soviet platforms to the mission prototypes. This would allow mission comparisons without imposing the unreasonable trade-off logic of percentage estimates.

Intention Categories

The classification of Soviet naval platforms provides a ready application of modern categorization methods to Soviet capabilities. The many overlapping and nonoverlapping features of ships provide an easily quantified basis for such analysis.

With appropriate adaptations the same categorization methods might be applied to issues of intentions as well. For example, naval analysts develop models of Soviet naval behavior in various situations and cases (e.g., McConnell, in Dismukes and McConnell, ch. 7). It should be possible to assemble lists of the features of these situations and cases. Analysis of these features could then determine the dimensions on which analysts perceive Soviet intentions as varying. The dimensional structures obtained by analysis of overlapping and nonoverlapping features could be evaluated by assessing the analyst's perceptions of similarity and differences between cases. These latter data can be used to create a dimensional space that should correspond to the dimensions obtained from features analysis.

In short, modern psychological techniques allow a quantitative assessment of semantic and perceptual dimensions or categories. The natural categories can be determined and then used as the basis for explicit coding and quantitative analyses. Because the dimensions and categories are obtained from the analyst's own cognitive relations, they tend to yield intuitively useful classifications, which, however, are not intuitively obvious and often cannot be obtained by other methods. An important consequence is that the quantitative coding based on these techniques is likely to be high in qualitative meaning to the analyst.

<u>Factor Analysis in Intelligence</u>

Several recent studies have attempted to quantify the factors analysts use in analysis. Dahlgren (in Heuer, 1978, ch. 5) translated a complex theory of international political violence into about fifteen separate social and societal factors, or variables, and various relationships among them. A panel of intelligence analysts assigned numerical values to each of the theory's variables and the median scorings across analysts for each variable were used to evaluate the theory. Note that Dahlgren did not derive the variables from the analysts' cognitive relations, but from a theory of political violence.

One intelligence application of the features similarity techniques described above was Kent and Wiley's (in Heuer, 1978, ch. 8) use of multidimensional scaling to determine voting blocs in the United Nations. While useful and suggestive, and well-received by the analysts, the results of this investigation are ambiguous largely because the wrong methodology was applied. Voting bloc analysis is inherently a clustering or grouping problem, rather than a dimensional problem. Rather than multidimensional scaling, Kent and Wiley should have applied a clustering or tree-building technique, and attempted to develop a voting bloc taxonomy. A cluster or tree diagram would be much more representative of the voting subgroupings, which Kent and Wiley set out to find, than are the scaling diagrams they derived from voting similarity data. The main point, however, is that intelligence analysts have used these psychological methods on features similarity data with some success.

Friedheim and Hehn (SNP, ch. 18) used content analysis of United Nations documents on the Law of the Sea to determine Soviet positions on various issues under negotiation in the Law of the Sea talks. For each issue (e.g., rights of transits through straits) the frequency of various themes mentioned (e.g., free transits with exceptions and limits) provided a score for that theme. By scoring themes, Friedheim and Hehn were able to compare the view of the United States, the USSR, and Admiral Gorshkov on several Law of the Sea issues. Friedheim and Hehn were also able to predict how often nations would favorably mention various themes by regressing characteristics of the countries (e.g., geographical characteristics, economic interest, etc.) against the thematic scores. This technique allows the analysts to determine factors which seem to make a difference in bargaining positions. For example, the important factors for the USSR position on various issues were (p. 354): membership in the East European caucusing group, major merchant fleet, Blue Water Navy, major fishing state, distant water fishing, straits state, broad shelf, major mineral producer, major oil producer, and offshore oil producer.

Several naval analysts have developed methods aimed at assessing the intentions of Soviet ship designers. These methods take an approach that includes explicit decomposition of design requirements and ship features and an attempt to logically relate one set to the other. While neither of the two methods described below used the psychometric features analysis methods described above, both suggest that systematic use of features data is not foreign to naval analysts.

Kehoe (SNI, ch. 19) presented a methodology for assessing the factors underlying warships design which is quite compatible with the psychometric approach outlined above. He determines the ship mission requirements for payload and performance in terms of various factors: hull size, seakeeping, propulsion speeds, cruising endurance range, habitability, payload, electronics, and weapons. These are the types of features we suggested above that analysts could assess and list for various naval platforms. In his chapter, Kehoe compares trends in these features for Soviet and U.S. ship types over time. Kehoe had experts evaluate the trend data and rank the major design characteristics into priority order, or as Kehoe put it (p. 380):

> ...which of these design characteristics appeared to "get the biggest piece of the cake" in the ship design process?

The experts determined that Soviet ship designers emphasize factors in the same priority as did the United States in building World War II vintage ships.

Another means of determining priority was not used by Kehoe, but has been used in other fields to determine critical features of technological change.

This is to regress various design features to predict some feature of technical merit. One such feature is initial operating capability date, which is a surrogate for modernity. The question then becomes which design factors seem to be driving Soviet ship developments? A similar technique was used by Alexander and Nelson (1972) to assess the factors influencing Soviet and U.S. aircraft turbine engine designs.

Kehoe evaluates his data on design factor trends in terms of various Soviet missions (e.g., sea denial). It would be interesting to use the features analysis methods outlined above to determine natural taxonomies or dimensions of Soviet ships, and to see if similar mission categories or dimensions emerged.

Meier (SNI, ch. 20) describes a different method aimed at the same objective as Kehoe's, determining the relative priorities given by the Soviets to firepower, sustained combat, command and control, speed, endurance, seakeeping, protection, and personnel support. Meier's method is a reverse engineering process which uses an iterative computer program that attempts to determine feasible design requirements, standards, and practices from the features of the finished ships.

D. ASSESSING COVARIATIONS

Stech (1981) noted the tendency of people to overestimate the strength of theory-driven covariations and to neglect or underestimate the strength of data-driven covariations. We also found this tendency seemed to extend to naval analysts. Obviously, the most direct remedy for these problems is to increase the use of quantitative measures of co-occurrences. Such measures should be habitually applied to any data that could conceivably be related. This will tend to reduce beliefs in spurious theory-driven correlations, because accumulating evidence will fail to support the theory, and to draw attention to unsuspected relationships. Only some of the observed correlations in data sets will be meaningful: some will be spurious statistical noise, and some will be statistically reliable but uninterpretable. None of these are major problems. Perhaps the main problems are to encourage analysts to give data-driven patterns of co-occurrence serious consideration, and to reconsider theories which are unsupported by data patterns.

Actuarial Models and Backcasting

Referring to Figure 13 we can note the optimal response, Y, given on the left side of the diagram. This is the best prediction possible given the objective values, X, of the stimulus inputs, S. For example, a naval analyst may be able to measure (X) several features (S) of the Soviets' building program for a particular ship class. The analyst may regress these against actual production figures in the past (criterion values) to determine the optimal combination rule. This rule can then be used to predict future output, Y. This process is labeled actuarial and as was outlined in Chapter 4, the actuarial prediction process is consistently superior to the process shown on the right in Figure 13, the clinical process. That is, when estimators attempt to intuitively integrate information on the inputs, S, using a psychological combination rule, they are less accurate predictors than an actuarial model of the same judgments.

This suggests that analysts make greater use of actuarial models in formulating and revising their theories about covariations. We will have more to say about these techniques below under "Improving Predictions". One of the advantages of actuarial methods is that they force the analyst to assess the relationships between criterion values and input data. Actuarial methods force the analyst to consider the strength of theory-driven covariation hypotheses in light of past data. If the theory-driven relationship is absent in the past, the analyst must reconsider using the theory as a basis for future predicitons. This use of actuarial techniques and past data to check suspected theoretical relationships is a form of "backcasting", i.e., an attempt to correctly estimate when true values are already known. If a theory cannot successfully fit past events, its validity for the future is questionable. If the theory consistently over- or under-predicts past criterion values, the analyst can made corrections to "debias" the error, and thus improve the theory's accuracy.

The actuarial formula provides the optimal predictions given the input data, but it may not be as useful or interpretable as the analyst's theory. In other words, the analyst may wish to retain and improve his or her theory (which is heuristic and interpretable), while using the actuarial formula for making predictions. As the theory comes to resemble the actuarial formula, the former becomes more accurate as a predictor, and the latter becomes more interpretable.

In effect, research on clinical and actuarial judgment suggests that estimators should attempt to assess covariations between input data and criterion values if an actuarial method is possible. When actuarial methods are possible they provide the optimum description of the covariations between input data and criteria, and greatly simplify assessment of covariation.

It is important to note that the expert judge must specify what the input variables are to be. Actuarial modeling itself cannot determine what input stimuli should be considered as predictive of the criterion values. Both the analyst's theory-driven hunches and empirical data-driven search for possible correlations can provide clues for the inputs to the actuarial methods. Only the naval analysts can know what variables are likely to be worth checking as possible predictors of future Soviet behavior. The actuarial method is the optimal means of assesing these hunches.

Bootstrapping

When criterion values are known or knowable, as in Soviet ship inventories, it is possible to fit regressions of input data on the criterion values to build actuarial prediction models. Analysts, however, are often called upon to estimate values for which there are no clear criterion values, e.g., the level of Soviet threat. In these cases analysts judgmentally provide the measure of the criterion values, deciding , e.g., whether the Soviet threat is high or low. No objective criterion values exist for such cases against which to compare predictions, i.e., the Y or R in Figure 13 do not correspond to any unambiguous measures.

This lack of criterion values does not eliminate the possibility, or the need, for systematic methods of assessing covariations. Psychologists have determined that linear models which fit regression equations to past clinical judgments (i.e., the R's in Figure 5.1), can be used to replace the clinical judgment process. Such models outperform clinical judges because they eliminate variations in human reliability. Reliability places an upper limit on validity, if reliability increases, greater accuracy is possible.

This technique of modeling the judge's judgments and then using the judgments of the model has been labeled "bootstrapping", i.e., pulling the judge up by his bootstraps (Dawes, 1971; Goldberg, 1970). Bootstrapping will improve judgments slightly under almost any realistic task conditions and it can be applied blindly, in cases where criterion information is absent or vague, with the expectation that the predicitions made will be improved (Camerer, 1980). Futhermore, as long as the regression model of the judge determines the input variables for the prediction, determining the exact weights used by the judge is not necessary; equal weights are about as good (Dawes and Corrigan, 1974). In other words, the weighting parameters of the bootstrapping model need not be very specific once the right variables are identified. The key again is knowing which variables to try and the only realistic source for these is the expert. Once the naval analyst identifies the variables that seem important for making an estimative judgment, the bootstrapping method can best determine the actual covariation between those variables and the expert's judgments. These variables can then be combined linearly with equal weights to estimate the expert's future judgments more reliably (and thus more accurately) than could the expert.

There are obviously implications for prediction in these actuarial and bootstrapping models, but the point for this section is that they provide systematic, data-driven means to appraise suspected covariations that eliminate the problems of theory-driven covariation appraisal.

Environmental Effects

Since expert analysts must play a central role in selecting the variables for use in actuarial or bootstrapplong models, it is important to reiterate the point that people are often insensitive to environmental effects such as regression or base rates. That is, analysts are unlikely to adequately attend base rate variables, and are likely to overattend case-specific variables (Einhorn and Hogarth, 1978). In particular, to improve covariation assessments, it is necessary to include data on disconfirming events as well as on positive hits. That is, analysts should be especially careful to collect and record data on (at least) all four cells of the four-fold present-absent cross-tabulations that determine the relation between an indicator variable and a predicted variables.

One means used frequently by naval analysts to display base rate data is the time trend line, i.e., a graph of data over time. For example, Kehoe (SNI, ch. 18) displayed data on various capabilities of Soviet and U.S. ship types over time to demonstrate changing trends in various features. On the other hand, merely displaying a base rate does not mean the analyst will use the base rate in estimates (e.g., see Ivanoff and Murphy, NPSP: 149, Figure 32).

O'Leary and Coplin (1975, ch. 4) presented a detailed case study of how careful cross-tabulations of data assisted State Department analysts to make judgements about covariation between military expenditures and other variables in Latin America. Although the State Department analysts made extensive use of quantitative data, they did not explicitly assess bivariate correlations to evaluate their theories or forecasts. O'Leary and Coplin showed that several postulated relationships between military arms acquisitions and other less quantitative factors (e.g., economic conditions, role of the military in government, need for internal defense) could be directly assessed with cross-tabulations or correlation analysis. These techniques were applied to cross-sectional data (i.e., a group of Latin American countries) as well as to data on individual nations. The covariation assessments were able to confirm several of the State Department analysts' theories, as well as showing that some theoretical relations did not exist in historical data. O'Leary and Coplin described (p. 148) this covariation "backcasting" exercise as:

> ...one viable way of incorporating clearly defined variables, quantitative techniques of analysis, and the important discriminating character of the Latin American nation, all of which appear to be essential to making accurate explanations and predictions about changes in Latin American military spending.

The main implication of the O'Leary and Coplin study is that quantitative covariation assessment methods can be applied even when the factors involved are largely qualitative.

E. CAUSE AND EFFECT ASSESSMENT

The problems analysts may encounter in generating cause and effect explanations of intentions were reviewed in Stech 1981. These problems generally fall into two groups: search and hypothesis generation. (This division corresponds to the two basic psychological models of problem-solving and thinking strategies: heuristic search and hypothesis generation and test, see Gerwin and Newsted, 1977.) Search problems interfere with the analyst's inductive search for causal patterns in data sets, or bias the search, so that certain erroneous or misleading patterns are easily found and labeled as causal. Hypothesis generation problems interfere with the analyst's deductive reasoning from data patterns and lead the analyst to generate false hypothetical explanations from available data sets. These false hypotheses may then mislead the analyst's subsequent perceptions and analyses.

Stech also noted that naval analysts themselves seem to be aware of the methodological difficulties of establishing cause and effect relationships (see, e.g., Booth, SNI: 470), but few of the sampled naval analysts made any use of sophisticated methodologies or causal analysis. Some of these causal methodologies have been applied to problems of naval intelligence and others to nonnaval intelligence estimation problems.

Causal Search

The sources of biases of causal search can be divided into (1) incomplete perceptions, (2) mistaken perceptions, and (3) overly simplistic perceptions. These problem areas overlap (incomplete perceptions contribute to overly simplistic perceptions) and are more than what is typically meant by "perception". We use the term "perception" here to emphasize an inductive search of data features for causal patterns which then modify hypotheses.

Incomplete perceptions. Incomplete perceptions of causal patterns are probably largely due to faulty covariation assessment. That is, an analyst may be over-attentive to positive instances of covariation of one event and another (++ events). This over-attention to positive cases can suggest a causal relation if the analyst neglects to search out evidence of +- or -+ events as well. That is, the suspected cause may be present when no effect occurs, or absent when the effect occurs. The analyst should also confirm the negative case, that the effect is absent when the cause is absent (-- events). There is a tendency, however, to not search beyond the ++ events for evidence of covariation. In the section preceding this one we outlined a variety of cross-tabulation and covariation assessment methods which help the analyst evaluate a suspected pattern of covariance.

Mistaken perceptions. Mistaken perceptions are those patterns which tend to be perceived as causal because of intuitive cognitive logic or perceptual bias regarding causal relationships. For example, things we focus our attention on tend to be perceived as causal. Outcomes which match the intentions of an actor are naturally assumed to have been caused by the actor. Events which share temporal or physical characteristics may be perceived as causally related. These intuitive logical relations and biases may lead the analyst to search for data to confirm these patterns, producing a belief in a false causal relationship, bolstered by selectively perceived data. These perceptions and the beliefs that result often seem intuitively obvious and amply confirmed by the data which are selected to "prove" them. It is quite difficult, if not impossible, to avoid such perceptions and beliefs altogether since they are based on highly efficient and logical concepts about causality. That is, for many or most causal relationships these perceptions and concepts are not mistaken at all.

Rather than attempting to avoid such perceptions or concepts, it is easier to attempt to establish their validity as systematically as possible.

The sampled naval analysts typically evaluated causal relationships narratively. That is, evidence in favor of a suspected causal relationship was listed and evaluated narratively to establish a coherent relation between the suspected cause and the effect. Evidence for or against (typically against) other possible causes would be listed and evaluated. Rarely was there any description of an effort mounted to make these evaluations systematically comparable, i.e., to give the favored and disfavored possible causes an equivalent assessment. that is, analysts typically reported what amounted to _results_: the supported hypothesized causes (and the disconfirming evidence). The strengths or equivalence of the assessment methods often could not be judged.

Causal analysis. Several social scientists have outlined methodologies for making systematic assessments of causal relationships in nonexperimental research (e.g., Asher, 1976; Blalock, 1964; Heise, 1975). These methods are labeled "causal modeling", or "causal analysis", and are techniques for selecting variables that are potential determinants of effects, and attempting to isolate the separate contributions to effects made by each suspected cause. Because all the candidate causes are evaluated within the same model, the assessment tends to be more systematic, equivalent, and transparent. As we noted earlier, "causes" can never be proved because they are mental rather than physical constructs. Causal models are merely analytic aids for evaluating and assessing hypothesized causal relationships.

Causal models typically use mathematical regression equations as causal equations. That is, a variety of suspected causes are combined in a linear equation in an attempt to determine the impact of each suspected cause on the effect. Causes with nonadditive, interactive effects (e.g., multiplicative ones) can often be easily modified into additive ones so that powerful linear mathematical methods can be applied (Blalock, 1964)). For moderately complex causal models a technique known as "path analysis" allows for the estimation of the magnitude of the linkages between causal variables, the possible causal relationships between variables, and the direct and indirect effects of variables on each other. Path analysis also allows for tests of the adequacy of the whole causal model (Asher, 1976).

The most sophisticated causal models include time as a major causal variable and consider the possibility of feedbacks in the system of causes. That is, the indirect effects of cause A on cause B may feed back on Cause A so that, at a later time, cause A has a different effect on cause B and on the overall effects in the system. These models are labeled "dynamic systems" and usually require extensive data on variables over time.

O'Leary and Coplin (1975, ch. 7) developed a series of quantitative causal relationships for State Department intelligence analysts in an effort to forecast the strength of evolving calitions among oil-exporting and oil-importing countries. They were attempting to translate State Department analysts' hypotheses into quantitative form, so this application is perhaps more relevant to analyst problems of hypothesis generation than to problems of causal search. However, because causal models can reveal unsuspected relations and invalidate intuitive relat. .ns they also serve to enhance the search for causal patterns.

O'Leary and Coplin developed quantitative indices of internal cohesion for the two groups of nations and of the bargaining between the two blocs. Data on oil and nonoil transactions (trade) and on votes on political issues were used to estimate future relations between oil-producing states and oil companies. These relationships were found to shift over time, suggesting a dynamic model of cohesion variables (causes) on the bargaining variables (effects).

Simplistic perceptions. The last set of problems analysts may have with causal search is related to overly simplistic perceptions. We use this term to reflect such problems as "minimal causation" (the tendency to search for the first plausible cause, see Stech 1981, Chapter 4), "causal hydraulics" (the tendency to perceive causality as fixed in amount), and reductionistic models (the tendency to accept too small a set of causes as responsible for an effect). These problems may lead the analyst to overlook valid causes, to ignore the possibility of multiple causes, or to see effects as coming about in only one way.

Search trees: To help overcome some of these problems, analysts might use "fault tree" and "decision tree" techniques. Fault trees are diagrams developed by engineers to determine how a particular event (a fault or failure) could occur in a system. The engineers reason backwards from effect to possible causes. For example, the engineers may want to determine all the possible ways an automobile might fail to start. This technique has been used extensively in nuclear safety analysis.

If a fault tree is an effort to answer the question "how could a state like this come about", a decision tree addresses the question "given this state, where can events go from here". In other words, a decision tree might be created for the possible future actions a mechanic might take if confronted with a car that refuses to start. Fault and decision trees are examples of what psychologists term problem-solving search trees.

Tree diagrams serve to systematize suspected relationships and the sequential nature of events and responses. They also increase the possibility that new relationships may be perceived that otherwise might have been missed, but they do not, in themselves, ensure that all possible alternatives are included. Possible casuses may be omitted from a fault tree and possible options may be left out of a decision tree. There are no methodological solutions to this incomplete specification of possible causes or effects, but some heuristic approaches may help the analyst fill out an initial causal search tree.

For example, the analyst might explore the possibility that the possible branches of the cause tree are limited by the nature of the cause-effect problems. For example, the analyst may be able to establish that there are only three groups in the Soviet Union capable of changing the design program for Soviet ships. That is, there may be a way to establish that the total possible causes of an effect in a causal search tree are limited to some finite number. This number (call it X) gives the analyst a "goal state" -- e.g., find X causes for the effect. When the analyst has compiled a list of X causes, the problem of completing the causal search tree is done. This strategy entails first examining the general cause-effect problem for the possibility of a boundary that defines and limits the number of possible causes, and then attempting to generate all specific cause-effect relations, rather than the more typical reverse approach. In the reverse approach, the analyst cannot know if he or she has reached the stopping point (i.e., has exhausted all possible causes).

Many cause-effect problems, however, do not have apparent boundaries around the possible branches, and the tree appears, a priori, unbounded. The analyst's stopping problem becomes, not "Are these all the possible causes," but rather "Are these all the possible causes worth investigating?"

Psychologists studying problem-solving behavior have identified two types of impediments to search tree construction. One impediment, "functional fixedness", entails representing objects by their conventional functions and failing to consider the objects' novel functions. An analogous block might be terms "event fixedness", the tendency to consider only conventional causes for events rather than novel causes. A related impediment results from "set effects", the tendency to attack a problem with one approach or method and not to change that approach.

An example of "set affects" might be a mechanic, attempting to determine the causes of an automobile's failure to start, thinking in terms of electrical problems (dead battery, wet spark plugs, bad distributor), but not in terms of fuel problems (empty gas tank, blocked fuel line, broken fuel pump). An example of "event fixedness" might be the mechanic who fails to realize that cars may not start for nonmechanical reasons (e.g., attempting to turn the ignition key the wrong way, attempting to start while the automatic transmission is set to "drive").

One escape from functional fixedness is to attach specific labels to objects and parts of objects. These labels allow people to mentally connect objects to different purposes that they might not consider when presented with the objects alone (see Posner, 1973; 154-158). This suggests that naval analysts should change perspectives and "disassemble" specific labels to the different perspectives or the component parts of the event. These labeled perspectives or parts would then perhaps suggest more possible causes for the event that would occur to the analyst who only considered the event as a whole entity from a single perspective.

This labeling hypothesis suggests that efforts by naval analysts to create event taxonomies which specify important features of events (e.g., McConnel's "rules of the game" effort, Dismukes and McConnell, 1979, ch. 7) are highly heuristic because the various features may suggest new perspectives and novel causes to the analysts building causal search trees. The more varied the ways analysts are able to label or code an event, the more varied the possible causes the analysts are likely to consider as related to that event. Similarly, schemes which classify the subevents of an event (e.g., pre-crisis, crisis-buildup, height of crisis, post-crisis, return to status quo) should facilitate building causal search trees. The more classification schemes the analyst can consider (e.g., dimensions of time, tension, geography, actors, institutions, technologies are just some of the possible bases for grouping subevents), the richer the set of causes the analyst is likely to consider.

Several of the mechanisms Ascher (1978) suggested for surprise-sensitive forecasting , e.g., suspension of plausibility checks, may facilitate expanding the branches of the causal search tree.

The phenomenon of "tunnel history", which Fischer (1970) lists as a problem with narrative analysis (see Stech, 1981, Chapter 4, "Narrative Logic"), is a set effect. That is, explaining naval events as due to naval causes limits the building of the causal search tree. The interdisciplinary approach taken by many of the sampled naval analysts is likely to reduce the tendency toward set effects, but analysts might enhance this positive effect by directly involving analysts from different disciplines in the causal tree construction and evaluation process.

Tree problems. Psychological research on fault trees (Fischhoff, Slovic, and Lichtenstein, 1978) shows that analysts will not typically notice the omission of important causal branches. For example, Fischhoff and his colleagues showed people (including auto mechanics) various versions of a fault tree for the problem of the nonstarting car, and asked for estimates of the proportion of no-starts caused by the category "all other problems". As various causal branches (e.g., fuel system) are deleted from the tree, the proportion of non-starts attributed to "all other problems" should increase. It did not. People seemed not to miss the absence of important causal branches, and seemed unable to appreciate how many causes had been omitted from the pruned fault trees. In other words, there was a faulty tendency to overestimate the completeness of incomplete causal diagrams. Neither self-rated knowledge nor years of mechancial experience improved the ability to detect this incompleteness.

Fischhoff and his colleagues found somewhat greater sensitivity to incomplete fault trees when people were explicitly urged to consider the tree's completeness and think about possible causes that might be left out. In this case people were more likely to estimate that causes were missing, but even these estimates tended to be low, i.e., there were more causes missing than even this sensitized group estimated.

These results suggest that analysts' efforts to fill out and elaborate a causal search tree are probably well-spent and will tend to pay off in possible causes that would otherwise be overlooked. The issue still open, however, is how the analyst can determine the point to stop searching for causes.

Stopping problems. While efforts to overcome impediments to causal search tree constructin will reduce the chances that naval analysts will overlook important causes, they do not solve the stopping problems -- knowing when the causal tree is sufficiently completed. This is essentially a problem of induction and information integration: given a rich causal tree, does the evidence suggest that some subset of the possible causal branches is an adequate or satisfactory explanation of the event.

Two methods which help analysts with this stopping problem were noted above under "Weighting Data": Bayesian techniques and linear modeling. Linear models are the basis for most causal modeling approaches and may be the most compatible method for assessing whether the cause tree seems sufficiently complete. The statistics of causal analysis techniques allow the analyst to estimate unexplained variance, which corresponds to the role of undetermined causes. If unexplained variance is too high (an analytic judgment must be made as to what "too high" means), the causal tree is incomplete and the causal search should continue. The use of Bayesian techniques allows the analyst to estimate the likelihood of an event given various causes, and to estimate the subjective probability that any causal explanation of the event is true. If the analyst finds his or her subjective probabilities for the various causal explanations are all too low (again "too low" requires an analytic judgment), the causal search is incomplete.

One of the more wide-spread uses of Bayesian techniques in causal analysis takes place in medical diagnosis (see e.g., Gorry, et. al., 1973); Lusted, 1968; McNeil, et. al., 1975; and Schwartz, et. al., 1973). Decision aided medical diagnosis makes use of decision trees which specify the possible actions and tests the physician can take and their possible consequences for a patient with one of several possible maladies. Bayesian techniques are used to evaluate the physician's subjective probability estimates that, given certain test results and symptons, the patient's malady is caused by one disease rather than another. As further information on tests or patient responses to treatment is acquired (i.e., as the physician and patient move to different nodes in the decision tree) Bayes formula can be used to update the physician's estimates of the probabilities of various causes. Bayesian techniques are especially useful in compensating for the tendencies to overemphasize case data (e.g., a positive test result) and to underweight base-rate data (e.g., the incidence of the disease in the population at large).

Bayesian techniques do not handle effectively situations where multiple causes are operating, e.g., a patient with two disorders of overlapping nature. However, it is for just such situations that causal modeling was developed. Similarly, Bayesian techniques may be less effective in cases where causes change over time. Dynamic systems modeling, however, is a means for attacking this problem. This suggests that the selection of a particular analytic aid for a causal analysis problem is likely to be a heuristic or experimental process. The "right" aid may not be the first one selected. An experimental approach to aids, i.e., a willingness to try various types and combinations may be required (Krischer, 1980).

Hypothesis Generation

Several biases exist in the process of causal hypothesis generation, i.e., the tendency to generate some types of causal explanations and not others. We do note the ease with which people (including scientists using sophisticated cause assessing methods) could detect a "presumptive agency" connecting a suspected cause with an effect and find confirming evidence for this hypothesis, even when it was false.

In contrast to the search approach to causal analysis, which emphasized recognition and organization of pattern features in data sets, the hypothesis generation approach focuses on the logical generation of hypotheses, their test, and subsequent revision. Hypothesis generation problems then are likely to occur at either the generation, test, or revision phases. We will discuss the latter two problem areas later, when we consider theories. In the present section we consider hypothesis generation problems.

The biases in causal hypothesis generation tend to fall into two groups: those that are largely due to cognitive processes and those that appear to be based on assumptions about causality in social relationships. The cognitive problems seem to occur because certain features of the information stored in memory about effects and possible causes tend to suggest certain causal explanations (which may be inappropriate). When these data features are largely social (e.g., a liked actor versus a disliked actor), the hypothesis generation problem can be labeled social, although it may have, at a more abstract level, a cognitive basis.

Cognitive problems. Three types of bias which may lead the analyst to generate faulty causal hypotheses: fundamental attribution error, representativeness, and determinism.

The fundamental attribution error is the tendency to attribute behavior to corresponding personal dispositions of the actor and not to environmental causes. In large part this tendency is due to focusing of attention and to the representativeness heuristic. The focusing of attention on an object (or actor) increases the likelihood that the object of attention will be perceived as causing events. The representativeness heuristic refers to the tendency to look for causes whose principal features match those of the effect. Determinism is the tendency to seek nonprobabilistic causes for phenomena and to not consider the possibility of causal forces that have probabilistic effects.

An actor is typically the center of attention, and thus is seen as causal, and the causes of the effects produced by the actor are typically deemed to have originated with the disposition of the actor rather than with the pressures of the environmental background.

This distinction between dispositional (or internal) sources of behavior and environmental (or external) sources is one of considerable importance for naval and military analysts. Did the Soviets become involved int his crisis because of their overall plan to destabilize the region (internal cause) or because their client state is threatened by a Western client (external cause). Did the Soviet naval construction plan change in response to Soviet geopolitical strategy (internal) or in response to naval threats from their adversaries (external). To what extent are Soviet actions due to a mixture of internal and external causes?

Social scientists have developed an extensive and elaborate set of quantitative methods to deal with one form of this causal problem; determining the causes of arms races. These methods could be extended to efforts to model crisis interactions, diplomatic penetrations, aid for conflicting client states, etc. In our sample of naval estimates none of the various dynamic arms models were used.

It is somewhat surprising for two reasons that none of the naval analysts sampled used quantitative arms race models in their analyses. First, the naval analysts themselves (e.g., Bowen, NPSP, ch. 4) occasionally drew parallels between previous naval races and the Soviet-American naval rivalry. For example, Bowen (p. 57) describes the circumstances of the current Soviet-American naval balance as

> ...similar to those that prevailed at the beginning of the century when Germany challenged the supremacy of Britain's navy.

Second, many arms rivalries in the 19th and 20th centuries have been naval. Huntington (1958) listed thirteen arms races in this period, of which seven and a third were naval (the third of a case is the nuclear competition of the United States and USSR). While Huntingtons list may not be exhaustive, it is quite likely that the high proportion of naval races would be found in a more complete list. Certainly, one could conclude that about half of the major arms races, for which social scientists have developed quantitative analytic models, have been naval.

Two main advantages of arms race models are that they (1) make more explicit and mathematically precise the analyst's implicit assumptions and intuitive hypotheses about arms competition, and (2) provide a clear distinction between foreign-induced and self-induced forces in a nation's arms program (Gantzel, 1973; Wallace and Wilson, 1978). Other aspects of the nature of arms competition can be included in these models, e.g., whether the states involved are competing in numbers or technology or both (e.g., Hollist, 1977; Huntington, 1958; Luterbacher, 1976). It is also possible to model an arms competition between two nations (e.g., the United States and the USSR) at the same time that competition between pacts and alliances are examined (e.g., NATO and Warsaw Pact, see, Rattinger, 1975; Wallace and Wilson, 1978). That is, there may be alliance or pact causal factors, as well as intranational and international causes for arms competition. Arms models also help the analyst avoid "mirror imaging". That is, while one state may be reacting largely to its competitor's behavior, the competitor state may be responding primarily to internal forces. "Mirror imaging" is the bias toward perceiving such situations as symmetrical, i.e., both competing states are reacting to the same kinds of factors. Several arms race modelers (e.g., Hollist, 1977; Luterbacher, 1976; Rattinger, 1975; Wallace and Wilson, 1978) using different models and various data sets all essentially concluded that while the USSR's arms and especially its strategic missile programs were reactive to Western (especially United States) arms, the United States' programs were not symmetrical: U.S. programs showed more reaction to internal forces (cost, tecnology, previous arms spending) than did the Soviet programs. It is a mistake to dismiss arms race models (as does Wohlstetter, 1975: 47) simply because the two nations involved do not behave in exactly the same (or a symmetrical) way. In fact, it is <u>because</u> the nations may not be reacting identically, or responding to forces which are the mirror image of each other, that arms race models are particularly helpful.

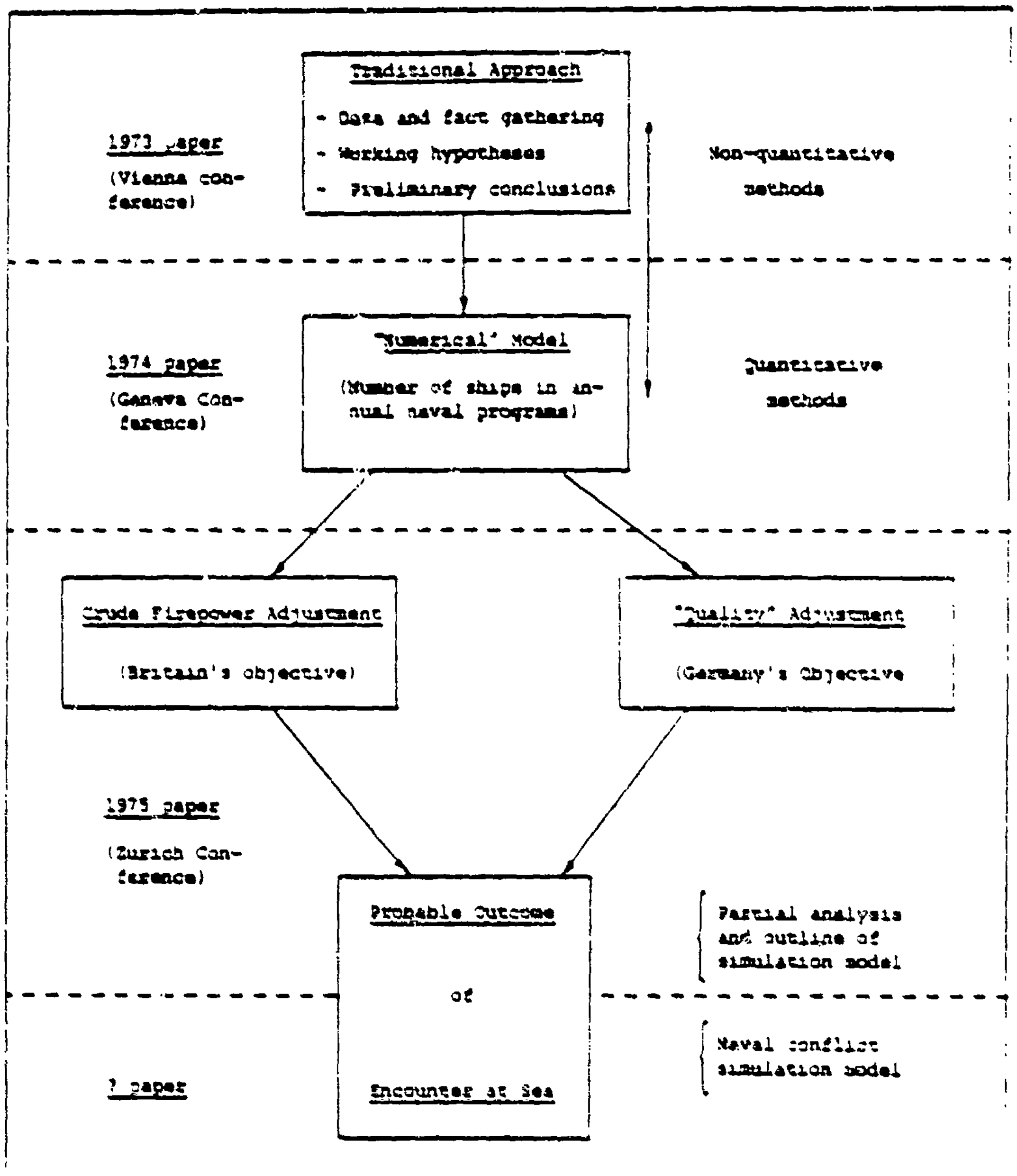
Explicit models of arms competition may yield some useful predictive indicators as well as systematize the analyst's reasoning on causal factors. For example, Wallace (1979) found that the product of the smoothed rate of arms increases for pairs of disputatious nations predicted whether war would follow the arms competition. Rapidly accelerated arms races escalated to war in 23 out of 28 cases, while disputes not preceded by accelerating arms competition resulted in war only 3 times out of 71 cases.

An excellent example of quantitative arms race modeling combined with detailed qualitative descriptive analysis is Lambelet's three-part series on the Anglo-German Dreadnought race (1974, 1975, 1976). Lambelet (1976: 50) presented a diagram (Figure 14) illustrating how his traditional analysis and his numerical methods were integrated in his study. His approach reinforces the point made here that quantitative assessments of causal factors can be a useful tool for determining the causes of a nation's naval objectives.

<u>Social problems</u>. Two types of causal problems can be related to the tendency to generate causal hypotheses on the basis of social intuitions. One of these problems was the tendency to attribute motivational causes when the consequences of actions are more foreseeable. Actors who are perceived as able to foresee the consequences of their acts were also perceived as more motivated (rather than pressured by external forces) to bring those consequences about.

A second problem of social perception of causes is the tendency to attribute the "good" actions (those approved of by the observer) of liked actors to dispositions and of disliked actors to luck, chance, or situation, on the one hand, and to attribute "bad" actions of liked actors to luck, chance, or situation, and of disliked actors to dispositions. Disliked actors, in other words, tend only to be seen as planning bad actions and as being forced (or stumbling accidently) into good actions.

Figure 14. Lambelet's Approach to Qualitative and Quantitative Aspects of the Anglo-German Dreadnought Race.*



*From Lambelet (1976: 50)

The methods outlined above in this section for systematizing causal hypotheses could assist the analyst to avoid these social cause biases. Another technique was outlined in Stech, 1981: Kelly's consensus, consistency, and distinctiveness method. This method can be viewed as a special form of causal model aimed at the statistical features of dispositional and situational causes of actions.

F. PREDICTIONS

> Forecasting would be an absurd enterprise,
> were it not inevitable.
>
> Bertrand de Jouvenel,
> *The Art of Conjecture*

Naval analysts seem to share many of the same problems experienced by other forecasters. The relative lack of specific predictions by naval analysts made it impossible in this study to systematically assess the prediction or forecasting track records of naval analysts. Such assessments are an excellent means for providing the analyst with self-correcting feedback. This suggests that naval analysts could improve their prediction and forecasting efforts by (1) frequently making specific and precise predictions, (2) regularly comparing predictions to outcomes and assessing the frequency, magnitude, and direction of errors, and (3) using this track record feedback to modify their prediction/forecasting methods.

Because analysts now tend to make vague, Delphic forecasts or Aesopian estimates, it is difficult or impossible to gauge the analyst's accuracy. This leaves anaysts with a very subjective impression of their estimation track record and the possibility that analysts, with the advantages of hindsight, will perceive their past track records as more precise and predictive than they actually were. A subjective, retrospective approach to self-appraisal is unlikely to yield the detailed feedback information on inaccuracy that analysts could use to systematically improve their methods.

Identifying Assumptions

A main problem preventing accuracy in forecasts and predictions is the central role of the forecaster's assumptions (Ascher, 1978). A careful and systematic prospective assessment of the prediction track record can assist the analyst to refine assumptions.

Backcasting

A related strategy is to employ frequent "backcasting exercises" to determine whether the assumptions underlying the forecast are true for past and present data (Ascher, 1978; Morgenstern, Knorr, and Heiss, 1973). As Ascher noted (p. 8) forecasters often fail to examine the biases of their predecessors and neglect to adjust their forecasts in a direction that would rectify earlier forecasts. He wrote (p. 110):

> ...the use of previous-error feedback is lacking in the forecasting efforts in every area except that of certain short-term econometric forecasting models.

This nonuse of error-correction feedback seems due to the forecasters' beliefs that their assumptions about the future already incorporate all the data pertinent to known trends. In many cases, however, the forecasters' methods cannot accurately predict known trends, let alone the unknown future.

The failure of intelligence analysts to backcast has been noted several times in connection with the underestimation of Soviet strategic missile inventories (Sullivan, in Godson, 1980; Wohlstetter, 1975a), i.e., estimators continued to underestimate capabilities even after evidence of past underestimates were noted, including a warning from Soviet leader Brezhnev that the West was undercounting! Among the explanations fo this persistent underestimation of capabilities is the hypothesis that U.S. intelligence analysts had misperceived Soviet strategic intentions and requirements, and had "mirror-imaged" the Soviet intentions to correspond to our own (e.g., see Sullivan, in Godson, 1980: 62).

Bootstrapping

A second technique that can help analysts to clarify their assumptions is bootstrapping. That is, a quantitative model of the analyst's predictions is constructed using the data inputs considered by the analyst and fitting these variables to the analyst's predictions. This model of the analyst will reflect the data variables that most influenced the predictions, thus providing the analyst with quantitative information on the ingredients of his or her predictions. By knowing which variables most heavily influence his predictions the analyst can better assess his prediction assumptions.

Two of the quantitative studies done by O'Leary and Coplin (1975) for State Department intelligence analysts used a quasi-bootstrapping approach toward improving predictive capability (although it was not identified as bootstrapping by the authors). In one effort ("Predicting political instability in tropical Africa", ch. 2, O'Leary and Coplin, 1975) they evaluated the views of State Department analysts on the causal variables leading to political instability in tropical African nations. From the analysts' narrative analyses, O'Leary and Coplin abstracted a variety of hypothesized causal factors. These factors were then quantified and regressed on measures of political instability. Had the study been a truly bootstrapping effort, the variables would be regressed on analysts' predictions of political instability, yielding a model of the analysts' predictings. The O'Leary and Coplin approach was to develop a model of the situation, but the same techniques could be used in a bootstrapping effort. Both model building techniques aid the analysts by explicitly relating variables to predictions, thus clarifying assumptions. O'Leary and Coplin (ch. 6) followed a similar approach in their effort to develop a quantitative model to predict violence in the Middle East. In this case the State Department analysts' projections (i.e., forecasts and predictions of Middle East Violence) were the predicted variables, and O'Leary and Coplin attempted to mathematically relate event data set measures to the analysts' projections. The analysts' projections of violence were closely related to the weekly and monthly frequencies of violent acts in the area. In effect, the analysts' projections of violence in the future were strongly influenced by the frequency of violence in the past week and month.

O'Leary and Coplin's analysis included an outline of an assessment technique to test the validity of the analysts' assumptions as revealed by the quantitative model, which seems to be one of the more useful consequences of bootstrapping the analysts' predictions.

Sensitivity Testing

Other problems include the tendencies to overweight case data and underweight base-rate data in predictions, and the tendency to overestimate the predictive validity of indicator variables. A partial antidote to these tendencies is sensitivity testing, that is, asking the analyst to consider how her or his predictions would be different if the base rates were vastly different, or if the predictive validity of the indicators were varied. These sensitivity exercises tend to make predictions more regressive (i.e., closer to the normative statistical predictions, see Fischhoff, Slovic, and Lichtenstein, 1979). Sensitivity tests seem to lead to greater attention to the predictive and diagnostic power of the variables used, and predictions more in line with this awareness.

Validity of Future Assumptions

The main conclusion of Ascher's (1978) and Morganstern, Knorr, and Heiss' (1973) evaluations of forecasting is that forecasting accuracy is dependent on the validity in the future of the central assumptions made by the forecaster. The variables selected by the forecaster may be adequate to explain past events, but the future structure of the problem may change, making these variables less important. The techniques described above, assessing the prediction track record, backcasting, bootstrapping, and sensitivity testing, help the analyst determine his or her assumptions, and assess their validity in the past, but they do little to aid the analyst to assess the validity in those assumptions in the future. To the degree that future trends tend to reflect past trends, these techniques help the analyst to make systematic projections. But if the future is unlike the past and full of major surprises, upheavals, and revolutionary event; the past, and methods which help the forecasting analyst make systematic use of the past, is of much less value.

Simple Models and Surprising Futures

The widespread finding that relatively simple actuarial and econometric models outpredict the forecasts of human estimators in the short and medium term suggests that analysts should, as a minimum, develop a simple, quantitative model of the phenomenon they are attempting to predict. This simple model should be validated on past evidence and used as a baseline against which the analyst can attempt to refine and "fine-tune" estimates. The baseline model would serve to highlight those particular facts and reasons the analyst believes the futures will not be a simple extrapolation from the past. By serving as a validated representation of the "surprise free" future, the simple model, in effect, enables the analyst to concentrate on those forces and developments that may cause surprises. To the degree that the future _is_ like the past, the simple model will do the best job of extrapolating the important variables. But the model lacks the analysts' insights, experience, and intuitions about new trends, incipient developments, sudden variations, changes in context or tone. These subtle cues can only be appreciated by the analyst, and the use of a baseline, "surprise-free" model may allow the analyst the opportunity to pursue these leads to the surprising future.

A Science Analogy

A useful analogy can be drawn between the situation described in the preceding section and Kuhn's (1970) sociological description of "normal" science and "revolutionary" or "paradigm shift" science. Normal science is the use of proved and accepted scientific methods to solve problems and puzzles which bear importantly on a scientific field. Problems are attacked that are widely believed to be solvable and, once solved, become important parts of a major scientific theory. Kuhn has much more to say about normal science, but the point here is that it is analogous to the development and use of simple quantitative models to deal with the "surprise-free" aspects of prediction. Such models provide a means to make accurate extrapolations from past evidence into the future using agreed-upon methods and data to solve important prediction problems.

A scientific revolution occurs when a major scientific theory is changed, i.e., the scientific paradigm shifts. This means that the interpretations of relationships observed in the past are changed, viewed from a completely new perspective, and given a different meaning. Such shifts come about, Kuhn believes, because, in the course of normal science, new, surprising and unexpected phenomena are constantly uncovered. Such anomalies, as Kuhn labeled them, are unexplainable or even uninterpretable within the context of the extant paradigm. For most scientists they are not important problems because they fall outside most of the categories and classification schemes of the paradigm. Either the anomaly seems unsolvable within the methodological or theoretical context of the paradigm, or, no interpretable solution seems possible, or the anomaly is interpreted as part of the paradigm after all.

Kuhn argued that anomalies tend to accumulate, however, and he wrote (1970: 52).

> ...the scientific enterprise has developed a uniquely powerful technique for producing surprises of this sort.

The buildup of surprises and anomalies leads to a blurring of the paradigm and a loosening of the informal rules for normal research.

Eventually, normal science experiences a crisis, and a fundamental shift of perception and belief occurs as scientists accept a new theory to interpret both the old paradigm and the accumulated anomalies. This new theory, developed in large part from the effort to explain the surprises generated from normal science, reorganizes how the old data and the new anomalies are perceived and organized. Scientific theory is transformed. Many features of the old interpretation remain, but the entire situation receives a new interpretation.

The use of simple quantitative prediction models in naval analysis is likely to "solve" many normal prediction problems as well as generate many anomalies and surprises. The naval analyst should be especially concerned with these unusual discoveries and should attempt to reformulate the concepts and categories used to explain the situation so as to incorporate the anomalies. These efforts may lead to a crisis between the new interpretation and the simple models, which may lead to a revolution and paradigm shift, i.e., a new theoretical view of Soviet naval intentions. If this new theory is successful in accounting for the anomalies, it can be the basis for a new set of simple quantitative models (although these will be significantly different from the succeeded models). The new models will, like the old, generate both solutions and new anomalies.

The predictive advantages of using a normal science/simple model and revolutionary scientist/analytic speculation system is that the analyst is focusing on surprises, anomalies, and the unexpected, but is "backed-up" by a reliable theoretical base. The normal theoretical base of the simple models will cope with "surprise-free" situations, freeing the analysts to develop new frameworks to incorporate these elements of the present (anomolies) that betoken the surprises of the future. Many anomalies will have no scientific relevance, they will be data collection errors, deceptions, accidents, noise in the analytic channels, etc. Distinguishing between the meaningless anomalies and those that signal the trends of the future is a major task. The use of simple modeling of normal theories and a "surprise-free" future might free the analyst to create the new theories and consider the uncertainties necessary to anticipate a surprising future.

G. THEORIES

> Life is the art of drawing sufficient conclusions from insufficient premises.
>
> Samuel Butler, Notebooks

> There is nothing like a theory for blinding the wise.
>
> George Meridith,
> The Ordeal of Richard Feverel

There is substantial evidence that people (including scientists) are extremely reluctant to give up a useful theory. This reluctance extends to situations in which considerable evidence discrediting the theory is available, and theories may survive even a complete disconfirmation of their evidenciary bases. This seems to be due to several factors. Theories serve functions other than accurate prediction. People are reluctant to dismiss a useful theory on the basis of evidence that might be erroneous. Negative evidence is difficult to integrate into a theory. Multiple hypothesis testing is extremely difficult to conduct. People tend to seek and accept confirming evidence more readily than refuting evidence.

Despite such difficulties a few of the sampled naval analysts did employ multiple hypothesis testing or disconfirmation strategies, and several analysts seemed alert to the significant of negative evidence.

If there was one main characteristic problem of theorizing among the sampled naval analysts it was perhaps the reluctance to make specific predictions on the basis of various theories and to systematically and repeatedly test theories against one another. In particular, there was too little explicit disconfirmation and too much confirmation. There was very little effort devoted to developing methods of testing theories which would reflect quantitatively the degree to which evidence supports or does not support any hypothesis. Although we found naval analysts updated their theories, the lack of explicit methods for evaluating the degree of support that evidence conveys on a given hypotheses made it difficult to ascertain exactly why the analyst changed the theory.

Although there are aids available to analysts which quantify hypothesis testing (e.g., Bayes Theorem is specifically formulated to update a prior hypothesis given new information), there are no quantitative aids that prescribe how or when scientists should reformulate theories. As Kuhn (1970) describes the structure of scientific revolutions, the process is largely a social one of scientists arguing, replicating, confirming, revising, and generally shaping and influencing each other's views and research. Scientific discoveries which fail to generate interest in other scientists die on the vine unless rediscovered in the wake of a scientific crisis and revolution. Similarly, an intelligence or naval analyst may reformulate a theory of intentions that produces better estimates, but unless the new theory is communicated to other analysts and to estimate users, it is likely to have very little impact, except on its originator. Furthermore, it is likely that the insights of one analyst are significantly sharpened and refined when they are shared with and examined by others. A theory developed in a community of analysts is likely to be better than the same theory developed only by its originator.

This suggests that there are important social and organizational dimensions to theory formulation, revision, and change; and that aiding these dimensions of estimation goes beyond the individual analyst (see Stech, 1979, for an appraisal of social, organizational, and political aspects of intention estimation). Just as there are weaknesses in estimation logic, there are weaknesses in the organizational processes of estimation. These social and organizational dimensions become important when the individual analyst is developing a new theoretical outlook that conflicts with or goes beyond the normal accepted theory, i.e., when the analyst's new theory conflicts with other analysts' theories.

Most of the advice given in this chapter involves greater precision of hypotheses, greater specification of variables and relationships, more use of mathematical and psychological techniques to ascertain and refine variables and relationships, and greater willingness to predict and check predictions. All of this advice amounts to saying that naval analysts might do intention estimation more like scientists do science. Quantification, measurement, specificity and precision, and prediction are means and ends of the methods and tools of science, he or she can find a large literature of helpful and useful advice on theory building and testing, scientific method, epistemology and the philosophy of science and scientific theory. There is no point attempting to review the pathways into that literature here, any reader who avails himself of the references in this study will quickly find many leads if such are needed.

A more important issue is whether and to what degree scientific methods are appropriate for the social, political, and naval science questions that concern the naval analyst. Science requires, in addition to the attributes noted above, control and experimental manipulation of variables. Control and experimentation are rarely possible in the social sciences to the degree that they can be exercised in physical sciences. Does this mean that the social sciences cannot be truly scientific and that attempts at scientific methods in these areas are fated to be little more than over-sophistication?

No clear-cut, general prescriptions can be offered to the analyst on this issue. It will be up to the analyst whether to adopt a scientific or quantitative technique, to attempt explicit theorizing and theory building, or to continue to rely on narrative explanation. Scientists and philosophers continue to wrestle with the question of whether social science can be truly scientific (e.g., see McClintock, 1981; Ziman, 1979). Just as unaided analysis and theory-building have their limitations, as outlined in Stech (1981, Chapter 4), scientific methods and theories have their limits, especially as applied to social phenomena. Naval analysts should be aware of the limits of both.

Whether a particular quantitative aid or theory-building approach will be useful or helpful is largely an empirical question. We have tried to outline how the naval analyst can detect and assess the shortcomings of his or her unaided estimation or narrative logic. The analyst will have to apply this general knowledge to the particular case to determine whether the shortcomings thus detected are serious enough to warrant the use of explicit quantitative aids. Similarly, the analyst will have to evaluate the advantages against the costs of developing a quantitative theory-building approach, the use of multiple hypothesis disconfirmation strategies or other theory-revision methods noted in the previous sections of this chapter.

In effect, the analyst faces an infinite regress: how to decide whether or not to try out a particular method or theory-building approach? If decision-aiding methods or judgment-enhancing approaches are recommended to help with this first question, the analyst is faced with the second question: how to decide whether an aided decision is better than an unaided one?

While it is a basic finding of this study that unaided estimation tends to be inferior to estimates that explicitly employ aids to information organization, integration, and inference, in general we cannot draw hard conclusions about <u>how</u> inferior unaided estimation might be. This makes the question of the costs and benefits of aided versus unaided estimation and theory-building an empirical one and one for which the analyst might want to seek out decision-aiding or judgment-aiding expertise, particularly if the costs and risks involved are very high.

The approach taken by O'Leary and Coplin (1975) seems an estimable one worth repeating (although perhaps with a more systematic appraisal methodology). They compared the qualitative analyses and forecasts of unaided State Department intelligence analysts to estimations based on quantitative approaches to the same intelligence issues. They also evaluated the costs and benefits of the quantitative methods. O'Leary and Coplin made an informal cost-benefit analysis, and it might be necessary for naval analysts to make such evaluations more explicitly and systematically (e.g., using cost-benefit techniques, decision-analysis, multiattribute utility methods, etc.) or to attempt a prospective rather than retrospective evaluation. The main point, however, is that the quantitative methods were tried to see if they work, with what limits, to determine how much improvement they could make, and at what expense. This is an approach we would recommend to any analyst who might be tempted to try an improved methodology for intention estimation.

IV. CATALOGUE OF AIDS AND PROCESS SUMMARY

The analytical aids discussed and evaluated in the previous sections of the report are catalogued for convenient reference in this section. The items covered in the one-page summaries correspond closely to the criteria outlined in Chapter III. Problems, pitfalls, and warnings related to specific aids are summarized in the "Comments" section of each summary.

No study of this nature and scope could produce an exhaustive survey of all possible analytical aids. Further, the state-of-the-art of information handling and interpretation is changing quite rapid.;. To accomodate future additions, updates, and enhancements to the catalogue, the one-page summary format has been employed.

The catalogue of aids is organized into four basic sections including:

A. Judgemental methods of both an explicit and implicit nature.

B. Specialized analytical aids to support specific elements of the estimation process.

C. Extrapolation methods and models that rely primarily on past events and experiences as an information base.

D. Structural representations and models which emphasize causal relationships and physical constraints.

A. JUDGEMENTAL METHODS

Methods catalogued in this section assist in the harnessing of individual and group judgement to deal with estimation problems. Emphasis is placed in these methods on the organization of individuals and groups to take advantage of overlapping capabilities, to fill in gaps in capabilities and experience, and to systematize the judgemental process.

Behavioral science studies have revealed several implicit patterns that arise in judgemental processes, while these often appear more in the nature of mental traps than aids, their identification in this catalogue can help to maintain an awareness of potential problems that may arise and to use these implicit patterns in a more constructive way.

ANALYTICAL AID: EXPERT OPINION

Category: Perceiving data, prediction.

Scope:

Method can cover any aspect of problem where expertise has been developed.

Logic Structure and Methodology:

Based on human perceptions, interpretations, and reliability. Most problems will involve group of experts with some overlapping capabilities and, probably, some gaps in coverage.

Comments:

Method suffers from difficulties in assembling and integrating opinions from group members dealing with complex problems. Severe difficulties in determing logical process and biases that affect opinion of individual or group.

ANALYTICAL AID: DELPHI TECHNIQUE

Category: Characterizing data, prediction.

Scope:

Represents structural approach to use of individual and group expert opinion with provision for feedback and learning.

Logic Structure and Methodology:

Responses to well defined questions or problems are integrated and returned to contributors with full information on the comparison of their estimates with others in group. Responses can be changed in proceeding rounds.

Applications:

Classifying Soviet Navy into component missions (Thorpe, NPSP, ch. 8).

Documentation, Validation, Review:

Morgenstern, Knorr, Heiss, 1973: 23-26

Comments:

Range of expert opinion generally converges sharply, thereby possibly masking real uncertainty. Method is costly and time consuming; reasons for shifts in responses are not usually apparent.

ANALYTICAL AID: SCENARIO METHODS

Category: Characterizing data, prediction.

Scope:

Involves complete description of outcomes that may be anticipated with explicit representation of all causal and contributing variables in a consistent framework.

Logic Structure and Methodology:

Scenario description usually includes quantitative and qualitative information. Explicit nature of description facilitates third-party review and comment.

Applications:

None Cited.

Comments:

Useful method to test consistency and plausibility of analysis. Complementary to expert opinion, Delphic, and most modeling efforts.

ANALYTICAL AID: AVAILABILITY AND REPRESENTATIVENESS HEURISTIC

Category: Preceiving data

Scope:

This heuristic involves intuitive attributes which exist in most perception and interpretation problems. While not really analytical methods, they are implicit in analysis and must be included here.

Logic Structure and Methodology:

The representative heuristic is the tendency to assume that a sample is representative of the population from which it is drawn, and to neglect features of the population that are not in the sample. The availability heuristic deals with the tendency to concentrate on information that is accessable and to limit the scope of search for complementary or supplementary information.

Applications:

Most judgemental estimates.

Comments:

This implicit behavior is a factor in most judgemental estimates and must be used constructively.

B. SPECIALIZED ANALYTICAL AIDS

This group of analytical aids deals with specific portions of the estimation process and emphasize the perception, characterization, and organization of information rather than prediction.

Some of the methods described in this section can be used to support and supplement the judgemental methods of the previous section, while others are intended to be used separately from judgemental methods.

ANALYTICAL AID: CODING OF EVENTS (EVENT CODING)

Category: Perceiving Data, Characterizing Data

Scope:

Method can be applied to any stream of information involving series of discrete action events.

Logic Structure and Methodology:

Coded events are amenable to psychological analysis and interpretation. Different coding categories may be used for expected and unexpected events. Method can capture gross relationships between events and finer-grained details of actions. Interruptions, ommissions, and non-occurrences may be detected agains an orderly background of actions. Requires varying level of input data and is totally dependent on that data.

Applications:

Conflict and scale of violence (Egypt-Israel) (O'Leary and Coplin, 1975). Crisis forecasting and prediction (March 1977 issue of International Studies Quarterly). Soviet naval operations (Dismukes and McConnell, 1979: ch. 2 by Petersen).

Comments:

Method is important in the organization of information based on observations in most estimation applications. Analyst coding events may be affected by a number of perception phenomena.

ANALYTICAL AID: CONTENT ANALYSIS (FACTOR ANALYSIS)

Category: Preceiving Data, Characterizing Data

Scope:

Generally applied to verbal or written statements, e.g., propaganda, speeches, memoirs.

Logic Structure and Methodology:

Method uses statistical techniques to determine objective frequencies for actions, events, or statements to determine such things as authenticity, trends in semantics or rhetoric, and/or shifts in interests. Method is systematic and is based on clear rules of coding and inference.

Applications:

Objective estimation of frequencies for actions, events, and/statements (George, 1959). Soviet Position in Law of the Sea negotiations (Friedheim and Jehn, 1977). Attitudes of Soviet elites (Heuer, 1978). Nazi intentions (George, 1959). Soviet political elite speeches (Kirk, 1980). Warship design (Kehoe, SNI, ch. 19).

Comments:

Method is very labor intensive and time consuming. Content can sometimes be quite subjective requiring careful training and monitoring of analyst.

ANALYTICAL AID: PARAMORPHIC REPRESENTATIONS

Category: Perceiving Data

Logic Structure and Methodology:

Method is idealized representation of group of experts assuming that a model capturing the essential elements of their background is more consistent than human integration. Intended to overcome human distortions and inconsistencies attributable to boredom, fatigue, distractions, and variable applications of integration rules. Mathematics usually involve simple linear combinations of predictive variables.

Applications:

None noted.

Documentation:

Hoffman, 1960. Slovic, 1969.

Comments:

Idealized model does not capture insights and innovative findings that may occur in human expert process. Not felt to be appropriate for complex, nonlinear, configural tasks.

ANALYTICAL AID: BOOTSTRAPPING

Category: Perceiving data. Assessing Covariation, prediction.

Scope:

Extension of paramorphic representation to use judgemental data as input to model also incorporating real data and observations.

Logic Structure and Methodology:

Linear models are fitted to past judgements. Idea of modeling the judge's judgements and then using the judgement of the model leads to term "bootstrapping".

Applications:

Instability in tropical Africa (O'Leary and Coplin, 1975).

Documentation:

Dawes, 1971. Goldberg, 1970.

Comments:

Provides systematic data-driven means to appraise suspected covariations that resolve some problems with theory-driven appraisal. Reliability of model must be considered in any application.

ANALYTICAL AID: METRIC AND NON-METRIC MULTIDIMENSIONAL SCALING, TREE FITTING, AND CLUSTERING

Category: Perceiving Data. Characterizing Data.

Scope:

Developed to deal with systemization of expert opinions on the features of events or other samples.

Logic Structure and Methodology:

Method is aid in constructing meaningful perceptual categories for analyst. Uses computer technique for representation of psychological or physical structure.

Applications:

Assess similarity between pairs of stimuli and sorting of classification stimuli into categories of Soviet general purpose platforms (Thorpe). Soviet naval behavior (Dismukes and McConnel). Identification of voting blocks in U.N. (Heuer, 1978, ch. 8).

Comments:

Appears to be more promising than Delphi because method attempts to make structure of problem and assumptions more explicit. See Shepard (1980).

C. EXTRAPOLATION METHODS AND MODELS

The methods described in this section use statistical and regression techniques to extrapolate future patterns and behavior based on past observations and experience. These methods are generally used on aggregated data.

The principal objective of these methods is to assist in the identification of cause-effect relationships, covariance analysis, and prediction. The methods are most applicable where it is difficult or impossible to arrive at an explicit representation of the underlying structure affecting a problem.

Both deterministic and probabilistic techniques may be employed with these methods.

ANALYTIC AID: LINEAR REGRESSION

Category: Weighting data. Cause-effect relationship.

Scope:

Method can test and/or establish relationships between sets of quantitative variables.

Logic Structure and Methodology:

Uses well developed statistical techniques to establish relationships and indicate error range of estimation. Related to, but more comprehensive than time-trend method.

Applications:

Variety of quantitative estimation problems.

Documentation, Validation, Review:

Slovic and Lichtenstein, 1971.

Comments:

Determination of causality is difficult due to problem of multicolinearity. Predictions can be reliable over the short term provided underlying structure and factors do not change abruptly.

ANALYTICAL AID: TIME TREND LINE

Category: Assessing Covariations

Scope:

Desc- - trend over time for any defined parameter or measurable event.

Logic Structure and Methodology:

Graphical representation of parameter or measurable event over time.

Applications:

Capability of U.S. and Soviet ship types (Kehoe, SNI, ch. 8). Latin America (O'Leary and Coplin, 1975, ch. 8) militatry expenditures and other variables.

Comments:

Useful display format. May lead to regression model based on actual relationship. Useful for identifying shifts in pattern of events.

ANALYTICAL AID: CAUSAL ANALYSIS

Category: Cause-effect Assessment

Scope:

Encompasses systematic assessment of causal relationships in nonexperimental situations.

Logic Structure and Methodology:

Method selects variables that are potential determinants of effects and isolates separate contributions to effects made by each suspected cause. Typically use regression analysis.

Applications:

Coalitions among oil-exporting and oil-importing countries (O'Leary and Coplin, 1975, ch. 7).

Documentation, Validation, Review:

Asher, 1976. Blalock, 1964. Heise, 1975.

Comments:

Similar to regression analysis and very data dependent.

ANALYTIC AID: ACTUARIAL MODELS

Category: Assessing Covariations

Scope:

Can be applied to any estimation problem involving relationship of outcome variables to measurable or observable variables.

Logic Structure and Methodology:

Uses regression analysis to relate response or outcome variables to measurable variables to determine combinational rule.

Applications:

None noted.

Comments:

Provides improvement over judgemental approaches to synthesizing input data to estimate outcome.

ANALYTICAL AID: BACKCASTING

Category: Prediction

Scope:

Validation method for model or mathematical aid based on regression, econometric, or structural model.

Logic Structure and Methodology:

Method determines whether assumptions underlying the forecast are true for past and present data. Forecasters often fail to adjust their forecasts in a direction that would rectify earlier forecasts.

Applications:

Several failures to use this technique, leading to underestimation of Soviet missile inventories, are noted (Sullivan, in Godson, 1980; Wohlstetter, 1975a).

Documentation:

Asher, 1978. Morgenstern, Knorr, and Heiss, 1973.

Comments:

Important to use this as method of validating and correcting any model.

D. STRUCTURAL REPRESENTATIONS AND MODELS

This class of analytical methods probes more deeply into causal and structural relationships of a physical and conceptual nature. The motivation for this is to develop a more representative framework within which the effect of a number of variables and uncertain assumptions may be tested.

These methods are most appropriate where significant patterns or paradigm changes mights be anticipated. Since physical relationships and constraints (force structure, weapons capabilities, logistic capabilities) are generally easy to capture in this type of framework, the methods tend to be applied in cases where physical capabilities have a strong influence, or are a constraining factor, on intentions and actions.

Analytical methods in this category may be deterministic or probabilistic.

ANALYTICAL AID: STRUCTURAL OR EXPLICIT MODELING (THEORIES)

Category: Cause-effect Assessment. Prediction

Scope:

Method attempts to represent real structural relationships involving behavioral and physical elements pertinent to estimation problem.

Logic Structure and Methodology:

Model representation may be theoretical (relationships are postulated and then studied by applying model) or based on understanding of actual structure. This class of modeling may be normative in form (indicates what intentions should be given constraints and physical capabilities) or simulative (what might be expected given certain assumptions). The normative form uses optimization techniques, while the simulation methods can involve a number of analytical methods including econometrics and system dynamics. Methods make extensive use of mathematical techniques. Feedback relationships may be modeled explicitly using system dynamics methods or other methods derived from control theory.

Applications:

Arms race models (Wallace, 1979). Instability in Tropical Africa (O'Leary and Coplin, 1975).

Comments:

Structural modeling is most applicable in dealing with combined physical and behavioral factors. It has strong promise in being able to deal directly with situations where physical capabilities (force structure, logistics, weapons capability) either may reveal intentions or place significant constraints on intentions.

ANALYTICAL AID: PARADIGM SHIFTS

Category: Prediction

Scope:

Any relationship, no matter how well it appears to be established may be subject to a paradigm shift and this fact must be considered in any changing environment.

Logic Structure and Methodology:

Paradigm shifts can affect the validity of any model or analytical method that does not explicitly take the possibility into account.

Applications:

None noted.

Documentation, Validation, Review:

Kuhn, 1970.

Comments:

No reliable methods have been developed to assist in the identification of such shifts; they may be dealt with in a fashion by expert opinion and time-trend methods.

ANALYTICAL AID: SEARCH TREES (FAULT TREE AND DECISION TREE)

Category: Cause-effect Assessment

Scope:

Describes any process with sequentially related elements.

Logic Structure and Methodology:

Sequentially related elements are represented explicitly in tree structure concerning dual outcomes or uncertain outcomes and consequences. Well developed mathematical techniques are available for probabilistic analysis, search and optimization routines.

Applications:

None noted.

Documentation:

Posner, 1973: 154-158

Comments:

Powerful technique for organizing relevant information on complex processes. May be used with Bayesian analysis and other probabalistic methods.

ANALYTICAL AID: BAYESIAN ANALYSIS

Category: Weighting Data

Scope:

Deals with any probabalistic problem where conditional estimates may be developed.

Logic Structure and Methodology:

Considers probabalistic relationships between event and causal or contributing parameters. Problems structured in this way can be dealt with by extensive and powerful mathematical tools.

Applications:

Estimation of probabilitions of Mid-East conflict (Heuer, 1978, ch. 2 by Schweitzer). Analysis of order of battle data (North Vietman, USSR-China, Arab Israel conflict). Integration of judgemental information (Morris, 1974, 1976).

Documentation, Validation, Review:

Slovic and Lichtenstein, 1971. Edwards, 1978.

Comments:

Method helps to integrate probabalistic data into judgements. Also useful in combining estimates of different experts and can be used with decision tree methods. Can estimate subjective probability that a causal explanation for an event is true.

V. BIBLIOGRAPHY

ALEXANDER, A. J. and J. R. NELSON (1972) "Measuring Technical Change: Aircraft Turbine Engines." RAND REPORT R-1017-ARPA/PR. Santa Monica CA: RAND Corporation.

ALEXANDER, C. M., R. LEPPER, and L. ROSS (1979) "Theory Perseverance." Unpublished. Stanford University.

ALLISON, G. T. (1971) Essence of Decision. Boston, Little, Brown.

ANDERSON, N. H. (1979) "Algebraic Rules in Psychological Measurement." American Scientist, 67, September-October, 555-563.

ANSCOMBE G. E. M. (1969) Intention. Ithaca: Cornell University Press.

ASCHER, W. (1978) Forecasting: An Appraisal For Policy-Makers and Planners. Baltimore, MD: Johns Hopkins University Press.

ASCHER, H. B. (1976) Casual Modeling. Beverly Hills, CA: Sage

BLALOCK, H. M. (1974) Casual Inferences in Nonexperimental Research. New York: Norton.

BRACKEN, P. (1977) "Unintended Consequences of Strategic Gaming." Simulation & Games, 8, 3, 283-318.

BROWNELL, J. R.; M. J., STOIL, and C. E. THOMANN, (1980) "Research Design for an Automated Behavioral Intelligence Product" Arlington, VA: Kappa Systems, Inc.

CAMERER, C. (1980) "General Conditions for the Success of Bootstrapping Models." Unpublished. University of Chicago Center for Decision Research.

CHAN, S. (1981) "Mirror, Mirror, on the Wall...Expert Judgement Under Uncertainty". Paper delivered atr the annual meeting of the International studies Association, Philadelphia, PA, March 18-21, 1981.

DAWES, R. M. (1980) "Confidence in Intellectual Judgments vs. Confidence in Perceptual Judgments," in E. D. Lautermann and H. Feger (. .) Similarity and Choice. Bern: Hans Huber.

-----(1974) "The Role of the Expert in Constructing Predictive Systems." Proceedings or the IEEE Man and Cybernetics Conference, 522-525

-----(1971) "A Case Study of Graduate Admissions: Application of Three Princples of Human Decision Making." American Psychologist, 26, 180-188.

DAWES, R. M. and B. CORRIGAN (1974) "Linear Models in Decision Making". Psychological Bulletin, 81, 2, 95-106.

DISMUKES, B. and J. McCONNELL (1979) Soviet Naval Diplomacy. New York, Pergamon.

EDWARDS, W. (1978) "Technology for Director Dubious: Evaluation and Decision in Public Contexts," in K. Hammond (ed.) Judgment and Decision in Public Policy Formation. Boulder, Col.: Westview.

FISHER, D. (1970) Historian's Fallacies. New York: Harper and Row.

FISCHHOFF, B., P. SLOVIC and S. LICHTENSTEIN (1979) "Subjective Sensitivity Analysis." Organizational Behavior and Human Performance, 23, 339-359.

-----(1978) "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation." Journal of Experimental Psychology: Human Perception and Performance, 4, 342-355.

-----(1977) "Knowing_with_Certainty: The Appropriateness of Extreme Confidence." Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

FOSTER, M. L. and S. H. Brandes (1980) Symbol as Sense: New Approaches to the Analysis of Meaning, New York: Academic Press.

FRIEDHEIM, R. L. and M. E. Jehn (1975) "The Soviet Position at the Third U.N. Law of the Sea Conference" in M. MccGwire, K. Booth, J. McDonnell (eds.) Soviet Naval Policy: Objectives and Constraints. New York, Praeger.

GANTZEL, K. J. (1973) "Armament Dynamics in the East-West Conflict: An Arms Race?" Papers, Peace Science Society, 20, 1-24.

GAZIT, S. (1980) "Estimates and Fortune Telling in Intelligence Work." International Security, 4(4), 36-56.

GEORGE, A. (1959) Propaganda Analysis: A Study of Inferences Made From Nazi-Propaganda in World War II. Evanston, Ill.: Row, Peterson.

GERWIN, D. and P. NEWSTED (1977) "A Comparison of Some Inductive Inference Models." Behavioral Science, 22, 1-11.

GODSON, R. (ed.) (1980) Intelligence Requirements for the 1980s: Analysis and Estimates. New York: National Strategy Information Center.

GOLDBERG, L. R. (1970) "Man Versus Model OF-Man: A Rationale, Plus Some Evidence for a Method of Improving On Clinical Inferences." Psychological Bulletin, 73, 6, 411-432.

GRIFFIN, D. R. The Question of Animal Awareness, New York: Rockefeller University Press.

HEISE, D. (1975) Casual Analysis. New York: John Wiley & Sons.

HEUER, R. J. (ed.) (1978) Quantitative Approaches to Political Intelligence: The CIA Experience. Boulder, CO: Westview.

HOFFMAN, P. J. (1960) "The Paramorphic Representation of Clinical Judgement," Psychological Review, 57, 116-131.

HOGARTH, R. M. and S. MAKRIDAKIS (1979) "Forecasting and Planning: An Evaluation." Unpublished. University of Chicago, Center for Decision Research.

HOLLIST, W. L. (1977) "An Analysis of Arms Processes in the United States and the Soviet Union." International Studies Quarterly, 21, 503-528.

HUNTINGTON, S. P. (1958) "Arms Races: Prerequisites and Results." Public Policy, 8, 41-86.

KIRK, E. J. (1980) Soviet Foreign Policy Perspectives in the 1970s: Group-and Individual Perceptions. Bethesda, MD: MATHTECH, Inc.

KRISCHER, J. P. (1980) "An Annotated Bibliography of Decision Analytic Applications to Health Care." Operations-Research, 28, 97-113.

KUHN, T. (1970) The Structure of Scientific Revolutions. 2nd Ed., Chicago: University of Chicago Press.

KUZMINSKI, A. (1973) "The Paradox of Historical Knowledge." History and Theory, 12 (3), 269-289.

LAMBELET, J. C. (1976) "A Complementary Analysis of the Anglo-German Dreadnought Race, 1905-1916." Papers Peace Science Society, 26 49-66.

LANGER, E. (1975) "The Illusion of Control." Journal of Personality and Social Psychology, 32, 322-328.

LANGER, W. C. (1972) The Mind of Adolph Hitler: The Secret Wartime Report. New York: Signet.

LAWRENCE, R. (1972) Motive and Intention. Evanston: Northwestern University Press.

LUSTED, L. (1968) Introduction to Medical Decision Making. Springfield, Ill.: Charles C. Thomas.

LUTERBACHER, U. (1976) "Towards a Convergence of Behavioral and Strategic Conceptions of the Arms Race: The Case of American and Soviet ICBM Build-up." Papers, Peace Science Society, 26, 1-21.

MccGWIRE, M. (1979a) "Naval Power and Soviet Global Strategy." Internal Security, 3 (4), 134-189.

-----(1979b) "Commentary: Soviet Intentions." Internal Security 3 (5), 139-143.

---- (ed) (1973) Soviet Naval Developments: Capability and Context. New York: Praeger.

McLACHLEN, D. (1968) Room 39: A Study of Naval Intelligence. New York: Atheneum.

MORGENSTERN, O., K. KNORR, and K. P. HEISS (1973), Long Term Projections of Power: Political, Economic, and Military Forecasting. Cambridge, MA: Ballinger.

MORRIS, P. A. (1974) "Decision Analysis Expert Use." Management Science, 20, 1233-1241.

MOSTELLER, F. (1981) "Innovation and Evaluation." Science, 211, 881-886.

O'LEARY, M. K. and W. D. COPLIN (1975) Quantitative Techniques in Foreign Policy Analysis and Forecasting. New York: Praeger.

POSNER, M. I. (1973) Cognition: An Introduction. Glenview, Ill: Scott, Foresman.

RATTINGER, H. (1975) "Armaments, Detente, and Bureaucracy." Journal of Conflict Resolution, 19, 571-595.

ROSCH, E. and C. B. MERVIS (1975) "Family Resemblances: Studies in the Internal Structure of Categories." Cognitive Psychology, 7, 573-603.

ROSENBERG, S. and A. SEDLAK (1972) "Structural Representations of Implicit Personality Theories." In L. Berkowitz (ed) Advances in Experimental Social Psychology, vol. 6. New York: Academic.

SATTATH, S. and A. TVERSKY (1977) "Additive Similarity Trees." Psychometrika, 42, 319-345.

SHEPARD. R. N. (1980) "Multidimensional Scaling, Tree-Fitting, and Clustering." Science, 210, 390-398.

SLOVIC, P. (1969) "Analyzing the Expert: A Descriptive Study of a Stockbroker's Decision Processes." Journal of Applied Psychology. 53, 255-263.

SLOVIC, P., B. FISCHHOFF and S. LICHTENSTEIN (1976) "Cognitive Process and Societal Risk Taking," in J. S. Carroll and J. W. Payne (ed.) Cognition and Social Behavior. Potomac, MD: Lawrence Erlbaum.

SLOVIC, P. and S. LICHTENSTEIN (1971) "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment." Organizational Behavior and Human Performance, 6, 649-744.

STECH, F. J. (1980a) "Self-Deception: The Other Side of the Coin." Washington Quarterly, 3, 130-140.

-----(1980b) "Intelligence, Operations and Intentions." Military Intelligence, July-September, 37-43.

-----(1979) Political and Military Intention Estimation: A Taxonometric Analysis. ADA 081486, Bethesda, MD: MATHTECH, Inc.

-----(1981) Estimates of Peacetime Soviet Naval Intentions: An Assessment of Methods. Bethesda, MD: MATHTECH, Inc.

TVERSKY, A. and D. KAHNEMAN (1981) "The Framing of Decisions and the Psychology of Choice." Science, 211, 453-458.

-----(1977) "Casual Schemata in Judgments Under Uncertainty," in M. Fishbein (ed.) Progress in Social Psychology. Hillsdale, NJ: Erlbaum.

WALLACE, M. D. (1979a) "Arms Races and Escalation." Journal of Conflict Resolution, 23, 3-16.

WALLACE, M. D. and J. M. WILSON (1978) "Non-Linear Arms Race Models." Journal of Peace Research, 15, 175-192.

WOHLSTETTER, A. (1975a) "Legends of the Strategic Arms Race." USSI Report 75-1. Washington, DC: United States Strategic Institute.

-----(1975b) "Optimal Ways to Confuse Ourselves." Foreign Policy, 20, 170-198.

-----(1974a) "Is There a Strategic Arms Race?" Foreging Policy, 15, 3-20.

-----(1974b) "Rivals, But No Race." Foreign Policy, 16, 48-81.

-----(1974c) "Legends of the Strategic Arms Race: Part I." Strategic Review, Fall, 5-30.