(12)

A098153

(6) ESTIMATES OF PEACETIME SOVIET NAVAL INTENTIONS: An Assessment of Methods

(10) Frank J. Stech

(9) Technical Report,

(12) 222

(11) March 1981

Prepared for:

Office of Naval Research
Department of the Navy
Contract No. (15) N00014-78-C-0727

MATHTECH, Inc.
4630 Montgomery Avenue
Bethesda, Maryland 20014
Phone: (301) 951-0808

393236

JOB

TABLE OF CONTENTS

PREFACE....1

Chapter 1. INTRODUCTION AND OVERVIEW....1

WHAT THIS STUDY DOES....1
Estimation Methods
Focus on Intentions....2
Focus on Biases....3
Peacetime Intentions....4
Origins....4
WHAT THIS STUDY COVERS....6
Sources....6
Selection method....6
Types....8
WHAT THIS STUDY DOES NOT COVER....8
Accuracy....8
Social Effects....8
Organizational Effects....9
Motivational Effects....9
Political Effects....9
HOW TO USE THIS STUDY....10
Naval Analysts....10
Soviet Analysts....10
Consumers of Naval Estimates....10
Cognitive Psychologists....11
Students and Managers of Intelligence....11

Chapter 2. CHARACTERISTICS OF ESTIMATES....12

METHODS OF PREDICTION....12
Sagacity....12
Control....13
Acumen....13
Prediction Requirements....14
Sagacity....14
Control....14
Acumen....14
SPECIFICITY OF INTENTIONS....15
Specificity and Precision....16
Forecasts and Predictions....16
Time Horizon....18
Situation Prediction....19
Target and Action Specification....20
Scoring Predictions and Forecasts....20
COMPLEX AND NONCOMPLEX ANALYSES....22
THE LOGIC OF ANALYTICAL PROCESSES....22

Chapter 3. CONFLICTS IN ESTIMATES....25

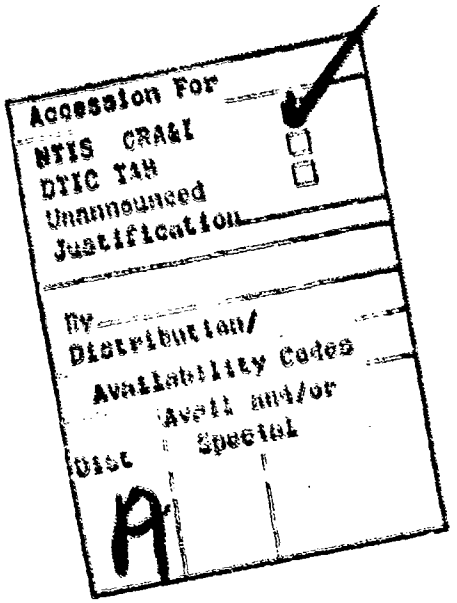
DIFFERENCES BETWEEN ESTIMATES--
PAST SOVIET NAVAL EVENTS....26

Case 1: Soviet Attitudes Towards Surface Ships, 1953-1961.....26
Case 2: Attitudes Toward Forward Naval Deployments in the 1950s.....29
Case 3: Soviet Naval Objectives in Egypt.....30
Case 4: Interpretations of the Gorshkov Series.....35
PRESENT SOVIET NAVAL EVENTS.....37
Case 5: Readiness for and Purposes of Forward Deployment.....37
FUTURE SOVIET NAVAL EVENTS.....41
Case 6: Trends in Anti-Carrier Warfare (ACW) Capabilities and Operations.....41
Case 7: Soviet Naval Support for Third World Coups.....44
Case 8: Soviet Presence in the Indian Ocean....45
Case 9: Future Soviet Objectives for Forward Deployment.....50

Chapter 4. EVALUATION METHODS AND APPRAISALS.....55

ESTIMATION LOGIC.....55
Perceiving Data.....56
The Availability Heuristic.....57
The Representativeness Heuristic.....64
Representativeness and Availability....65
Assigning Weights to Data.....67
Knowing more than you can know--introspection on data weights..68
Redundant cues.....71
Cue variability.....74
Number of cues.....74
Cue-response compatibility.....75
Cue salience and vividness.....75
Nonevents.....76
Case cues versus base-rate cues.....76
Illusory correlation.....77
Vivid cues.....78
Characterizing Data.....81
Problems caused by characterizations...81
Prototypes and schemas as categories...83
Fuzzy categories.....84
Probabalistic boundaries.....84
Categories and memory.....84
Taxonomy biase.....86
Stereotyping.....87
Extreme cases.....88
Natural taxonomies.....89
Generalizations and representativeness.....92
Hindsightful categorizations.....94
Biased sampling.....94
Assessing Covariations.....96

Twofold tables.....98
Neglect of negative evidence.....99
Theory-driven and data-driven covariation estimates.....99
Perceiving nonexistent correlations...101
Illusory naval correlations.....102
Assessing Causes and Effects.....104
Causal meaning.....105
Causal errors.....105
Causality and chance.....106
Inhibitory causes.. 107
Causality and representativeness.....108
Foreseeability.....108
Magical thinking.....110
Capabilities and intentions.....111
Causal search biases.....111
Minimal causation..... 111
Causal hydraulics.....112
Parsimony, misguided parsimony, and indiscriminant pluralism..113
Calling a cause a cause.....114
Fundamental attribution error.....115
Assessing dispositional causes.....117
Causal warnings.....120
Prediction 120
General and specific causes of failure.....120
Illusions of control.....122
Illusion of order..... 122
Problems with long range forecasts....123
Medium-term forecasting.....124
Short-term forecasts.....125
Clinical versus actuarial judgment....127
Predictions and representativeness....127
Regression effects.....128
Regressive predictions through dilution 128
Diagnostic and undiagnostic data.....129
When base-rate data are used.....129
Sensitivity testing induced base-rate use..... 129
Predictions, scenarios, and compound probabilities.....130
Theory-driven overpredictions.....130
Illusion of validity.....131
Surprise and sudden events.....131
Forming, Maintaining, and Changing Theories...133
Theory conservatism.....135
Scientific theories.....135
Theories and evidence.....137
Maintaining existing theories.....137
False consensus bias.....138
Theory formation.....138



Perseverance of discredited theories.....141
Overconfidence.....142
Data-driven and theory-driven overconfidence.....143
Learning from experience.....144
NARRATIVE LOGIC.....145
Historical Perspectives.....147
Fallacies of history.....147
Inquiry.....148
Explanation.....150

Chapter 5: LITTLE USED METHODS AND EXPERIMENTAL TECHNIQUES FOR BETTER ESTIMATES.....155

PERCEIVING DATA.....155
Events and Nonevents.....155
Data on Frequencies.....157
Features of Samples and Populations.....158
WEIGHTING DATA.....161
Policy Capturing.....161
Bayesian Techniques.....163
Sample Size and Base-Rates.....163
Redundancy and Variance.....163
CHARACTERIZING DATA.....164
Aiding Memory.....164
Assessing Prototypes, Categories, and Dimensions.....165
Intention Categories.....168
Factor Analysis in Intelligence.....168
ASSESSING COVARIATIONS.....170
Actuarial Models and Backcasting.....170
Bootstrapping.....171
Environmental Effects.....172
CAUSE AND EFFECT ASSESSMENT.....173
Causal Search.....173
Incomplete perceptions.....174
Mistaken perceptions.....174
Causal analysis.....174
Simplistic perceptions.....175
Search trees.....176
Tree problems.....177
Stopping problem.....178
Hypothesis Generation.....179
Cognitive problems.....179
Social problems.....181
PREDICTIONS.....183
Identifying Assumptions.....183
Backcasting.....183
Bootstrapping.....184
Sensitivity Testing.....185
Validity of Future Assumptions.....185
Simple Models and Surprising Futures.....185

A Science Analogy.....186
THEORIES.....187
Theory and Naval Analysis:
An Endnote.....190

BIBLIOGRAPHY.....191

APPENDIX: HISTORICAL FALLACIES.....203

TABLES

Table 1. Sources of Naval Estimates Assessed and Consulted. (Abbreviations used in this study in parentheses.)..... 7
Table 2.1. Characteristics of Complex and Noncomplex Analysis.....21
Table 4.1. Correlations Between Data Sets on Trends in Diplomatic Port Visits.....73
Table 4.2. McConnell's Classification of Cases in Soviet Third-World Diplomacy of Force 90
Table 4.3. Analysis of McConnell's Classification of Cases in Soviet Third-World Diplomacy of Force in Terms of Concensus, Consistency Distinctiveness.....119
Table 4.4. Booth's Causal Warnings to Naval Analysts and Related Biases.....121
Table 4.5. Types of Historical Fallacies.....149

FIGURES

Figure 2.1. Examples of Six Levels of Specificity in Intentions..... ..17
Figure 4.1. Schematic Diagram of Availability Heuristic.....58
Figure 4.2. Hypothetical Relationship Between Indicator Data and Performance Data..... ...82
Figure 4.3. Examples of Biased Sampling.....95
Figure 4.4. Number (and Percentage) of Diplomatic Port Visits in the Indian Ocean and Mediterranean by the Soviet Navy 1967-1969, 1970-1976.....97
Figure 4.5. An Example of Oversophisticated Forecasting.....126
Figure 4.6. Responses to Delphi Questionnaires on the ASW Mission of the KRESTA II Cruiser.....134
Figure 5.1. General Estimation Diagram Illustrating Prediction Analysis (on the left) and Process Analysis (on the right)..... ...162
Figure 5.2. Lambelet's Approach to Qualitative and Quantitative Aspects of the Anglo-German Dreadnought Race.....182

PREFACE

The task of estimating the intentions of both friend and adversary has always been critical to statecraft and military affairs. The Chinese strategist, Sun Tzu, wrote long ago that it is foreknowledge which allows the statesman or general to achieve things beyond the reach of ordinary men.

But intention estimation has not been an altogether respectable occupation, involving as it does concerted snooping, prying, and spying on the one hand, and the possibility of disasterous blunders if the wrong estimate is made, on the other. This shady image is often reinforced by the remarks of journalists and even intelligence agency leaders which equate the gathering of information on intentions with the stereotyped image of the spy, cloaked and daggered, suborning and corrupting informers and insinuating his way into the enemy's secret councils. Nor are frequent reminders needed that it is easy to go wrong; intelligence failures are uncomfortably common now as in the past. Since the echoes of the CIA's scandals and abuses recently faded rarely has a month gone by without the appearance of an article on the "op-ed" page of a major newspaper (usually adorned with an illustration of a suitably shady and hooded figure) reminding the public that, while spy satellites may be able to count the Russians' missiles and forces, only the old-fashioned agent, out in the Moscow cold or some other hostile capital, can inform us what our adversaries intend to do with their forces. Although this equating of intention estimation with traditional spying is inaccurate, it serves to reinforce the public perception of intention estimation and intelligence gathering and analysis as, if not respectable, certainly necessary.

It has not always been so. As recently as the 1950s military intelligence analysts were explicitly prohibited from estimating the intentions of an adversary and required to estimate only capabilities (Stech, 1980b). As the nature of military and naval conflicts and the political relations of the world's nations grew more complex and swift-paced, the requirements for intelligence estimates of intentions changed. Explicit attention to intentions as well as capabilities is now required of military analysts (Stech, 1980b) and it has been suggested that no intelligence estimate is complete which fails to give attention to intentions and the integration of intentions and capabilities into a decision-making analysis of the adversary (Sullivan, in Codson, 1980: 51).

It is not likely that a nation's intelligence organizations can suddenly shift from a philosophy that prohibited intention estimation to one that requires it and just as suddenly begin getting first-rate estimates of intentions. There are many who believe the United States has not recently had foresightful intelligence estimates of intentions and there are even some who seem to believe foresight is largely impossible in the realm of intelligence (see Stech, 1979). Nevertheless, there have been some remarkable success stories of intention estimation foresight (see Stech, 1979, and Stech, 1980b) and it seems well worth the effort to try to improve the process of intention estimation by examining what is required, how it is done, the nature of the problems, and the possible solutions.

An initial step in this direction examined the phenomenon of intention estimation, whether it could or should be done, how failures

came about and why successes occurred (see Stech, 1979). An important part of that survey related the general problem of intention estimation to recent research in cognitive and social psychology on human information processing. The present study takes this research a step further and investigates how naval analysts estimate the peacetime naval intentions of the Soviet Union and how this task is affected by our human limits as information processors and problem-solvers. In effect, we asked and attempted to answer the following questions: What are the essential processes of naval analysis of intentions? How do the problems and limits of human information processing affect those processes? How do these problems appear in actual naval estimates? How serious and wide-spread are such problems? What might be done to aid and reinforce human information processing and problem-solving to possibly strengthen naval estimates of Soviet intentions?

This study was prepared for the Office of Naval Research under ONR contract N00014-78-C-0727 with Mathematica, Inc. I am deeply grateful to Commander Ken Hull and Mr. J. R. Simpson for their advice, guidance, and support in this research effort. I am also greatly indebted to many colleagues, friends, and advisors who guided me to the hearts of the intention estimation and human cognition problems. To the extent that I have aimed in the right directions, they are due the credit. If I have missed the target, I am to blame. Drs. Richard Betts, Garry Brewer, Judith A. Daly, Robyn Dawes, Baruch Fischhoff, Michael Handel, Edward Jones, and Barry Turner, and Major Generals Doyle Larson and Schlomo Gazit provided me valuable comments on intention estimation, cognition, and intelligence analysis. Mr. Richards Heuer shared with me his many articles and thoughts on the psychological elements of intelligence analysis. Ms. Cynthia Grabo provided many insights into indicators and the problem of warning and allowed me to delve into her extensive writings on these subjects. The following intelligence veterans and scholars shared their experiences and reactions to my efforts: Dorwin Cartwright, Ray Cline, William E. Colby, Matthew Gallagher, Raymond Garthoff, Arthur Hulnick, Klaus Knorr, William Kaplowitz, James McConnell, Michael McGwire, Dan Maxim, Walter Pforzheimer, Sayre Stevens, and Barton Whaley. Ms. Alma Hall once again rose above and beyond the call of duty and produced the typescript, tables and figures. I am also grateful to Ms. Emma Davis for her assistance.

Most of the prefatory quotations used in this study can be found in Bartlett's Quotations or any similar guide. Some of the more lighthearted aphorisms come from A. Bloch, Murphy's Law: Book II (1980 Los Angeles: Pricer, Stearn, Sloane) and P. Dickson, The Official Rules (1978, New York: Delacorte Press).

The views expressed here are those of the author and do not necessarily represent views held by The Office of Naval Research, Mathematica, Inc., or MATHTECH, Inc.

CHAPTER 1. INTRODUCTION AND OVERVIEW

> Before a war military science seems like a real science, like astronomy, but after a war it seems more like astrology.
>
> Rebecca West

WHAT THIS STUDY DOES

This report examines a sample of publicly available estimates of the peacetime naval intentions of the Soviet Union written by experts on the Soviet Navy. It does not examine the Soviet Navy itself, except through these sampled estimates, because the objective of this study is to determine how such estimates are made, where the estimation methods might be weak (or particularly strong), and to make some recommendations for the authors and users of such estimates. Because of this focus on methods, the images of the Soviet Navy conveyed here are apt to be misleading -- the reader should consult the original estimates for more faithful images. But the reader seeking a means by which to weigh and judge conflicting estimates, and the estimator seeking to improve the craft, may find some helpful analysis here.

Evaluations of the methods of estimators have been conducted before and these prior efforts have influenced the approach taken here. For example, George (1959) assessed the methods and predictive accuracy of propaganda analysts in World War II. More recently Archer (1978) assessed the methods and accuracy of forecasters predicting trends in population, national economic data, energy, transportation, and technology. Cockle (1978), Lee (1977), and Holzman (1980) have assessed the methods and accuracy of CIA estimates of Soviet defense expenditures. O'Leary and Coplin (1975) compared the methods of State Department intelligence analysts to quantitative techniques addressed to the same issues. Albert Wohlstetter (1974a, b, c; 1975a, b) evaluated the accuracy of U.S. Defense Department estimates of Soviet strategic capability and criticized what he saw as the fallacious basic assumptions of the estimators. In addition, as is done in this study, previous writers have recommended greater use of systematic, quantitative, or mathematical aids to intelligence and defense analysis (e.g., Heuer, 1978; O'Leary and Coplin, 1975; Saaty, 1968).

Estimation Methods

What is the point of assessing the methods of estimators or their processes of judgment, rather than just assessing the estimates and judgments themselves? This is a reasonable question but it reflects a belief that somehow the facts speak for themselves and that the methods of estimators have only an unimportant impact on those estimates (Garthoff, 1978). The idea that intelligence can somehow emerge from raw facts, untouched by the sullying fingers of method, is part of what Wasserman (1960) labeled "inductionism" and "naive realism" -- beliefs that all the facts need to turn into truth is unbiased observation, and that the "unvarnished" facts will admit only one interpretation.

Both the analysis of intelligence production, and the study of human judgment in general, have demonstrated that the methods (be they

intelligence analysis methods, or just the usual patterns of everyday thought) influence the judgments (see e.g., Steeh, 1979; Einhorn and Hogarth, 1981). The effect of such a seemingly trivial methodological issue as how a question is asked can have a major influence on the answer. A recent review of research on how people make decisions (Einhorn and Hogarth, 1981) argues that (emphasis added):

> . . . the most important empirical results in the period under review have shown the sensitivity of judgment and choice to seemingly minor changes in tasks (p. 72).

If minor, easily-overlooked aspects of judgment tasks can significantly change judgments, even though no data are changed, then it seems worthwhile to examine the estimation task and assess which details of method might be reducing the reliability or validity of estimates. A recent conceptual paper on estimates based on uncertain data (Howell and Burnett, 1978: 46-47) makes the related point:

> Man is both an able "intuitive statistician" and a pragmatic rule maker depending on what he is asked to do with uncertain information . . . (this) implies that description of task parameters deserves more attention than it has thus far received.

This study is an in-depth examination of the "task parameters" of the analyst estimating the peacetime naval intentions of the Soviet Navy.

Focus on Intentions

> Always pray that your opposition is wicked. In wickedness there is a strong strain toward rationality. Therefore there is always the possibility, in theory, of handling the wicked by outthinking them.
>
> Marion J. Levy, Jr., 1978

Why look at estimates of intentions? This is not an easy question to answer, and it conflicts with a long tradition that has dictated that military intelligence should estimate capabilities, not intentions (e.g., see, Steeh, 1980a; Garthoff, 1978). It can be argued that intentions must be estimated along with capabilities (Garthoff, 1978), that intentions can and should be estimated (Steeh, 1979), and that military doctrine has evolved from the traditional view to one which requires estimates of intentions (Steeh, 1980a). There is also the problem that estimates which appear to deal only with capabilities often conceal (not intentionally) an implicit estimate of intentions, reinforcing the need for methods which assess intention estimates.

Focus on Biases

> If one views his problem closely enough he will recognize himself as part of the problem.
>
> Duchan's Axiom

Why focus on estimation biases, weaknesses, and fallacies? The search for methodological weaknesses is as old as science. The first scientific encyclopedia, Roger Bacon's Opus Majus, delivered to Pope Clement IV in 1268, devoted the first of its seven major parts to human failings and their causes. Classic examples of analyses of faults are Thomas Hobbes' Leviathan and Behemoth, narrations of the follies and shortcomings of man. Swift's Gulliver's Travels satirized political faults; Voltaire's Candide, social flaws; Bentham's Book of Fallacies, historical and economic errors. This report makes no pretense to join these ranks, but they serve to illustrate a warning. If Bacon's man was excessively sinful, Hobbes' man a brutish beast, Swift's men villainous knaves, and Voltaire's man an ignorant naif, the pitfall of the present study is that it might be read as characterizing the naval analyst as a clever but self-befuddled fool. This is so far from the aim of this study that its opposite is far closer to the truth: it is my opinion (subjective and biased perhaps) that naval analysts are so careful and sophisticated that any problems remaining with their estimates are extremely subtle and nonobvious.

It is a very great mistake, a mistake Bentham and others have made, to believe that errors of logic or reasoning, fallacies of method, or failures of forecasting accuracy are prima facie evidence that the estimate is false in all respects and utterly useless. It is for this very reason, the need for analysis which can chop estimation logic into the weak and strong parts, that this study was undertaken. If naval analysts were fools, this study would be unnecessary, and a schoolboy could do for naval analysis what Bacon and his heirs have done for other realms. It is because naval and intelligence analysts tend to do their work so well that careful assessments of their methods are needed to determine any subtle and nonobvious biases that may remain.

Peacetime Intentions

Why focus on estimates of peacetime Soviet naval intentions? Is it really necessary to assess Soviet peacetime naval intention? Should not the naval analyst concentrate on the wartime intentions of the Soviet Navy? Up until about 1965, it was quite likely that any major contact between the U.S. or Western fleets and the Soviet Navy would occur during a war. Since 1965, however, contacts between the Soviet and Western navies are more frequent and extensive and are apt to occur in crisis situations that could lead to war. How the navies of the West and the Soviet Union conduct themselves in these peacetime and crisis encounters has a considerable impact on the likelihood that such crises might evolve into war. It is useful to review the origins of this interest in the peacetime Soviet Navy.

Origins. Intense Western concern over the forward deployments and fleet modernization of the Soviet Navy began in 1967, when a Russian "blue water" capability had become a credible reality. In 1967 the Soviets launched the first Y-class nuclear ballistic missile submarines (SSBNs), with 16 tubes for 1300-nautical mile SS-N-6 ballistic missiles. The first Moskva-class helicopter carrier became operational, carrying the fir

Soviet naval avaiation units afloat. In May and June, 1967, during the Arab-Israeli War, the Soviets deployed two anticarrier task forces to the Eastern Mediterranean, matching the two U.S. carrier task groups, marking the beginning of an expanded Soviet permanent presence in the Mediterranean. Between 1965 and 1976 Soviet Mediterranean warship operating days quadrupled from 4 500 to 18,600, with the largest jump coming in 1967. In October 1967, during the Arab-Israeli War, Egyptian forces manning a Soviet-built Komar-class missile patrol boat sunk the Israeli destroyer Eilat with a surface-to-surface Styx cruise missile.

These events had an immediate and continuing impact on Western appreciations of the goals of the Soviet Navy. They also seemed to many analysts to reflect a sharp change in Soviet intentions and led to various interpretations (as we shall show in Chapter 3) of why the Soviets had altered their naval policies, of what they could and would do with their new forces and forward deployments, of the risks the Soviets might run in these activities, and how they would employ these capabilities in peacetime and crises. Furthermore, a variety of methods were developed by analysts and employed to investigate these intention questions.

Despite the differences of opinion on long-run Soviet naval intentions, there was considerable agreement on the short-run significance of these events among Western observers. For example, the Y-class/SS-N-6 submarine, roughly comparable to the U.S. Ethan Allan-class/Polaris SSBNs, allowed the Soviets to begin closing the margin in offensive sea-based strategic missiles which the United States had enjoyed from 1961 to 1970.

The Moskva-class represented a significant departure from a Soviet policy dating from the post-Stalin era which deprecated the survivability and utility of aircraft carriers in nuclear war and stressed the superiority of the Soviet anticarrier defensive task forces. While some Western analysts quickly pointed out that "the Soviets (never) questioned the value of diversified carrier capabilities in a variety of situations short of (nuclear) war" (Wolfe, 1972: 37), others (e.g., Herrick, 1968; Smolansky, 1977) note that the Soviets have still to launch their first fleet aircraft carrier. Still others have noted the Soviets act as if their anticarrier warfare task forces are a significant hindrance on Western use of carriers in crises (Dismukes and McConnell, 1979: 22), implying a continued vulnerability of carriers. Analysts agree that the new Soviet naval aviation capability poses a new dimension to the problem of estimating Soviet naval capabilities and intentions, but disagree on just what these developments portend.

The expanded Soviet out-of-area deployments to the East Mediterranean foreshadowed the commencement in 1968 of regular deployments to the Indian Ocean, and ended what McConnell and Dismukes (1979: 16) noted was "a full decade (1957-1967) (in which) no recognized instances of Soviet coercive diplomacy based on forces in the forward area occurred " The Eilat sinking demonstrated that conventional gun-armed ships could be outranged and sunk by much smaller missile-armed craft. Wolfe (1972: 23-24) asserts that this event "served perhaps more than anything else to sensitize Western naval circles to the threat implicit in the Soviet Navy's adoption of anti-ship missile armament . . . the Soviet potential to challenge Western surface supremacy came to be taken far more seriously than hitherto." Analysts agree that these developments are perceived in the West as "coercive diplomacy" and a potential challenge to Western surface supremacy, but they disagree on how realistic these perceptions

are and whether these are the reasons the Soviets have undertaken forward deployment and adapted missile and rocket technology to naval weaponry.

The combination of these events made several conclusions about Soviet naval intentions fairly widespread. The explansion of the Soviet strategic offensive capability represented in the Y-class program, it was reasoned, would "draw" other Soviet units into blue water areas, especially ASW forces to protect the Y-class boats from Western hunter-killer submarine and surface forces. The Moskva-class ASW heliocopter carriers could serve this function, could operate against U.S. Polaris boats, or could be adapted to other roles, for example, as a vertical envelopment platform to support amphibious landing operations. Any of these missions could be expected to draw Soviet forces out of home waters into areas previously dominated exclusively by Western navies. The equipping of Soviet surface and submarine units with long-range surface-to-surface missiles gave the Soviet blue water task forces a credible capability previously lacking against Western forces that were out of range of Soviet air power.

The Soviet Navy, without increasing its forces (indeed in some respects, e.g., cruisers and destroyers, with fewer forces) had been transformed, in the eyes of the West, by its construction program and its expanded deployments, from a coastal auxiliary, into a potent competitor of the West on the high seas.

Furthermore, the Western perception of an upswing of Soviet "coercive diplomacy" (McConnell and Dismukes, 1979: 14; Dismukes and McConnell, 1979) using these transformed capabilities was compared to pronouncements (made in 1967) by Admiral Sergey Gorshkov, Commander-in-chief of the Soviet Navy, that in the mid-1950s Moscow had embarked on a new naval policy "aimed at building up capabilities not only for nuclear (world) wars, but also for conventional (local) wars and the 'protection of state interests' abroad in peacetime" (McConnell and Dismukes, 1979: 15). The fruits of this new policy, the naval units which appeared in the late 1960s, were seemingly put to use projecting Soviet power and "protecting Soviet state interests," beginning in 1967 and continuing up to the present.

In addition to this supposition that the Soviets have intentionally designed their forces for traditional diplomatic power projection and the capability to influence or engage in limited conventional wars (as well as to undertake other missions, e.g., strategic offence or defense), Western analysts have reached other conclusions that make the Soviets' intentions for these forces of great interest. These capabilities make Soviet involvement in Third World crises more likely and more frequent than would superpower crises between the USSR and the United States alone. Similarly, the United States, as a consequence of Soviet inducement, is more likely to become involved in smaller crises than in bigger ones which force it to directly confront the Soviets. This strong and increasing possibility that local crises may draw together both superpowers (as happened in the Middle East in 1956, 1967, and 1973, and in the Indian Ocean in 1971) makes any Soviet intention regarding local involvements, power projection, coercive diplomacy, or conventional capabilities of interest to the United States, both in the political sense that Soviet influence may thereby be expanding, and in the conventional and strategic military sense because U.S. military forces (especially naval ones) may become enmeshed in conflict with Soviet units. In

addition, the Soviet Navy's expanded forward presence increases its daily interaction with Western fleets since both sides detail units to observe each other's exercises and operations. In the past, harrassments were not uncommon. Although the 1972 U.S.-Soviet naval agreement to prevent naval incidents reduced the dangers of accidents and conflict by mishap, the frequent intermingling of forces in the various areas where large deployments overlap (the Mediterranean, Norwegian Sea, the Baltic and North Seas) made interactions with the Soviet Navy common and unavoidable as Soviet out-of-area deployments increased. Thus, with respect to Soviet world influence, to strategic deterrence, to Third World crises, and to daily peacetime noncrises contacts, the Soviet peacetime intentions for its naval forces are of great military and political significance to the West.

WHAT THIS STUDY COVERS

This study examines in detail the methods of naval analysts as these relate to different steps of intention estimation. Although the main purpose of this study is to assess methods, some analysis of the characteristics of intention estimates is necessary. In Chapter 2 the various means of prediction are described along with their requirements, the issues of specificity and precision are discussed, and the possibility of assessing the complexity and logic of estimates is suggested.

In Chapter 3 nine cases are reviewed which reveal differences in analysts' conclusions on important issues regarding the peacetime intentions of the Soviet Navy. These cases provide a mini-review of the methods and conclusions of many of the naval analysts in our sample, provide a short survey of some critical issues, and introduce the nonnaval analyst to some important controversies. The main purpose of this chapter, however, is to demonstrate the range of possible error in current naval estimates.

Chapter 4 contains the main analysis of this study, a dissection of estimation into component parts and an assessment of these parts from the perspective of cognitive psychology, logic, and information processing. This assessment is independent of the content of the estimates and aims at a deeper understanding of the estimation process.

Chapter 5 provides a variety of recommendations for strengthening the estimation process and offsetting many of the information processing problems noted in Chapter 3. Many of the methods recommended have already been tried or tested in other areas of intelligence estimation.

Sources

The analysis in this study is based on a sample of naval estimates widely available to the public. These sources are listed in Table 1. (Table 1 also notes the abbreviations used in this study for the more frequently used volumes in lieu of the regular referencing format.)

Selection method. These estimates were selected in the following manner. A literature survey was made of recent papers directly or peripherally related to the peacetime intentions of the Soviet Navy. This led to a list of twenty-plus volumes and papers. Many of these overlapped and eliminating duplications led to a list of thirteen titles, very

Table 1. Sources of Naval Estimates Assessed and Consulted. (Abbreviations used in this study in parentheses.)

Estimates Assessed:

BLECHMAN, B. (1973) The Changing Soviet Navy. Washington, D.C.: The Brookings Institution.

BOOTH, K. (1974) The Military Instrument in Soviet Foreign Policy 1917-1972. London: RUSI.

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

HERRICK, R. W. (1968) Soviet Naval Strategy. Annapolis, MD: Naval Institute Press.

HUDSON, G. E. (1976) "Soviet naval doctrine and Soviet politics: 1953-1975" World Politics, 29, 1, 90-113.

McCONNELL, J. M. and B. DISMUKES (1979) "Soviet diplomacy of force in the third world" Problems of Communism (Jan.-Feb. 1979) 14-27.

McGWIRE, M. (ed.) (1973) Soviet Naval Developments Capability and Context (SND). New York: Praeger.

-----(1979a) "Naval power and Soviet global strategy" International Security, 3, 4, 134-189.

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

McGWIRE, M., K. BOOTH and J. McDONNELL (eds.) (1975) Soviet Naval Policy Objective and Constraints (SNP). New York: Praeger.

McGWIRE, M. and J. McDONNELL (eds.) (1977) Soviet Naval Influence: Domestic and Foreign Dimensions (SNI). New York: Praeger

MURPHY, P. J. (ed.) (1978) Naval Power in Soviet Policy (NPSP). Washington, D.C.: U.S. Government Printing Office.

NITZE, P. H., L. SULLIVAN, et al. (eds.) (Atlantic Council Working Group) (1979) Securing the Seas: The Soviet Naval Challenge and Western Alliance Options (STS). Boulder, CO: Westview.

RUBINSTEIN, A. Z. (ed.) (1975) Soviet and Chinese Influence in the Third World. New York: Praeger.

-----(1980) "The evolution of Soviet strategy in the Middle East" Orbis, 24 (2), 323-337.

Sources Consulted

BATHURST, R. S. (1979) Understanding the Soviet Navy A handbook. Newport, R.I.: Naval War College Press.

FAIRHALL, D. (1971) Russia Looks to the Sea A Study in the Expansion of Soviet Maritime Power. London: Andre Deutsch.

HUSSAIN, M. M. (1979) "Soviet perceptions of U.S. positions-of-strength diplomacy" World Politics, 31 (4), 495-517.

JONES, C. D. (1975) "Just wars and limited wars: restraints on the use of the Soviet armed forces" World Politics, 28 (1), 44-68.

-----(1977) "Soviet hegemony in Eastern Europe: the dynamics of political autonomy and military intervention" World Politics, 29 (2), 216-241.

PIPES, R. (1977) "Why the Soviet Union thinks it could fight and win a nuclear war" Commentary, 64 (1), 21-34.

POLMAR, N. (1972) Soviet Naval Power: Challenge for the 1970s. New York: National Strategy Information Center.

similar to that shown in Table 1. This tentative list of titles was shown to two leading figures in naval analysis, Michael McGwire of Dalhousie University and James McConnell of the Center for Naval Analysis (CNA), as well as to technical officers in the Office of Naval Research. The list was also discussed informally with naval analysts at CNA, CIA, and in DoD. In all cases we asked if there were titles or papers not on the list that should be, or titles on the list that could be dropped. The resulting recommendations led us to add several papers and to consult, but not analyze several more.

Types. It is not likely that this list of estimates is highly representative of any one school or type of naval analysis. Both quantitative and qualitative papers from a variety of disciplines other than naval analysis are included. Analysts from government and academia are also included. Not all the papers in these collections received equal attention; the major emphasis was put on those most clearly related to peacetime naval intentions.

WHAT THIS STUDY DOES NOT COVER

Several important issues regarding the methods of naval analysts are not treated in this study. These are all important dimensions of naval analysis methodology and deserve detailed assessment, but such assessments require resources that exceed those of this study.

Accuracy

We do not attempt to comment extensively on the accuracy of the content of the sampled estimates of Soviet intentions. While it is possible to demonstrate occasional instances where an analyst has been proved right or wrong by events, in general, naval estimates are too vague or imprecise to allow us to assemble a useful track record of accuracy In many cases accuracy could only be appraised by resorting to classified material. A partial record of "hits" and "misses" would be highly misleading. Some analysts, for example, may handle only the hardest topics (in the sense of predictions), and have a large number of obvious misses and an unmeasurable number of hits. It would be inappropriate to compare these analysts with those predicting other issues.

Social Effects

Each estimate is assumed in this study to be the product of its authors. This assumption neglects the social facts that shape the planning, conduct, and drafting of intelligence products. Analysts are not immune from social pressures or influences during these stages, and these social effects can perhaps be detected in the finished estimative products. To assess this social influence adequately, however, would require either a participant-observer or direct observation of the analysts at work. These are both feasible methods but were impracticable in the present study.

Organizational Effects

This study makes no effort to determine whether naval estimates tend to reflect the organizational interests of the institutions that sponsor them. Lacking any clear metric for prediction accuracy we cannot meaningfully determine if analysts for one organization typically over- or underestimate the Soviet Navy Nor can we meaningfully assess what an organizational bias might be through simply examining the estimates of analysts from that organization. Even if they all "sang the same song," it could simply be that they were all accurately perceiving the same data. It is also possible that analysts in an organization came to the same conclusions because of information processing biases. That is, the organization may predispose the analysts to use certain methods (not to reach certain conclusions), these methods may tend to bias information processing in certain directions, and analysts come to share conclusions because they share methods. While these are intriguing possibilities, they could not be addressed here.

Motivational Effects

The common understanding of the term "bias" gives it a motivation meaning: "I am biased toward a particular viewpoint because that is how I want things to come out." An alternative to this "wishful thinking" definition is one which sees bias as serving instrumental ends: "I am biased toward a particular viewpoint because it will suit my ends to have that view accepted."

Note that these meanings do not imply any conscious deceptive intent on the part of the analyst. Deception might be a hoax, or a lie, but it cannot be a bias. These motivational biases are unconscious self-deceptions.

This study does not deal with motivational biases for two reasons; philosophical and psychological. A motivational bias implies that the analyst tends to think (is biased) toward a conclusion that somehow he or she could know is not accurate. It implies that the analyst could control the effects of motivations on perceptions and thinking if he tried hard enough or struggled to remain objective. In effect, motive one (wishful fulfillment) struggles with motive two (objectivity) and if it wins, the analyst is biased because motive one then struggles with perception and again wins. This logic introduces several layers of regress from what can be meaningfully observed and measured.

Psychologically we can demonstrate several means by which bias in analysis can occur simply through information processing mechanisms (Chapter 4) without recourse to motivational mechanisms. These biases are strong and pervasive, and can be assessed directly. While we cannot deny that motivational biases exist, it seemed more profitable to assess the more accessible cognitive biases. Since these are less intuitively obvious than motivational biases, they are probably more insidious and difficult to guard against. In fact, most people are unwilling to accept that, without explicit help, they cannot logically think through complex problems.

Political Effects

This study does not address the charge that intelligence estimates are often wrong because they are political creations with little relation

to reality. While some estimates have sufficient policy import to become political footballs (e.g., the Team A-Team B episode regarding Soviet strategic intentions in the mid- to late-1970s), most estimates probably never reach important politicians and are protected from political pressures by sheer neglect. Political pressures may be brought to bear on any forecasting or estimative effort that has potential policy relevance, but in a pluralistic society with a highly decentralized bureaucracy, as in the United States, a variety of differing political viewpoints are likely to be represented in any pressures that are brought to bear.

This does not mean that political competition can substitute for estimation objectivity. The interaction of these two tendencies is an important component of the analysis of intelligence production but beyond the scope and competence of this study.

HOW TO USE THIS STUDY

The following remarks suggest how various readers might peruse this study other than reading it cover to cover.

Naval Analysts

Chapter 4 describes a variety of nonobvious problems with estimation logic and narrative logic. The examples demonstrate how these information processing problems can occur in naval analysis. Naval analysts may find reasons to reflect on their own methods in this chapter, and many new ways to question and challenge the conclusions of rival analysts. This chapter provides an objective means of assessing the methodological validity and logic, and the cognitive process, of intention estimation, independent of data or conclusions.

Chapter 5 reviews a variety of methods and techniques which have been tried in other intelligence fields and which might improve naval analysis and intention estimation. While not a handbook or cookbook of methods, it suggests ideas for new methods in naval estimation.

Chapter 3 briefly reviews some major controversies among naval analysts regarding the future peacetime intentions of the Soviet Navy.

Soviet Analysts

Chapter 3 briefly summarizes differing opinions on the future directions of the Soviet Navy. These open questions suggest some important gaps in our understanding of the past, present, and future of this increasingly important Soviet institution and the need for nonnaval perspectives.

Chapter 4 suggests how some of these differences of opinion arise from the way analysts process information on the Soviet Navy. These biases may affect any analysis. They also suggest some possible directions to go for remedies. These directions are outlined in Chapter 5.

Consumers of Naval Estimates

The problematic characteristics of current naval estimates (Chapter 2) make them less useful than more precise, specific, and

predictive estimates would be. But precision and predictiveness must be traded off for uncertainty. At present, uncertainty is not handled explicitly or quantitatively by naval analysts. They are well aware of uncertainty but seem not to know what to do with it except acknowledge it and skirt it. This leads to many biases in the processing of uncertain information (Chapter 4). Naval analysts, like most people, seem largely unaware of these biases. Problems with the content of estimates may actually be due to the methods used to process information rather than to problems with the analysts' data, perspective, or conclusions.

While a range of analytic opinion probably exists on any subject of policy importance, this range itself can be an important gauge of analytic error and uncertainty (Chapter 3), and a sign that (1) the differences are due to methods as well as opinions, (2) other methods may be useful (Chapter 5).

Cognitive Psychologists

Examples of cognitive and information processing biases are usually the product of laboratory demonstrations. Many "real world" examples from naval analysis are presented and assessed in Chapter 4. Chapter 5 suggests many methods for improving estimation and judgment, most of which have had field tests in intelligence analysis, but which deserve greater investigation in controlled laboratory settings.

Students and Managers of Intelligence

Serious scholars of intelligence may find the analysis of the characteristics of estimates in Chapter 2 of interest; in particular the discussion of prediction, specificity, and precision.

The biases described in Chapter 4 are not unique to naval analysis or intention estimation and may weaken intelligence analysis of all forms. Similarly, the suggested methods for improvement (Chapter 5) have had some favorable reception in intelligence production and deserve further investigation and application.

CHAPTER 2. CHARACTERISTICS OF ESTIMATES

METHODS OF PREDICTION

The prediction of the future behavior of another actor, such as the Soviet Navy can rely on three modes or methods of analysis, which Scheibe (1979) has labeled sagacity, control, and acumen.

Sagacity

Sagacity depends on the understanding of an analyst or estimator of relationships between subtle cues and behaviors. A simple example would be the analyst who uses the number of submarines now in various stages of completion in Soviet yards as an indicator of future Soviet submarine strength. A more subtle example is the relationship analysts perceive between a few marker words and Soviet military doctrine (McConnell, SNI: 605-614; Dismukes and McConnell, 1979: 314; Gallagher, SNP, ch. 3; McCGwire, SNI, ch. 2; it is sometimes argued that these perceived relationships are, in fact, illusory). By observing which features are correlated with a behavior, or precede a behavior regularly, the analyst becomes able to forecast the behavior by watching the indicator cues, much like a doctor can forecast the future course of an illness by making an appropriate diagnosis and using it for prognosis. The key to this prediction process is association, either immediate, or distant. Immediate association occurs when the indicator cues are perceived as closely related to the behavior to be predicted, for example, when the indicators occur just before the behavior. An important class of these associations are those the analyst views as causal -- the occurrence of indicator cues are perceived as necessary and sufficient for the occurrence of the behavior. For example, the analyst might perceive a strategic threat to the Soviet homeland as the cause of a Soviet Navy fleet attack on Western forces, believing that such an attack would occur only when such a threat was perceived by the Soviets.

Distant associations might depend on data that are diagnostic rather than immediate or causal. The fact that a particular Soviet task group includes ships with surface-to-surface missiles may suggest a tendency for that task group to behave in particular ways that it would not if it were, say, composed of surface-to-air or antisubmarine warfare ships. In another example, Petersen (in Dismukes and McConnell, 1979: 46, 98) has noted that the age of Soviet surface ships is inversely related to their out-of-area deployment and the frequency of their port calls in the Third World -- the Soviets tend to call with their newest ships. Age is thus somewhat diagnostic of mission, and hence, future behavior.

Obviously, whether an empirical relation between indicator and behavior is immediate or distant depends on factors other than causality and diagnosticity, and is in part subjective. Furthermore, an association may be very strong but not very predictive. For example, the probability that a Soviet ship making a port call will be modern is quite high, but the probability that any given modern Soviet warship will make a port call is probably quite modest. Port call predicts age better than age predicts port call.

In addition to an observed empirical association between indicator and behavior, the analyst may rely upon a theoretical relationship or

association between indicator and behavior. This theoretical association may result from some insight of the analyst into the dynamics of the behavior, as when the analyst notes that the increased strategic radius of U.S. Navy units created by the A-3 strike aircraft and the Polaris system would "draw out" the Soviet Navy into a forward deployment. Predicting such a relation between indicator (increased range of strategic threat) and behavior (forward defense) could not be deduced from empirical observation, except in hindsight. It could be inducted, in foresight, from a theoretical analysis of the causes of the dynamics in Soviet naval defensive strategy.

Control

The second predictive method Scheibe noted depends upon control of behavior. By controlling situations, or rewards and punishments, it is often possible to accurately predict behavior. Analysts of the Soviet Navy rarely exercise any meaningful control over the subject of their study. Such control is not impossible however. The "intelligence experiment" can provide predictive information by controlling the information available to an adversary. An historical example of an intelligence experiment occurred in World War II when U.S. naval intelligence had messages sent in the clear about Midway Island being short of water. Japanese intercept stations monitored this traffic and U.S. intercept stations in turn monitored Japanese traffic that revealed that the target of the forthcoming Japanese attack was short of water. Through this experimental control the U.S. analysts confirmed the identity of the Japanese target.

While the analyst may not control the situation or the reward structure, he or she may be able to predict the behavior of an agent who does have such control. It may be possible to predict the [illegible] Admiral Gorshkov, and since he exercises control over certain [illegible] the Soviet Navy, predict what those aspects will be in the [illegible]. Similarly, by assessing the state of the art of Soviet hydroacoustics the analyst may be able to predict future Soviet acoustic ASW capabilities and tactics. At times the agent of control is the analyst's own country. The analyst who was aware that the Polaris system would impose new defense requirements on Soviet naval forces could predict the Soviets would react to this requirement. While it is not always easier to predict the future actions of one's own country than those of an adversary, the patterns of action-reaction that often characterize military moves provide the analyst a useful control prediction tool, at second hand as it were.

Acumen

The third method of prediction which Scheibe (1979) describes is acumen, or psychological penetration:

> . . . an appreciation of the other's point of view in a specific place and time . . . a particularized act of identification with the other . . . penetration to the specific thoughts and intentions of a specific person (p. 40).

Recently an historian recommended such acumen as a major element of effective intelligence analysis. Richard Pipes (in Godson, 1980) argues (p. 177) that historical scholarship was one of the "keys to political intelligence analysis" and that the other was

> . . . a deep knowledge of the country, of its general culture and its political culture in particular, from what the Germans call Fingerspritzengefuhl, the feeling 'on the tips of your fingers' for a given culture where you know that some things are more probable and others less so.

In many respects acumen is the result of pure reasoning, developing an appreciation of the other's options and accurately assessing the probability that a particular option will be chosen. James March (1978) characterized rational choice in these terms: "a guess about uncertain future consequences and a guess about uncertain future preferences" (p. 587). Decision analysis (which is just applied rational choice) offers one means by which an analyst might clarify his or her thinking about how the adversary is making these guesses. "A player with greater acumen," Scheibe wrote (p. 52), "will be able to penetrate a game tree more deeply by accurately eliminating from consideration the range of moves which the opponent is not likely to exercise." Decision analysis could help the analyst outline the game tree and enumerate uncertain future consequences and uncertain future preferences of an adversary.

Prediction Requirements

Given these three methods of prediction, what are their requirements; what must the analyst be able to do to effectively employ them?

Sagacity. At a minimum sagacity requires the analyst to (1) categorize the behavior and the indicators and measure them, (2) detect correlations between the two sets of categories, (3) distinguish causal from noncausal relationships, (4) detect diagnostic information in the categories, (5) formulate theoretic relationships between categories, (6) inductively and deductively reason from the indicator categories to a prediction of the behavioral categories, (7) adjust reasoning in the wake of confirming or disconfirming experiences (see Kinhorn, 1974, for comparable analysis of the psychometric criteria for expertise).

Control. To employ control for prediction the analyst must (1) characterize the relationships of the control situation (i.e., know what is controlling what), (2) be able to influence the controlling agent (directly or through guile, as in the intelligence experiment), or (3) be able to use sagacity or acumen to predict the controlling agent.

Acumen. The requirements for insight are by no means clear (see, e.g., Ornstein, 1977). Imagination and empathy are undoubtedly useful in creating a representation of the other's decision process. Historical scholarship, cultural experience, and decision analysis are probably useful in validating that representation although warnings abound for each enterprise (e.g., see Fischer, 1970, or Fischhoff, 1980a, on history).

But the problem of acumen is an extremely difficult one to solve even when the decision-maker is oneself (e g., Howard, 1980; Fischhoff, 1977, 1980; March, 1978). For example, what is "optimal" is highly subjective (Einhorn and Hogarth, 1981), influenced by the context (or framing) of the problem (Tversky and Kahneman, 1981), and unstable (March, 1978). Each of these difficulties is even greater when the subject of acumen analysis is not oneself or an accessible client, but a remote and possibly wary adversary, also capable of acumen.

Scientists have just begun to understand how people develop insights into such problems as physics and chess (e.g., Larkin, et al., 1980). For example, people who have "physical intuition," or insight into the means of solving physics problems ("experts") tend to have a greater body of knowledge about physics than do those with few insights ("novices"). In addition, experts have that knowledge efficiently indexed by patterns that, when recognized, guide the expert to the relevant parts of his or her stored knowledge. Larkin, et al. (1980: 1336) characterize "physical intuition" as knowledge organized into complex schemes that guide a problem's interpretation and solution. Similar conclusions have been reached about mathematical and engineering problems (e.g., Wickelgren, 1974).

It should be noted, however, that physical and mathematical problems have optimal solutions. It is easy to determine whether the problem-solver chose the correct answer or took the right steps towards the solution. There is no uncertainty associated with these judgments either. Many decisions that are of interest to the naval analyst are not so easily scored. One reason for this is that the Soviets, like others, must make "trade-off" decisions -- benefits must be weighed against costs and one "good" weighted against another. Rather than the best decision, the problem-solver or decision-maker seeks the best within constraints, and such decisions are made in an uncertain environment where future values and outcome can, at best, be known only probabilistically.

The story is oft repeated that Soviet Navy commander Gorshkov has in his office a large sign to the effect that "better is enemy of good enough" (Kehoe, SNI, p. 386). Evidence from Soviet shipbuilding is entirely consistent with the notion that the Soviets are very good ship engineers and make rational tradeoffs between ship qualities and costs (e.g., Maier, SNI, ch. 20). It is less clear why they perceive various aspects of ships as qualities or give them the weight they do (e.g., Kehoe, SNI, ch. 19; Thorpe, NPSF, ch. 5).

Despite these differences, decision insights, like physical or mathematical intuition, probably rely on complex schemes of knowledge which the "expert" uses to recognize types of problems. The naval intelligence analyst attempting to use acumen to predict Soviet behavior needs to construct, if only implicitly (but preferably, explicitly), a model of Soviet decision-making schemes.

SPECIFICITY OF INTENTIONS

Intentions involve four different elements: an action, the target at which the action is directed, the situation in which the action is to take place, and the time at which the action is to be performed. Each of these elements varies along a dimension of specificity. At the most specific level is the plan, a detailed set of instructions describing

all phases of action, target, situation, and time. Even a very specific plan may fail to foresee every contingency in the situation, or provide all the details needed to effectively accomplish the action. At the most global (least specific) level intentions may lack detail on each of the four elements and may reflect only the general orientation of one actor toward another. The following diagram follows the outline of Fishbein and Ajzen (1975: 296) and uses entries from McConnell (in Dismukes and McConnell, 1979) to demonstrate the specificity of intention.

As Figure 2.1 suggests, estimates of Soviet naval peacetime intentions may be highly specific. On the other hand, many estimates provide no more than a global intention and a cluster of actions, and fail to predict specific actions, situations, and times these general actions will occur.

Specificity and Precision

The specificity of an intention estimate is closely tied to the possible precision of that estimate. It is impossible for a vague, global intention estimate to be precise. The minimum level of specificity for intention estimates is specific action. This at least permits precise measures of behaviors. This level lacks precision on when, where, or in what circumstances the specific action will occur. An estimate limited to intentions for specific actions may have great utility nonetheless, especially for such tasks as engineering countertactics.

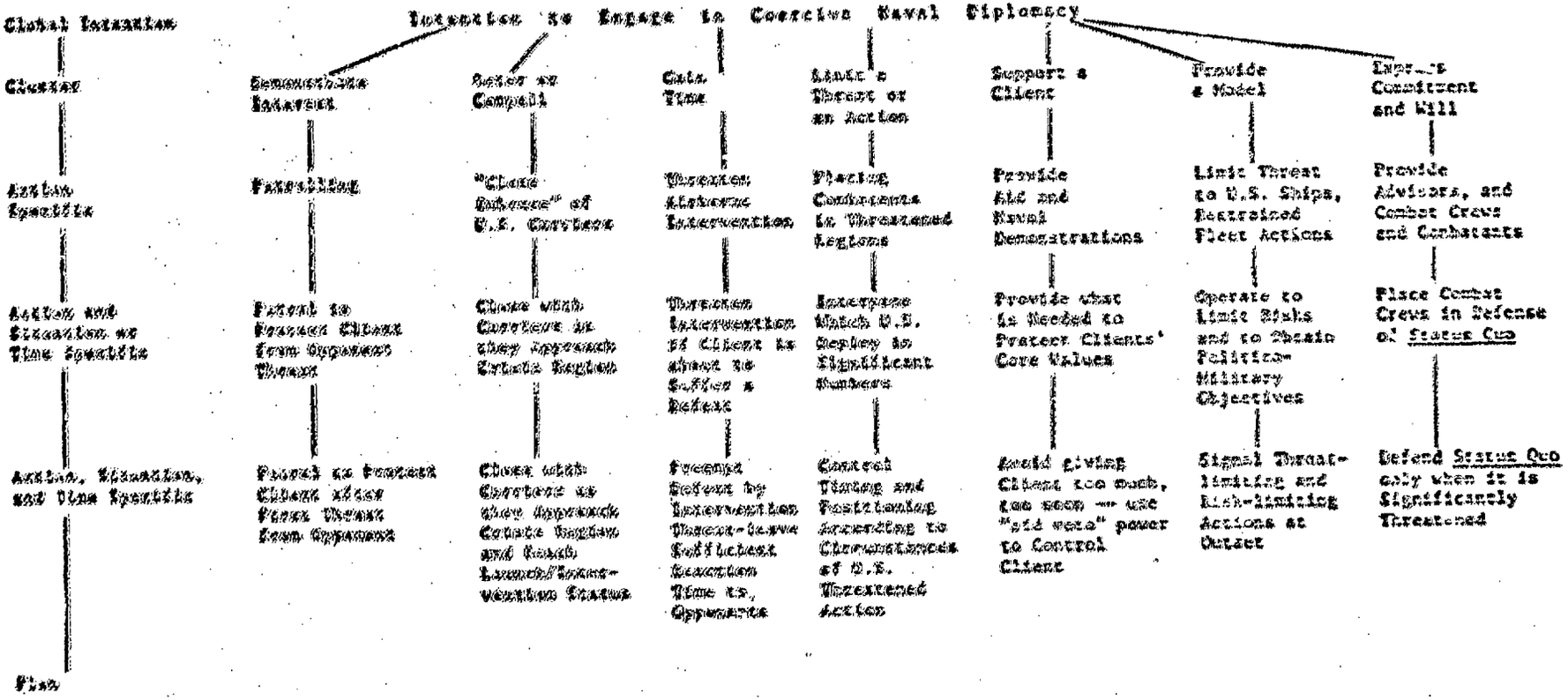
Forecasts and Predictions

Estimates of intentions may be either highly precise (i.e. predictions) or less precise (i e., forecasts). The distinction between forecast and predictions is by no means a settled issue (e.g., see Freeman and Job, 1979: 115 ff), but we will take prediction to mean an estimate of all four specifics of intuition: the action, the target of the action, the situation in which the action is to take place, and the time at which the action will be performed. If any of these four specifics is omitted, we term the estimate a forecast so long as at least one specific element remains.

Several readers will take exception to these definitions, especially that of forecast. Others (e.g., Freeman and Job, 1979: 117) define prediction much as we do, but define a forecast as a prediction without specification of situation. This seems too precise to us. Some estimates go no further than saying, "There is a small likelihood that the Arabs will mount a dramatic initiative against Israel in the next few months" i.e., only the time and target factors are specified (this and other examples are from State Department intelligence reports, see O'Leary and Coplin, 1975: 173). This vagueness is not unusual in actual estimates, and should be labeled as something other than predictions, which implies validation or falsification. Many intelligence forecasts (or projections, cf. O'Leary and Coplin, p. 173) cannot be verified or falsified, e.g.:

> Progressive deterioration of the situation over time could lead to major hostilities. (Specifies situation.)

Figure 2.1. Examples of the Levels of Specificity in Intentions.



> Nasser may decide that some sort of military action, regardless of the consequences, is the least risky course left to him. (Specifies action.)
>
> Egyptian suspicion/hostility toward Israel is unlikely to diminish under Nasser or his successors. (Specifies target.)

Despite being specific on one factor, the lack of specificity on any other factors prevents any useful disproof of these estimates (e.g., can anyone actually determine if "Egyptian suspicion/hostility toward Israel" has or has not "diminished"?).

Time Horizon

The expressions short-range, short run, or short-term and long-range, long run, or long-term are often used by estimators. Rarely do estimators specify what chronological periods are meant by these expressions. This vagueness is not uncommon:

> Rarely, if ever, do [forecasters] specify where the boundary between short-range and long-range lies or what exactly is meant by short and long. While there is a traditional requirement in science to replace qualities by quantities this tradition, apparently, is not applied to the realm of prediction and forecasting (Taschdjian, 1977: 41).

Both historical explanations and future predictions have time horizons. This horizon is the boundary which separates what can be explained (in the past) or foreseen (in the future) from what cannot. An historical explanation can be extended into the past only so far before its important variables become anachronisms (e.g., nuclear deterrence cannot be applied to explain events prior to 1945).

Historians sometimes do attempt to explain history in terms of First Causes, just as some analysts of the Soviet Union resort to an image of primal Russian man as an explanatory device. Such explanations are hardly worthwhile since they usually fail to link events with proximate causes and they offer feeble links between events and distant causes. The greater the distance in the past the lower the relevance of the connections, because the number of contributory causes increases tremendously as the explanation reaches farther into the past.

When estimators use the recent past and the present for predictions and forecasts they assume past trends will stay valid through the time horizon of the forecast. As that horizon extends there is less probability that the ceteris will remain paribus, i e., the present and past trends, extended into the future, are more likely to change. The immediate past and the immediate future are approximately related in a straight, linear fashion. The multitude of starting points in the distant past converge on the present. The possible courses of future events diverge from the present. As events become more distant from the present, either in the past or the future, they are less likely to be related in a straight, linear manner to present events.

For example, Ascher (1978) has found, for a variety of forecasts

(energy, population, economics, transportation), that the more distant the period forecast, the greater the error, although the relationship between distance and error is by no means linear.

In contrast to the forecasters Ascher studied, naval analysts rarely specify the date for which they are making a prediction. Instead, naval analysts, when they consider a time horizon at all, usually do so within a situation-contingent context. That is, they tend to estimate how rapidly or slowly the Soviet Navy might act within the context of a specific situation, e.g., crisis. This form of contingent time prediction implies that time predictions are only accurate if situations can be correctly identified, and their onset accurately marked.

It is interesting that naval analysis of Soviet capabilities has highly specific time horizons, but that analysis of Soviet naval intentions does not. That is, analysts of capabilities realize that, for example, ships must be planned, designed, constructed, and delivered on schedules, and that the "physics" of these schedules allows the analysts to project when systems will pass various stages. However, even those naval analysts who believe Soviet naval behavior is planned seem not to find evidence of the time factor in those plans. In fact, such analysts tend to explicitly avoid the concept of a Soviet timetable. This implies that the Soviets have only flexible, contingent plans, although the Soviets may have a general framework for coercive naval diplomacy, for example. While an estimate of capabilities which lacked an explicit time horizon would be automatically judged incomplete, estimates of Soviet Navy plans for future peacetime behavior that included an explicit consideration of time would probably be judged highly as improbable by other analysts. On the face of it, however, there is no less reason why the Soviets should not change their planned capabilities than they should their planned behaviors.

Situation Prediction

The situation, or context, in which action takes place tends to be the framework of many naval estimates. The two most common and general situations considered are wartime and peacetime missions. Wartime missions are generally subdivided into nuclear, general conventional, and limited war. Peacetime naval tasks include such subdivisions as strategic deterrence, diplomatic presence, power projection, crisis influence, client support, demonstrations, exercises. Specification of situations dictate the conditions and circumstances under which the Soviet Navy would take or avoid a particular action.

Two means of specifying peacetime situations seem to be used by naval analysts in our sample: scenarios and "rules of the game." The authors of Securing the Seas, for example, outline several scenarios which might lead to a U.S.-USSR naval confrontation. Scenarios usually specify a particular geographic region (e.g., the Middle East), specific actors (e.g., Persian Gulf naval states, the U.S. and Soviet Navies), and particular actions (e.g., attempts by Persian Gulf states supported by the Soviets to block the flow of Western oil). In contrast, McConnell (in Dismukes and McConnell, 1979, ch. 7) attempts to specify the various diplomatic security missions the Soviet Navy might undertake in a variety of regions, for different Third World actors, and which might involve several different actions. Clearly, McConnell's "rules" are far more

flexible and able to cover more situations, but are much less precise, than particular scenarios. On the other hand, a particular scenario is not likely to occur, i.e., it is likely to be wrong in one or more particulars. McConnell's "rules," by attempting to specify future relationships, rather than specific future events, are less empirically testable, but perhaps more theoretically heuristic as a result. That is, his rules do not specify which particular actions, actors, or targets will be present in any given situation (although the rules restrict the range somewhat for each situation). Instead, the rules generate a limited number of situation types which are loose enough to include most or all anticipated Soviet naval diplomatic activity. The "rules" do not predict intentions so much as suggest the likely bounds on various situations (e.g., the Soviet Navy would not risk confrontation with the West in a particular situation, but may risk increasing tension, etc.).

Target and Action Specification

Naval analysts are often quite precise in specifying the targets and actions of the Soviet Navy in certain elaborated scenarios. For example, the frequent contacts between Soviet Navy anti-carrier warfare (ACW) ships and U.S. aircraft carrier task groups during Middle Eastern crises (e.g., 1967, 1970, 1973) have led to fairly precise estimates of how the Soviets will organize and deploy their ACW forces against the carrier targets. In general, to the degree that the targets are naval systems and capabilities (e.g., aircraft carriers and task groups), or naval actions (e g., amphibious intervention ashore), naval analysts seem able to estimate specific Soviet naval actions and targets of action. As the targets become less naval in nature the estimates become less specific and differ more from each other. For example, when the general target is Third World client states, naval estimators are much less precise in specifying the naval actions and specific targets of Soviet naval action. Analysts tend to differ on which actions the Soviet Navy will try, on the targets the Soviets will try to influence, and the goals the Soviets seem to be trying to achieve. When targets are more naval in nature (e.g , U.S. carrier groups) analyst disagreements on Soviet actions and goals diminish (but do not altogether disappear).

Scoring Predictions and Forecasts

It is quite straightforward to evaluate the accuracy and validity of a prediction; one simply compares it to reality. If the action, target, situation and time coincide with events, the prediction was confirmed. It is far more difficult to score forecasts. An analyst, for example, forecasts an increase in military action but fails to specify the time horizon. How does one evaluate this openended statement -- the forecast becomes true just as soon as the projected action takes place, and stays open until it becomes true. Some forecasts can be more easily scored than others, e.g., those that specify action and time. For example, projections of Soviet naval capabilities may only specify what the Soviets will build and the dates, without reference to targets or situation. These projections can be scored straightforwardly. Forecasts of Soviet naval intentions, particularly those cast in situationally dependent terms, are more difficult to score correctly, particularly if

Table 2.1. Characteristics of Complex and Noncomplex Analysis

Noncomplex Analysis	Complex Analysis
1. Vague goals for analysis and no clear criteria for specificity of conclusions.	1. Explicit goals and criteria for specifying content of conclusion.
2. Limited or nonexistent plan of analysis, no schematic blueprint for sequencing the inference and deduction process.	2. Clear methodological outline, explicit statement of relations between methods and inferences.
3. No clear relation between (a) methods and data, and (b) conclusions or recommendations.	3. Explicit criteria for weighing methods and evidence against conclusions.
4. Little consideration of dynamic processes in time; primary focus on the status quo.	4. Attempt to account for past trends and tendencies and relate them to the status quo.
5. Repeated use of same analytic method, or same data base, with little increase in new information.	5. Progression from conclusions, to new data, to methods, to new conclusions, to new data...etc.
6. Minimal critical reflection on methods employed, no use of devil's advocacy, multiple methodology, hypotheses disconfirmation.	6. Explicit consideration of method strengths and weaknesses, attempts to offset biases and weaknesses through multiple methods, use of disproof as well as confirming evidence.
7. Heavy reliance on unproven hypotheses, assumptions, conventional wisdom, stereotypes, doctrinal "givens."	7. Reliance on empirical testing, parsimony, and Occam's Razor rather than convention.
8. Focus on small details.	8. Emphasis on major processes.
9. Use of highly-global, single-factor theories for large scale phenomena.	9. Use of multiple hypotheses, multiple data sources, multiple methods to building blocks of theory.
10. Testing to confirm hypotheses (fortifying and entrenching).	10. Testing to disprove hypotheses (process of elimination).
11. Thematic vagabonding: rapid, superficial shifting from concept to concept, insufficient development of events or explanations.	11. Explicit links between events and explanations, with causal explanations of input-process-output relationships.
12. Explanations in terms of causal series, single causes and single effects.	12. Explanations in terms of multiple causes, causal networks and multiple effects (both main effects and side effects).
13. Surprising, large, or sudden changes are not related to the past.	13. Explanations in the past are sought for dramatic or exponential changes.
14. Unwillingness to reach decisions, make predictions or forecasts, and tendency to draw contradictory conclusions.	14. Effort to forecast and predict, attempts made to verify predictions, internal consistency of conclusions.

the situational specifications are vague or loose. Furthermore, situational forecasts are contingent predictions (i.e., the Soviets will do X if conditions A, B, C obtain, but not otherwise) which may require complex scoring rules for cases which approach but do not actually reach the exact specifications.

COMPLEX AND NONCOMPLEX ANALYSES

Doerner's (1980) study of decision-making in complex environments provides a list of characteristics for what might be termed "complex analysis," or estimates of complex systems. These characteristics can be contrasted to noncomplex, or simple, analysis, which will be inappropriate for the estimation of complex problems. The fact that an analysis is complex does not imply that the conclusions that emerge must also be complex -- the conclusions may be simple. As the list of characteristics tends to reflect, it is the analytic process and method which is complex.

These two lists cannot be used to assess estimates' accuracy or logic. They are only guidelines to evaluate the appropriateness of an estimate given the subject of the estimate. In general, good estimates of complex phenomena will tend to have many of the characteristics found under "complex." Characteristics listed under "noncomplex" are those Doerner observed when individuals were unsuccessful in coping with complex problems.

If an estimate of something complex does not have many of these "complex" characteristics, it may still be accurate, logical, and appropriate. For example, scientists often employ an analytic process much like that labeled "complex" here. This may lead to discovery of a very strong, lawful relationship. In reporting a scientific law, scientists traditionally resort to a style whose character is somewhat like what we have labeled "noncomplex." Because the scientist has succeeded in turning a highly complex subject into a simple law, the noncomplex report is accurate, logical, and appropriate (but perhaps misleading as to the real requirements for that discovery). If the newly-discovered law is challenged, the scientist may resort to a full "complex" report on its discovery, detailing the analytic process.

Although these characteristics are only loosely diagnostic of whether an estimate is sufficiently complex for its subject matter, taken together they offer an additional useful assessment test for estimates. If a large number of the "noncomplex" characteristics are noted in an estimate, it may imply that the estimator, like Doerner's decision-makers, has not adjusted his or her thinking to the complexity of the problem.

THE LOGIC OF ANALYTICAL PROCESSES

While complexity may be diagnostic of analytic appropriateness, it is a poor index of analytic predictive success. Probably the best indicator of predictive success is the analyst's track record; i.e., has this analyst (or this analytic technique) successfully predicted similar events in the past? Such a track record can only be meaningful for precise predictions, and since most naval analysts make forecasts rather than predictions, such a track record scoring exercise is largely impossible or useless.

For example, in our sample of naval estimates we found some

forecasts and predictions about Soviet behavior vis-a-vis Afghanistan which turned out quite wrong, e g.,

> . . . the USSR has never been in a position of overwhelming military strength vis-a-vis the rest of the world; and this point alone induced caution. In other words, although the Soviet Armed Forces are obviously capable of winning a swift and convincing victory over, for example, Afghanistan, the repercussions in the rest of the world would be such as to more than counteract the kinds of gain that, in the foreseeable future, the USSR might win from such an invasion (Vigor, SND: 19).

Without a means of scoring an analyst's overall successes, as well as failures, the existence of a stray miss, or a stray hit, serves no purpose in assessing the analyst's estimation methods. In short, the best means of assessing the methods of naval analysts, i.e., examination of their prediction track records, is unavailable because the forecasts and predictions of naval analysis of Soviet intentions are far too imprecise and unspecific to be scored meaningfully or consistently.

How then can we assess these analytic methods? Although the procedure adopted in this study is inferior to assessing prediction track records, it may be the best possible given the nature of current estimates. This study dissects the intention estimation process into distinct elements; the separate mental and logical actions typically taken in the course of an estimate. Each of these elements is then evaluated from the perspective of cognitive psychology, logic, and information processing. That is, we ask how well do people generally perform such mental and logical steps? Are there widespread weaknesses or biases in information processing at that stage in estimation? Can we demonstrate that naval analysts actually experience these problems? In other words, we chop up the estimation effort into its necessary pieces and assess the likely weaknesses of each piece. Problems that are invisible from the perspective of the entire estimation process become visible in these individual pieces.

Having found that various pieces of the estimation process may be weak, we cannot argue that any particular estimate is weak. However, the possibility of weakness should motivate analysts and the consumers of estimates to look very closely at the methodological components of estimates, not at just the estimate's contents. If such examination reveals a bias actually present, this suggests at least some greater measure of uncertainty regarding the estimate's conclusions. It may imply the conclusion is invalid. Thus, while we cannot in this general study pass judgments on individual estimates, we do outline a procedure for assessing intention estimation methods, and thereby assess the logic of their conclusions. Illogical methods can produce valid predictions, but only by chance. Valid methods stand a better chance in the long run of successfully predicting future actions.

The main result of this study is to demonstrate that analysts and consumers of estimates can evaluate analytic methods objectively, i.e., independently of the contents of the estimates, by focusing on the individual logical steps in the estimation process. There is, however, a much quicker means of assessing analytical predictive accuracy for a group

of estimates. That is, if several analysts make forecasts or predictions about an event that differ significantly from each other, they cannot all be correct, and it may turn out that they are all wrong. But the fact that the predictions or forecasts differ significantly implies some (maybe all) of the analysts are going to be wrong, and that the group of estimates probably contains considerable error.

In the next chapter, several case studies offer "mini-reviews" of the estimates by our sampled naval analysts on significant questions regarding the future peacetime intentions of the Soviet Navy. Wide differences of opinion exist on these issues, implying some of these estimates will turn out incorrect in some particular. These cases serve to demonstrate this range of difference (and thus the inherent range of error), as well as introducing the nonnaval analyst to some major controversies in this field.

In Chapter 4, the estimation process is dissected and evaluated. Examples of information processing biases and methodological weaknesses are taken from the sampled naval estimates to demonstrate some problems in methodology and logic. Estimation is viewed from two perspectives, as a process of judgment and decision, and as a process of historical reconstruction and narration. In both cases the procedure is the same, to separate the process into steps and evaluate the strengths and weaknesses of these separate pieces. The main characteristic of estimates with which we concern ourselves is the logic of the methods. First, however, we must look at the content of some estimates.

CHAPTER 3. CONFLICTS IN ESTIMATES

> I cannot forecast to you the action of Russia. It is a riddle wrapped in a mystery inside an enigma; but perhaps there is a key. That key is Russian national interest.
>
> Winston Churchill, 1 October 1939

> There's a lot of wish fulfillment involved in our view of Soviet political intentions.
>
> Richard Helms, 1978

One means of appraising estimates is to assess the dispersion of (or differences between) estimates regarding a particular event. For example, suppose different predictions of the number of Soviet submarines of a certain type in the year 1990 varied from eighty to one hundred. The minimum error in this set of predictions is 10 which occurs if the 1990 figure is exactly 90. That is, we know now that our set of estimates in 1990 will have an error factor of more than ten percent, with the possibility that the error may be much greater. If the 1990 figure was 70, or 110, for example, the present estimates are in error by 30 units. If the dispersion of current estimates were made greater, say 75 to 105, the minimum error increases from 10 to 15 units. The dispersion of a present set of estimates thus provides an index of the minimum error (or maximum accuracy) possible in the future, but gives no indication of the maximum error that is possible. The wider the dispersion of current estimates, the greater the future error.

Estimates of Soviet naval events rely on explanations of past events, interpretations of current operations, and expectations regarding future operations. In this chapter we examine some differences between estimates of various Soviet naval events in each of these periods. The differences between (or dispersion of) these estimates provides a rough measure of the minimum error in these examples. Since each case of differing estimates was arbitrarily sampled from a much larger set, and no effort was made to select those estimates which differ most, the range of dispersion, and hence the minimum error, would increase as more extreme estimates were added to the current set.

As this selection of different estimates will indicate, in each of the three periods, and for each issue, the dispersion of the sampled estimates is considerable, but owing to the universal lack of specificity in all the estimates examined, it is possible to reach only a qualitative estimate of the minimum error in each case. In most cases, however, even these qualitative assessments reflect what seem to be serious minimum errors. It is beyond the competence of this study to judge how serious.

Finally, the sampled cases are by no means the only issues for which there seem to be large differences between estimators. Were issues regarding wartime Soviet naval intentions included in this study, the number of cases of serious differences between estimates might easily double. Even in the realm of these estimates of peacetime intentions, there are few issues for which the range of estimates is not considerable.

To summarize, just as differences between estimates betoken differences in data bases, and methodologies, they also reflect minimum expected error of estimation. The consumer of estimates can reduce minimum expected error only by narrowing the range of estimates; i.e., by disregarding extreme estimates. The criteria for selecting estimates to disregard should include the potential weaknesses in the estimation methods. This analysis of estimation errors and fallacies allows the consumer of estimates to judge which estimates rely on vulnerable methods, and thus to disregard those extreme estimates which are most probably weakest. The estimate consumer can thus reduce the dispersion of estimates in a sensible manner and thereby reduce minimum error.

DIFFERENCES BETWEEN ESTIMATES — PAST SOVIET NAVAL EVENTS

Case 1: Soviet Attitudes Toward Surface Ships, 1953-1961

Estimators differ regarding Soviet attitudes towards the surface fleet in the decade following Stalin's death. On the one hand, the statements of party leaders, the building program, and the abolition of the naval ministry and change of naval commander are used as evidence of a shift from Stalin's vision of a well-balanced, ocean-going fleet of large surface vessels (including aircraft carriers) to a fleet based heavily on submarines, small ships and land-based aircraft (e.g., McGwire, SND: 189-92, 203). On the other hand, the statements of Admiral Gorshkov, the deployment of naval units in forward areas (e.g., the Mediterranean), and Gorshkov's "salvaging" of the Soviet cruiser program are used as evidence that Soviet plans for a balanced fleet of surface vessels for forward deployment can be traced back to 1954 (e.g., Herzog, MPSP: 39; Herrick, 1968: 71-2; Jameson, STS: 41; Dismukes and McConnell, 1979: 14).

This difference is neither superficial nor merely historical. The former view suggests that the structure of the Soviet fleet was determined in large part by economic stringencies and the mission requirement to keep Western aircraft carriers and amphibious task forces from Soviet coasts. In other words, the major causes of Soviet decisions were internal economics and defensive needs in reaction to Western threats. The latter view implies execution of a long-term, well-coordinated plan for forward deployment laid more than a decade prior to realization, and followed consistently after Stalin's death. Obviously, whether treated as trends, or as theoretical models of relationships, these two viewpoints offer widely different explanations of the course and determinants of Soviet surface fleet developments.

After Stalin's death in March 1953, the naval ministry was abolished, and Stalin's vision of a big, balanced fleet was abridged to a plan to build submarines, land based aircraft, and light, fast surface forces (Herrick, 1968: 62-66). Admiral Kuznetsov, who shared Stalin's outlook, was replaced in 1955 by Admiral Gorshkov, who was noted for his interest in naval missile technology and his World War II leadership of Black Sea naval operations in support of the Red Army. Herrick (p. 71) credits Gorshkov with saving the surface fleet, including the heavy cruisers, from Khrushchev's and the 20th Party Congress's moves to cut these construction plans drastically. Herrick (p. 71) writes:

> To prevail on Khrushchev not to carry out his publicly-announced intentions of scrapping 90 percent of the Soviet Navy's cruisers, Admiral Gorshkov conducted a well-conceived and adroit but cautious, step-by-step campaign over a period of more than two years. Gorshkov s goal, politically speaking, was an ambitious and audacious one -- not just to complete and retain a large percentage of the cruisers but to gain practical, if not theoretical, acceptance of the continuing importance of large warships in the nuclear era.

McConnell (in Dismukes and McConnell, 1979: p. 2) argues that "naval diplomacy had no secure home in Stalinist Russia" (a judgment Herrick, 1968, ch. 4 and 6, seems not to share) and that naval diplomacy bears the marks of long-term Soviet commitment commencing with Khrushchev and Gorshkov. He writes (p. 10)

> . . . by the mid-1950s, Moscow appreciated the need for coercive naval diplomacy in the Third World, that it perceived its current capabilities to be inadequate for this role, and had already adopted a long term construction program that would create credible capabilities.

In support of this judgment McConnell cites Gorshkov's 1967 retrospective commentary that the decision for a diplomacy of naval force was made in 1954 and aimed at creating a balanced, ocean-going fleet capable of nuclear and non-nuclear wars, and protecting state interests abroad in peacetime. McConnell then observes

> Gorshkov's claims of a consciously planned development, of course, may have been a rationalization after the fact to show the party's prescience and control over events.

McConnell nevertheless rejects this possibility, instead arguing that by 1955 all the ingredients, except coercive naval diplomacy, of present Soviet Third World policy were in action. (This judgment seems inconsistent with several lines of evidence, such as the Soviet naval construction program in the 1950s and the Soviet's relinquishing in 1955 of its only two naval bases abroad.)

Immediately after mustering evidence in support of Gorshkov's claim that 1954 marked the decision for a patient and restrained move toward a navy of open-ocean surface vessels, McConnell turns (p. 15) to the 1954 decision, revealed by "Khrushchev himself," to "shift from surface ships as the main arm of the fleet to a navy based primarily on submarines and secondarily on land-based aviation and light surface ships." McConnell explains (p. 18) the discrepancy between Gorshkov's claims and Khrushchev's decision as due to the Soviet's having "aimed developmentally at a Third World diplomacy of force [over the decade 1955-65], [but] they were probably not yet operationally committed." As late as 1960-65, he writes (p. 20), "there was still no decision in favor of a diplomacy of naval force" and only in 1965, with the Party adoption of a local war doctrine, and the 1967 involvement of the Soviet Navy in the Six-Day War crisis, does McConnell finally see positive evidence of "a

real change of policy" built on the 1954 decision to "create credible capabilities," such as surface vessels, for coercive naval diplomacy.

Khrushchev's frequent, strong deprecations of surface ships, in addition to the 1954 plan to cut their construction, are obviously embarrassments for McConnell's (and Gorshkov's) thesis that the 1954 plan envisaged balanced, ocean-going surface vessels. McConnell asserts (p 13) that "in general, too much has been made of Khrushchev's deprecation of surface ships." To minimize this problem, McConnell argues first that the surface ships Khrushchev criticized were "gun cruisers and destroyers," not the surface ships now used in Soviet anticarrier task groups, which rely on surface-to-surface and surface-to-air missiles, and which, by implication, Khrushchev would not find so objectionable. Second, he argues that the modern surface vessels were "the result of a production decision made under either Khrushchev or his predecessors." Herzog (NPSP: 39) echoes this assessment.

It is hard to reconcile McConnell's judgment that Khrushchev had a well-hidden tolerance for large surface ships with either his statements or Soviet naval construction under him. McConnell's first argument implies that the anticarrier task forces used by the Soviets in crises during their coercive naval diplomacy period (circa 1967 to the present according to Dismukes and McConnell, 1979) "typically" do not include the gun cruisers and destroyers Khrushchev saw as outmoded. This is not the case, however; gun cruisers and destroyers are often components of Soviet anti-carrier warfare (ACW) task groups (Dismukes and McConnell, 1979: 173-5, 179, 211-12). The fact that ships, which even McConnell seems to grant were labeled by Khrushchev as obsolete, are still to be found in service alongside more modern vessels, hardly seems compelling evidence that Khrushchev's frequent caustic remarks on surface ships were exclusively deceptive propaganda. If the core of Moscow's Third World diplomacy of force lies in the Soviet's anticarrier task groups, as McConnell asserts (p. 21), the continued presence of gun vessels can only be an embarrassment, given Khrushchev's statements; an embarrassment the Soviets seem to have minimized by retiring Khrushchev more quickly than they have retired the gun cruisers and destroyers he deprecated.

McGwire's (SND: 203) assessment of Soviet construction suggests that very large cutbacks in surface ship building were made as a consequence of decisions in 1954-55 and 1958-59, which he interprets as major steps away from the goal of a well-balanced fleet with world-wide capabilities -- the goal Gorshkov in 1967 (and McConnell later) assert was the basis of the 1954 decision. McGwire argues that not until 1961 were construction shifts made that reflect a move toward forward deployment (SND: 189-92), and a concerted build-up of surface fleet units. This shift he links to the deployment by the United States of the A-3 attack aircraft on aircraft carriers and the Polaris submarine and missile combination. These two systems gave the U.S. Navy the capability of attacking the USSR from over 1200 miles at sea, and this "drew" the Soviet Navy out to a forward defense.

McGwire interprets these surface ship building patterns as indicating little Soviet interest in surface vessels until the defensive mission required new surface capabilities to deal with the Western threat. McGwire has noted (SND: 139-41) that one of the major consequences on ship building of the 1954 decision was to shift shipbuilding capacity from naval surface capability to the construction of merchant, fishing, and

other civilian-oriented shipping. He characterizes (SND: 190-1) the 1954 decision as follows:

> What is perhaps less well known, and is certainly not made clear by Gorshkov's [1967] article, is that the 1954 decisions . . . involved wholesale cancellations of building programs and a new concept of operations that tied the fleet even more closely to the range of shore-based air support . . . To the extent that the Soviet Union can claim to have a "balanced fleet," this stems from . . . about 1961 . . . These 1954 decisions . . . had the effect of taking the Soviet Navy several steps further away from being a well-balanced fleet with a world-wide capability.

Even the 1961 decisions, which McGwire sees as directly related to the U.S. Navy carrier and Polaris threats, he labels (SND: 191) as "exploratory and making do with what was available, rather than the culmination of some long-planned and well-prepared policy." The evidence of Soviet shipbuilding, McGwire writes (SND: 194), "argues against the contention that between 1956 and 1962 Gorshkov fought a skillful and partially successful rearguard action to save the surface ship for the Soviet Navy."

Case 2. Attitudes Toward Forward Naval Deployments in the 1950s

As was discussed in Case 1, McConnell (in Dismukes and McConnell, 1979) accepts Gorshkov's claim that the policy of protection of state interests was formulated in 1954. McConnell argues that this decision formed the basis for the gradual development of a capability for coercive naval diplomacy by 1967.

McConnell links (p. 28) the forward deployment of Soviet naval forces in the 1960s to an overall pattern of "diplomacy of force" which includes trade, economic and military aid, cooperative naval diplomacy as well as coercive naval diplomacy. He argues (p. 3) that diplomacy of force plays a greater role in Soviet than in Western calculations. McConnell describes (p. 28) the intentions behind this diplomacy as

> . . . the credible intention is not to prevent U.S. intervention [in the Third World] entirely -- Moscow would not have the strength of will for that -- but to confine U.S. intervention to defensive ends. The aim is not to deter, but to limit -- a limiting show of force.

Herzog (NPSP: 39) also suggests that exerting influence in areas of unrest through naval power was a Soviet goal of the mid-1950s, which was hindered by inadequate surface ships. The decision to deploy forward based, open-ocean naval forces he attributed to a Soviet desire to influence events in the Third World through seapower. Herzog (p. 39) dates the formation of this objective as no later than the 1956 Suez crisis and characterizes it as follows:

> The Soviet Union undoubtedly saw in this unrest the opportunity to extend its influence southward and perhaps

> eventually realize a dream that has been held by Russians since Peter I and Catherine II, the extension of Russian control over the Near East and the acquisition of warm-water ports.

This is not to say that the strategic function of these forward deployments is neglected. McConnell, for example, writes (p. 28):

> To be sure, the focal point of the Soviet naval presence is within the arc of strategic concern often drawn at Polaris or carrier-air range around the USSR, but it is also on the water's edge of a charmed political arc; in crisis after crisis in the Middle East and on the Indian subcontinent, Moscow has reiterated that it cannot be "indifferent" ...

McGwire (SNP: 506, SND: 191) traces the Soviet policy of forward naval deployments to the evolving Western strategic threats of the late 1950s and early 1960s -- in the naval case carrier air threat and Polaris. On the one hand, to counter these threats the Soviet fleet would have to venture out of home waters in strategic defense. On the other hand, the success of Polaris highlighted the potential of a Soviet SSBN force, and the Yankee program was set in train. The short-term requirements, to which McGwire sees (SND: 508-510) the 1958-59 and 1960-61 decisions responding, were the need for counterforce strikes against carriers and U.S. SSBNs before these units could complete launching their strategic strikes. One estimator, who is otherwise an advocate of a coercive naval diplomacy interpretation of Soviet forward deployment, writes (Weinland, SNP: 380)

> The Soviets appear to have adopted a "forward deployment" posture for . . . strategic defensive purposes -- to be in position to "counter" U.S. and NATO sea-based strategic offensive capabilities . . . This probably remains the principle raison d'etre of the deployments. Since the late 1960s, however, deployed forces have been utilized in an additional capacity; as active instruments of Soviet foreign policy, protecting and promoting Soviet overseas interests.

Case 3: Soviet Naval Objectives in Egypt

Assessments of Soviet tactics in the Middle East range from "increasingly bold" to "cautious" to "prudent." Evaluations of its behavior vary from "running higher risks" to "low risk" to "circumspect." Estimates of Soviet goals range from "alarming strategic outreach" to "a relatively narrow agenda." Since these extremes related to past Soviet behavior (for which historical data are available) and to present and future events (for which more hypothetical bases are required for estimating), it is not unreasonable to suggest that the range of minimum error may be large in estimates of Soviet naval behavior in the Middle East.

The differences among analysts' explanations of Soviet objectives in Egypt offer an example of the range in estimates for a case in the past.

Rubinstein (SNP: 156) suggests that the returns the Soviets received for two decades of economic and military aid invested in Egypt between 1955 and 1975 were largely strategic: naval facilities and air bases useful in countering the U.S. Sixth Fleet. Some years later, Rubinstein (1980) concluded that Moscow had broadened its military probes beyond its essential security belt (which might include the eastern Mediterranean) and now demonstrates "alarming strategic outreach" (p. 323). Rubinstein in 1975 assessed Egypt as having considerably more influence over the USSR than vice versa, and judged Soviet efforts as aimed at obtaining needed strategic support facilities in the eastern Mediterranean. In 1980 Rubinstein lists five Soviet objectives in the Middle East: 1) undermining U.S. alliances, 2) weakening Western influence and position, 3) expand USSR influence and position, 4) obtain military privileges in order to support projection of Soviet power on behalf of friendly states and in pursuit of strategic advantages, and 5) thwart a regional Pax Americana. In his 1980 article Rubinstein notes the "stunning strategic setback" Moscow experienced in Egypt but against this he observes (p. 334) the changes since October 1973 in Soviet behavior: "an increasingly bold and confident forward policy . . . [and] ability and readiness to project power into areas of opportunity." He perceives Moscow as pursuing a "more venturesome policy" and "running higher risks for regional gain." He concludes that Moscow's "military prowess has encouraged diplomatic assertiveness;" an assertiveness Rubinstein sees (p. 335) as increasingly unchecked by the West:

> In no crisis in the Middle East since 1967 has the Soviet Union deferred to American preferences. Increasingly Moscow determines the course and outcome of regional crises.

[It is difficult to see the Soviet reaction to the U.S. strategic alert in October 1973 as anything but deference to American preferences. It is also difficult to detect Moscow's hand in the Iran-Iraq and Syrian-Jordanian, let alone the continuing Arab-Israeli crises, or to detect the Soviets determining the courses and outcomes of Middle Eastern events, but perhaps it has kept its hand well hidden.]

These two analyses by Rubinstein, separated by five years, clearly reflect an image of Soviet diplomacy and military policy in dramatic flux. Earlier Soviet actions were far more cautious and directly tied to strategic military requirements. This gave the Soviets little flexibility or leverage with clients such as Egypt. Now, Rubinstein argues, the Soviet military requirements are largely satisfied, and the Soviets can use their many tools of influence (aid, trade, naval diplomacy, etc.) for their main geopolitical purpose, i.e., eliminating any American role in Middle Eastern security affairs, and enhancing the Soviet role.

In marked contrast to Rubinstein's estimate of changing Soviet objectives in Egypt and the Near East, Ginsburg (SNI) offers a view much more pessimistic for the Soviets. Ginsburg suggests that the Soviets seek influence in order to convert it into "material gains": military installations, service facilities, or bases. Because of a variety of limits on their behavior and success, Ginsburg believes the Soviets "will, now, and in the foreseeable future, pursue a relatively narrow agenda and direct their attention to a few choice locations instead of spreading

themselves thinly" (p. 446). He argues that the Soviets will "gravitate toward low risk projects" and adhere to a cautious search for military installations. Ginsburg sees (p. 458) Third World states turning to the Soviets if they are rebuffed by others and if the Soviets can provide needed services, but any Soviet "success" today may be easily undone tomorrow. In other words, the Soviet-Egyptian and Soviet-Somalian episodes may easily reoccur.

Both the Ginsburg chapter and Rubinstein's 1975 (SNP) chapter analyzed the interaction between the Soviet Union and client states. Both examined the various limitations that clients are able to place on Soviet action, and the degree to which such limits circumscribe Soviet flexibility and influence. The analyses highlighted Soviet dependence upon clients for strategic bases -- bases in Egypt (to replace those lost in 1961 in Albania) to meet the strategic threat in the Mediterranean, and in the Indian Ocean to forestall a strategic threat from that area. Rubinstein's later (1980) paper implies that the Soviets are far less dependent on foreign basing. He seems to imply that improved Soviet military capabilities, by reducing Soviet reliance on foreign bases, have reduced the degree to which client states can hold Soviet strategic capability (represented in foreign bases) ransom.

Dragnich (SNP, ch. 13) focuses on the Soviet's efforts to secure naval facilities in Egypt, until recently the centerpiece of Soviet influence in the Middle East. He notes the relation between, on the one hand, Soviet naval visits, diplomacy, and strategic needs, and, on the other hand, Egyptian interests. Up until 1967 the Soviet strategic need was growing for air and naval bases in the eastern Mediterranean but Soviet efforts to acquire them were completely unsuccessful. Only Egypt's disastrous loss in the 1967 war, and the destruction of Egypt's armed forces opened the way for a quid pro quo with the Soviet Union: Soviet arms for Egyptian bases.

Dragnich rejects the hypothesis that political and diplomatic influence-seeking motivated Soviet aid before and after the 1967 war. Rather, defensive strategic deployments against Mediterrannean Polaris and carrier-borne nuclear weapons were "sufficient" (p. 268) to justify a swap of aid for bases. Forward basing of naval defenses against the U.S. strategic deployments offered the Soviets, according to Dragnich (p. 269), "a make-shift alternative to a larger and more capable navy."

Ra'anan (SNP, ch. 11) offers a novel account of Soviet decision-making in the Middle East. Ra'anan argues that Soviet military personnel do not come and go at the demand of client states, and Soviet presence in client countries is a consequence of Soviet decisions. With respect to Soviet bases in Egypt, Ra'anan rejects (p. 187) any notion that Egyptian initiatives governed Soviet presence. In particular, he reasons that the Soviets foresaw an Egyptian defeat at Israel's hands in the 1970's and initiated withdrawal to avoid a second embarrassment. The switch in Soviet policy from pro-intervention in 1970 to withdrawal in 1972 Ra'anan attributed (p. 199) to waxing and waning influence of two opposed Kremlin factions. He gives Egyptian actors virtually no role in these events. Ra'anan outlines (p. 205-6) a Soviet "plan" for the 1973 Middle East War, which he suggests was formulated by Moscow in 1972 and carried out by Egypt and Syria.

Ra'anan acknowledges (p. 26) that his "reconstruction of developments" is "speculative." One commentator labeled his analysis "one

of the best short stories of the year" (Kerr, 1975: 99). Ra'anan disregards the role played by the Egyptians in, first, seeking Soviet aid and advisors, and second, expelling the Soviets in 1972. He also attributes remarkable foresight to Soviet leaders. For example, "Brezhnev's plan," according to Ra'anan (p. 205), included foreseeing in 1972 all of the following: the coincidence of the Arab attack in October 1973 and a "domestic or other crisis" which would distract the United States; the Israeli false alerts of 1973; the Egyptian-Syrian surprise; Israeli counterattack successes; U.S. reactions to Soviet moves toward intervention; Egyptian territorial gains in the Suez; and American post-war diplomatic moves. It is doubtful that this clarity of foresight existed in the Kremlin or for Kremlin scholars. These elements of "Brezhnev's scenario" are barely plausible even in hindsight.

Kerr (1975) disagrees with Ra'anan's analysis of the Soviets' influence on Egypt. Kerr notes that, like the United States, the Soviets have difficulties trying to translate aid into influence. The Egyptian conflict with Israel gave the Soviets a powerful tool for influence in the form of military aid and assistance. In the wake of the 1973 war Egypt turned to the United States for diplomatic aid which Moscow could not deliver. Kerr says virtually nothing about the role of strategic defense or Mediterranean bases in shaping Soviet policy toward Egypt. The primary Soviet goal in Egypt, according to Kerr, was to maintain their presence and influence in the region.

Kerr rejects Ra'anan's thesis that the Soviets foresaw the events and outcomes of the October 1973 war and benefited from the war's results. In contrast, Kerr argues (p. 107) "it is clear that the Soviets badly miscalculated in supposing they stood to gain anything from it . . . their client's relative success redounded not to their benefit but to that of the United States." Since war with Israel stood as the primary basis for Egypt's relationship with the USSR, Sadat's shift from war footing to peace offensive in the wake of the October War undid the Soviets' long-term effort to prevent an American-backed status quo.

Freedman's (SNP, ch. 12) analysis of the Soviet Union and Egypt addresses the role of both Egypt and the United States in altering the influence and presence of the Soviets. Freedman assesses (p. 215) Soviet reluctance to aid Sadat before the 1973 war as due to unwillingness to "let themselves be further exploited" or to be dragged into confrontation with the United States. Soviet preferences were to maintain its presence in Egypt but to prevent its Egyptian hosts from seeking a military "solution" to the Arab-Israeli problem. The "no-war, no-peace" situation which the Egyptians found so repellent was responsible for the Soviets' military presence and naval bases in Egypt. It was a situation very favorable for the Soviets as well as Israel and the United States, but not Egypt. The Soviet expulsion by the Egyptians was a heavy strategic loss to the Soviets, according to Freedman (p. 218). Freedman concludes (p. 230) that "All in all, the Soviet relationship with Sadat's Egypt has not been a pleasant or profitable one for the Soviet leadership." Like Rubinstein (SNP: 179), Freedman observes (p. 230) that Egypt influenced and exploited the Soviets more than the reverse.

Weinland (NPSP, ch. 15) argues that "there is little doubt" that the Soviets' immediate objective in the Mediterranean was to improve its military capabilities in that region. But beyond that, Weinland writes (p. 260)

Why they wanted that force in the Mediterranean, and what they planned to do with it, cannot be specified with any certainty.

Weinland prefers the explanation (SND: 292-305) that the Soviet forces in the area were needed for strategic defense of the USSR against seaborne NATO nuclear weapons, and could also serve secondarily to protect Soviet interests in the region.

With respect to the Soviet Navy and Egypt, Weinland (NPSP: 267 ff) offers what amounts to the converse of Ra'anan's thesis: a description of Sadat's long-term campaign to improve Egypt's strategic and geopolitical power in the region. Rather than a "Brezhnev Plan," as Ra'anan offers, Weinland outlines the Sadat Plan for manipulating the Soviet presence to achieve Egyptian goals. Weinland describes Sadat as using 'carrot and stick' tactics on the Soviets to obtain the military aid and independence from Soviet diplomatic restraints that he needed to reopen the Arab-Israeli war. For example, the offer in December 1972 to renew the 1968 naval support agreement, which allowed the Soviet Navy extensive access to Egyptian support and repair facilities was a "carrot." The ejection of Soviet advisors and air defense personnel in July 1972 was a demonstration of the "stick."

In Weinland's account, the Soviets' Mediterranean squadron was a Soviet vulnerability with respect to Egypt-Soviet relations. Its reliance on shore support enabled Sadat to "hold it hostage" and threaten to reduce or eliminate that support if the USSR failed to meet Egyptian demands. In terms of making points with the Egyptians by using the squadron to protect Egypt from the intervention of the U.S. Sixth Fleet, Weinland suggests (p. 269) the effort failed: "Egypt was dissatisfied with Soviet performance during the war."

There is little question that after 1973 Soviet influence in Egypt declined precipitously. Weinland's (NPSP) analysis shows that the loss of Egyptian shore support facilities had a dramatic impact on the Soviet Mediterranean squadron: the squadron's strength declined, transits of the Turkish straits declined (although transits of naval auxiliaries remained constant), ship day totals leveled off and then declined, Soviet naval port calls increased to possible alternatives to Egypt (Libya, Syria, Yugoslavia).

At least with respect to the Mediterranean littoral states of the Middle East, Weinland's analysis suggests that Soviet defensive missions led to the need for influence ashore because the fleet needed support ashore. When this support was removed (as in Albania and Egypt), one important component of Soviet influence, the Soviets' Mediterranean naval presence, contracted markedly. Why the Egyptians removed that support involved far more than the actions of the Soviet Squadron, but Weinland seems to argue that the Squadron's presence did little to facilitate Soviet influence in the first place, offered a point of vulnerability when that influence was weakened, and was a significant early victim when the influence was lost.

Using the case of Soviet-Egyptian relations to draw inferences about Soviet-Middle Eastern affairs, suggest several conclusions about estimates:

1. Estimators who emphasize the influence of Egypt on the Soviets

(Dragnich; Rubinstein, 1975; Freedman; Kerr) all note the Soviets defensive strategic needs for Mediterranean naval and air bases as major motivations for Soviet efforts at influence. Those estimators who emphasize only the Soviet influence (Ra'anan, Rubinstein, 1980) find Soviet naval needs only a minor factor in either involving the Soviets with the Egyptians or shaping Soviet influence. These latter estimators emphasize the Soviet objective of acquiring influence in the Arab world.

2. The assessment of whether the Soviets exercise influence on client states through their naval presence, or at all, varies with the degree to which the estimator focuses on the influence of the client state: the greater the focus on the actions and goals of the client state, the less the perceived influence of the Soviets or their navy; the greater the focus on the Soviets' goals and efforts, the greater the influence role perceived of the Soviet Navy and other diplomatic or military instruments. That is, influence follows a simple hydraulic relationship; if one party has more, the other has less. All estimators seem to view influence as somehow constant and variously divided.

3. Estimators who focus on the Soviets tend to perceive the Soviet Navy as having posed an effective counter to the U.S. Navy in the 1973 Middle East war. Estimators who focus on the client states (especially Egypt) raise the possibility that clients saw the Soviet Navy as having only limited influence on the 1973 crisis, and as playing a minor role in the crisis despite its dramatically expanded presence in the Mediterranean.

Case 4. Interpretations of the Gorshkov Series

Few bodies of Soviet naval literature have generated as much disagreement as have Admiral Sergey Gorshkov's unique publications (see McConnell, SNI, ch. 29 for a summary of the publications and the debate). There seems to be agreement that writings with these details and scope by the serving Commander-in-Chief of a Soviet armed service are unprecedented, that they are extremely important for understanding both Gorshkov and the Soviet Navy, and that the debate over the meaning and significance of these unique writings shows no sign of abating. The differences among interpreters of the Gorshkov articles and book are extensive and can only be briefly suggested here: the analysis of the Gorshkov literature has every indication of becoming a major branch of U.S. naval intelligence on the Soviet Navy.

Commentators on the Gorshkov writings include Herrick (SND, ch. 23), Hibbitts (NPSP, ch. 1), Hudson (SND, ch. 1, 1976), Jameson (STS, ch. 2), McConnell (SNI, ch. 29), McGwire (SND, chs. 34,36; SNP, chps.28, 33; SNI, ch. 30), Thompson (NPSP, chp. 2), Vigor (SNP, chs. 25, 32), Weinland (SNP, ch. 29). Differences between commentators range from global issues, such as the authoritativeness of the Gorshkov series as Soviet naval doctrine (e.g., McGwire, SNI, ch. 30; McConnell, SNI, ch. 29); to interpretations of particular phrases and expressions (e.g., Vigor, SNP, ch. 32; McGwire, SNP, ch. 33).

Naval analysts do not even seem agreed on as simple and straightforward an issue as the frequency of Gorshkov's publications (e.g., "Admiral Gorshkov publishes infrequently," Weinland, SNP: 547. "Gorshkov has published one or more articles in a Soviet military journal almost every year since at least 1963," Hibbitts, NPSP: 1. Jameson, STS:

31, characterizes Gorshkov's writings as "voluminous").

Several commentators assert the hypothesis that the Gorshkov book, The Sea Power of the State, reflects Soviet naval doctrine (Hibbitts, Jameson, McConnell, Vigor) while others interpret Gorshkov as advocating a new doctrine (Herrick, Hudson, McGwire, Thompson, Weinland). Some commentators see parts of Gorshkov's writings as doctrinal and other parts as advocacy (e.g., Hibbitts, NPSP: 4, 21). The latter analysts suggest Gorshkov is espousing his side of an ongoing debate within the highest Soviet military and political councils on the mission and role of the Soviet Navy. The former analysts suggest the debate has been resolved, Gorshkov won his points, and his series is an authoritative "concrete expression of doctrine" (McConnell, SNI: 566).

While both sides on this issue use an internal analysis of Gorshkov's writings, the authoritative doctrine school emphasizes his use of "doctrinal authenticators": key words that tend to be associated in Soviet writings with military policy and expressions of military doctrine (McConnell, SNI: 566, 604-612). For example, McConnell (p. 605) emphasizes the importance of the words "unity of views" in the editors' introduction to the Gorshkov series of articles, and that "this expression is a telltale indicator of military doctrine (or military policy) . . .", Hibbitts (NPSP: 4) concurs in this interpretation of "unity of views." McConnell notes several of these telltale indicators at critical points (e.g., introductions of Gorshkov's writings).

Analysts supporting the view of Gorshkov as advocate or lobbyist for naval interests rather than as doctrinal spokesman tend to point to the context of Gorshkov's remarks as against the content, frequently noting the discrepancies between what Gorshkov obviously wants (e.g., "balanced fleet") and what he has been able to get in his 25 years as head of the Soviet Navy.

Since all readers of Gorshkov agree that he strongly supports a balanced fleet, and most agree that the Soviet Navy has not been balanced in the past, the different interpretations of Gorshkov have different implications for Soviet shipbuilding. If Gorshkov is a spokesman for doctrine, the Soviet fleet should be tending toward a balanced fleet. If Gorshkov is merely an advocate, the Soviet fleet may remain unbalanced into the future. Or one might agree with Gallagher (SNP: 56) that military doctrine, while authoritative, is highly generalized and ambiguous and a poor key to Soviet strategic policy. In this case, whether Gorshkov is doctrinal spokesman or advocate has little to do with Soviet naval decisions.

The interpretation of Gorshkov's views on specific concepts may have important implications for estimating Soviet naval peacetime intentions. For example, Gorshkov's concept of "command of the sea" has been viewed in narrow terms by McConnell (SNI: 599-601) as "creating a favorable operational regime" for wartime strategic naval offensive and defense (e.g., SLBM and anti-ASW operations). Vigor (SNP: 605) gives the expression broad meaning: complete removal of the enemy's fleet from the naval theaters of war. Vigor specifically interprets Gorshkov as using this expression to mean the Soviet Navy "must aim to acquire superiority of force over the enemy in the principle theater of operations and be strong enough in secondary areas to prevent him from interfering." Vigor sees this usage as "a valuable clue to some likely paths of future development of Soviet naval strategy."

McGwire (SNP: 631-635) rejects these interpretations of Gorshkov since the phrase is rarely used, used in a context that suggests Gorshkov is defending himself from charges of being a Mahanist (i.e., advocate of bourgeois doctrine), and irrelevant to the Soviet Navy, which historically has always been concerned with the problem of how to conduct naval operations without command of the sea. The concept of sea denial holds a more central role in both Soviet naval capabilities and doctrine as well as in Gorshkov, according to McGwire (p. 634). It has greater relevance to future Soviet naval strategy than sea command as well, because the sea denial strategy "relies on the use of overseas bases to discharge . . . war-related tasks in peacetime."

PRESENT SOVIET NAVAL EVENTS

Case 5. Readiness for and Purposes of Forward Deployment

Two related questions deal with Soviet Navy forward deployment: how ready were the Soviets to deploy forward when they did, and what purposes were served by forward deployment? The first is largely an historical issue while the second involves past as well as present events (i.e., the Soviet Navy is still deployed forward, and the question can be raised as to how the purposes of that deployment have shifted or remained constant).

The various explanations of why the Soviet Navy deployed forward also bear on whether the Soviets prepared for forward deployment (and were ready to do so), or whether they were unready for any of several reasons (unprepared and unwilling, prepared but unwilling, prepared and willing but unready through chance circumstances -- e.g., loss of bases, etc.). Griffiths (SND, ch. 1) categorizes these explanations into three schools. The first sees Soviet forward naval deployment as a bid for control of the seas and the result of an overall plan for global projection and intervention capability. This school holds an image of Soviet foreign policy behavior as motivated by militant territorial expansionism aimed at world rule.

The second school Griffiths perceives explains forward expansion as an emerging interest in a limited capability for projecting Soviet influence in the Third World, which developed along with the defensive forward deployment aimed at countering Western sea-based strategic strike forces. As the Soviet Navy emerged from home waters to attempt the latter task, the utility of the former task grew apparent and capabilities were first reoriented and then perhaps specifically developed to deal with the protection of Soviet state interests in the Third World. This goal of limited projection Griffiths associates with a motive of limited political (rather than territorial) expansionism -- "Moscow can be expected to advance its power and influence whenever given the opportunity; but when the opportunity is denied, the Soviets are likely to behave with prudence" (p. 13).

The third school Griffiths terms "protective reaction" -- Soviet forward deployment is seen as an essentially defensive response to the Western sea-based strategic forces. This school sees neither a Soviet interest in nor any considerable capability for an offensive political use of seapower. The Soviet motive is reluctant expansionism, the result of strategic insecurity and fear rather than ambitions for power and

influence. Western moves force the Soviets to take reluctant and cautious moves in self-defense.

Each of these schools is excessively narrow and simplistic, according to Griffith, and each is inadequate to explain Soviet forward deployment. Griffiths believes this is particularly so of the first school. In lieu of any one school, Griffiths suggests Soviet foreign policy (and naval forward deployment) results from the competition within the Soviet Union of "limited projection" and "protective reaction" interests. These "conflicting trends in Soviet behavior" (p. 14) as well as situational (i.e., non-Soviet) variables influence Soviet foreign policy and forward deployment.

The first school implies a readiness for forward deployment: as part of the overall plan for control, the Soviet Navy would be prepared for its role. The Navy would not move forward until ready because of the possibility that a lack of capability or will on the naval front might upset the overall plan. Inasmuch as this viewpoint envisages no pressures on the Soviets to deploy before they are ready, and since premature deployment might produce penalities for the overall plan, it is likely that this school would require that the Soviet Navy be fully ready for all major missions when it emerged on the high seas. Any reflections of major unreadiness to execute essential missions would be inconsistent with a Soviet overall plan for control of the seas. Inasmuch as the Soviet Navy could (perhaps) be optimally designed against the existing Western fleets now in being on the world seas, as U.S. Chief of Naval Operations Zumwalt once hypothesized, there would be little reason, given the first school argument, for the Soviet Navy to move forward before it was ready.

In neither the "limited projection" nor "protective reaction" schools is it necessary that the Soviet Navy be ready prior to forward deployment or even now. Both schools presuppose the requirement for forward deployment was imposed by the Western strategic threat to the USSR, and required some demonstration by the Soviets even if they were unready to fulfill all the necessary missions.

Dismukes and McConnell (1970, ch. 8) offer a variant of the first and second schools outlined by Griffiths. While Dismukes and McConnell do not see Soviet ambitions as extending to world sea control, they do attribute Soviet forward naval deployment to an overall plan aimed at coordinating Soviet foreign policy instruments (trade, aid, diplomacy, etc) for expanded Soviet influence in the Third World. This plan (or "development strategies for long-term goals" as Dismukes and McConnell characterize them), formulated between 1951 and 1955, laid down the guidelines for "basic continuities of policy," including naval construction and operations, that extend up through today. The decisions for construction of an "ocean-going" fleet were made in 1956 to provide the capabilities for coercive military and naval diplomacy. The initial target of this plan was to develop a counter to the U.S. attack aircraft carriers, which posed both a strategic threat as well as a counter to Soviet activism in the Third World.

Dismukes and McConnell seem to have in mind a hybrid of the first two schools Griffiths outlined. They see forward deployment as part of a long-term Soviet strategy (first school) rather than as an interest that emerged with experience (second school). On the other hand, they see the Soviets as interested in projection of power in the Third World (second school) rather than world-wide sea control (first school). Finally, they

see the Soviets as interested in expanding, defending, and preserving "state interests" abroad (second school) rather than aimed at territorial expansion and world rule (first school).

They also argue that the Soviets were quite unprepared for their first foray into coercive naval diplomacy, exercised on behalf of Syria in 1957 (Dismukes and McConnell: 7-10, 282). The Soviets quickly recognized their unreadiness, according to this account, and foresook any subsequent attempts at coercive naval diplomacy until the mid-1960s, when the surface ship capabilities dictated by the 1954 decisions emerged from the long-term construction programs.

Obviously, Dismukes and McConnell's case for a long-term plan for a coercive naval diplomacy in the Third World rests in part on the exact nature of the 1954 naval ship-building decisions. The differences of opinion on this issue are discussed above under Case 1. It is even more important to their argument, however, that the Soviet Navy be ready in the mid-1960s for forward deployment. Since this forward move is seen as part of the long-term plan, it would be inconsistent for the Soviets to make the mistake twice of deploying forward for coercive naval diplomacy without having the necessary capabilities. The first test of the Soviet Navy's readiness for this long-planned role came in June 1967 with the Six-Day War.

Were the Soviets ready in 1967 to counter the Sixth Fleet's task groups and prevent them from intervening in the Six-Day War (as the Soviets and their friends consistently claimed after this interpretation was raised in the West in the aftermath of the war)? Dismukes and McConnell (pp. 158-168) make the case that the Soviet deployments were significantly different in the crisis period than in the previous three years of Soviet Mediterranean operations in the following four ways: the size of the Mediterranean squadron increased dramatically (ship days and daily average strength in 1967 were nearly double the 1966 levels, Weinland, NPSP: 262); for the first time Soviet combatants (destroyers) were given the "tattletale" mission of observing and following U.S. attack carriers (this task had heretofore been the duty of slower, noncombatant intelligence auxiliaries); the Soviets mounted an air and sealift to the Arab states; and they indicated they were considering direct military intervention (probably with Airborne Troops) if the Arab capitals were threatened by Israeli ground forces.

Although this evidence makes an excellent case for a change in Soviet naval crisis behavior, it does not reflect a readiness to actually counter the U.S. Sixth Fleet. Several pieces of evidence suggest the Soviet Mediterranean units were not ready or capable of preventing intervention by the U.S. carriers.

Although Soviet ships elsewhere in the world were intensively harrassing U.S. Navy ships (e.g., in the Sea of Japan on May 10 and 11 1967) the Soviet ships avoided any actions of this type in the Mediterranean during the crisis. The combatants on tattletale duty tended to stay beyond their weapons' ranges from the carriers, and only once did a Soviet destroyer take up a parallel (i.e., firing course) near a carrier. Soviet naval contact with Western naval forces was limited almost entirely to these prudent and circumspect tattletale operations. Most of the Soviet combatants remained anchored well away from the Western carriers or the Arab-Israeli fighting. The Soviet squadron in the Mediterranean throughout the crisis was small relative to the Black Sea

Fleet (e.g., the Soviets never had more than two surface to surface missile [SSM] destroyers or one cruise missile submarine in the Mediterranean during the crisis, although there were four SSM destroyers in the Black Sea Fleet and the Fleet's one SSM cruiser did not leave the Black Sea during the crisis (Dismukes and McConnell: 161-2). In the wake of Soviet threats of 10 June to intervene, President Johnson sent U.S. carriers toward the Syrian coast. Soviet tattletales followed but no action was taken to interpose Soviet warships between this intervention threat and the Soviets' clients.

Dismukes and McConnell (p. 167) conclude of this initial trial of Soviet coercive naval diplomacy that "The Soviets probably came away from the 1967 war with opinions confirmed about how to deploy effectively to the Mediterranean during crises." In contrast to this favorable (to the Soviets) assessment, MccGwire (SND: 344) concludes the Soviets were unready for this forward deployment to the Mediterranean:

> . . . it is by no means clear that the political gains from this Soviet presence outweigh the various unfavorable reactions to the introduction of a superpower confrontation to the area, to shortfalls in the support of client states, or to the imperialist overtones of naval power. Meanwhile, the early deployments were dangerously exposed, and in hostage to the West.

If Dismukes and McConnell represent a hybrid of Griffiths' first two schools, MccGwire (SND, ch. 25) represents a "pure version" of the second school. He perceives (p. 350) the Soviet Navy as drawn out into forward deployment by initially the increased range of U.S. carrier-borne strike forces and secondarily by the Polaris deployments:

> This generated the third and most far-reaching change in Soviet naval policy, involving the extension of maritime defense zones and a shift to forward deployment.

Neither the use of Soviet naval presence to influence Third World events, nor the use of naval capabilities to inhibit intervention by the Sixth Fleet were sufficiently important, or guaranteed of success to have been major causes of the shift to forward deployment in the Mediterranean, according to MccGwire (p. 351). He argues that the 1954 decisions did nothing to enhance capabilities for these missions, and favorable opportunities for adopting these policies were neglected in 1954-1960. He sees the Mediterranean deployment, when it came in 1963-64, as placing a considerable strain on Soviet naval resources just to accomplish the strategic defensive missions and leaving little or no surplus for naval diplomatic missions. As the numbers of units on forward deployment increased in 1967-68 "the opportunities for the political exploitation of their presence" also increased.

MccGwire views the Soviet Navy as having deployed forward basically unready to execute its strategic defensive mission. The need to counter the Western threat forced the Soviets to "forego . . . the basic naval requirements of survivability" and the only counter the Soviet Navy posed for Western strategic systems in this initial phase of forward deployment required the protection of peacetime. The Soviet Navy was

incapable of surviving war beyond the first salvo during this first forward deployment.

The thesis that the Soviet Navy was drawn into the Mediterranean by the carrier and Polaris threat is also advanced by Smolansky and Joynt (SND: 364). The coincidence of the loss of submarine facilities in Albania (1961) and the deployment of U.S. SSBNs (1960) posed a severe problem for the Soviets. The initial Polaris deployment in the Mediterranean in 1963 was followed by Soviet proposals to make the area a nuclear-free zone and, in 1964, by the first continuous deployment of Soviet naval ships in the Mediterranean. In June 1963 the Soviets upgraded their six-year old military aid agreement with Egypt and began deliveries, for the first time, of modern first-line equipment. Between 1963 and 1967 the Soviets brought intense diplomatic pressure on Egypt to allow Soviet naval ship access to Egyptian facilities. Smolansky and Joynt conclude (p. 364) that Khrushchev's 1964 tour of Egypt was "prompted in part by his determination to obtain a naval base for use of the Soviet Mediterranean squadron."

Khrushchev's involvement in the Middle East despite difficulties and setbacks (e.g., loss of support facilities in Albania, Egyptian unwillingness to grant wide access to facilities) is seen by Smolansky and Joynt as responding to "the imperative of U.S. naval deployment in the Mediterranean." The "local war doctrine" adopted by Brezhnev and Kosygin represented a continuation of Khrushchev's policy, perhaps in a lower key in the Mediterranean, but in a more vigorous form elsewhere (e.g., the greatly expanded Soviet aid to North Vietnam in 1965).

Khrushchev's actions are consistent with the thesis that he was forced into forward deployment by the Western strategic threat and was unwilling to commit the Soviet Navy to the mission of supporting Third World clients because of its unreadiness. Khrushchev turned down such a request from Nasser during the 1958 crisis following the Iraqi revolution with the explicit explanation that the Soviets were not ready for a clash with the West (see Dismukes and McConnell: p. 11-12). As late as 1967 the legacy of Khrushchev's policy could be seen in the failure of Soviet naval units to do more than follow U.S. carriers and patrol the supply lines that delivered Soviet equipment after the fighting to replace the Arab losses. The Soviet explanation that their naval units inhibited Western intervention in 1967 was constructed post hoc in the wake of Western reactions to Soviet forward deployment (cf. McGwire, SND: 193, 203-4, 539-40).

FUTURE SOVIET NAVAL EVENTS

Case 6. Trends in Anti-Carrier Warfare (ACW) Capabilities and Operations

The analysis of Soviet naval diplomacy by Dismukes and McConnell (1979) views the Soviet Navy as increasingly engaged in coercive power projection in the Third World. At the center of this strategy is the need to prevent an effective peacetime counter to U.S. aircraft carriers. McConnell (in Dismukes and McConnell: 21) writes:

> The core of Moscow's Third World diplomacy of force lies in its capabilities for countering U.S. carrier task groups.

These ACW forces, according to several writers, are comprised of a cruise-missile submarine, several torpedo submarines, a surface-to-surface missile (SSM) ship and a surface-to-air missile (SAM) ship. McConnell views (p. 13) surface ships, in particular SSM and SAM ships, as "integral parts of Soviet anticarrier task groups." In fact, as noted above, gun cruisers and destroyers are also found in ACW task groups.

During crises, the Soviet Navy appears to deploy a number of ACW task groups equal to the number of Western carrier groups in the area. These deployments seem to be oriented more to crisis than war: e.g., while ACW units are introduced to the crisis area, on occasion ASW units have withdrawn (Dismukes and McConnell: 291). The ships the Soviets introduce to the crisis area are usually those designed for the anti-carrier task, according to Dismukes and McConnell (p. 294) and

> . . . are intended to counter the carriers' psychological impact and to reduce U.S. freedom of action in employing its carrier forces.

Soviet naval authorities make similar claims for their intentions during crises.

An extreme case of this deterrent role for Soviet ACW occurred during the Angolan crisis, argue Dismukes and McConnell (pp. 148-50, 259), when a Soviet ACW task group in the South Atlantic anticipated a possible move southward of a U.S. carrier group and moved to interpose itself on the carrier's route to the crisis area. Dismukes and McConnell view this as a unique instance of the Soviets' use of an ACW group to discourage deployment of USN forces in anticipation of U.S. moves (Herzog, NPSP: 41, concurs in this assessment).

Dismukes and McConnell acknowledge (p. 244) that ACW units are now some of the older elements in the Soviet fleet, but view this as a virtue, making them cheap counters to the carriers in crises, and expendable units for the wartime mission of destroying the carriers in the opening salvo. The ACW task groups are still the "core of Moscow's Third World diplomacy of force" (p. 21).

Other naval analysts seem to agree that while the Soviet surface units formerly played the ACW role outlined by Dismukes and McConnell, the wartime anti-carrier task now is performed by a coordinated submarine-aircraft team. The use of surface ships in the ACW role in the early 1960s provided Soviet designers and shipbuilders the breathing space needed to develop space-borne surveillance and communications platforms, Soviet Naval Aviation (SNA) aircraft, and nuclear submarines for the ACW task, while surface design and construction concentrated on upgrading the ASW capabilities in new major types (Herzog, NPSP: 43-49). While Herzog agrees that the Kotlins, Kyndas, and Kresta Is were designed for the ACW role, he writes (p. 49) that in the mid-to-late-1960s

> The Soviet Navy had effected the transfer of the antiship mission [including ACW] . . . to the submarine force.

The heart of the new ACW team is the C-class nuclear submarine with the SS-N-7 cruise missile and the Backfire bomber. All of the Soviet Navy's new major surface units (Kresta IIs, Karas, Krivaks) are ASW rather than ACW ships, and writes Herzog (p. 53), "designed to operate under the

umbrella of SNA, i.e., in the [Soviet] SSBN sanctuaries." (Herrick, NPSP, ch. 9, advances a similar concept.)

Analysts of Soviet warship construction trends and designs concur in Herzog's assessment of the shift of the ACW mission from the surface units to a submarine-SNA-satellite team (e.g., MccGwire, NPSP: ch. 5). For example, MccGwire notes (p. 98) the increasing clear separation of Soviet naval construction into those units designed for distant water operations and those intended to operate within the home fleet areas. From 1966 onwards the surface programs "have been tailored to the antisubmarine role" (p. 100).

These shifts leave the Soviets' "core" naval forces for peacetime power projection in an anomalous position. On the one hand the older Kotlins, Kyndas, and Kresta Is, are designed for ACW and still effective, but are aging counters to U.S. carriers. On the other hand, the new Soviet surface units are less effective counters to carriers in combat, being designed for ASW and requiring SNA air cover, but these are the most effective new Soviet surface ships for distant deployments. While these units may have been designed with Third World political missions in mind (cf. Herzog, NPSP: 55), their capability as ACW forces is probably more psychological than tactical; possibly impressing the naive, but otherwise more hostages than counters to U.S. carriers. The fact that the Soviets withdraw these newer forces into sanctuaries as crises evolve, and replace them with older ACW units, implies that the Soviets themselves are aware of these anomalies.

Friedman (NPSP, ch. 11) points out that the evolution of the Soviet Navy's sea denial doctrine dictates the eventual assignment of the ACW task to more efficient and mobile air and submarine forces. The drawback of this strategy is that, for application at great distances from the Soviet homeland, foreign air and submarine bases are required. Foreign bases, however, pose problems of sea control which are not easily solved with a sea denial force. Sea denial doctrine also prohibits surface concentration of forces: any geographical concentration provides a tempting target for a preemptive strike. "As tension increases," Friedman writes (p. 219) "the [Soviet] ships must disperse: they cannot present neutrals with an image of increasingly concentrated strength." An impressive "crisis presence" is inconsistent with a sea denial doctrine, risking too many forces in one area to a preemptive blow.

The analysis of Soviet open-ocean naval exercises by Daniel (NPSP, ch. 12) also tends to reflect the shifting role of the surface units from ACW to ASW tasks. Up until 1971 ACW strike activities were the major portion of these exercises. Since 1971 the emphasis shifted from the ACW strike phase to strike coordination, and from 1974 on, to the reconnaissance and surveillance phases and a greater concentration on a "buildup of tensions" scenario. In the Okean 75 exercise, ACW strikes were conducted primarily by aircraft, possibly aided by submarines. In 1973 ASW became the main theme of major Soviet naval exercises, and seems to have figured as the main theme since. Daniel concludes (p. 231) that "Soviet commitment to the ASW problem is just as strong if not stronger than commitment to ACW." Since the Soviets take no special pains to conceal the surface and air aspects of their major open-ocean exercises, and even draw world attention to them, it can be assumed that they see no need to conceal the shift of the wartime ACW mission from the surface forces. Where this leaves the Soviet Navy's Third World diplomacy of

force capability relative U.S. carriers is still an open question.

Case 7: Soviet Naval Support for Third World Coups

Naval analysts differ significantly on the role played by the Soviet Navy in Third World coups: some seeing the Soviets as behaving very circumspectly, aiding only the status quo powers; while others see the Soviet Navy as having actively assisted in the overthrow of incumbent regimes. One example of such differences involves the role of the Soviet Navy in the 1969 Libyan coup.

Jameson (in STS: 68) claimed a central role for the Soviet Navy in that coup:

> The [Soviet] Mediterranean Squadron was notably employed in an exercise of Soviet diplomatic support (or so it appeared) . . . during the . . . coup against King Idris. Soviet, Syrian, and Egyptian military units were engaged in maneuvers involving a practice landing on the Egyptian coast close to the Libyan border. The Mediterranean Squadron was strung out between the Egyptian coast -- very close to the Libyan border -- and the Island of Crete in a disposition apparently aimed at preventing the penetration from the Western Mediterranean of a hypothetical U.S. carrier task force. The precise relation of the timing of this exercise to the coup is impossible to identify, but it seems highly likely that some sort of coordination took place. The most significant force in Libya to protect the King's government . . . was a British tank detachment. . . . The crews of these tanks, however, were in Cyprus. To fly them to Libya would have required that the transport aircraft overfly a substantial part of the Soviet squadron, which was bristling with surface-to-air missiles.
>
> On September 26, 1969, Radio Cairo commented on the Libyan coup as follows: "The presence of the Soviet fleet at this strength is a guarantee for neutralizing the effectiveness of the Sixth Fleet and deterring it from carrying out new imperialist adventures."

Herzog (NPSP: 40) agrees with Jameson's view.

Blechman (1973: 23) suggests that Soviet intervention in this coup existed only in the eyes of the Arabs:

> Whether the United States and its allies were actually constrained may be less important than whether other nations believed them to have been so hampered. For example, Radio Tripoli claimed that Soviet naval operations deterred British intervention following the coup d'etat in Libya in 1969. In point of fact, Britain and the United States controlled air bases in Libya at the time and would not necessarily have had to intervene by sea even if the Soviet Union had interposed a force. Moreover, the Soviet Union

itself has not claimed credit for this "defeat" of the imperialists.

McConnell and Dismukes (1979, 24) reject not only a Soviet naval involvement in the Libyan coup but in any coup:

> The USSR has never made a show of force in support of a successful coup against an established government before time has proven the new regime's control and thus the emergence of a new status quo.
>
> Despite charges to the opposite effect, we find no evidence of a Soviet show of force on the morrow of the overthrow of King Idris in Libya in 1969, nor during the same year's successful coup d'etat in Somalia.

In their book on Soviet naval diplomacy, Dismukes and McConnell (1979) include a special case study by Roberts of the "non-case" of Soviet coercive diplomacy in the Libyan coup. Roberts reviews the evidence for the case that the Soviet naval presence off Libyan shores was coup related and the evidence that these activities were independent of the coup. For example, the exercise planning preceded the coup by months; the coup itself occurred during a standdown in the exercise; there were no political connections between the Libyan plotters and the Soviets; Libyan and Soviet relations remained cool after the coup.

Roberts proposes the counterhypothesis that the Libyan plotters may have taken advantage of the Soviet exercise to shield the coup attempt from possible intervention by the United States or Britain. This counterhypothesis also depends, however, on the notion that Soviet ships posed an effective SAM barrier between Libya and a British airlift or the U.S. Sixth Fleet. The British and American airfield forces in Libya could have assisted the King, however, with no intervention from the sea. Further, the Soviet "exercise" screen would have stopped a British airlift or a Sixth Fleet sortie only if the Soviets were willing to go from an "exercise" to a war footing on a moment's notice. Since Roberts is arguing that the Soviets had no inkling of the coup, it is incredible that during an exercise Soviet forces would destroy British aircraft or U.S. ships. In any event, Roberts suggests the plotters may not have even been aware of the Soviet naval activities before the coup. Qaddafi's group may have exploited the Soviet presence only after the fact, as Blechman suggested.

Dismukes and McConnell (1979) review several episodes in which Soviet naval forces apparently acted in support of post-coup regimes (e.g., Somalia, 1969; Guinea, 1970; Sierra Leone, 1971). There is no convincing evidence of Soviet naval intervention prior to a coup in the Third World. However, the active role of the Soviet Army in the 1978 and 1979 Afghanistan coups may indicate a change that will affect Soviet naval actions in the future.

Case 8: Soviet Presence in the Indian Ocean

Virtually all naval analysts mention political, economic, and diplomatic factors as motivating the Soviet presence in the Indian Ocean.

However, analysts give sharply different emphases to these factors, and include different behaviors in these broad categories. For example, some analysts stress strategic military defense as the primary reason for the Soviets' entry and presence (Jukes, SNP, ch. 16; MccGwire, SND, ch. 30; Smolansky, SND, ch. 29; SNP, ch. 14). Other analysts stress diplomatic and political factors (Calhoun and Petersen, NPSP, ch. 13; Dismukes and McConnell, 1979; Graham, NPSP, ch. 16; Kelly, SNP, ch. 15; McConnell, SND, ch. 28; McConnell and Kelly, SND, ch. 31). Most analysts mention trade and economics.

Differences among analysts include such basic questions as whether the Soviets' move into the Indian Ocean is a "drive for warm water ports" (Graham, NPSP: 275; Herzog, NPSP: 39). Herrick, (1968: 143), in contrast, claims:

> . . . no evidence has been uncovered to support (but much to controvert) allegations that the Soviet Russians, like their Tsarist predecessors, are subject to an historic and lemming-like urge to the sea to obtain warm water ports.

Jukes, MccGwire, and Smolansky all emphasize the military purposes served by the Soviet Navy in the Indian Ocean, especially with respect to the strategic ASW and ACW missions. The Polaris submarines, with the A3 missile introduced in 1964, can strike most of the Soviet Union from the Arabian Sea in the Northwest Indian Ocean. Similarly, the possibility of strikes from U.S. aircraft carriers in the Persian Gulf against Soviet central Asia poses a strategic defensive problem.

Jukes and MccGwire also note the role of Soviet space systems in motivating a Soviet presence in the region. Jukes (SNP: 315-16) raises the connection between the Soviets' SS-9 and SS-18 heavy intercontinental ballistic missiles, which can attack the United States by flying the "long-way," i.e., around the South Pole. Accuracy along this trajectory would require some form of mid-course guidance. This might be supplied from a well-surveyed ground station in the Indian Ocean area. MccGwire (SND: 427, 436) notes that the polar orbit from the Soviets' space center passes over the Indian Ocean and Soviet space support ships have been frequent visitors to the region.

The primary factor behind the Soviet presence in their views, however, is the threat of Polaris operations. MccGwire (SND: 441) writes:

> All the available evidence . . . points to the high priority accorded by the Soviet Union to the threat from U.S. ballistic-missile submarines and measures toward countering it . . . it is hard to escape the conclusion that the very clear indication that the United States was developing the capability to operate ballistic-missile, submarines in the Indian Ocean, must have been a major (and perhaps the determining factor) in deciding the Soviet Union that she, in turn, must embark on the costly process . . . of developing the capability to operate in the area.

Jukes writes (SNP: 315) that the two Soviet purposes in the region are "area familiarization with the Arabian Sea against a postulated future

Polaris/Poseidon deployment," and maintaining a position to continue influence on Arab and Middle East developments. With respect to the latter he writes (p. 316):

> But from "protection of state interests in time of peace" as a by-product of a combat role to installation of forces primarily in the interest of such a peacetime function is a long step . . . this step has not been taken yet in general terms, and the Indian Ocean deployment presents no evidence with which to contradict this interpretation.

Smolansky (SND: 421) views Russian entry as motivated by several factors, but primarily by

> Western actions (specifically the introduction of the Polaris/Poseidon fleet into the Indian Ocean) have been a major consideration impelling the establishment of a Soviet naval presence in the area. Thus, far from the aggressive intent so frequently ascribed to these recent Russian moves, Moscow's main concern seems to have been military defense.

Smolansky sees the Soviets as having no vital interests in the region with te exception of the strategic defensive mission of preventing SLBM strikes from the Arabian Sea. While Soviet interests are important, they are not vital to the USSR's security.

Nevertheless, Smolansky predicts the Soviets will continue to conduct a contest for influence in the region while attempting to neutralize the U.S. ocean-borne strategic threat. This contest is likely to accelerate and intensify, he predicts (p. 422). Smolansky (SNP, ch. 14) does not foresee any significant Soviet influence developing in the Persian Gulf, nor does he view domination of the Gulf as a Soviet goal. Nor is interference with the shipments of Persian Gulf oil to the West a likely Soviet move in the future, since any interference would prompt a confrontation with the United States, something the Soviets have consistently avoided. Moscow will pursue "low risk" opportunities but the Soviets' general lack of influence and inability to control events will yield only marginal chances for improving their diplomatic and political presence.

In contrast to Jukes, McGwire, and Smolansky, several naval analysts disagree that the Soviets' foremost purpose in the Indian Ocean is strategic defense. McConnell (SND: 390) explicitly disputes the Polaris threat as a major motivation for the Soviet Indian Ocean presence. He notes that the United States has never announced a Polaris deployment to the Arbian Sea or the Indian Ocean. The submarine tenders that typically support U.S. ballistic missile submarines have not appeared in the Indian Ocean. McConnell also claims there is no need for a U.S. deployment of SLBMs to this area since the targets are otherwise in range, and the transit time to the Arabian Sea would be inefficient and atypical in comparison to known deployments from the U.S. submarine bases at Guam in the Pacific, Holy Lock in the North Sea, and Rota in the Mediterranean. McConnell notes the lack of any comparable facility to service U.S. deployments in the Indian Ocean. He argues (p. 391) that a U.S. Polaris deployment would be counterproductive since

> There is probably no easier way to get Russian ships steaming all over the Indian Ocean than to introduce a strategic threat.

Nevertheless, Soviet ships are "steaming all over the Indian Ocean," especially in the northwest quadrant that would serve best as an SLBM launching point. McConnell says nothing about the U.S. naval communications station that was constructed on Australia's Northwestern Cape in the 1960s and 70s, the communications facility that was formerly maintained at Asmara, Ethiopia, or the U.S.-British naval facility developed at Diego Garcia, at the southern margins of the Arabian Sea. MccGwire (SND) argues that the capability provided by these facilities to support U.S. Polaris operations are sufficient to produce a Soviet defensive reaction, even if the Polaris submarines themselves do not practice deployment. McConnell acknowledges that this threat may be taken serious by the Soviets and he quotes a Pravda reporter as saying the Americans intend to expand Polaris combat patrols to the northern Indian Ocean.

McConnell concurs (p. 391) with Smolansky that a Soviet blockade of Persian Gulf oil is "scarcely credible" since it would harm the Arabs, whose favor the Soviets court, but more importantly, it would lead to war with NATO. Nor does McConnell see Russian planners as much concerned with protection of sea lines of communication through the Indian Ocean, since their main concern is general nuclear war, and such logistics would be unimportant. Consequently, McConnell writes (p. 391), "the evidence more and more suggests peacetime rather than wartime missions for the Soviet Indian Ocean detachment." The peacetime protection of state interests, of which Admiral Gorshkov writes, includes, in the Indian Ocean, protection of Russian shipping, fishing, and scientific activities; influence-peddling through goodwill visits; and protection of client states.

In the last role, McConnell sees Soviet ships as providing a deterent to Western intervention and a maintenance of the status quo, rather than commitment to violent offensive action or confrontations. He does not view (p. 400) the Soviet Navy as likely to contest issues which were important to Western interests. McConnell suggests the Soviets' long-range objective in the region is the negotiated mutual withdrawal of U.S. and Soviet navies:

> Basic to Soviet reasoning would be the calculation that there is more to be gained by the removal of a Western naval presence than the maintenance of a competing Soviet presence.

Calhoun and Petersen's (NPSP, ch.13) analysis of Soviet views on naval arms limitation reiterates the strong Soviet concern over U.S. ballistic missile submarine deployment in the Indian Ocean and the Diego Garcia base. However, they note changes in Soviet naval operations in the Indian Ocean that are not merely defensive and which suggest the Soviet Navy may seek a limitation only on strategic naval capabilities in the Indian Ocean, leaving the Soviet general purpose ships free to conduct naval diplomacy. Namely, Soviet naval forces were involved in four cases

of naval diplomatic initiative between 1973 and 1975: mine sweeping in Bangladesh, supporting Iraq's attack on Kuwait, patrolling the Strait of Hormuz, and development of naval facilities at Berbera, Somalia, and Aden, in South Yemen, astride the Bab el Mandeb straits to the Suez Canal. Calhoun and Petersen see (p. 242) these as instances of "direct intervention to shift the balance (rather than to maintain the status quo) militarily and politically in favor of the preferred side and -- more significantly -- away from the West (and China)." Calhoun and Petersen view (p. 240) these changes as revealing "something about Soviet intentions:" (a) a capability and willingness to match the increased force levels deployed to the region, and (b) a heavier commitment of Soviet military power to nations in the region.

Graham (NPSP: 287) also seems to see no Soviet interest in withdrawal from the Indian Ocean:

> What Moscow pursues is the chance to exploit any weakness, any possibility that offers itself a chance to increase its influence, to bring adversaries into disarray, and to create a clientele among the Indian Ocean regional states.

Graham categorizes these as political, economic, and military efforts aimed at an "outflanking strategy" and a "chokepoint strategy." With respect to the first, Graham writes (p. 281-2) "the Soviets perceive a need . . . for naval capabilities in order to outflank from the sea both China and NATO . . .. The prospect of being able to deny [Persian Gulf] oil to the West may be major impetus to Soviet global strategy." The Soviet naval forces in the Indian Ocean are seen as achieving these ends as well as protecting Russia's sea lines from the Black Sea to the Soviet Far East. Graham does not note that Soviet forces are insufficient for any of these missions in either war or peace.

Graham describes (p. 282) the "chokepoint strategy:"

> Soviet strategy in the Indian Ocean is to establish a naval presence . . . through which Moscow can influence the policies of the Indian Ocean littoral states as well as control the approaches . . . should the need arise, such as in wartime.

This strategy is aimed at depriving the West of strategic materials and oil. Again, the Soviet Navy, lacking any air capability afloat, could not execute this strategy, a major weakness in Graham's thesis. Furthermore, the Soviets have experienced repeated setbacks from Indian Ocean littoral states (e.g., Somalia, Indonesia). Nonetheless, Graham sees (p. 298) the Soviets as persisting in the pursuit of influence and facilities to achieve these strategic and political missions.

Graham seems to see fewer problems in store for the Soviets in seeking these objectives than actually exist. The Soviet Navy's inability to close off chokepoints or outflank both NATO and China has already been noted. Graham sees Ethiopia as ample compensation for the Soviet naval facilities lost in Somalia. This overlooks the Eritrean rebels, who are fighting the Soviet-backed Mengistu-regime, control much of the coastline, and have prevented any extensive Soviet development of bases to replace those lost in Somalia at Berbera and Mogadiscio.

The Atlantic Council review of Soviet naval developments, Securing the Seas, also asserts some of Graham's hypotheses. The Soviet quest for warm water ports is seen (p. 441) as a possible Soviet objective in the Indian Ocean. Soviet blockades during another Arab-Israeli war or Soviet support for an Arab blockade of the Strait of Hormuz are suggested (p. 419) as "plausible scenarios" of Soviet interpositioning and resistance to U.S. involvement in the area. However, the STS review does not see (p. 440) direct Soviet harrassment or interruption of Western maritime trade as likely, even though Soviet naval and political capabilities for such interference have increased, because the Soviets' own maritime units are as vulnerable (or more so) to interception. The Soviets are more likely to attempt to stem the flow of resources or trade at the source, through strikes, embargoes, sanctions and other means that do not put the Soviet merchant fleet in jeopardy.

Case 9: Future Soviet Objectives for Forward Deployment

Regardless of differences among analysts over the causes of Soviet naval forward deployment, the level of out-of-area naval activity has remained high. An important analytic issue is the future use of this naval force. What are Moscow's intentions governing the future use of forward deployed naval forces?

Booth (1977) notes the differences of emphasis between two groups of analysts on this question. One viewpoint emphasizes the possibility that the Soviet Navy will continue to be and may increasingly act as an instrument of general Soviet foreign policy. This group of analysts (the "political school" will serve as an inaccurate but short appellation) sees the Soviet Navy as becoming more active, aggressive, and expansive in peacetime missions. At the extreme Soviet Navy interventions against Western clients and Soviet deterrence of Western intervention in crises are seen as possible uses of growing capabilities for power projection. In effect, the possible use of Soviet naval power and other capabilities of force projection to upset the status quo and the superpower balance in the Third World are no longer viewed as too risky for the Soviets.

On the other hand, a second group of analysts (the "military school") views the forward deployed Soviet Navy as primarily still concerned with wartime naval missions, with only a secondary emphasis on enhancing diplomatic influence or prestige. This group views the Soviets as still not having solved its problems in achieving general war missions. While political gains can be secured as by-products of continued naval efforts to solve the war mission problems, "the surplus for projecting power by military force has been," Booth writes (p. 58), "at most, small." Thus, while the Soviets will exploit opportunities to upset or challenge the status quo, they cannot afford to risk much to do so.

Booth (1977: 49) observes that both groups agree that the Soviet Navy requires maritime stability to fulfill its foreign policy tasks, whether narrow or expansive. A direct clash with the West would remove the freedom of the seas and risk escalation to general war, which would eliminate much or all of the Soviet's capability to project diplomatic power or influence. Booth writes (p. 49) that "for these reasons the Soviet Union is unlikely to engage in limited hostilities at sea beyond occasional maritime truculence."

The authors of the Atlantic Council Working Group (Nitze, Sullivan, et al., STS: 418) seem to question Booth's assumption:

> Because of the increasing strength of their conventional deep sea forces and the relative decline of Western naval strength, the Soviets may elect to resist U.S. involvement in some "local war."

They also view the Soviets as having "plainly demonstrated their intentions" to supply client states with arms and proxy forces in local wars. This intention to become involved in local conflicts could touch off a direct U.S.-USSR confrontation in their views, and two "plausible scenarios" are suggested. During another Arab-Israeli war the Soviets might attempt a blockade of Israel. Alternatively, if a Mideastern nation attempted to block the Strait of Hormuz, the Soviets might resist U.S. attempts to thwart the blockade.

The Atlantic Council Working Group also viewed (p.421) the Soviet Navy as increasingly likely to attempt to inhibit "the free movement and deployment of Western naval forces" during Third World crises. This effort in the past has been "a clearly inhibitory impact" on the West, and, they imply, this impact will probably become greater. Inasmuch as they view (p. 423) other elements of Soviet diplomatic projection (e.g., the Merchant Marine) extending into traditionally Western areas and regions, the geographical realm for U.S.-Soviet naval confrontation may also increase. An important element of Soviet maritime expansion is access to defensible, warm-water ports for naval forces. The working group views (p. 410) the Soviets as likely to attempt to gain such access politically, especially in the Indian Ocean, but by this they imply a major role for the Soviet Navy in the effort.

Graham (NPSP, ch. 16) also seems to view the Soviets as increasingly willing to risk confrontations with the U.S. Navy in the forward area. He writes of an evolving Soviet strategy in the Indian Ocean of chokepoints, and implies (p. 290) a Soviet willingness to interfere with Western oil supplies and sea passage in conventional war. He writes (p. 289) that the Soviets "recognize that some capability to interdict Western oil supplies gives [the Soviets] great political leverage, even if they never exercise that capability in an operational sense." This clearly implies a Soviet effort to win such control in peacetime as well. The Soviets' goal (p. 288) in the Indian Ocean is "to become the dominant external power," and their major objective (p. 275) is to secure warm water ports.

Some members of the "political school" of analysts see the acquisition of overseas bases as a major objective of Soviet naval diplomacy (e.g., Blechman and Levinson, SNI, ch. 22; Kelly, SNI, ch. 26), but so do some spokesmen of the "military school" (e.g., Booth, 1977: 62; McGwire, SNI: 634). The two groups differ in what they consider the motivation for these efforts. Kelly (p. 513) views Soviet Navy use of Third World facilities as increasingly for "practical, direct support for specific foreign policy initiatives and campaigns," e.g., support of local wars. The military school stresses the economic and strategic advantages which bases in the forward area give the Soviet Navy (e.g., McGwire, SND, chs. 18, 25) and raises the question of whether those bases the Soviets acquired in the past, e.g., Egypt, at great expense yielded sufficient political gains (McGwire, SNI: 634-37) to tempt the Soviets to quickly repeat the experience. The strategic military need for bases is another matter, one which may justify heavy costs and may require diplomacy in aid

of military purposes (Booth, 1977: 62). Kelly also sees (p. 520 ff) the Soviets as seeking submarine bases for strategic deterrent purposes and possible conventional purposes, i.e., near sea lane chokepoints. Nevertheless, she gives equal emphasis to the Soviet Navy's diplomatic and political missions that these bases facilitate. Blechman and Levinson see (p. 426, 439) the efforts to establish a submarine base in Cuba as "an almost textbook case of Soviet political-military tactics" and suggest that major goal in this effort is to challenge U.S. security politically, by showing U.S. resolve and credibility are weak. They argue that the Soviets seek the base primarily as a symbolic demonstration of an American lack of will. In contrast, McCGwire (SND, ch. 33) outlines several naval tasks that Soviet submarines could perform more efficiently and effectively from Cuba, e.g., against U.S. carriers and Polaris/Poseidon submarines.

Dismukes and McConnell (1979) offer the most extensive analysis of the use of the Soviet Navy for diplomatic and political objectives. They write (p. 295) "political considerations have dominated Soviet motivations for operating their general-pupose navy in the forward area." A variety of "magnets" in the Third World will continue to attract superpower naval involvement and yield returns to coercive naval diplomacy; "there is not the slightest sign of official [Soviet] disenchantment," they write (p. 303) "with naval diplomacy."

Dismukes and McConnell believe that a defensive, status quo principle governs the use of the Soviet Navy for political and diplomatic purposes and will tend to do so for the future (p. 245-46). Most Soviet objectives will be obtained without violence, through naval presence or the augmentation of area forces as needed (pp. 284, 287). Their crisis behavior will be cautious and restrained and Moscow may vary the risks by its encouragement to clients, but will be reluctant to increase risks itself. However, Dismukes and McConnell note (p. 308) that

> . . . in general, should the sovereignty of a client be jeopardized, [they] would expect the Soviet politico-military instrument to be made available . . . Nothing else would maintain Moscow's credibility as a patron.

Seemingly, the scenarios outlined by the Atlantic Council working group would seem less plausible to Dismukes and McConnell.

In contrast to the favorable outlook for the continuation and expansion of Soviet naval diplomacy which Dismukes and McConnell (1979) perceive, other analysts place greater weight on the limitations the Soviets face in this cause. McCGwire (1979a) questions Soviet political commitment and naval capabilities for this course, Friedman (NPSP, ch. 11) notes the limitations imposed by the Soviets' "sea denial" philosophy, Ginsberge (SNI, ch. 23) observes the limits imposed by Soviet influence mechanisms, and Herzog (NPSP, ch. 5) describes the impact of the strategic mission.

McCGwire (1979a: 177-182) views Soviet commitment to peacetime naval operations as a continuum ranging from "protecting Soviet lives and property" at the low end of commitment, to "establishing a strategic infrastructure to support war-related missions" at the high end. McCGwire views this latter as the primary motive for Soviet naval involvement in

Third World diplomacy. Because this mission touches directly on the security of the Soviet homeland, analysts of the "military school" view it as the foremost motive for the Soviet Navy's peacetime actions (cf. Booth, 1977; McGwire, 1979a: 179). Naval involvement in aid of this goal serves other political and diplomatic functions.

In between these extremes McGwire (1979a) views "increasing Soviet prestige and influence" and "countering imperialist aggression" as objectives for which Soviet political commitment is low. He perceives (p. 179-80), for example, no evidence of "Soviet readiness to actually engage Western naval forces, in order to prevent them from intervening against a Soviet client state." McGwire does see a greater Soviet willingness to supply and support clients during third party conflicts and a policy of incrementalism and opportunism; probing Western responses and creating precedents. McGwire sees the Soviet Navy's long-term interests as only partly served by the foreign policy of increasing influence in the Third World, and conversely, Soviet foreign policy is sometimes ill-served by the foreign entanglements and commitments required by naval security concerns and the requirements for a distant strategic infrastructure These conflicts of interst are likely to continue and remain limits on Soviet peacetime naval diplomacy in the forward area.

Herzog (NPSP, ch. 3) describes how the Soviet Navy's strategic peacetime deterrent forces have begun to retreat into ballistic missile submarine (SSBNs) sanctuaries near Soviet coasts. To control these areas and protect these peacetime strategic deterrent and wartime strategic reserve nuclear forces, the Soviet surface fleet is being reconfigured for ASW purposes, to operate under the land-based air cover of the SNA. These newest surface ships, however, are also those most capable of operating in distant waters and exercising peacetime influence missions in the Third World. In effect, the peacetime mission of SSBN sanctuary protection conflicts with the naval influence mission. If newer ships are risked in Third World area crises, they cannot defend SSBNs near Soviet coasts. In fact, as crises develop, the Soviets seem to remove these newer ASW units from the danger area. This means the newer Soviet surface ships might only be commited to low risk diplomatic missions. Herzog writes (p. 55)

> As long as the Western fleets are built around the CV (i.e., the attack aircraft carrier), there is no way the Soviet surface units, Kiev included, could assume an offensive role.

Under the defensive umbrella of the SSBS sanctuaries these units can perform their "strictly strategic defensive role," but this limits their capability and constrains Soviet intentions to risk them in Third World naval diplomacy.

Friedman (NPSP, ch. 11) points to the inconsistencies between the Soviets' sea denial philosophy and Soviet need for sea control forces, which the Soviets have never shown any interst in building, and have spent much effort attempting to circumvent. Sea control demands greater capability for flexibility and sustained operation than the Soviets have yet designed into their ships or supply facilities. Either the Soviets must suffer limited endurance from their units, or improve afloat replenishment and repair capabilities, or seek a network of foreign bases and facilities. Foreign bases, however, impose sea control requirements.

Replenishment and repair ships require fleet concentrations which are inconsistent with sea denial dispersal tactics. Greater individual unit endurance implies greater costs, nuclear propulsion, and violates the economic and technical bases of the sea denial philosophy. These inconsistencies pose serious limits on Soviet efforts to project power in peacetime with naval forces designed for sea denial.

Ginbergs (SNI, ch. 23) observes that the political instruments available to the Soviets for advancing Third World diplomacy offer few assurances of success. Although the Soviets may labor long and hard for influence, presence, and bases in the Third World, all these can be lost quickly and suddenly, leaving the Soviets no alternative but to start over elsewhere. In effect, Moscow cannot count on allies in the Third World. Consequently, Ginbergs sees (p. 446) the Soviets as following a "narrow agenda of bases to acquire" in a "few choice locations," and tending to "gravitate toward low-risk projects." However, Soviet successes will be confined to a few Third World states with a peculiar constellation of features: those with home-grown socialism, that have fallen out with the West, that are in or face military conflicts, or which require Soviet protection to avoid military defeat. These factors can offer the Soviets worthwhile opportunities, but they considerably narrow the choices on the menu.

CHAPTER 4. EVALUATION METHODS AND APPRAISALS

This chapter is divided into two main sections, Estimation Logic, and Narrative Logic. The first section treats intention estimation as a judgment process. It decomposes intention estimation into seven main tasks and examines in detail the analytic processes within each task. A variety of cognitive biases are noted that tend to distort or weaken analytic accuracy. Examples, taken from our sample of naval estimates, are used to show how these cognitive biases actually occur in the work of naval analysts.

The second main section, Narrative Logic, views intention estimation as a form of historical narrative rather than as a problem of judgment or decision. This section divides historical narration into inquiry and explanation, and examines the problem of fallacious historical reasoning in these two areas. Many of the historical fallacies reviewed in this section are directly related to the cognitive biases described in the section on Estimation Logic.

ESTIMATION LOGIC

> There is no such thing as a classification of the ways in which men may arrive at an error: it is much to be doubted whether there ever can be.
>
> Augustus de Morgan, 1847

In appraising estimates of Soviet naval intentions one can examine an analyst's argument, compare the analyst's logic to the data and to other analysts' arguments, and evaluate the arguments' logical structure, i.e., how analysts connected together hypotheses, data, conclusions and theories. In this section the logical process of estimation is broken into parts and the weaknesses of these parts are assessed. Throughout this section an effort is made to use examples from our sample of naval estimates. These are used to suggest a possible weakness and to illustrate how it may occur in naval estimation; our goal is not to prove that the particular analyst's argument is logically weak -- it may or may not be.

We break down estimation logic into seven processes:

- perceiving data
- weighing data
- characterizing data
- assessing covariations
- assessing causes and effects
- prediction
- forming, maintaining and changing theories

Not every estimate explicitly includes each process, e.g., some estimates studiously avoid making any predictions. We would argue, however, that some form of each of these processes is involved implicitly

in every estimate. This conclusion is based on the plausible notion that all judgmental thinking and reasoning (not just naval estimation) entails these processes. A recent explication of this general notion about thinking, Nisbett and Ross's Human Inference: Strategies and Shortcomings of Social Judgment (1980), provided this seven-part taxonomy and reviewed literature that relates these processes to all forms of inference.

This seven-part breakdown is not immutable, other analyses of reasoning use more or less than seven categories to break down the thinking and judgment processes. For example, Einhorn and Hogarth (1981) use four: information acquisition, evaluation, action, feedback and learning. Because their taxonomy was derived specifically for decision-making, rather than inference and judgment in general, it differs from Nisbett and Ross's seven-part process in collapsing several of the latter's categories (i.e., perceiving, weighing and categorizing data generally equate to information acquisition and, in part, evaluation), and in putting more stress on action and learning (which in the Nisbett/Ross scheme are part of the prediction and theory processes).

In this section each of the seven processes is described in general and in terms of naval analysis. The biases and weaknesses which influence that process are described and related to instances selected from our sample of naval estimates. If examples could not be found in the naval estimates sampled, a hypothetical naval example is described. (The failure to find a "real" instance of a particular bias in no way means that that bias is absent in naval estimation -- only that it may be rare in our sample, or overlooked by me.)

Before beginning the description of the parts of estimation logic it is helpful to relate these processes to Scheibe's (1979) trichotomization of prediction. Recall that Scheibe attributed prediction to sagacity, control and authority, and acumen. Sagacity, or the ability to organize perceptions of past behavior and abstract from these a set of predictive indicators, is related to the seven-part estimation process examined in this section. In contrast, Scheibe's acumen is insight or empathy into the actions or reactions of another rather than analysis or judgment. Acumen can, however, be thought of as an estimation ability, most commonly found as narration, or explanation. That is, many estimates are basically coherent narrative explanations of events which attempt to organize the salient facts into a wholistic, meaningful story (Pipes, 1980; Sarbin, 1980). In a later section estimation-as-narration is examined. In the present section the estimation process is viewed as one of logic, judgment, and decision-making.

Perceiving Data

One of the oldest and most reliable generalizations in psychology is that perception is not merely the passive receipt of information but is instead a process of active interpretation, selective attention, deduction, and inference. The "facts" never "speak for themselves" in forming perceptions and perceptions are always influenced by "knowledge structures," or general information processing schemes, that people use to organize the stream of data they perceive. Two of the better-known knowledge structures which affect the perception of data have been labeled the "availability" and "representativeness" heuristics (or judgment strategies; Kahneman and Tversky, 1972, 1973; Tversky and Kahneman, 1971,

1973, 1974).

These heuristics are simple, generally automatic, strategies people use in a variety of inferential and judgment tasks, including data perception. These strategies are often efficient and effective in organizing data into useful perceptions. Because they are often applied automatically (that is, without conscious deliberation) they may be misapplied or misused in cases where more appropriate heuristics would yield more accurate perceptions. The representativeness and availability heuristics are not always involved in perception, and in cases where their application might generate errors, people sometimes avoid using them. In other cases, however, they do not, and the heuristics lead to erroneous data perception.

The Availability Heuristic

In general, the more often and the more frequently events are perceived, the more easily they are encoded and remembered and the more likely it is that the perceiver will form a standard or stereotyped construction for them. "Availability" refers to the ease with which memories and stereotyped constructions of events can be retrieved by an individual. The "availability heuristic" occurs when an individual, attempting to judge the relative frequency of an event, or the likelihood of events, is influenced by the ease of accessibility of memory or reconstruction (Tversky and Kahneman, 1973). In other words, memorability, or ease of imaginative reconstruction may influence the estimation of frequency or likelihood of events. Since frequency of events is correlated (but not perfectly) with memorability and construction, this heuristic is generally efficient, more memorable events often are more frequent or more likely. There are, however, a variety of psychological factors that influence memorability that are not associated with frequency, e.g., vividness, personal or perceptual salience, completeness and ease of encoding. To the degree that factors such as these, and not frequency, are the basis for memorability, the use of the availability heuristic as a guide to estimation of frequency or likelihood will lead to errors.

This heuristic is diagramed in Figure 4.1. The impact of frequency and other factors affecting memorability is shown in part A. In part B, when an estimate is made, the memorability of events is attributed only to frequency and likelihood (solid arrow), rather than to salience, vividness, or the other factors which might also have caused greater memorability. Although the diagram does not show this, the availability heuristic may work in converse: less memorable events are judged as less frequent or likely than more memorable ones, although several factors other than infrequency can reduce memorability.

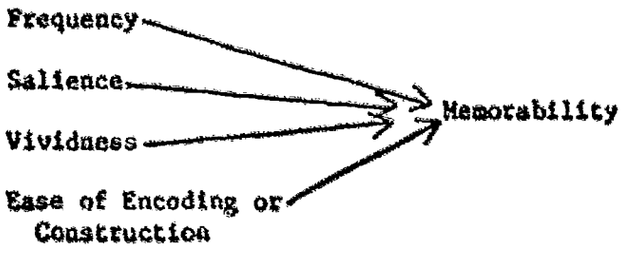
Among the factors known to influence memorability other than frequency of events are vividness, salience, encoding ease and opportunity, and causality. Vivid events, those that appear as distinctive against a uniform background are more likely to be remembered, but are not necessarily more frequent.

Several aspects of Soviet naval operations tend to make them more vivid to Western observers. The deployment forward of the Soviet Navy out of Russian coastal waters, even though initially restricted to small numbers of units, attracted considerable Western attention, chiefly

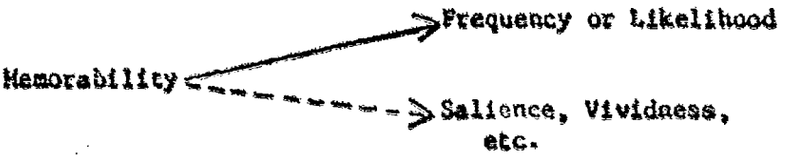
Figure 4.1. Schematic Diagram of Availability Heuristic

Perceiving Encoding Data:

A. Data Characteristics:



B. Availability Heuristic:



thr gh its novelty. Similarly the dramatic increase in the level of these out-of-area deployments attracted attention in part because they were atypical of past Soviet naval behavior. The vividness of these sudden and dramatic changes in Soviet behavior elicited significantly increased Western comment. Such attention was not given to equally dramatic changes in the Soviet Navy that occurred against a more varied background or at greater distance, and were thus less vivid. An example is the Soviet Naval Aviation (SNA) which underwent significant changes at the same time the Soviet Navy deployed forward. With a few exceptions (e.g., Murphy, NPSP, ch. 10) these changes received very little Western attention.

A second aspect of recent Soviet naval operations which increases vividness is the tactic of "close embrace" of U.S. naval ships by Soviet combatants in crisis areas. Routinely, U.S. warships are "shadowed" by Soviet auxiliaries, usually intelligence collectors. While this tactic is effective near ports and chokepoints or in narrow waters, in the open ocean U.S. task forces can generally outrun the intelligence collectors or can "flush" (sail off in different directions) and lose them. During crises, the "tattletale" role of the Soviet auxiliaries is taken over by warships, some of which seemed to have been equipped with special weapons (e g., rear firing missile systems) for the purpose. The switch-over from auxiliary to warship tattletales coincident with rising crisis tensions, and the close shadowing of U.S. ships all combine to create a particularly vivid event for Western observers. (The U.S. Navy adopted a parallel strategy in the 1973 Middle East crisis.)

Because "close embrace" tactics are particularly dangerous and could lead to inadvertent escalation of crises, they came under diplomatic attention and a U.S.-Soviet agreement was negotiated to prevent, hopefully, collisions at sea. While the vividness of close embrace events is not correlated with fr[illegible]cy, it does parallel the serious consequences of accidents produced by these events. On the other hand, such events are not common except in crises, and even then, only one or two Soviet warships are usually given this mission, and these ships are usually older, ACW types, rather than the Soviets' newer ASW vessels.

A second feature of data which can influence its memorability and thus its availability is salience. Salience is a broader concept than vividness and includes any distinctive, physically prominent, easily perceived or visible feature, novelty, uniqueness, or singularity. An event may be salient if it relates specifically to some aspect of the perceiver, e.g., naval analysts who are also aviators would probably find data on Soviet naval aviation more salient than analysts with no aviation background. Such data are more distinctive and easily perceived by someone with aviation experience than to someone with none. Vivid events are salient but salient events are not necessarily vivid.

Soviet naval activity is more salient to naval analysts than, say Soviet merchant marine activities, and has received the overwhelming attention of naval analysts, although some naval experts believe the Soviet Merchant Marine is the more effective seagoing arm of Soviet diplomacy in the Third World (cf. MacGwire, 1979). Even though the Soviet merchant fleet is specifically designed for the peacetime roles of expanding Soviet influence, increasing Soviet ties, and exporting the Soviets' goods and views, the primary interest of Western naval analysts is in its capabilities for power projection (e.g., Soviet roll-on,

roll-off ships as floating military hardware warehouses). Analysts who are quite adept at discerning peacetime roles for Soviet warships (which are designed primarily for wartime missions) spend little time examining the less salient peacetime missions of the Soviet merchantmen, and instead, accentuate their more salient wartime and crisis capabilities (there are some exceptions to this: Ackley, SNI, ch. 14; Davidchik and Mahoney, Appendix A, in Dismukes and McConnell, 1979; Fairhall, 1971; Herzog, NPSP: 38)

Naval analysts themselves acknowledge on occasion that salience may have an undue influence on their estimates. Dismukes and McConnell (1979: xv), for example, caution that

> The principle research technique employed [in their book Soviet Naval Diplomacy] is the intensive case study, focused primarily on politico-military actions. This approach has inherent risks: principally, the danger that the importance of naval power as a direct instrument of policy will be unconsciously exaggerated.

A more striking instance of the impact of salience on naval analysis is the near absence of any treatment of the Chinese in accounts by U.S. analysts of the peacetime intentions of the Soviet Navy. This is not true of other U.S. analysts, e.g., political scientists (although some political scientists can no doubt explain all Soviet behavior without reference to the Chinese). For example, a recent volume on Soviet influence in the Third World by nonnaval analysts, i.e., political scientists, (Rubinstein, 1975), explicitly considers how Soviet behavior is shaped by considerations of the Chinese. U S. naval analysts are nearly unanimous in ignoring the impact on Soviet naval activity or intentions of the Soviets' on-again, off-again competition with China in the Third World. An exception is the realization by some analysts (e.g., Graham, NPSP: 281-83, 290; Murphy, NPSP: 127,132) that, if the Chinese in wartime cut the Trans-Siberian rail lines, the Soviets' sole year-round supply line to the eastern half of Russia is through the Indian Ocean. Thus the Soviet peacetime efforts to establish footholds in that region is partly a response to China.

The Chinese lack an open-ocean naval capability and show few signs of developing any. Soviet naval activity may directly aid the Soviet goal of limiting Chinese influence in the Third World, but that Chinese influence does not now involve naval forces or sea power. For these two reasons the "China factor" has extremely low salience for U.S. naval analysts. This does not mean the Chinese have no influence on peacetime Soviet naval intentions or actions, although that influence is generally less significant than the U.S. role. In specific cases, however, the Chinese may be a significant cause of Soviet behavior.

While historical analogies are dangerous, it is interesting that prior to World War II the British, Dutch, French, and Americans all tended to assess Japanese naval intentions specifically in terms of their own Pacific and Southeast Asian interests. With the partial exception of the French, the Chinese war and the Korean and Manchurian Japanese colonies had very little place in Western appraisals of the Japanese Navy. The Japanese Army's need for petroleum to prosecute its war in China provided a strong motivation for the southward deployment of the Japanese Navy.

While China had virtually no navy, the "China factor" played an important role in the pre-1941 intentions of the Japanese Navy, a role Western analysts tended to underrate.

A fascinating case study of Soviet and Chinese competition and its impact on Soviet Navy peacetime intentions could be made using Indonesia as a focus. Soviet aid to Indonesia in the late 1950s was massive, exceeding Soviet aid to Egypt at that time (a case much favored by U.S. naval analysts). Including aid from other communist countries, Indonesia received more than $1.5 billion by the end of 1961, more than any other country up to that time, and more than the $1.3 billion which the USSR had provided China between 1949 and 1957 (van der Kroef, 1975: 55). Since Soviet aid was concentrated into a brief decade, this investment in Indonesia is particularly impressive.

Soviet supplied ships gave Indonesia the largest and most modern navy in the Indian Ocean in the early 1960s. The Soviets may have run the Sverdlov cruiser production beyond its planned termination explicitly to deliver promised ships to Indonesia (McGwire, SND: 190, 194, 209).

Uprisings and military rebellions by anti-communist, Indonesian officers and troops in the late 1950s led to near civil war. Soviet arms aid greatly strengthened the role of the Communist Party in Indonesian politics. However, these arms also strengthened the most threatening anti-communist group in Indonesia, the armed forces. Even more curiously, Moscow continued the arms supply while Sukarno's policies leaned increasingly toward Peking. Peking supported Sukarno's campaigns against Malaysia and the Philippines diplomatically while Moscow temporized and supplied the arms. This anomalous state came to an end in the 1965 Chinese-backed coup attempt by leftist officers against the anti-communist officers. In the aftermath of that unsuccessful attempt both Moscow and Peking lost virtually all their gains and Indonesian relations with Moscow remained cold until the mid-1970s and even now remain cool.

It is interesting to speculate on why Moscow, which was beginning to deploy its Navy forward in the Mediterranean in 1964, did not project its naval power in 1965 and 1966 toward Indonesia to retard the rapid erosion of Russian influence that followed the 1965 coup. Given Moscow's tremendous investment of aid, prestige, and diplomacy, and the strategic significance of the entrances to the Indian Ocean, the Soviets would seem to have had more than sufficient motivation to attempt naval diplomacy in this case. Possibly the Soviets lacked the capabilities (Soviet deployments to the Indian Ocean did not begin until 1968); or they lacked the will to begin their first exercise in coercive naval diplomacy so close to the U.S. involvement in Southeast Asia; or perhaps they were not interested in Third World naval diplomacy per se, but only as such efforts accomplished war-related ends (e.g., obtaining bases for forces which could deal with U.S. carrier and submarine nuclear threats to the USSR). In any event, the Indonesian case (or perhaps nencase) has been largely neglected by analysts of the Soviet Navy.

Certain events are easier to recall than others, or can be recalled more completely, not because of differences in encoding, but because of decoding (or retrieval) factors. One factor that aids in memory retrieval is hindsight; knowing how events turned out. Knowledge of outcomes facilitates reconstruction from memory of events consistent with those outcomes. Events which were not consistent with outcomes may

not be as effectively recalled.

This hindsight bias has been detailed experimentally by Fischhoff and his colleagues (Fischhoff and Beyth, 1975; Fischhoff, 1975a, b, 1977, 1980a; Slovic and Fischhoff, 1977). Knowing how an event turned out increases the perceived likelihood of the reported outcome and changes the perceived relevance of event descriptive data. People are largely unaware of the effect that outcome knowledge has on their perceptions, and they consequently overestimate what they would have known without outcome knowledge. They also overestimate what others actually did know without outcome knowledge.

It is not uncommon for analysts to be unaware of the hindsight bias, and also to believe that greater knowledge of outcomes will improve the analysis of how those outcomes came about. In fact, the hindsight bias may impede the analysis of events by making important descriptive data seem inevitable, trivial, or irrevelant. Knowing how events turned out may make analysis less, rather than more, insightful. Nevertheless, recommendations by analysts to defer analysis until events turn out are occasionally encountered, like the following observation by MccGwire (SNP: p. viii) on the Gorshkov series of articles:

> A definitive assessment of the articles would have to wait on greater hindsight and more external evidence.

The problem with this is that the "definitive assessment" of the series will probably perceive a greater degree of coherence and clarity in Gorshkov's writing than actually exists. The analyst's knowledge of outcomes and the changes that take place in "external evidence" subsequent to Gorshkov's writings will perhaps assist in understanding Gorshkov's meanings, but will not make those writings any clearer a key to future developments. As Fischhoff (1980a) observes

> . . . it is generally assumed that the past will readily reveal the answers it holds . . . One can explain and understand any old event if appropriate effort is applied . . . we should hold the past in a little more respect when we attempt to plumb its secrets. While the past entertains ennobles, and expands quite readily, it enlightens only with delicate coaxing.

In a later section of this chapter we will describe at some length a variety of historical fallacies and show how they occur in naval analysis of Soviet peacetime intentions.

The more easily data can be encoded into memory the more readily it can be recalled. Similarly more completely encoded data, i.e., data that are fitted into an existing mental framework, are more easily remembered. Neither ease of encoding nor completeness are indicative of frequency. If easily encoded or completely encoded data are easily recollected it is erroneous to allow this availability to influence judgments of frequency or likelihood.

Psychologists have evolved many labels for encoding processes which aid in the perception and storage of data: schemas, stereotypes, scripts prototypes, and personas are some of the terms used. These concepts differ slightly from each other but they all represent concepts

familiar to the perceiver which enable the efficient categorization of an event or stimulus as the member of a class.

There are several heuristics peculiar to concept formation and encoding which will be reviewed in the section below on characterizing data; the main point here is that the availability of well-used schemes or concepts for perceiving and encoding data may facilitate the operation of the availability heuristic and thus bias frequency or likelihood judgments.

International crises provide a familiar concept for the analysis of peacetime Soviet naval intentions. Soviet naval events during these periods, especially those which take place in proximity to the crisis area, may be encoded as crisis-related. An example of this, reviewed in Chapter 3, was the proximity of a Soviet naval exercise in 1970 to the Libyan coup. This was perceived by some analysts as indicating a Soviet intention to support the coup makers.

Similarly, a Soviet exercise held in proximity to the "Cod War" between Britain and Iceland in 1973 has been interpreted as related to that crisis (Dismukes and McConnell, 1979: 355; Dismukes, SNI: 493-4) even though no Soviet goals for involvement in the crisis were evident, and the exercise took place in a traditional Soviet exercise area in the traditional exercise season. The exercise had no clear relationship to the British or the Icelandic activity. During the "second Cod War" the Soviets conducted no exercises in the area, but this occurred in November rather than the spring exercise season.

Events which seem to be causes of other events are likely to be more easily perceived and memorable than events which have no causal significance. In a later section evidence will be reviewed that suggests causal scenarios are readily imagined to explain effects, and are readily perceived in an array of data. This ease of generating causal explanations may be taken as an index of the aptness or likelihood of such scenarios as explanations.

The variety of explanations proposed by naval analysts for Soviet peacetime naval developments (Chapter 3) suggests how easily plausible causes can be perceived in complex data. Many of these proposed explanations are partially or wholly inconsistent with each other. It seems likely that excessive causality is perceived in the Soviet Navy's influence on the U.S. Navy during crises (e.g., the 1967 or 1973 Middle East Wars). There is a strong tendency by some naval analysts to accept the Soviets' own claims that Soviet naval presence prevents U.S. naval intervention during these crises. Some analysts even suggest that Soviet naval forces off Angola in 1975 anticipated a U.S. naval intervention and forestalled it before it was even attempted. To be sure, the Soviet Navy may be the "cause" for U.S. nonintervention, or it may make nonintervention (decided on for other reasons) an easier policy to follow (i.e., the Soviet presence facilitates, but does not cause, nonintervention). But the Soviet presence may have no effect on the decision not to intervene, or even a negative effect (i.e., some may see the Soviet presence as an unjustified prohibition that should be challenged to demonstrate U.S. resolve and freedom of action). It cannot be demonstrated on the evidence to date that the Soviet Navy has caused U.S. nonintervention in any crisis, yet this perception persists.

The Representativeness Heuristic

This second judgment strategy described by Kahneman and Tversky (1972, 1973; Tversky and Kahneman, 1974) involves the use of salient features of a sample of data to infer its membership in some data category. This strategy is a relatively simple application of resemblance or "goodness of fit" criteria to categorization judgments. A particular case is examined for salient features, which are used to infer the membership class of that case.

The use of the representative heuristic leads to accurate judgments so long as the salient features of the sample are true reflections of the population from which it was drawn. However, many samples from a population will be atypical on the salient features, and virtually all samples will be atypical on at least one feature (i.e., the sample case that is exactly average on all dimensions; like the "average" scientist who is 33.3 years old, married to .95 of a wife, with 2.1 children, is an extremely improbable, if not impossible, case).

The problem with the use of the representativeness heuristic is the tendency it produces to overlook sample information that is unrelated to sample features, but is related to the base-rate of that sample, i.e., the likelihood of that particular sample in the population. For example, the salient features of iron pyrite are quite similar to those of gold. People who mistake "fool's gold" for the real thing are placing too great a weight on the specific features of the sample (its representative gleam, color, shape, etc.) and too little weight on the base-rate data (the relative abundance of iron pyrites and the scarcity of gold) which is unrelated to the sample features. The representative heuristic applied in this case leads to the conclusion that "what glitters is gold." Sometimes this is true, but as the wise know, not often -- appearances can be deceiving.

An example of the representative heuristic can be found in Vigor's (SNP: ch. 32) analysis of Admiral Gorshkov's discussion of "command of the sea" [gospodstvo na more]. Vigor writes (p. 619)

> . .. the way in which he elaborates it . . . seem(s) to me to provide a valuable clue to some likely paths of future development of Soviet naval strategy.

In other words, the salient features of Gorshkov's elaboration of "command of the sea" appear (to Vigor) representative of future Soviet naval strategy. Vigor assesses Gorshkov as supporting "command of the sea" for the Soviet Navy and advocating command over principle theaters of operations, and an ability to stop the enemy's navy from carrying out operations.

MccGwire (SNP, ch. 33: 633-34) questions Vigor's reading of Gorshkov primarily on base-rate grounds:

> The term is used relatively infrequently [by Gorshkov and others] and almost wholly in its prejorative (Mahanist) sense, or else descriptively . . .. In Soviet usage, "command of the sea" does not carry the emotive baggage that it has in the English language. . . the Soviet Navy is more likely to speak in terms of "defending the country's maritime frontiers," "concentration of forces," or "combat stability."

MccGwire concludes (p. 632 33) that the Soviets fail to see any practical relevance to modern warfare in the concept of command of the sea. Nor do they see it as realistic given modern weapons that can create a vast "no command" zone. Since command of the sea "is not a major factor in Soviet naval strategy," according to MccGwire, there is little point to consulting elaborations on it for future Soviet naval strategy developments. In MccGwire's view the concept is uncharacteristic of the Soviets, and Soviet strategy of the future is unlikely to include it. It appears representative only if one overlooks the general character (base-rate) of Soviet strategy.

Naval analysis provides a strong opportunity for the misguided application of the representative heuristic because of a prevalent tradition of case study and class analysis. Case studies include detailed analyses of crises, theaters, client-patron state relations, etc. Class analysis focuses on Soviet submarines, aircraft carriers, surface classes, aviation, etc. The danger is that salient features of a particular case or class will be taken as indicative of the membership of that case, and the base-rate likelihood of such membership will be neglected.

Crisis analysis may be extremely susceptible to the representative bias. Crises are rare and infrequent and have very little in common with either day-to-day peacetime missions or wartime missions. Naval forces designed for flexible execution of peacetime and wartime missions may perform crisis tasks quite well, generating the perception that the forces were designed for crisis purposes. (For example, U.S. helicopters and aircraft carriers in the South China Sea performed the evacuation of Saigon quite well but it would be a mistake to attribute this successful flexibility to some premeditated design of forces for just such crises.) Crises always carry a heavy burden of the unknown and unanticipated; neither of the great powers can be said to have effectively anticipated or planned for any major Third World crises.

Since many naval forces never practice their wartime missions realistically even in training (e g., only one live warhead has ever been fired from a U.S. ballistic missile submarine), and since peacetime missions are often inconspicuous, crises performances may be highly salient but greatly misleading in the analysis of the intentions guiding those naval forces.

It has been argued that surface anticarrier warfare (ACW) is a diminishing feature of Soviet naval forces, in terms of construction, exercises, and doctrinal writings. The use of surface ACW units still seems to occur, however, during crises. It is unlikely that surface ACW is representative of the Soviet Navy in any tactical sense. Surface ACW may be unique to great power naval crisis demonstrations and may no longer signify a Soviet resolve to prevent U.S. intervention in crises (if it ever actually did).

Representativeness and availability. The availability and representative heuristics may combine, as Nisbett and Ross (1980: 26) write:

> . . . a person who is required to account for some observed action or outcome may search the list of available

antecedents for those that seem to be the most representative "causes" of the known "consequences."

There is a very strong tendency to match causes with consequences: motivational causes are given for events with strong motivational or emotional consequences; complicated, multifaceted causes for complicated events. The fallacy of identity, that causes must resemble effects, is described in a later section. This tendency has also been termed "magical thinking" by Shweder (1977).

The combination of these two heuristics may arouse mental theories and schemas to help organize perceptions. Theories are fairly explicit propositions about objects and classes (e.g., Soviet forward deployments are part of Soviet naval diplomacy). Schemas are less propositional and more schematic (e.g., the meaning of "defensive"). Perceptions are readily assimilated into pre-existing impressions, beliefs, and theories. In later sections we discuss the impact of these preconceptions on estimates of covariation, causality, and prediction. At the stage of perception, the assimilation effects of theories and schemas means that data that do not fit the theory may not even be perceived, and are thus unavailable at later estimation stages. The lack of much analysis by the U.S. naval analysts of the Indonesian case and the impact of China on the Soviet Navy is consistent with the lack of theories or schemas that encompass these data.

The arguments among naval analysts on whether Khrushchev secretly favored surface ships and whether Admiral Gorshkov saved surface ships from the breakers' yards seem to reflect the combination of availability and representativeness heuristics evoking a causal theory which affects the perception of data. As discussed in Chapter 3, Case 1, the claim was made in 1967 by Gorshkov that the Soviets decided in 1954 to create a balanced "ocean-going" fleet capable of protecting state interests abroad in peacetime. This "theory" serves Herrick (1968: 71) as an explanation for Gorshkov's cautious campaign to save the surface ships despite Khrushchev's distaste for them. The same theory serves McConnell (in Dismukes and McConnell, 1979: 13) as a basis for the thesis that Khrushchev was less deprecating of surface ships than he appeared to be, and that in 1954 Moscow set in train the plan for power projection in the Third World through naval diplomacy (i.e., Gorshkov can be taken at his word).

McGwire (SND: 189-194) challenges both Herrick's and McConnell's conclusions, citing inconsistent data -- the Soviets stopped large surface ship programs in 1954 and did not resume them for some years. The decisions for forward deployment were made for strategic purposes, not as part of a naval diplomacy plan. Gorshkov had little visible influence over the Party's decisions to cut back surface ships and even today he is still attempting to create a balanced, ocean-going fleet. McGwire interprets Gorshkov's 1967 remarks as hindsightful, post hoc rationalizations designed to demonstrate Communist wisdom in foreseeing the future, and to put a favorable face on the open-ocean missions thrust on the Soviets by Western strategic capabilities. In effect, McGwire is arguing that the perception of Gorshkov as the savior of the surface ships, and Khrushchev as a crypto-supporter of them, is due to Gorshkov's conveniently providing (in 1967, after the Admiral knew how events turned out) an available and representative explanation that seemed to fit events

(as they appeared in 1967) but which is actually inconsistent with the course of events from 1954 through the early 1960s.

Assigning Weights to Data

Analysts and estimators not only perceive data, they must decide which data are significant and important, and which to ignore. As was noted earlier in this chapter, a major source of error in forecasting, perhaps the major source, is the set of key assumptions made by the forecaster (Ascher, 1978: 199). Key assumptions can be thought of, in part, as decisions the forecaster makes about important dimensions or variables. Once the key variables are decided, the forecaster collects data on these variables, and not on others. Ascher writes (p. 199)

> The core assumptions underlying a forecast, which represent the forecaster's basic outlook on the context within which the specific forecasted trend develops, are the major determinants of forecast accuracy.

There is more to the concept of "core assumptions" than just weighing data; e.g., the forecaster must also decide which variables are related to each other and in what ways. But weighing data and deciding on which variables and dimensions to attend are important elements of the key assumptions analysts and estimators make.

As our "mini-reviews" in chapter 3 suggested, naval analysts weigh variables differently. Some analysts (e.g. MccGwire) emphasize the Soviet Navy's ship construction practices and use the size, shape, and capabilities of the vessels as clues to Soviet plans and purposes. To the degree that Soviet ships, submarines, and aircraft are constructed for narrowly defined tasks, and have limited mission flexibility, these patterns and trends provide a useful indicator of at least the major anticipated wartime missions of the Soviet fleet. The "hardware" method is less predictive when the Soviets use their ships to accomplish missions for which they were not designed, or for missions that were not anticipated when the ships were built. Similarly, if the Soviets engineer more ships for greater mission flexibility (e.g., the Kiev class), the hardware methodology will be less successful in helping rule out possible missions.

In contrast to the "hardware" analyst, who puts great weight on the physical stuff of the Soviet Navy, "software" analysts (e.g., Hudson, SNO: ch. 21, 1976) examine the verbal behavior of the Soviet Navy's spokesmen, namely their naval doctrine, for indicators of future naval operations. Whose statements actually constitute doctrine is a major analytic question, as the debate over the authoritativeness of Admiral Gorshkov's writings demonstrates (see chapter 3, case 4 for a review). Analysts who have decided Gorshkov is expressing doctrine tend to take him at his word with respect to the future directions of the Soviet Navy. Estimators who read Gorshkov as an advocate take the view that Gorshkov's prophecies and intimations can be expected to occur only if he wins his case with the USSR's defense and political leadership. In other words, actions and events, not Gorshkov's statements, will determine whether the statements were predictive. Some analysts (e.g., Gallagher, SNI: ch. 3) view doctrine as a very slippery guide to the future, even when it is

known to be authentic. Gallagher (p. 56) sums up the problems with weighing doctrine too heavily as an indicator of future Soviet naval behavior:

> [doctrine] is a charter attesting the military's right to participate in policy formation and general authorization for a broad and comprehensive development of the armed forces. To interpret it as a key to Soviet strategic policy would be both to misread the doctrine and to underestimate the dynamics of the Soviet decision-making process.

There are a variety of nonobvious psychological problems that may occur when analysts assign weights to data. Analysts may believe they weight heavily data which they actually attend to very little. Analysts may over-weight data which are (a) from small samples; (b) apparently relevant (e.g., causal), vivid or highly salient; (c) based on a case (rather than on a base-rate); (d) from certain evaluative (e g., good-bad) dimensions; or (e) single dimensional. In this section we review these problems and examine how they apply to Soviet naval analysis.

Knowing more than you can know -- introspection on data weights. People who solve problems, make judgments or decisions, or form estimates can be asked to report what factors or data dimensions most influenced them. Alternatively, while people are solving a problem, making a decision, or forming an estimate, they can be asked to "think out loud," and thus provide a record of the factors they considered. These techniques are known as "process tracing." Psychologists have also developed the technique of presenting people with a problem, decision, or estimation task with carefully controlled variations in input data. For example, judges may be asked to evaluate the nutritional quality of breakfast cereals with carefully varied levels of vitamins, minerals, calories, fats, etc. By comparing the judgmental outputs with the data inputs, psychologists can determine the degree to which judges weighed the data, without asking the judges. This technique is termed "policy capturing," "decision modeling," or "linear regression" (after the mathematical technique used to determine how the data relate to the judgments).

A significant difference emerges from studies of judgment when the process-tracing technique is used and when linear regression techniques are employed. The data that the linear regression method shows are most heavily weighted are sometimes not the data people report as most heavily weighted. In fact, people occasionally report being influenced by data which are mathematically unrelated to their decisions, or, alternatively, people report ignoring a set of data that, in fact, can be shown mathematically to be the most important component of their judgments. In other words, people may have poor insights into their own problem-solving, judgment, or estimation processes. This problem becomes critical when the person attempts to improve these processes. Exactly when and how these discrepancies between (a) judgments and (b) reports on judgments occur are not yet well known (Einhorn and Hogarth, 1978) although they are under intensive discussion and investigation by psychologists (e.g., Nisbett and Wilson, 1977a, b; Smith and Miller, 1978; Wright and Rip, 1980).

Many studies have found discrepancies between "objective" or

"computed" weights given to data by judges (as determined by linear regression), and the "subjective" weights reported by the judges when asked to describe the data they used in the task (see Slovic and Lichtenstein, 1971: 683-84 for a review). Judges strongly overestimate the importance they place on minor variables (in terms of computed weights) and underestimate their reliance on a few major data dimensions. People are generally quite unaware that their judgments in many multi-cue estimation and decision-making tasks can be predicted using only a few dimensions of the input data. Slovic and Lichtenstein (1971: 684) report:

> Across a number of studies, varying in the number of cues that were available, three cues usually sufficed to account for more than 80% of the predictable variance in the judges' responses. The most important cue usually accounted for more than 40% of this variance.

The tendency to believe that more data dimensions or cues were used than actually were to make an estimate or judgment may contribute to overconfidence, a phenomenon discussed later in this chapter under theory maintenance and change.

An interesting sidelight on the lack of self-insight on data weighting policies was provided in a study of 13 stockbrokers (Slovic, Fleissner, and Bauman, 1972). The accuracy of these brokers' self-insights into their policies for recommending investments was measured by correlating a broker's subjective weights with his computed weights across eight cue factors. The correlation between this self-insight accuracy score and stockbroker experience (number of years) was -.43; i.e., brokers with greater experience had less insight. Slovic, et al., suggest that less experienced brokers may be following the evaluation strategies they were taught more self-consciously, and giving greater attention to the mechanics of judgment. With more experience, these skills become more automatic, demand less attention, and may become harder to describe accurately. On the other hand, experience should also increase the availability of plausible decision strategies which the brokers may report (inaccurately) as their own.

Nisbett and Wilson (1977a) labeled this lack of self-insight on judgment policies "telling more than we can know." They suggest that there may be little or no direct introspective access to higher order cognitive processes: people may be unaware that certain data have an important influence on estimates. Instead of reporting what actually influences judgments, they may report a priori, implicit causal theories or judgments about the extent to which a particular fact is plausible evidence for a given estimate. While people may have no direct access to their estimation cognition, they may nevertheless accurately report about them when the "social theory" conforms to the "cognitive process." As Nisbett and Wilson wrote (p. 231):

> Accurate reports will occur when influential stimuli are salient and are plausible causes of the responses they produce, and will not occur when stimuli are not salient or are not plausible causes.

(We will have more to say about salience and data weighting later.)

Nisbett and his colleagues (Nisbett and Bellows, 1977; Nisbett and Wilson, 1977a, b) reported a variety of psychological experiments that demonstrate how people's judgments can be influenced significantly by data without their awareness of this influence. In some cases the direction of influence may be opposite to that believed by the person making the judgments. Clearly, if people do not accurately know why they reach an estimate it will be difficult for them to change or improve their estimation process.

The Nisbett and Wilson (1977a) thesis remains controversial (e.g., Smith and Miller, 1978) and seems only to occur sometimes, i.e., sometimes people do seem to report more than merely a "social theory" (e.g., Wright and Rip, 1980). Psychologists who utilize the process-tracing method are particularly reluctant to accept the idea that they are investigating social theories rather than studying estimation or judgment processes (e.g., Svenson, 1979: 97-8).

A second factor determines the ability of process-tracing to capture the actual cognitive process of the judge: the ability of the process-tracing experimenter to adequately translate the judge's verbal reports into a process-trace (Einhorn, Kleinmuntz, and Kleinmuntz, 1979: 476). The task of process-tracing itself involves judgment and possible error, and while the judge's reports may faithfully reflect mental processes (contrary to Nisbett and Wilson's thesis) experimenters may be unable to translate these statements into a process model that accurately predicts judgments. Einhorn, et al., note that highly predictive process-tracing models of a judge's estimation process will occur only if the judge's insight is accurate and the process-tracing experimenter is able to accurately capture that insight. If either insight or process-tracing ability are low, the process-trace will have only partial predictive success. If both are low the process-trace will not predict judgments.

Nevertheless, the concept of inaccurate self-insights into data weighting has experimental support and serves as a partial explanation for conflicts in naval estimates. That is, naval analysts who examine similar or identical data dimensions (e.g., Admiral Gorshkov's writings, or Soviet deployments in the Indian Ocean) may reach very different conclusions and predictions as we saw in Chapter 3. A partial reason for these disagreements may be the analysts' inability to determine exactly which cues they weighted most heavily in their estimates. Two analysts, who both believe they gave the same weights to the same cues, may reach different estimates because their self-insights into cue weightings are inaccurate.

In addition, a Soviet Navy analyst may be influenced by the increased salience of a cue that previously seemed unimportant, but may be unaware of his new attentiveness and sensitivity to the cue. The analyst may thus perceive a "change" in Soviet behavior where none in fact exists.

The structure of the data an analyst uses to make an estimate may influence how those data are weighted; even when these structural characteristics should be irrelevant to cue utilization they seem to have strong effects on cue weights. These structural aspects of data include correlations between cues, cue variability, number of cues, and cue-response compatibility (see Slovic and Lichtenstein, 1971: 685-688 for

experimental studies and references).

Redundant cues. When people have important cues which agree, they tend to weight them equally and use both. When the cues disagree people tend to focus on only one or to use other cues to resolve the conflict. High agreement between cues is a major factor in increasing confidence in predictions based on these inputs. For example, people express more confidence in predicting the grade point average of a student who received all B's than in predicting the average of a student whose record included many A's and C's (Kahneman and Tversky, 1973). However, highly consistent data patterns are most often observed when the data input variables are highly redundant or correlated. If two data variables are perfectly correlated they are totally redundant and knowledge of one is sufficient. If two perfectly redundant variables are weighted equally and are both used in making predictions, the estimator has, in effect, counted a single piece of information twice. People tend to have great confidence in predictions based on redundant variables. However, a prediction based on several variables can achieve higher accuracy when the data variables are independent of each other than when they are redundant and correlated. Redundancy among data cues decreases accuracy as it increases estimator confidence (Tversky and Kahneman, 1974).

Relying on highly correlated cues seems to be a widespread problem in Soviet Navy analysis. We have already referred to the tendency of estimators to rely heavily on a particular set of cues, e.g., Soviet "hardware" construction. Such reliance is necessary, to some degree, because of the perception and coding of any set of cues (hardware, software, crisis behavior, etc.) requires specialized data gathering and recording, interpretation and analytic expertise, and so on. The advantage of this specialization is that the estimator gains a firm understanding of the meaning of the data set, and of its strengths and weaknesses. The danger of the specialization is that the estimator is likely to develop a great degree of confidence in judgments which rest on highly correlated and redundant data cues. That is, data on various dimensions of Soviet naval construction, for example, are likely to be highly correlated and consistent (e.g., they may be largely determined by economic factors), and the analyst will tend to give these cues equal weights and use them all in making predictions. This confidence in prediction is inappropriate, however, since the use of independently generated data sets would lead to more accurate predictions (e.g., if "hardware" factors resting on a single generating factor such as economics were integrated with crisis behavior resting on a generating factor such as diplomacy).

Analysts take two major steps to offset this shortcoming. First, they often go outside their specialities and examine independent data sets (i.e., the specialist in crises behavior will consider construction patterns). This can (and does) lead to accusations that the interloper is insensitive to the nuances of data outside his speciality. It can also lead to the following judgment process: a tentative hypothesis is formed based on one's own speciality, and then independent data sets are consulted for confirming evidence. The shortcomings of this process are discussed later in more detail; the major problem being that the independent data have not been appropriately integrated.

The second step naval analysts take to offset the limitations of specialization is to exchange work and to comment on each other's data, methods, and conclusions. This exchange seems exceptionally high for a field of intelligence (which is often characterized by tight compartmentation) and seems to produce a remarkable level of technical sophistication among naval analysts. The shortcoming of this step is that much of the exchange focuses on conclusions rather than on methods and data. Even here, however, there is a high level of sophistication; naval analysts are not unwilling to undertake a complete replication of another's analysis and to announce contrary findings as a result (e.g., McConnell, SNI: 612-13; and Weinland, SNP: 547ff).

[I do not mean to suggest the sophistication implied in the above is unique to naval intelligence. In part, this observation is due to the biased sample of estimates examined. Most of the authors and all of the editors of these estimates have close ties or direct affiliations to the Center for Naval Analysis or to Dalhousie University's Centre for Foreign Policy Studies. These two institutions in turn are closely connected through personal and semi-official ties, leading to a close-knit community of analysts, who have frequent and strong disagreements, and who undertake careful and scholarly debates. This situation and process are undoubtedly the best antidotes for overconfidence in improperly weighted data, as well as many of the other biases discussed here.]

A simple step which analysts should take but rarely do, is to estimate data set correlations, either mathematically in the case of quantitative data and data that can be scaled, ranked, or rated; or impressionistically for qualitative data. These correlations offer an index of the redundancy in the data sets, and provide a signal that excessive confidence in predictions may be inappropriate. [There are problems with impressionistic correlation estimation that are discussed later.]

For example, ship days in a region, port visits, number of ships in region, transits through straits, support ship days, are data sets related to "showing the flag" and naval diplomacy. These variables often tend to be correlated (e.g., Weinland, NPSP). For example, the pattern of correlations in Table 4.1 shows that, for the Indian Ocean, the three data sets are highly correlated (and hence redundant): port visit length and number are redundant, and the more total ship days the fewer and shorter the port visits. Half of the twelve variable pairs in Petersen's Table are statistically correlated.

This should not be taken to suggest that Petersen errs in reporting correlated data sets, only that an explicit awareness of redundancy would help offset inappropriate data weightings. For example Petersen writes (Dismukes and McConnell, 1979: 91) that

> . . . even though . . . Indian Ocean deployments (as measured in ship days) have been consistently smaller than those to the Mediterranean, the effort devoted to diplomatic visits . . . has, just as consistently, been proportionately higher. The relative intensity of the Indian Ocean effort . . . would appear to confirm the prominence of political concerns in the Indian Ocean squadron's mission structure.

This observation is not consistent with Petersen's data, which show that

Table 4.1. Correlations between Data Sets on Trends in Diplomatic Port Visits, 1967-1976[a,b]

	Atlantic			Mediterranean			Indian			Pacific			All		
Data Set	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
a. Number of port visits	—			—			—			—			—		
b. Total ship days in region	.71	—		.34	—		-.65	—		.40	—		.49	—	
c. Ship days in port per 1000 ship days in region	.44	.22	—	.21	-.77	—	.64	-.83	—	.88	.35	—	.00	-.74	—

[a] Computed from Patterson (in Blanchard and McConnell, 1979: 92) Table 3.2.

[b] r >.63, p <.05
>.77, p <.01

the more total ship days spent by the Soviet naval squadron in the Indian Ocean, the fewer the diplomatic port visits it made and the shorter these visits were. As the Soviet Navy spends more time and effort maintaining a presence in the Indian Ocean, it spends less time and effort on diplomatic missions. In the other three oceans, and overall, Soviet total ship days are positively related to port visits.

[It is interesting that this pattern is unique to the Indian Ocean, although Petersen implies that similar patterns are observed in the Atantic, the Indian Ocean, and the Mediterranean. In the Atlantic, more Soviet ship days are positively related to longer port visits. In the Mediterranean greater total ship days are related to shorter port visits, but total ship days has a weak positive relation to number of port visits. Across all oceans, there is a strong negative relationship ($r = -.74$) between total ship days and ship days in port, and a weak positive relation between number of visits and total ship days ($r = .49$).]

Cue variability. Cues which vary more tend to be weighted more, relative to less variable cues. Similarly, an increase in the variance in a salient cue tends to produce a heavier weighting of that cue. This increased weighting of the cue may persist even in subsets of data for which the cue was not varying in any unusual way. Cue variation is an excellent basis for making differential predictions, but if carried over to subsets which no such variation is present, constitutes a serious bias.

There is a strong tendency for naval analysts to emphasize out-of-area deployments by the Soviet Navy (e.g., Petersen in Dismuke and McConnell, 1979; Murphy, NPSP: ch. 6) as evidence of Soviet naval diplomacy (e.g., McConnell and Dismukes, 1979; Dismukes and McConnell, 1979). Between 1964, when this forward deployment began, and 1974 these out-of-area ship days increased more than 3.5 times (from 13,758 to 50,700 ship days; Petersen in Dismukes and McConnell, 1979: 92). Since 1974, however, the total number of out-of-area ship days worldwide has remained constant and the total in each of the four ocean areas (Atlantic, Pacific, Indian, Mediterranean) has also stayed roughly constant (Murphy, NPSP: 127-134). This cue is still highly salient and while it demonstrated high variability from 1964-1974, the lack of variability since 1974 has not reduced the weight given this cue by many naval analysts.

Number of cues. Increasing the number of cues available to people tends to increase their confidence in the accuracy of their predictions, with no corresponding increase in actual accuracy. For example, Oskamp (1965) had 32 clinical judges read background information on a psychological patient. The information was divided into four parts and, after reading each part, the judges answered 25 questions about the attitudes and beliefs of the patient. They also gave a confidence rating with each answer. The correct answers were known to the experimenter. As the amount of information on the patient increased, accuracy on the questions remained constant, while confidence increased dramatically and out of all proportion to accuracy.

We have no means to judge the confidence of naval analysts in their estimates, short of a content analysis of their use of "probably" and "maybe" versus their use of "certainly" and "surely," a questionable technique at best. We can observe, however, that naval analysts in our

sample use a very large number of variables in making their judgments. As we noted in Chapter 3, the minimum error in their estimates on several crucial issues is large (i.e., wide disagreements imply that some of the analysts must be wrong, and, of course, they may all be wrong). Analyst confidence in judgments should not be based on the amount of information related in the estimate, but on the probative value of the information used. Neither should an estimate consumer take the amount of information to be an index of the estimate's accuracy.

Cue-response compatibility. The greater the compatibility between a cue and the required reponse (e.g., estimate or prediction), the more important that cue is likely to be in determining the response (see, e g., Hamilton and Fallot, 1974). Cues which require complex transformations to be useful will be less heavily weighted than cues which are readily related to the estimation response. This tendency reinforces the operation of representativeness heuristic, since cues will be weighted heavily which are representative of the estimation response. For example, if the estimation required is to predict the future number and tactics of Soviet attack submarines, the cues that will be most heavily weighted will tend to be those most compatible and representative: data on present and past attack submarine numbers and tactics. In general, this heuristic will provide adequate, but may lead to significant biases if cues which are unrepresentative in character are also important for the estimation response. For example, space surveillance of submarines and surface targets, and real-time retargeting of ballistic missiles may reduce the requirement for a large fleet of attack submarines; their missions may be taken over by aircraft and land-based systems. Alternatively, if SNA aircraft were to become more vulnerable, more attack submarines might be required. Neither the space surveillance cue nor the SNA cue are as compatible with the estimation response as the more representative submarine cues. There may be the danger of underweighting the former, although their predictive value may be high.

Cue salience and vividness. Perhaps the most important determinants of cues' weights are the vividness and salience of the data they contain. Repeatedly psychologists have found that vivid, concrete information is weighted more heavily than abstract, statistical data (e.g., a single case study is more influential than a statistical summary of a large series of cases). This tendency means that some types of data are almost never weighted heavily, e.g., nonevents and nonoccurrences. It also leads to the over-use of "representative" rather than statistical samples, the reliance on too-small samples, and the use of inappropriate "relevance" criteria for data weighting (Bar-Hillel, 1978; Tversky and Kahneman, 1974). These problems all contribute to overweighting of data which are less reliable than other data that could be utilized.

Psychologists have only recently become aware of the degree to which salience influences the weight given data by judges and estimators (see Hamilton, 1979 and Taylor and Fiske, 1978, for reviews of the psychological literature). Data are salient which are distinctive, physically prominent, easily perceived or available, highly visible, novel, unique, singular or unfamiliar. None of these features of the data by themselves should lead the estimator to give them more weight than other, pallid evidence, yet there are strong tendencies to do so. Salient

data is given more attention (this was discussed above under "perceiving data"), salient events produce more extreme judgments, salient factors may be credited with playing a special or causal role in events, and estimators tend to make confident and extreme inferences on the basis of salient data (Hamilton, 1979: 59-64).

Nonevents. The use of nonoccurrence and nonevent data is uncommon, perhaps for two reasons. First, there is the tendency for behavior to "engulf" the perceiver, as Heider (1958) suggested, so that nonbehavior is just not as easily perceived, or categorized as behavior, and is thus not available to the estimator. Second to use nonbehaviors and nonoccurrences as inferential data, the observer must have a mental theory relating event A with event B. That is, only if event A is likely to occur in conjunction with critical event B (e.g., mobilizations, A prior to attacks, B), can A by its absence, provide diagnostic data as to B. However, people are not highly skilled at detecting covariations, as will be reviewed later, they tend to estimate them on the basis of positive instances only, thus they will tend not to be aware of the inferential utility of nonoccurrences, and will not weight them enough. Nonevents lack salience because they fail to fit into the estimator's mental theory of the covariation situation.

There have been several Soviet nonevents, however, which have prompted considerable discussion among naval analysts, especially after the nonevents became events. Three examples are aircraft carriers, forward deployment, and foreign bases. Each of these aspects of the Soviet Navy received some attention when they were nonevents, but received much more attention when the Soviets showed signs of changing their behaviors, i.e., by building their versions of carriers, by keeping ships out of home waters, and by seeking and building bases abroad. It is difficult to judge the degree that naval analysts as a group neglect nonevents; some weight nonevents very heavily. For example, Herrick (1968) saw the lack of Soviet aircraft carriers in the late 1960s as a clear sign that Soviet naval doctrine was highly defensive, with no presumptions or aspirations of sea control. On the other hand, the decade of Soviet Navy nonsupport for North Vietnam during its war with the United States (including the U.S. Navy) has prompted relatively little analysis and mostly puzzlement among naval analysts.

Case cues versus base-rate cues. One implication of the "representativeness" heuristic (reviewed above) is that estimators may weight data on specific cases heavily while paying far less attention to abstract statistics which cover many cases (Bar-Hillel, 1978, 1979; Kahneman and Tversky, 1973; Nisbett and Ross, 1980; Ross, 1977; Tversky and Kahneman, 1974).

For example, in outlining the "rules of the game" followed by the Soviet Navy in peacetime, Dismukes and McConnell (1979: ch. 7) weighted the Middle East crises of 1967, 1970, and 1973, quite heavily. Less heavily weighted it seems were the 1968 and 1969 crises off Korean waters, and the 1972 noncrisis in the South China Sea. In the latter cases the U.S. Navy posed a direct threat to Soviet clients and allies, but produced little or no Soviet naval confrontation. The former cases of naval confrontation (on behalf of much more tenuous Soviet clients) seem to be the primary basis for the "rules" that Dismukes and McConnell suggest the

Soviet Navy observes. They classify the latter crises as "Support of a Client against Western Great Powers" (p. 260) and the former as "Support of a Client Against a Western Client" (p. 267). However, their "rules" clearly place the Middle Eastern crises as cases of superpower confrontation as well (e.g., ch. 5, p. 288-89), while the Asian crises are not considered superpower confrontations (e.g., ch. 5). If Dismukes and McConnell selected as most representative those crises which are apparently confrontational (but which might lack much incentive or motivation for confrontation), while tending to weight less heavily crises where no confrontation occurred (but where the incentives and motivations to counter a Western threat or attack on a client or ally was much greater), as we are suggesting, then their "rules of the game" will tend to suggest the Soviets are more likely to risk a clash with Western navies than is implied by the total data base on Soviet Navy crisis behavior.

It is impossible to assess quantitatively whether naval analysts have overweighted specific cases and underweighted the total data base on Soviet crises. Nor is it possible to determine if it is better to weight heavily the worst cases (i.e., those most suggestive of U.S.-USSR naval conflict) which pose the most serious threat, or to weigh all crises more equally and thus determine the more probable nature of Soviet crisis behavior. Since there are significant dangers in either course of action (e.g., see Wohlstetter, 1979, on the dangers of underestimation of threats, and Stech, 1980a, on the dangers of overestimation), perhaps naval analysis would be better served by more explicit attention to the role of the representativeness heuristic and the tendency to underweight base rate, and an explicit assessment of both the worst (but unlikely) cases and the more likely (and less serious) base rates.

Illusory correlation. One important consequence of these salience effects is "illusory correlation," an overestimation of the frequency of co-occurrence of distinctive stimuli (Chapman, 1967; Chapman and Chapman, 1967; Hamilton and Gifford, 1976). Salient and distinctive events may seem to co-occur simply because they are both distinctive, rather than because they are actually correlated. For example, both laymen and psychologically sophisticated clinicians believed that psychiatric patients who exaggerated or distorted the eyes in drawing were likely to be paranoid; that those who emphasize the mouth or drew feminine or childlike figures were dependent; that those who drew muscular, broad shouldered, manly figures were impotent, and so forth. In fact, none of the "symptoms" noted in the drawings has any valid clinical correlation with the disorders mentioned. The widely perceived correlations were empirically invalid.

In researching this phenomenon, Hamilton and Gifford (1976) found that uncommon group members (e.g., minority members) tend to be associated with uncommon behavior, even when no empirical correlation was present. Salient stimuli tended to be psychologically associated and people overestimated the frequency of their co-occurrence when no correlation existed. Data salience may generate the perception of illusory relationships.

The illusory correlation bias would be less troublesome if people were able to efficiently evaluate the strength of a correlation. That is, while an estimator might mistakenly perceive a relationship between two highly salient events, no damage would be done if the estimator could

accurately and quickly assess the strength of the relationship. Finding a low correlation, the estimator could modify his/her perception. Unfortunately, the estimation of correlation is a highly biased and difficult task, as is discussed below.

Furthermore, people do not seem to be able to efficiently check their hypotheses against the available data, and are able to find support for wrong hypotheses in almost any set of data. This weak hypothesis testing ability, also discussed in detail below, will tend to allow perceptions of illusory correlations to persist despite data which would invalidate them.

The hypothesis that the Soviet Navy knowingly supported the Libyan coup, which has been discussed above and in chapter 3, seems to be an illusory relationship prompted by the proximity of two highly salient events, a Soviet naval exercise and the successful coup. It is also apparently quite resistant to contradictory and inconsistent evidence.

More generally, the perception among Western analysts that the Soviets actively foster and support coup attempts by socialist, auti-Western factions seems just as strong as the communist perception that the United States sponsors coups by rightwing, anti-communist factions. While this perception may be true in specific instances (e.g. Afghanistan, 1978; Iran, 1953) it is probably an illusory correlation in general. The glee and support with which both sides greet coups they favor does not imply, although it is taken to by the other side, complicity before the fact.

Vivid cues. Vivid, concrete information, as is commonly found in case studies, tends to be used to a greater extent and is given greater weight than statistical or abstract data (Kahneman and Tversky, 1973; Nisbett, Borgida, Crandall, and Reed, 1976; Nisbett and Ross, 1980). This bias is dangerous because it amounts to weighting a sample of one heavily and a sample of many not at all or only lightly. Only in the rare event that the case study is a perfect predictor of the entity being estimated should case data be utilized to the neglect of data based on many cases.

Several characteristics of data can make them more vivid, none of which add to the diagnostic, predictive, or probative value of the data. Nisbett and Ross (1980: 45-51) note three: emotional interest, concreteness and imagery provocativeness, and spacial, temporal or sensory proximity. Taylor and Fiske (1978: 259-51) note brightness, motion complexity, and novelty.

The emotional impact of data often influence estimations based on them. For example, Walster (1966) found people attribute more responsibility to a person involved in a serious accident than in a trivial one, although behavior was identical in both cases.

Case studies of U.S.-Soviet naval confrontations that almost led to conflict are likely to be far more emotion-laden than cases with insignificant outcomes. In both instances the naval behaviors may be quite similar, and the different outcomes the result of diplomatic factors, client behaviors, or other aspects of the nonnaval context.

Concreteness refers to the level of detail and specific information about actors, actions, and contexts. Greater detail aids in the "imaginability," i.e., the tendency to prompt sensory images. Events may be highly concrete, while statistical data are far less so, and nonevents (see above) have virtually no concreteness at all.

Information proximity increases its vividness. Firsthand observation is more vivid and weighted more than data from second- or thirdhand sources. Events which are 'close to home' are more vivid than more peripheral occurrences. Recent events are more vivid than the 'distant past.'

There are several psychological reasons why vivid data receive more weight, none of which have anything to do with such data being more diagnostic or predictive. Nisbett and Ross (1980: 51-59) discuss four: memorability, redundancy, recruitment, and rehearsal.

Concrete and imaginable data are more easily stored and recalled from memory than abstract information. Imagery in particular seems to accentuate memorability. More memorable data are more available for recall, which may facilitate the operation of the availability heuristic. That is, more vivid events should be recalled more easily and judged as more frequent than less vivid events. This is what has been found in studies of the perceived frequencies of causes of death and lethal events (Combs and Slovic, 1978; Lichtenstein, Slovic, Fischhoff, Layman and Combs, 1978). Causes of death which were easy to imagine, easy to remember, well publicized, or contained features which increased their salience were overestimated (i.e., perceived to cause more deaths than they actually did). Less vivid causes were underestimated. Because the factors increasing vividness of causes were unrelated to the actual lethality of the causes, the impact of vividness on data weights was to increase estimation errors.

Carroll (1978) has shown that imagining a specific case in detail tends to increase the perceived likelihood of the case and may make the specific case seem more likely than the class from which the case was drawn, a mathematical impossibility. That is, imagining an automobile accident on a nearby street corner tends to increase the estimated likelihood of the accident, and may increase that likelihood beyond the perceived probability of an accident on the entire street.

Vividness implies greater detail and hence more information. While more facts may be presented in a vivid, concrete case study than in a statistical abstract, many of the facts in the former are highly redundant and add little or nothing to the estimation or prediction task. Such redundancy, however, facilitates encoding and recall from memory. Consequently, redundant information may be more available from memory than more relevant facts, but far less useful for estimation and prediction.

Because vivid, concrete information is stored in memory on a multitude of dimensions (due to its rich detail), and accessible from memory via many pathways, on recall it is more likely to recruit mental theories or schemas, as well as specific details. These mental theories themselves are linked to additional images and stored episodes which can be readily recalled to reinforce the specific details of the vivid case. These theories may provide a convenient "explanation" of the case, further reinforcing the conception of the case as a valid, confirming instance of the theory. Abstract information is less likely to recruit an organized mental theory, or fit so readily into an explanatory framework.

Vivid data are likely to remain in thought for longer periods after it is received. Tesser (1978) has found that the longer a person thinks about a case, the more extreme the evaluation of the case becomes (see Taylor and Fiske, 1978: 265, for additional evidence). Thus, vivid cases by remaining in thought longer, will tend to prompt more extreme

evaluations. Nisbett and Ross (1980: 55) suggest people may weight vivid information more heavily simply because they find themselves thinking of it more: 'if it weren't important, why would I keep thinking about it?' To the extent that estimators and analysts assume that what occupies their thoughts must be important for their estimates and predictions, they will be vulnerable to an overweighting of vivid information.

The impacts of vivid cases on analysts of Soviet peacetime naval intentions are present throughout the sampled estimates. For example, analytic concern with the peacetime role of the Soviet Navy can be dated fairly precisely; prior to June 1967 there was little concern over the peacetime functions of the Soviet Navy. After the Six-Day War, and the Soviet Navy's involvement in the crisis, analytic concern developed and flourished. In fact, the sinking in that war of an Israeli destroyer by a Soviet antiship cruise missile fired from an Egyptian patrol craft can be viewed as the single, vivid event which jarred Western analysts into reconsidering the threat posed by the Soviet Navy in war and peace, and the threat of Soviet-equipped client states.

[A 'nonevent' of a similar nature makes an interesting comparison. In the 1973 October War, Egyptian and Syrian craft repeatedly fired Soviet-supplied antiship cruise missiles at Israeli vessels without as much as a serious hit, much less any sinkings. Israeli missile boats sank 19 Arab vessels, including ten Soviet-supplied Arab missile boats, without a single Israeli loss (Safran, 1978: 166). The Soviet-built weapons produced no positive effects and this nonevent, lacking the vividness and emotional saliency of the 1967 sinking, largely escaped attention, at least in the West. On the other hand, the success in 1973 of Soviet-built anti-tank missiles against Israeli armor, and of Soviet supplied surface-to-air missiles against Israeli jets, two extremely vivid events provoked even more Western comment than the 1967 sinking.]

Taylor and Fiske (1978: 251) provide a cogent summary of the psychological difficulties of weighting data:

> . . . instead of employing base rate . . . information logically, people are often more influenced by a single colorful piece of case history evidence . . . Instead of using correlational evidence appropriately, subjects' subjective estimates of correlation magnitude are often determined largely by positive instances . . . Instead of reviewing all the evidence that bears upon a particular problem, people frequently use the information which is most salient or available to them, that is, that which is most easily brought to mind.

Having opened this section with Ascher's observations on the centrality of core assumptions to the accuracy of forecasts, it is fitting to close with his remarks (1978: 160) on a major source of consistent inaccuracy in forecasting:

> . . . forecasters believe their assumptions about the future already incorporate all the data that is pertinent to known trends. It is another case of forecasters being locked into a scientific outlook, which acknowledges only information

relating to the phenomena studied rather than information on the behavior of the experts studying them.

Characterizing Data

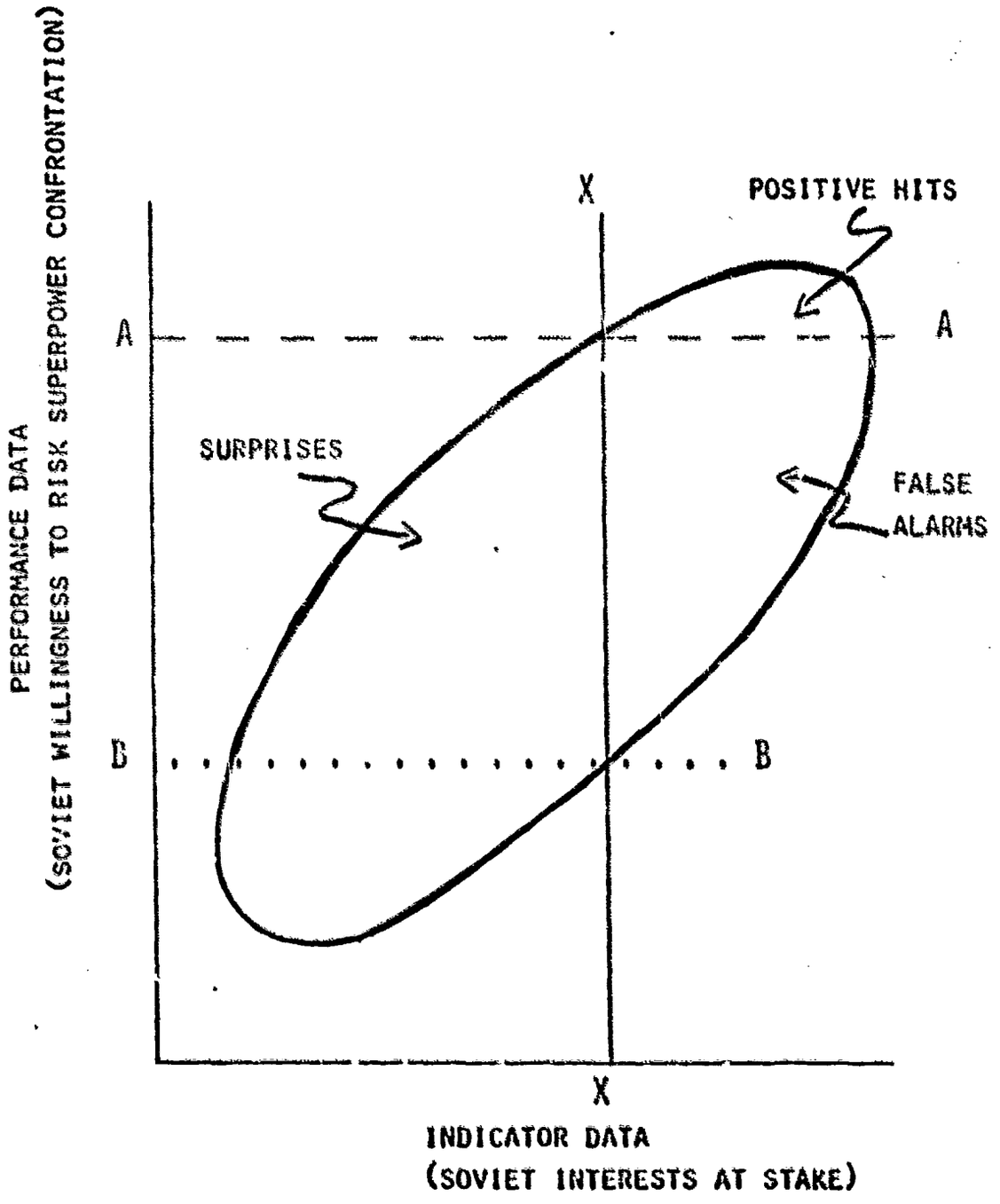
As data are perceived and weighted, they are characterized perceiving involves an act of categorization. That is, people attach a description to an individual piece of information to help code it for memory and later inferential work (e.g., such a categorization might be Soviet "tattletale" shadowing of U.S. ships). They characterize data aggregates or samples (e.g., Soviet naval behavior in the 1973 Middle East crisis), i.e., grouping facts together. They generalize from samples to make inferences about a population (e.g., Soviet naval crisis behavior). Each of these data characterizing tasks is a necessary step in inference. The first two steps are basically descriptive while the last is inferential, but all steps in categorizing or characterizing data involves some inference -- the data never 'speak for themselves' even in so simple a matter as description; some inference about categories is also necessary.

Perhaps the oldest idea in psychology and epistemology is that mental theories (or preconceptions) influence the coding and interpretation of data. One of the commonest and most dangerous misconceptions people have is the assumption that they approach data without preconceived notions, without theories. Wasserman (1960) labeled this tendency in intelligence analysis "naive realism." the belief that knowledge and intelligence consists of "unvarnished" facts which allow only one interpretation. Halsman's (1956) analysis of strategic intelligence found the widespread notion that knowledge was "facts divorced from thought or interpretation." There are a variety of damaging consequences of this naive faith in "plain facts" (see Steele, 1979: 54-58 for an analysis of these consequences and intelligence failures), one of which is the vulnerability of believers to their own and others' preconceptions and mental theories. Those who believe the "raw" facts speak for themselves will tend not to check carefully how those facts were categorized or interpreted. As we will demonstrate in this section, data can be characterized improperly, or organized into categories that make subsequent inferences difficult to make.

Problems caused by characterizations. Some of the difficulties that result from characterizing data can be demonstrated with Figure 4.2. Suppose there is a strong positive relationship between Soviet diplomatic interests in the Third World and Soviet willingness to risk a superpower naval confrontation in crises when those interests are threatened (as has been suggested by naval analysts, e.g., Dismukes and McConnell, 1979: ch. 7). The horizontal axis shows the indicator variable, Soviet interests increasing from left to right. The vertical axis shows the performance variable, Soviet risk-taking, increasing from bottom to top. The strong (but not perfect) relationship between these two variables is indicated by the ellipsoid shape; all Soviet behaviors fall inside the ellipse.

Suppose an analyst characterizes all instances to the right of the vertical line, X, as "vital interests at stake " those to the left as "nonvital," and the analyst predicts that "the Soviets will risk confrontation when vital interests are at stake." The analyst's prediction success depends on the characterization of the performance

Figure 4.2. Hypothetical Relation Between Indicator Data and Performance Data



data. If the criteria is set high, say at line A, then only extreme Soviet behavior is characterized as "risking confrontation" (e.g., when Soviet ships fire on U.S. ships). The implication of this extreme characterization is that the analyst is usually wrong; in most cases in which Soviet interests are "vital" (greater than X) the Soviets do not fire on U.S. ships (the area labeled "false alarms" is much larger than the area labeled "positive hits"). The analyst faced with this state of affairs might question the strength of the relationship (i.e., the validity of the theory), or may question the characterization of risk-taking.

Suppose the analyst redefines risk and characterizes the Soviets as "willing to risk confrontation" if they merely increase the total number of Soviet warships in the crisis area (denoted in the Figure by line B). Now the analyst is always correct that when Soviet interests exceed the X level, the Soviets will move more ships into the crisis area (i.e., exceed the B level). Now, however, the analyst is plagued by surprises; the Soviets often increase the number of warships in crisis areas when their interests are below the X level (the area labeled "surprises" is larger than the area in the ellipse to the right of the X line). Again the analyst may question the validity of his theory, or his characterization.

This hypothetical example demonstrates the difficulty analysts have with verbal characterizations of data when their theories are strong but not perfect. In the absence of a quantitative assessment of the theory, the analyst appears as either an alarmist pessimist (the A line) or an inefficient optimist (the B line), and in both cases, wrong more often than right. Because of the sensitivity of his predictions to the characterization of data, the analyst may reject an accurate (but imprecise theory) and accept his experiences as evidence that the theory is wrong, when in fact the error is in the reliance on inaccurate and inappropriate verbal characterizations of a relationship (the theory) that cannot be specified clearly in verbal terms.

This example assumed that the analyst is sensitive to surprises and false alarms as well as to positive hits in assessing the strength of the relationship between the indicator data and the performance data. In fact, people are not good judges of correlation, primarily because they attend only positive hits and ignore surprises and false alarms. This bias is covered below under "assessing covariation." The implication of this insensitivity is that the analyst may retain an inaccurate theory or not question his characterizations of data.

Prototypes and schemas as categories. The traditional view of categories and classification schemes assumes that all members of a category possess a small set of critical features. This suggests an all-or-none criterion for category membership. If an object has all the critical features, it belongs in the category, if not, it does not. Such distinct, necessary and sufficient critical features are not common in everyday categories, e.g., those applied to events and complex objects. The example above demonstrates how everyday categories tend to be located on a continuum of shared family resemblance. All members of a natural category (e.g., Soviet risk-taking behavior) do not share a set of singly necessary and jointly sufficient features defining membership. Instead the members of a category share a pattern of overlapping and crisscrossing

similarities. Cantor and Mischel (1979) have reviewed how natural everyday categories are used in the perception of people. Their analysis is also relevant for the characterization of complex events, such as those examined by naval analysts.

Fuzzy categories. In contrast to the traditional view of categories as marked by clear boundaries, psychologists have come to view natural semantic categories as "fuzzy sets" that lack the classical all-or-none, either-or characteristic (see Cantor and Mischel, 1979: 8-10, for the philosophical and psychological roots of this evolution). Rather than being characterized by necessary and sufficient features, natural categories seem to be organized around prototypical or focal stimuli; exemplars that are the best examples of the concept. Less prototypic members form a continuum away from the more central prototypic exemplars. All category members possess some of the critical features of the category, enough to be judged as members. But category members may differ in how well they fit the abstract concept represented by the category name. That is, they may be more or less "good" cases of the prototypic exemplar.

Cantor and Mischel (1979) point out that a category continuum is translated into linguistic hedges: "Soviet ships firing on U.S. ships are a true case of Soviet risk-taking" or "Soviet expansion of crisis naval presence may be taken as an instance of risk-taking." Because "risk-taking" is an ill-defined concept, membership may be a matter of degree, rather than of sharp logical boundaries.

Probabilistic boundaries. The "fuzzy set" approach suggests that a categorization decision is probabilistic in nature. The hypothetical analyst in the example above would have been better off attempting to determine the correlation between the members of the two continua rather than attempting to observe clear boundaries between "risk-taking" and "nonrisk-taking" behaviors. Because members of natural categories vary in their degree of membership (prototypicality) there are many ambiguous borderline cases and overlapping and fuzzy boundaries between categories. An explicit treatment of this probabilistic character is likely to be more realistic than an absolute pigeonholing approach.

The difficulty of categorizing complex stimuli is suggested by several of the methodological devices developed by naval analysts to categorize such things as the missions of the Soviet Navy (e.g., Thorpe, NPSP, ch. 8) or the effectiveness of Soviet ships (e.g., Kehoe, SNI, ch. 19; Meier, SNI, ch. 20). These methodologies assume that a concept like "mission" is difficult to define and divide and may have only fuzzy boundaries. Thus, complex methods are needed to abstract judgments of the prototypic meaning of such categories as "the antiship mission" so that it has some wide agreement among analysts. Another analytic technique is the explication of taxonomies; classification schemes to organize objects (e.g., ship classes and types) or events (e.g., superpower crises versus local crises).

Categories and memory. Since the early work of Bartlett in the 1920s and 1930s, psychologists have been aware that complex events are reconstructed in memory rather than recollected (Norman, 1969). Bartlett found that accurate reports were the exception when people attempted to

recall stories, arguments, or drawings. What typically occurs in memory is a reconstruction from a general prototype or "schema," or an active characterization, of the original data. This schema depends heavily on whatever the person perceives to be the isolated, striking, or salient details in the original data. Consequently, recall of complex data tends to be shorter than the original, more modern, more coherent and consequential, and these errors increase with time.

With increased time there is more "constructive remembering," or invention, and people are often more pleased about and certain of those items they invented than those they accurately recalled. Posner (1969), Franks and Bransford (1971) and Bransford and Franks (1971) found that people abstract prototypical or average schema from complex visual stimuli or sentences and use these for subsequent recognition judgments. Further, subjects rated themselves most confident of having seen schematic sentences even though such sentences were never actually shown the subjects in the original sessions. Related experiments (Barclay, 1973; Bransford, Barclay, and Franks, 1972) found that subjects store not only information from complex data but also implications and inferences from the data. These latter may, however, be recalled as having been in the original data. In other words, schemas are abstracted from complex data and stored in memory, while other details of the stimuli are lost or stored less efficiently. During recall, the schema is used to recreate the stimuli or to decide if new stimuli are from the same category (Tsujimoto, 1978). Schemas capture the typical, average, general properties of the complex data and then bias subsequent recall or recognition toward the schema.

The debate between Vigor and MccGwire (SNP, chs. 32, 33) on Admiral Gorshkov's use of the phrase "command of the sea" may be a possible example of bias-to-schema. Vigor's reading of recent Soviet doctrine and activities allows him to view the concept of command of the sea as indicative of future Soviet naval strategy. MccGwire seemingly argues that the main features of "command of the sea" are not actually present in Gorshkov's advocacy or in Soviet naval doctrine, although they may be in a Western prototypical view of Gorshkov's writings or Soviet actions. MccGwire implies that Vigor's schema is too abstracted from Soviet doctrine and practice, and the apparent resemblance between "command of the sea" concepts and Soviet doctrine is due to the absence of details of that doctrine that were lost in the schematic abstraction. When those lost details are considered, MccGwire implies, "command of the sea" no longer appears representative of Soviet naval doctrine or strategy (see MccGwire, 1979: 161-63, for further discussion of Gorshkov and "command of the sea").

Naval analysis, like other forms of military intelligence, requires extensive recall of past events and data. Analysts seem to develop the ability to recollect great amounts of information. It seems quite likely that this experience parallels the highly accurate recall and rapid memory storage feats noted in expert chess players. Such experts can study a chessboard for only a few seconds and recall each piece's location, and retain these memories for weeks or months. Such recall is possible, however, only for meaningful board positions; chess pieces which are placed randomly are no more readily remembered by the experts than they are by novices. Studies of the eye movements of the chess experts show they literally focus on the most important strategic relations

between the pieces (Klatzky, 1975). Similarly, the excellent recall of naval analysts probably results from their abilities to fit data into a meaningful context, or to determine its promising, potentially relevant features, so that when a new context is formed they can recall the earlier data which could now be fit into the new context.

Demonstrations of this context-triggered reorganization of memory and previously stored data abound in naval analysis (indeed in all forms of research and scholarship). For example, Herrick's (NPSP, ch. 9) analysis of the USSR's "blue belt of defense" concept of a sea area defense zone required the integration of a variety of comments on the "blue belt" made by communist spokesmen in several Warsaw Pact nations over many years with indirect evidence from Soviet doctrine and exercises. Disparate data which had in common the mention of the "blue belt" were integrated with data that implicated a system of sea area defense to yield a concept of a unified, multi-service military plan for defense against seaborne nuclear attack by Western aircraft carriers and Polaris/Poseidon submarines.

In summary, memory scholars (e.g., Klatzky, 1975; Norman, 1969; Posner, 1973) find that for material to be stored in long-term memory, it must be integrated with existing concepts and fit the schemas created by previous data. This schema-linked storage may bias retrieval, however. For example, Funkhouser (1968) had people classify objects according to color or shape. Subsequently, they were better in speed and accuracy of recall when allowed to use the original classification scheme than when forced to use a new set of categories. In other words, memory is searched effortlessly provided that the context at the time of retrieval (problem context) matches the organizational classifications made at input, i.e., if the data at retrieval are "context addressable" (Shiffrin and Atkinson, 1969). If the context at retrieval requires a new classification, memory search is more difficult. Typically, however, the problem context matches the input schema only for easy problems. When the two differ, inappropriate data may be recalled (those with input schema similar to the problem context but not necessarily relevant to the problem), or appropriate data may not be recalled, or the problem context may be skewed to match what seem to be appropriate data in a different schema.

Taxonomy biases. Two organizing mental structures commonly used in memory are ordered lists and hierarchies, i.e., taxonomies. These structures can bias recall. Many personal experiences appear to be organized in memory as lists, however multidimensional events tend not to be ordered in memory on lists of several dimensions but rather along a single dimension. DeSoto (1961) noted that peoples' impressions of others on a variety of dimensions, such as voice quality and intelligence, tend to be highly correlated; i.e., those judged high in one quality are also judged high on others. This tendency to see various qualities as co-occurring has been termed the "halo effect" by social psychologists. The implication is that, for certain discriminations and judgments, people tend to collapse various dimensions into one, and to generalize from that one dimension to others.

There is a tendency for naval analysts to view Soviet doctrine and Soviet naval behavior as moving in parallel (e.g., Hudson, SND). For example, Hudson views Navy and Party views on the Soviet fleet as differing in the late 1950s and early 1960s on the role of the Navy in

nonnuclear wars, with Gorshkov backing a balanced fleet, and the Army and Party as seeing ocean-going surface ships as obsolete for nuclear war and too expensive for other tasks. Hudson sees the Party views as going together with "the foreign policy initiatives of the Soviet Union and the generally relaxed international atmosphere during this period [1953-1957]" (p. 279). Gorshkov's views he perceives as going together with various naval activities, e.g., maintenance of a surface fleet, and development of balanced naval forces. The naval evolution since 1953 is seen as due to "unremitting pressure" for change from Soviet Navy officers. A closer examination, however, of the events of the international atmosphere at that time might not suggest that it was relaxed, as Hudson believes. And the analysis of the Soviet construction (McGwire, SND: 187-92) casts doubts on the "balance" of the fleet or an institutional commitment to surface ships. The correlations between Party views and atmosphere and between Gorshkov's statements and naval programs may be due in part to a "halo effect."

Stereotyping. There is strong evidence that memory of concepts has a hierarchical structure (e.g., Warren, 1972). More importantly, statements which are not represented by hierarchical structures tend to be altered in memory so that they can be (Dawes, 1966). For example, Dawes compared memory of statements of the form "Some X are Y" with those of the form "All (or No) X are Y" and found that "some" statements are more likely to be recalled as "all" or "no" statements than the reverse. This preference for absolute and concrete concepts ("all" or "no") relative to relativistic or ambiguous concepts ("some") pervades other analytic tasks as well. A further characteristic of hierarchical organization is that properties of classes are often represented at higher levels (e.g., "birds fly") and these properties are then applied to lower levels (e.g., "eagles fly"). This can lead to misapplication of properties (e.g., "ostriches fly"). Further, properties are often assigned a class on the basis of observations of a subclass or single member (e.g., "eagles are killers," therefore "birds are killers"). Such thinking is termed "stereotyping" and may be due in part to the organizational structure of memory. The implications of hierarchical memory structure are that class assignments are sometimes inappropriate, and inferences from properties of individuals to properties of classes and vice versa are not always thought out. People assume natural categories have the properties of classical logical categories (e.g., rigid boundaries) when they do not.

For example, naval analysts have noted that the Soviets designed several classes of surface ships for anti-surface warfare and used such ships in task groups during crises, apparently to counter U.S. aircraft carrier task groups. There is a tendency for analysts to treat all Soviet ships in ASW task groups as being specially designed missile-armed types (e.g., Herzog, NPSP: 39-40). For example, Roberts (in Dismukes and McConnell, 1979: 160) writes of the

> More or less standard anticarrier task groups . . . composed of SSN and SAN surface units and cruise-missile and torpedo-attack submarines.

In fact, only some of the ships that undertake the ASW task in crises are missile-armed; others are older, gun-armed vessels.

Extreme cases. Because isolated and striking details of data are important in creating prototypes and schema in memory, there is a strong tendency to recall or recognize unique or striking information with little loss of detail even over long periods of time (Rock and Englestein, 1959). Mnemonists (individuals with greater than usual memory capability) attribute their prodigious memory, in part, to their ability to experience each particular datum as a unique instance (Luria, 1968). There are several implications of this characteristic of memory. First, nonoccurrences of events and negative instances are rarely as striking or unique as occurrences or positive instances, and will tend to be less well-remembered, as was noted above. Second, because unique and striking events tend to be well-remembered, relative more commonplace events, they tend to be over influential when estimates are made, because of the availability heuristic. Third, the perception of physical stimuli tends to be affected by the so-called "central tendency of judgment," that is, smaller stimulus values are overestimated and larger ones are underestimated. These inaccurate sensation judgments are compounded when psychophysical sensations are stored in memory, i.e., smaller, weaker values are further overestimated and larger, stronger values underestimated. Since "smallest, least, weakest," and "strongest, largest, most" are striking and unique, they tend to be well-remembered in terms of detail but not in terms of actual magnitude. Further, the "law of sense memory" suggests that the more extreme the stimuli, the more distorted the memory of its magnitude relative less extreme stimuli. In short, perception compresses magnitude, memory does so even more, and compression is most severe for those stimuli most likely to be recalled. Fourth, the distinctiveness of striking and unique data tend to isolate them from other data, thus improving their memorability, but also reducing the ability of the individual to integrate these data with others. In effect, the classification of data as unique protects it from forgetting or interference while in memory, but may also isolate it from further cognitive integration, unless such integration explicitly involves other distinctive cases.

Rothbart and colleagues (Rothbart, Fulero, Jensen, Howard and Birrell, 1978) presented people with information on individuals and asked them to make judgments about the characteristics of the group composed of those individuals. In one test, the heights of two groups of individuals were given, and in both twenty percent were over six feet tall. However, in one group those over six feet were moderately tall (6' 1" to 6' 4") while individuals in the other group were very tall (6' 5" to 6' 11"). People estimated the percentage over six feet in the first group at about twenty percent, but at thirty percent for the second group. Similar results were obtained for two groups for which the severity of crime was manipulated. In other words, the frequencies of extreme cases were overestimated when these cases were more salient.

To summarize, the following features of memory will tend to bias data characterization:

- Memory reconstructions streamline, condense, and modernize input data, fitting it to preexisting schema.
- "Constructive remembering" (invention) increases with time and

people are often most pleased with and confident of the fidelity of such memories. These constructions tend to reflect the characterization schema used by the individual to organize the input data.

o Recall and recognition tasks which match in context the organization schema of memory inputs can be performed relatively effortlessly. Recall contexts which do not match memory schema require greater mental effort.

o Memory lists fail to capture the multidimensional attributes of stimuli, producing the "halo effect " and the assumption that favorable or unfavorable qualities co-occur.

o Hierarchical organization of memories is often misapplied to nonhierarchical inputs, or applied too rigidly, producing stereotypes, the assumption that attributes of members of a class extend to the entire class. That is, data that are stored as "fuzzy sets" may be recalled in the belief that they are all-or-none sets.

o Striking details tend to be best recalled but also tend to be most compressed in magnitude toward less impressive data. The distinctiveness of striking details may prevent their integration with other data in memory and lead to overestimation of their frequency. Since negative instances are generally less striking than positive instances, the former are less memorable and less available for further cognitive work. Striking details will be readily integrated only with other striking details

The tendency for extreme or striking details to be integrated only with other striking details is an instance of the segmentation of events in such a way that maximizes commonality among elements within the category and maximizes differences between categories. This is entirely reasonable, but does not help determine which dimensions are actually used to distinguish categories. Rothbart (1980: 47) suggests that predictive value is one criteria for categorization. The Soviet Navy is not usefully categorized as "a navy" because that fails to distinguish it from other navies. However, the categorization of it as "a navy for coercive diplomacy" offers predictive value by distinguishing it and suggesting what it can or might do. The selection of categories thus reveals something about the analyst's prior expectations about meaningful dimensions. In Figure 4.2, the categories of "interests" and "risks" were chosen because the analyst perceived the possibility of predicting one from the other.

An example of categorization in naval analysis is McConnell's classification of cases in Soviet Third-World diplomacy of force (in Dismukes and McConnell, 1979: 252), which has three case divisions and seven case categories within divisions (see Table 4.2). The fact that McConnell's divisions refer to the high seas and the Third World, for example, implies that he does not see a naval diplomacy of force as like , against the First World

Natural taxonomies. It is noteworthy that McConnell's classification is similar to those observed by psychologists in natural

Table 4.2. McConnell's Classification of Cases in Soviet Third-World Diplomacy of Force*

Case Divisions	Case Categories	Types of Display
1. Security on the High Seas	1. Demonstration of intent to protect USSR assets, or the assets of clients, at sea.	Attentional
		Purposeful
		Temporizing
2. Third World Domestic Security	2. Demonstration of support for the domestic authority of an established government	Limiting
		Supportive
	3. Demonstration against an established government in defense of Soviet citizens and property.	Exemplary
		Expressive
	4. Military support to a domestic faction during any interregnum when the U.S. is inhibited from counter-involvement.	
3. Third World International Security	5. Demonstration of intent to protect a client in a confrontation with a state that, held in odium by the international community, does not enjoy U.S. patronage.	
	6. Demonstration of support for a client threatened (or that might be threatened) by a Western great power or in actual conflict with such a power.	
	7. Demonstration or actual intervention against a U.S. client that is defeating a Soviet client.	

* From Dismukes and McConnell (1979: 252-253).

semantic taxonomies; namely people tend to apply three levels of classification, with the optimal level of abstraction at the middle level (see Cantor and Mischel, 1979: 13-15 for review). At the middle level people use broad, inclusive, but still rich and distinctive categories, inclusive enough to cover many kinds of objects or events in a category, but also detailed and vivid enough to allow a description in great detail of the prototypical category member, and to distinguish a category from others at the middle level. Objects or events categorized at the middle level are really different and share few features, while objects and events at lower levels (e.g., instances of defense of Soviet assets on the high seas) may have many common features. Note that the categories at the same level in a taxonomy tend to be psychologically continuous -- without clear-cut boundaries -- but clearly separable on the basis of clear prototypical cases of each. It is the estimation of the degree of prototypicality, rather than a search for necessary and sufficient features, that psychologically marks peoples' judgments of how well various objects or events fit a particular category label.

The "fuzzy set" nature of complex natural categories and the difficulty in specifying all the features which give members of a natural category a family resemblance contribute importantly to the conflicts between naval analysts. For example, there are characteristics of Soviet ships that suggest they would be effective deterrents to interventions during crises by Western aircraft carrier task groups. Overlapping characteristics of these ships give them the appearance of effective elements in a "blue belt defensive plan" against nuclear strikes. Since these diplomatic and strategic missions share many features, deciding whether a particular event is best characterized as one or the other may be difficult. Often analysts can do no more than to conclude that a particular event was more like the 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, ship construction, or deployment patterns; 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. To characterize any of the complex elements of Soviet naval behavior as belonging to only one of these categories is to overlook the psychological nature of the analytic process and the richness and ambiguity of all natural phenomena.

The best naval analysts recognize this and treat their categorizations as matters of degree (although they often neglect this probabilistic nature when explaining events or drawing conclusions). For example, Weinland (SNI: 507) writes.

> There is a continuous dimension to Soviet naval-diplomatic activity. The low intensity continuing impact of naval presence has eluded measurement by students of naval power . . . the Soviets apparently find crisis management a more or less continuous task.

The major challenge to these naval analysts is tracing and weighing these overlapping and criss-crossing categories of Soviet intentions and behaviors. There may be little point to carrying on the tradition of viewing Soviet naval events in "either-or" categories. A more profitable approach may be to determine prototypical cases and attempt to measure the

distances between prototypes and specific events. An explicit effort to measure "family resemblances" rather than to debate pigeonholes may be more meaningful both psychologically and analytically.

Generalizations and representativeness. Among the most important characterizations of data made by analysts are generalizations, inferences about a population on the basis of evidence on a sample from that population. Generalization is a central aspect of scientific inference and rules for valid generalization have evolved. These rules, however, are rarely characteristic of natural generalization (cf. Rothbart, 1980). The rules for scientific generalization are (1) define the appropriate sampling domains relevant to the issue, (2) sample representatively (fairly) from the appropriate domains, (3) code data into categories using unbiased rules, (4) compute contingency between measurements, (5) retain or discard hypotheses about relationships. People do not naturally follow these rules or apply them validly when generalizing. For example, they neglect the size and variance of the sample and generalize from very small samples. They allow their hypotheses to influence the definition of sampling domains, sample selection, and data coding.

People making judgments are insensitive to the statistical "law of large numbers." This law describes how large-sample statistics approach corresponding population statistics. The larger the sample the more accurately its statistics will reflect the population. On the other hand, small samples may differ markedly from the population, and small samples from different populations may be very similar. Generalizations based on small samples will tend to be less accurate than those based on larger ones.

This tendency to neglect sample size and variance may be particularly strong in naval analysis, with its heavy reliance on case and class analysis. A particular single case study or ship class may seem highly representative of some naval behavior or mission. The single case or class may then form the basis for analytic generalizations about future Soviet behavior. Such small samples, however, are likely to vary considerably from the population of cases or classes.

For example, both the Kiev and Moskva aircraft carriers seem somewhat representative of Western aircraft carriers, however, both have been designed for a much different set of missions than those of Western carriers. The Soviet carriers are apparently what the Soviets have labeled them: large anti-submarine warfare cruisers (cf. Herzog, NPSP 47, 49-50, 55; McGwire, NPSP: 85, 91-92). Similarly, the use of the Kildin and Kashin classes to shadow U.S. carrier task groups during Mediterranean crises and the presence of a rear-firing missile system on these ships seemed representative of anti-surface ship classes (e.g., Roberts in Pikmukes and McConnell, 1979: 219). It now seems, however that these classes carry only ASW missiles in the aft-facing launchers (McGwire, NPSP: 94-95). In terms of their overall surface construction and conversions, large Soviet surface ships are designed for ASW missions (Murphy, NPSP: 112), however, small, vivid samples of behavior, such as shadowing of U.S. carrier groups, may lead analysts to ascribe anti-surface capabilities to Soviet ships in excess of what really exist.

In addition to an insensitivity to sample size people tend to be insensitive to sample bias. Unless a sample is unbiased (selected randomly), or the biases are known and can be compensated for, a

generalization based on the sample will also be biased. One often overlooked bias is that produced by role contexts. For example, Ross, Amabile, and Steinmetz (1977) found that certain social roles lead to biased behavior samples, e.g., a person asking hard questions appears knowledgeable (knowing all the answers) while the person being asked the questions seems less intelligent (getting questions wrong). The roles, rather than the peoples' activities predispose these perceptions. Observers of individuals in these roles tend, however, to attribute their different perceptions to abilities not roles: the questioner is believed to be more knowledgeable than the person being questioned.

A similar phenomenon has occurred repeatedly in naval analysis of superpower crisis behavior. The Soviet Navy is often perceived as attempting to counter or deter U.S. naval intervention in the Third World. Since the Soviet Navy began such demonstrations in 1967, the U.S. Navy has not launched air strikes or landed Marines in any of the crises which also involved the Soviet Navy. The Soviet role of "deterrer" conveys a highly biased sample of behavior, since the U.S. Navy may have been just as likely to not intervene if the Soviets had not been present. The deterrence role appears successful if nothing happens: if the party being (supposedly) deterred does nothing, the party doing the deterring (supposedly) is obliged to do nothing, except make tacit or overt threatening gestures. So long as the "deterred" party does nothing (for whatever reasons), these threats need not be executed; they can actually be bluffs. Only if the supposedly deterred party ignores the threats (i.e., turns out not to be deterred after all) is the deterrer forced to act and make good the threats. If the deterrer's bluff is not likely to be called, the deterrer can make repeated deterrent threats which appear successful and effective. It may well be, however, that in each case there was little or no likelihood of U.S. intervention, or of Soviet willingness to back up the threat. The role advantage, however, goes to the deterrer, i.e., the Soviets.

In other words, the Soviets can easily gain the appearance of a successful, effective counterforce by their efforts to deter what the United States was unlikely to do anyway. Unless the U.S. Navy is willing to call their bluff, the Soviet Navy obtains a role biased advantage at relatively little cost. The perception of observers, or analysts, is likely to be that the Soviet Navy caused the U.S. Navy not to intervene. The multitude of other factors which might have caused U.S. nonintervention (e.g., a lack of any U.S. incentive to intervene) are overlooked, and the biasing role of deterrer is neglected. The estimate is that the Soviets have succeeded in preventing U.S. intervention.

There is some evidence that the Soviet Navy itself has not fallen victim to this bias. In situations during which the U.S. Navy was likely to call the bluff, e.g., in port demonstrations off Vietnam and Korea, the Soviet Navy has tended to stay clear and has made little pretext about deterrence (e.g., see McConnell in Dismukes and McConnell, 1979: 262). Even in those cases where the Soviet Navy is ostensibly deterring or countering U.S. task forces (e.g., in the October 1973 crisis) the Soviets seem to have taken pains to stick to anti-carrier exercise routines so that U.S. Navy officers could recognize them as exercises rather than actual threats (cf. Roberts in Dismukes and McConnell, 1979: 219), and the Soviets even went so far from actual operational practice as to operate submarines extensively on the surface, a foolhardy move for anyone really

expecting trouble.

Hindsightful categorizations. One of the strongest sampling biases tends to occur via hindsight. Once a person forms a tentative impression of a set of data, in looking back over past events, he or she may make systematic retrospective distortions to bring those data into conformity with impressions (Snyder and Uranowitz, 1978). For example, confirming instances may be selected to bolster beliefs and disconfirming instances may be overlooked. Current beliefs can exert a powerful channeling effect on attempts to reconstruct or remember the past. Events that confirm beliefs are readily recalled, while other facts are erroneously recalled in ways that "confirm" the belief.

Perhaps the most dramatic instance of this phenomenon in naval analysis is the acceptance by Western analysts of Admiral Gorshkov's assertion in 1967 that

> . . . by 1954, Moscow realized both the need for naval diplomacy, to compliment the other instruments of its new Third World policy, and the need for patience and restraint until the proper capabilities in this role could be acquired through long-term construction programs (McConnell, in Dismukes and McConnell, 1979: 14).

There is little evidence to support Gorshkov's claim and much to refute it, as has been noted above in Chapter 3. His claim does provide a simple explanation for the Soviet Navy's progress toward naval diplomacy -- it was planned. One wonders, however, how it is that the Soviets could be so foresightful as to recognize a decade in advance the political importance of the Third World, let alone the need for naval capabilities for diplomacy there, and yet be so shortsighted as to construct forward-projection forces which were heavily dependent on bases and support facilities in unreliable noncommunist countries. One would expect, for example, such a foresightful adversary to spend its $6 billion on a more profitable relation than that between the USSR and Egypt.

Biased sampling. Rothbart (1980) has outlined in general terms how biased sampling can lead to seemingly irrefutable (but nonetheless biased) evidence to support a belief.

Suppose for the sake of this argument that Soviet naval behaviors can be hypothetically categorized as "diplomatic forward" or "strategic forward" as in Figure 4.3. Panel A shows the total data set: there is no change in behavior between the two periods except a doubling of activity in the forward area.

Suppose, however, that a naval analyst with a diplomatic (or strategic) perspective samples data from the past on Soviet naval diplomacy in the forward area (Panel B). The analyst finds the Soviets have doubled their diplomatic (or strategic) activity.

On the other hand, an analyst concerned with the "new" Soviet Navy (Panel C) or the overall pattern of Soviet naval behavior in the forward area (Panel D) would find that Soviet strategic naval activity in the forward area is twice as likely as diplomatic activity

In other words, one set of data and different sampling biases allow naval analysts to arrive at different perceptions of Soviet naval

Figure 4.3. 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

activity. Several widely different generalizations about Soviet naval actions in the forward area are drawn, none being entirely accurate, but all completely consistent with the data sampled. If all four cells of the contingency table in Panel A are utilized, half-right conclusions are avoided. However, people typically look only at confirming data, and rarely consider all cells. For example, "diplomatic forward" is confirming data for the diplomatic analyst. But "strategic forward" is nonconfirming and may not be viewed as relevant by the analyst interested in naval diplomacy.

An example of the sampling bias can be found in Petersen's analysis of Soviet Navy diplomatic port visits (in Dismukes and McConnell, 1979: 91-92). Petersen argues against McGwire's thesis that, after an initial survey of ports in the Indian Ocean, very few visits were made to other Indian Ocean countries, and diplomatic visiting there fell off after 1969. Petersen writes

> While it is true that a drop in diplomatic visits 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 1976, 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.

Petersen presents (p. 92) tabulations of diplomatic port visits for the Indian Ocean and the Mediterranean for the years in question (see Figure 4.4). Petersen overlooks two important base rates: first, the total levels of port visiting in the two areas; second, the disproportionate time periods. Indian Ocean diplomatic visits were, over the total time period 1967-1976, twice as frequent as Mediterranean visits. Proportionately, Indian Ocean visiting in the 1970-1976 period (57%) was much less than Mediterranean visiting (80%). Secondly, while the year ratio between the two time periods is 3:7, and the ratio of visits in both oceans in the two periods is 1:5.4, the ratio of Indian Ocean visits in the two periods is 1:3.9, compared to the Mediterranean ratio of 1:12. In other words, while Mediterranean visits per year increased over the two periods (2.3 and 4), in the Indian Ocean they decreased and there were almost twice as many visits per year in 1967-1969 (10) in the Indian Ocean as in 1970-1976 (5.4). Petersen's sample statistic of 30 to 28 visits is biased and misrepresents the patterns of diplomatic visits in the two regions, in particular, the sharp decline of visiting in the Indian Ocean to a level comparable to that in the Mediterranean.

Assessing Covariations

The ability to discern relationships between sets of data is a fundamental skill for all people as well as analysts. Learning the relationships between warnings and danger or between tracking signs and prey is a basic part of survival.

Similarly, naval analysts must assess the covariation between Soviet strategic needs and naval activities, forward deployment and naval diplomacy, political commitments and crisis behavior, capabilities and

Figure 4.4 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.)

actions, etc., to understand and predict Soviet naval intentions. Naval analysts are sensitive to these covariance assessment tasks and on occasions have identified covariations of great relevance to the Soviet Navy which have been inadequately addressed; e.g., the relationships between shifts in naval policy and in foreign and defense policy, the relationships between naval policy and technical developments (McGwire, SN&S: (x).

Two-fold tables. The simplest representation of covariation is the four cell presence-absence table:

		Behavior X	
		Present	Absent
Indicator A	Present	20	10
	Absent	80	40

Even with this simplest manner of representing covariation between two events, people have difficulty detecting covariations (Smedslund, 1963; Ward and Jenkins, 1965).

Among the common problems in detecting covariation is an almost complete reliance on the "present/present" (or ++) cell in the four cell table. People will tend to see indicator A as associated with behavior X simply because sometimes A is present when X is present. Some people will attend two cells (++ and +-) and conclude that a relation exists because there are more cases of X with A than of no X and A. Others may attend cells ++ and -+ and decide the relationship is negative, behavior X is indicated by the absence of A. Without considering all four cells, no judgment of covariation can legitimately be made, but people do not intuitively consider more than one or two of the cells. When the two variables are continuous rather than dichotomous the covariation estimation problem becomes even harder.

There is some evidence that naval analysts occasionally do estimate covariation on the basis of one or two rather than all four cells in the four fold table. For example, Dismukes and McConnell (1979: 307) wrote

> The variations we have noticed in Moscow's willingness to exchange silent breaches of the status quo are linked to variations in Soviet tactical views on the exacerbation and relaxation of tension.

To support this they mention one case of encouragement and exacerbation (++ cell) in 1967, and two cases of nonencouragement and relaxation (-- cell) in 1971-1972 and 1973. They note neither the encouragement and relaxation (+- cell) situation (Soviet support for Ethiopia might be an example), nor the nonencouragement and exacerbation (-+) situation (Soviet

nonencouragement of Iran or Iraq in their recent conflict may be an example). While the existence of cases in the ++ and -- cells is necessary to establish a positive covariation, it is not sufficient. A relative absence of cases in the +- and -+ cells is also necessary.

Neglect of negative evidence. In part, the difficulty people have with covariation estimation may derive from the general failure to use negative evidence. The absence of indicators tends to be overlooked and people form concepts of negative instances only with great difficulty. The heavy reliance on positive instances (++) may reflect people's inability to learn readily from negative instances (see "Nonevents" above, and Nisbett and Ross, 1980: 49; Mason and Johnson-Laird, 1965).

Naval analysts have not commented on the lack of a "diplomacy exercise" by the Soviet Navy, i.e., naval exercises and maneuvers which are explicitly and exclusively concerned with a peacetime crisis role and missions well short of general war. Such scenarios are often selected in Western exercises (especially intervention in Third World crises, or on behalf of Third World allies, e.g., South Korea). The resulting "crises" phases in Soviet exercises are extremely brief, and lead quickly into general or nuclear war missions. A Soviet exercise that included naval, airborne, or amphibious intervention in the Third World but which did not include general or nuclear war aspects would be extremely strong evidence for the thesis that the Soviet Navy has been designed for a course of coercive naval diplomacy in the Third World.

On the other hand, negative evidence is sometimes seen as significant by naval analysts, as we noted above under "Nonevents." For example, Thompson (NWSP: 32) asks why Admiral Gorshkov's book Sea Power and the State had not been reviewed (as of February 1977) by either Pravda or Kommunist. These nonevents suggest that the Admiral's views are not completely accepted as doctrine in the Soviet Union, i.e., Thompson suggests this is a -- instance, no review, and no Soviet decision on authoritativeness.

Theory-driven and data-driven covariation estimates. Nisbett and Ross (1980: 97ff) have made a useful distinction between "data-driven" and "theory-driven" judgments of covariation. When people have no a priori theory or expectation of a relationship, strong correlations (e.g., r = .6 to .8) tend to be underestimated, i.e., judged to be weak relationships. Only when the relationship is extremely strong (r = .8 to 1.0; r = 1.0 is a perfect relationship, so one variable changes the other changes by a proportional amount) do people perceive it as a strong covariation, and they still tend to greatly underestimate the strength of the relationship. In other words, data-driven covariations are difficult to recognize, and when recognized, are underestimated. Strong, but unexpected, relationships will tend to be missed. Furthermore, "decoy features" in the data, which bring to mind incorrect theories about covariation, may conceal strong correlations.

This relatively weak perception and judgment of "natural" (rather than anticipated) covariation has important implications for naval analysis. Unless an analyst expects to find a relationship, he or she will tend not to notice one unless it is very strong. This means many strong covariations (e.g., r= .6 to .8) will go unnoticed, or will be underestimated. Potential indicators of future Soviet activities may be

overlooked and important patterns in the naval data may be neglected. Similarly, an analyst may fail to note the success of an indicator in predicting future events.

In this respect it is very interesting that one of the best quantitative naval analysts sampled, Petersen, notes correlations in his data only when they are extremely strong (e.g., $r > .8$, Figure 2.3, Table 3.3., Figure 3.2, in Dismukes and McConnell, 1979). For example, Petersen's Figure 2.3 shows an extremely strong negative relationship between the mean age of a class of Soviet ships and the percentage of that class deployed out of area in the year 1975 ($r = -.91$). Describing this figure Petersen writes (p. 45, italics added):

> . . . the number of ships in a class that operate beyond home waters seems to be related to the age of that class.

Similarly, in describing a nearly perfect covariation between the age of Soviet ships and the number of diplomatic visits they make ($r = -.98$, Figure 3.2, p. 98), Petersen writes (p. 96) "the Soviets apparently prefer to use their newer large surface combatants for diplomatic visits."

On the other hand, there are important covariations in Petersen's data on which he does not comment. For example, the negative relationships noted earlier in this chapter between the total Soviet ship days spent in the Indian Ocean and the number and length of diplomatic port visits (see Petersen, Table 3.2, p. 92) seems unnoticed; instead Petersen writes (p. 91)

> The relative intensity of the Indian Ocean effort is particularly striking and would appear to confirm the prominence of political concerns in the Indian Ocean squadron's mission structure.

It would seem that Petersen did not expect this negative relationship and, despite its strength ($r = -.65$ and $-.83$), overlooked it and implied the converse. Similarly, he takes no note of the tendency across all oceans for total Soviet out-of-area ship days to covary negatively ($r = -.74$) with length of diplomatic port visits.

[I again apologize to Petersen for this apparent overattention to his work. Being one of the few analysts who adequately presents his data, he becomes most vulnerable to this form of methodological analysis. Petersen is probably less guilty of biases than others, but easier to catch. His extensive use and reporting of quantitative evidence is highly commendable.]

On the other hand, if a theory or expectation implies a relationship, it becomes much easier to recognize covariation between variables which are expected to be related.

For example, Weinland noted (NPSP: 263)

> Since all but a fraction of the [Soviet] surface combatants and auxiliaries that operate in the Mediterranean come from the Black Sea . . . changes in the pattern of Soviet transits through the Turkish Straits provide rough but generally reliable reflections of changes in [Mediterranean] deployments.

Weinland is entirely correct. Soviet surface combatant transits of the Turkish Straits correlated very highly (r = .91) with Soviet Navy ship days in the Mediterranean, 1964-76 (see Weinland, Tables 10 and 11, p. 262-63). Auxiliary transits also covary strongly (r = 90) with Mediterranean naval presence. Earlier in this chapter we warned analysts against the temptation to treat highly correlated (redundant) variables as independent evidence or to weight them equally with truly independent variables. Weinland seems to avoid this problem by noting the correlations in his data.

In short, if a particular relationship is not predicted by an a priori theory or expectation, or if the person holds a theory which incorrectly presumes a weak or null relationship, there is a tendency to miss or underestimate covariations in the data. Theory-driven covariations are much more easily perceived.

In fact, the perceived covariations and estimates of association based on theory are often far higher than are justified by the data. Theory-driven covariations are often overestimated, to such an extent that variables with no association or a negative covariation may be estimated to have a strong positive correlation. This phenomenon of "illusory correlation," already mentioned above under characterizing data, requires further explanation.

Perceiving nonexistent correlations. The psychological analysis of "illusory correlation" began with a series of elegant studies by Chapman and Chapman (1967, 1969; Chapman, 1967). They investigated the puzzling finding that expert clinical psychologists reported observing associations between certain projective test responses and particular clinical symptoms (e.g., parnoia), although repeated validation studies of the tests found no associations with symptoms. In other words, the expert clinicians were inferring personality and psychopathic characteristics from invalid indicators. The Chapmans reasoned that beliefs about covariations often result from semantic associations (rather than from frequencies of co-occurrence). Thus clinicians might believe that paranoids would exaggerate the eyes in the Draw-a-Person (DAP) test, even though the DAP has virtually no predictive validity. In effect, common semantic associations, being readily available mentally, might be taken as indicative of a high frequency of co-occurrence, in keeping with the availability heuristic. If so, these semantic associations should apply equally to nonexperts. This is what the Chapmans found: naive judges reported the same co-variations in random pairings of signs and symptoms that clinicians claimed to see in practice. When people were asked to simply rate the tendency of a given symptom to "call to mind" a given body part, the ratings were highly predictive of the reported (illusory) correlations. The body part most often called to mind by the symptom was the same that expert and nonexpert judges perceived as co-varying with the symptom.

In a cognitive test of the "illusory correlation" principle Chapman (1967) presented pairs of words to people. The left-hand words were semantic associates of the right-hand words (e.g., lion-tiger). The pairings were equally frequent (e.g., lion-eggs appeared as often as lion-tiger and bacon-eggs), thus there were no true correlations between the appearance of a left-hand word and a right-hand word. Nevertheless,

the observers reported retrospectively that each right-hand word appeared most often with its semantically associated left-hand word.

Using the Rorschach test, the Chapmans (1969) provided an even more conclusive demonstration of illusory correlation. Two responses to the Rorschach consistently differentiate male homosexuals from heterosexuals. In addition to these two valid indicators, there are several signs that seem intuitively (but not in fact) to differentiate between heterosexuals and homosexuals. The Chapmans found that practicing clinicians commonly reported five intuitive indications (which are all empirically invalid) as most characteristic of homosexuals, while the two valid signs were rarely reported. When laypeople were asked to rate the degree to which homosexuality called to mind the various valid and invalid signs, homosexuality called to mind the five plausible but invalid signs but not the two valid signs. In other words the experts and the laypeople shared the semantic associations between the invalid indicators and homosexuality, but were unaware of the valid indicators.

In the crucial phase of the study, the Chapmans showed nonexperts pairings of the valid indicators and the homosexuality condition which covaried either 50 percent (i.e., the valid indicator co-occurred half of the time homosexuality appeared), 67 percent, 83 percent, or 100 percent. The invalid signs were also presented but did not covary with homosexuality. The manipulation of covariation of the valid signs did not influence perceived covariation, i.e., even when the relation between the valid signs and the homosexual condition was perfect, it was not recognized. However, if all the invalid signs were deleted, the nonexperts were slightly sensitive to the increased covariations between the true signs and homosexuality. Further, when the Chapmans created a series of massive negative covariations between the invalid signs and the symptoms, the nonexperts reported slightly reduced, rather than very strong, positive correlations. True, but unexpected, covariations were unperceived even if perfect when they appeared along with the plausible, invalid signs, and were only partially recognized when they appeared alone. Invalid, but plausible, covariations were perceived even if the data strongly indicated an opposite relation. Theories or preconceptions of what "should" covary consistently overwhelmed data-driven covariations.

The tasks performed by the Chapmans' subjects were considerably easier than the tasks of naval analysts estimating intentions. The data were well-organized, clearly and concisely presented in rapid succession indicators were presented at the same time as symptoms, there were fewer demands on memory, etc. Yet their subjects consistently failed to detect data-driven covariations and repeatedly perceived nonexistent but plausible theory-driven covariations. This bodes ill for the naval analysts, whose tasks are far less clear cut.

Illusory naval correlations. Two seemingly clear cases of illusory correlation will be briefly noted, both having been mentioned before. The perception that Soviet Navy presence co-occurs with Third World leftist coups is a notion that seemingly persists, and generates perceptions of nonevents (e.g., Soviet Navy support of the Libyan coup), despite several analysts rejection of it (e.g., McGwire, 1979; Dismukes and McConnell, 1979: 257-59).

The second case of illusory correlation is much clearer (but

perhaps less significant). Petersen (in Dismukes and McConnell, 1979: 91) describes "the expansion in the number of Soviet diplomatic visits [in the Indian Ocean] between 1967 and 1976" as "impressive." In fact, Petersen's data (p. 92) show a negative correlation (r = -.38) between year and number of diplomatic visits by the Soviet Navy in the Indian Ocean. Rather than an "impressive expansion" in visits over this period there has been a contraction. The Indian Ocean is the only region where the trend in number of visits is so markedly downturning; in other oceans the number of visits has been constant or slightly increasing. It is thus not surprising that Petersen expected a positive relationship.

Psychologists have proposed four different reasons for the tendency of people to overestimate the degree of covariation in theory-driven associations and to underestimate the covariation in data-driven associations for which there are no clear cut expectations. The Chapmans (1967, 1969) explain the occurrence of theory-driven illusory correlation in terms of the semantic, associative connections between the variables which are mistakenly perceived as correlated. This explanation is similar to Tversky and Kahneman's (1974) availability heuristic; i.e., an associative, semantic connection is readily available in thought and is thus suggestive of a statistical association.

Taylor and Crocker (1979: 48) offer an explanation in terms of the operations of the theory itself (they use the term schema). Theories make some types of evidence more available and influence the type of evidence the person is likely to use in judging covariations. Specifically, people search for data that are consistant with the theory, rather than irrelevant or inconsistent information. People consequently over-rely on the number of "positive hits" (or ++ occurrences) which confirm the theoretical relationship they are assessing. Secondly, the judgment of how frequently two events co-occur could be based on the associative bond between them, as Tversky and Kahneman suggested. When the association is strong the person is likely to conclude that the events are frequently paired. Strong associates are likely to seem to co-occur frequently.

Hamilton and Gifford (1976) offer an explanation that relies on a general tendency to see rare events as co-occurring. That is, events which share the same characteristic, infrequency, may be perceived as being associated, and thus as co-occurring. Rather than a semantic association being taken for correlation, they suspect a statistical characteristic (infrequency) as giving rise to an association, which in turn suggests a correlation, e.g., by means of greater availability of associates.

Rothbart (1980) offers a fourth possible explanation for the occurrence of illusory correlations in cases when two events are split into frequent and rare groups. In this case, even when no covariation at all exists, the number of cases will be greatest in the ++ (frequent, frequent) cell. Simply because most of the data fall into this one cell, it becomes the most obvious and conspicuous, and people may perceive a covariation merely because they tend to attend this cell most closely. Rothbart suggests pairs of rare events and of very common events will both tend to be perceived as co-occurring, because of the great inequality in the size of the cells in the four-fold table. The ++ case becomes the most obvious and available explanation of the relationship between the events.

Covariation assessments are important in their own right, but also

because they are essential for the estimation of cause and effect relationships. Correlation does not imply causation, but, as will be discussed next, causation implies correlation.

Assessing Causes and Effects

> Causation is an idea that is generally used because it is generally useful.
>
> D. H. Fischer, 1970

Assessing causes is one of the most difficult analytic tasks. Scientific methods and inferential statistics were developed in part to reach sound conclusions about causes. Causal reasoning without such aids is subject to bias and inaccuracy.

Analysts of the Soviet Navy often describe the difficulty of determining causes and assessing effects in their field. For example, Booth (SND, ch. 19) summarizes the discussion on whether or not the Soviet Navy's forward deployment has made any real difference in Soviet foreign policy. He wrote (p. 368)

> . . . this question could not be easily resolved . . . It was surrounded by an inherent obscurity . . . because of the elusiveness of the cause-effect relationship.

In a similar vein, McGwire (SND: vii) wrote:

> There was no agreement [among naval analysts] on the impact of the Soviet Navy's forward deployment on Soviet foreign policy, and to what extent the one led to the other . . . it was difficult to identify specific instances where a Soviet naval presence had "caused" particular developments or had produced shifts in Soviet policy . . .

A fundamental difficulty is that causes can never be observed, they are always inferred from either (1) experiment (the soundest basis for inference), (2) experience and observation (an imperfect but often adequate basis), or (3) theory (often an unsound basis). There is an inherent gap between the first two of these bases and the third which cannot be bridged in an entirely satisfactory way. People think in terms of a theoretical language based on such notions as causes, forces, systems, and properties. These are attributes or relationships with no physical, tangible equivalents. People test reality with covariations, operations, and measurements. Science deals with this basic distinction by separating the language of theory from the language of operations (i.e., experimental and observational methods). "Concepts in the one language are associated with those in the other merely by convention or agreement between scientists" wrote one scholar of causality and science (Blalock, 1974: 6).

The distinction between theory and operations is often elusive, even for scientists. There is a very strong psychological tendency to think and reason in terms of causes and forces as if these were real, rather than mental, creations. While it is possible to find operational

evidence that supports a causal concept, it is impossible to prove a causal theory; causal laws can never be demonstrated. A correct causal law will, however, allow scientists to make accurate empirical predictions. This accuracy is the basis for the agreement between scientists which Blalock mentions. But the correctness of prediction does not demonstrate the correctness of the causal theory. Theories, like people, are often right for the wrong reasons.

Scientists use theories for the same reasons the rest of us do, they need to simplify reality. By making untestable assumptions, a few key assumptions can be indirectly tested. In reality no two events or objects are ever identical, but in science and everyday life it is necessary to act as though events can be repeated and objects do have properties that remain constant for some period. Without these important, simple assumptions it is impossible to generalize beyond single, unique events.

Theoretical models of reality help in making these assumptions so that other assumptions about causes and systems can be tested. An ability to predict events is supporting evidence for a causal theory (but not proof). Prediction is tangible, measurable, and "real" while "causes" are not; the two should not be confused.

Causal meaning. What is generally meant by "causality?" Scientists require three things in tests of causal laws. Each is necessary and none alone is sufficient as evidence for causality (Blalock, 1974). First, if X is believed to cause Y, a correlation must exist between them. If no regular correlation exists, no regular causal relation can exist. Second, there must be an appropriate temporal relationship, i.e., X must occur before Y occurs. While the theoretical need for causes to precede effects is obvious, the operational problem of determining which events occur first has no obvious solutions (e.g., unknown aspects of Y may cause X, which changes at the same time changes in Y become measurable or observable). Third, a "presumptive agency" must exist to connect X and Y. That is, some operator exists which generates Y and corresponds to X and which is organized so that a connection between X and Y can be separated into a sequence of compatible components that can be reasonably expected to overlap (cf. Fischer, 1970; Heise, 1975). In some respects, the third condition would seem to be sufficient to believe X causes Y. However, as we discuss below, the human mind is an extraordinarily fertile field for "presumptive agencies" -- even random events can be readily connected in the mind. Confirming evidence is easily retrieved from memory or created by the imagination to demonstrate causal relations which do not, in fact, exist.

Causal errors. In this section we discuss major psychological sources of error in causal thinking. People are imperfect causal analysts under the best conditions; even scientists may hypothesize and find nonexistent causes for events. For example, early in this century French scientists discovered a new form of radiation, the N-Ray, and other scientists confirmed this discovery. The radiation was entirely imaginary, however, the seeming evidence for it was totally useless, and the compelling concept of an N-Ray turned out to be completely wrong (Klotz, 1980).

We have already described how biases in the perception, weighting,

and categorization of data can produce what are, in effect, "illusory data bases" (Taylor and Crocker, 1979). Even with accurate data, people perceive illusory correlations and may fail to perceive perfect but unexpected covariations. In other words, up to this point, before any assessment of cause and effect is attempted, we have noted a variety of data problems which would bias even accurate causal assessments. As we will see however, causal estimation itself is beset with problems, which, along with those that precede it, can severely impair causal inferences.

These problems cascade as more information is integrated. There are more opportunities in more complex, causal estimates for bias and error to occur. It becomes more difficult to determine from an estimate itself just where such problems started or ended or to assess the magnitude of bias at various stages.

Covariation assessment and causal estimation serve to integrate and summarize data and facts into information and knowledge. It is difficult to separate items of information and knowledge into the various elements and processes that composed them. As we go higher in the analytic estimation process it becomes more essential to examine these components and processes of analysis as well as the finished products. However, even though the analytic processes and ingredients cannot be entirely deduced from finished estimates, an outline of the potential problems with the "higher-order" processes (causal estimation, prediction, theory testing) will help both the analyst and the estimate consumer to detect vulnerabilities and errors.

Causality and chance.

> Chance is a word void of sense; nothing can exist without a cause.
>
> Voltaire, A Philosophical Dictionary

> Chance is perhaps the pseudonym of God when He did not want to sign.
>
> Anatole France, Le Jardin d'Epicure

There is a strong tendency to look for and find causal relationships even in events that occur together by chance. Because people are highly sensitive to the ++ co-occurrences and relatively insensitive to +- and -+ instances (which invalidate the co-occurrence hypothesis), people tend to overestimate the causal links between events which are randomly connected.

Several factors increase the likelihood that a pair of events which randomly co-occur will be perceived as causally linked. First, if a strong a priori theory exists to link the events, an observer will be especially alert for ++ instances, which are then taken as evidence in support of the theory. Even more fundamental than this first bias, however, is the tendency to perceive any event, object, or case on which we focus our attention as playing a causal role. Third, the pairing of an intention and a favorable outcome is very likely to lead to the estimate that the actor who intended the outcome actually caused the outcome, although the favorable outcome may have resulted from chance. The

contingency between an actor's apparent intentions and outcomes is regarded as persuasive evidence of causality. This perception of cause-effect is greatly strengthened if the actor is also seen as capable of producing the desired outcome. The result is that the actor's good fortune will be misperceived as the result of the actor's intentional and successful action. Fourth, temporal and physical contiguity alone can suggest both covariation and causality. Actions by the actor with a certain pace or momentum that are followed by events with similar pace or momentum are perceived as causally related. Events which co-occur in proximity, which share features of size, shape or form, temporal pacing, and duration, etc., tend to be perceived as causally linked and the hypothesis that the co-occurrence was coincidental is underweighted. Fifth, observers attribute identical "good" actions to the dispositions, intentions, and motives of liked actors but attribute them to the situation and environment for disliked actors. Similarly, "bad" actions by liked actors are attributed to the situation and environment, but the same bad actions by disliked actors are perceived as reflecting intentions, motives, and dispositions. In other words, disliked actors intend only bad actions, and perform good actions only when forced by circumstances to do so. Liked actors are disposed to "do good" and "do bad" only when pressured by the situation. Preconceptions about the targets of observation may fill in causal relationships that reflect the preconceptions more than any causal reality. In general, observers attempting to estimate the causes of actions greatly overestimate the impact of intentions, dispositions, and motives, and underestimate the role of situational, environmental, or contextual pressures. Observers are highly likely to infer intentions from actions, and to deduce that intentions caused actions, when, in fact, the data do not warrant such conclusions.

One of the most salient differences between analysts of the Soviet Navy hinges on this distinction. One group of analysts perceives the Soviet Navy's forward deployment and efforts at naval diplomacy as dispositional: as actions intended, planned, and consistent with other Soviet foreign policy behaviors (e.g., Dismukes and McConnell, 1979). Other analysts (e.g., McGwire, 1979) see the forward deployment as forced on the Soviet Navy by situational and environmental pressures, namely the strategic threat posed to Russia by long-range U.S. Navy weapons. The forward deployment and much of the subsequent Soviet naval diplomacy is viewed as caused by this Western threat, not by Soviet intentions or dispositions. Obviously, the first group of analysts will tend to attribute much more expansionistic and aggressive intentions to the Soviet Navy, while the latter analysts will see Soviet intentions as being much more defensive and limited.

Inhibitory causes. While there is a psychological tendency to overestimate causal relationships, this will tend to be true only of positive instances, e.g., if A, then B (i.e., A causes or facilitates B, if A is absent, B is presumably also absent). A second form of causal relation is much harder to detect, that is, when A inhibits B (if A absent, then B present; or if A present, then B absent). These negative relations, or inhibitory causes, will tend to be underestimated.

Consistent with the analysis of negative events in this chapter, observers have difficulty detecting negative causes, i.e., effect B occurs

only in the absence of A, and never in A's presence; A is an inhibitory factor which suppresses B. Because A and B vary inversely, rather than directly, their relationship is more difficult to perceive than facilitory causes which vary positively with their effects. Inhibitory causes are cases of what cognitive psychologists term "exclusive disjunctions" (e.g., Posner, 1973: 76, 115). People find such concepts more difficult to code into memory, and to identify correctly, than positive concepts.

The role of inhibitory causes becomes important in naval analysis as estimators attempt to assess such things as deterrence and the "denial" element in Soviet motives (see Booth, SNP: 365), i.e., Soviet efforts to deny Western influence in Third World countries. As was noted above, it is quite easy to fall victim to a "deterrence trap," and assume that a self-proclaimed deterrent is the cause for a nonevent. Similarly, the Soviets can easily claim causal credit for a lack of Western influence in a region. Such theory-driven inhibitory relationships may be overestimated, while data-driven inhibitory relationships may be underestimated.

Causality and representativeness.

> In war events of importance are the result of trivial causes.
>
> Julius Caesar, De Bello Gallico

A person may use the representativeness heuristic when seeking causal explanations of events. Causes may be sought which resemble the effects observed. Great causes may be sought to explain great events, complex causes to explain complex events, etc. Rothbart and Fulero (1978) term this the "profound motive fallacy," the tendency to implicitly match causes and events. Shweder (1977) uses the term "magical thinking" the belief that resemblance implies causality. Intuitive perceptions of the causes of events differ depending on whether the consequences are important or trivial. Severe and bad outcomes are attributed to malevolent motives, while large good, outcomes are attributed to good intentions (Nickel, 1974; Rothbart and Fulero, 1978).

Foreseeability.

> Never attribute to malice that which is adequately explained by stupidity.
>
> Hanlon's Razor

Rothbart and Fulero's study suggests that "foreseeability" may play an important role in causal attributions. If an actor is perceived as being able to foresee the consequences of his actions, then the actor's motives are perceived as more profound and important the more serious the outcomes of the actions. Good outcomes suggest good motives and bad outcomes bad motives if the actor could anticipate his actions' consequences.

While the tendency to match causes and effects may be strong, it is not overwhelming. However, it does lead to an overestimation that

representative causes are responsible for effects and an underestimation of the importance of less representative but valid causes. In particular, a case of the "profound motive fallacy" seems to occur in analyses of current Soviet foreign activity, including naval behavior.

Analysts of Soviet foreign policy motives have long realized that Soviet or U.S. actions could lead to nuclear war. It was often argued that neither side wanted such a war and that it could only come about through a miscalculation; neither would ever take actions that could explicitly lead to nuclear war. In effect, the real possibility of nuclear war was explained by the belief that neither side wanted it but neither side could foresee how it might occur, or predict it, or always be sure of avoiding it. This appreciation of Soviet and Western outlooks was fairly common among Western analysts as long as Soviet strategic capabilities were inferior to those of the United States. This outlook seems to have been a major contributor to the Western mirror-image perceptions of the West and the Soviets: that the Soviets, like U.S. leaders, did not believe that nuclear wars could be won, and that nuclear war deterrence, rather than its conduct, was of the greatest importance.

As the Soviets first matched and then seemed to surpass Western strategic systems levels, data accumulated which suggested the Soviets did not perceive nuclear war as a hopeless situation, and that they were perhaps explicitly considering strategies to fight and "win" nuclear war. At any rate, their military planning and doctrine included much greater attention to the conduct and outcome of nuclear war than was present in Western strategy, and much less concern with the deterrent aspects of nuclear systems.

This Soviet "warfighting" outlook implied foresight; i.e., the Soviets had anticipated the wartime implications of nuclear systems and were designing their own not just for a peacetime deterrence mission (if at all) but also (or perhaps exclusively) for a warfighting mission.

This "warfighting" attribution seems also to have led [illegible] analysts to a "profound motive" attribution -- if the Soviet[illegible] foresee nuclear war to the extent of planning for it, they must have a profound motive for their behavior with respect to these profound consequences. If they could anticipate such a war, even plan for it, they must intend in some way to use such war preparation, if only as a threat to gain their objectives. The Soviets were thus seen as more willing to intentionally manipulate the threat of nuclear war to get their way, because they were seen as foreseeing nuclear war.

A Soviet willingness to prepare for nuclear war (rather than to merely prepare deterrent systems, which has been the Western course, at least until recently), obviously does not necessarily imply that the Soviets are willing to risk nuclear war, or to use strategic systems as instruments of political blackmail or coercion. Some Western analysts seem to believe, however, that this is the case, (e.g., Pipes, 1977). It seems at least possible that these analysts are attributing profound motives (willingness to risk war or use nuclear blackmail) to account for profound effects (preparing for war) and Soviet foresight (planning for war). Certainly the most profound form of defense is defense from nuclear attack. Soviet efforts to that effect seem to have led some Western analysts to attribute the Soviets with motives more profound than simply defense, i.e., with motives that are offensive as well. The fact that the Soviets are unwilling to see Western political leaders as forsaking the

threat of nuclear war is underweighted, while the Soviet's planning for nuclear war is overweighted in assessing the causes of Soviet behavior. One consequence of this profound motive fallacy" is that the forward deployment of the Soviet Navy and its use in diplomacy is perceived by some Western analysts as suggesting the Soviets are willing to risk nuclear war in confrontations with the U.S. Navy (see McGwire, SND, prologue for a similar conclusion).

Magical thinking. In "magical thinking" the salient features of effects are used as indices to guide the search for causes. Those events" that share the features of the effects (or have directly opposite features) are perceived as the most likely causes.

A stress on military or naval doctrine as the cause of naval actions, may be seen as, in part, magical thinking. Doctrine shares many of the features of naval actions and thus forms a ready explanation of naval behavior. However, doctrine is ambiguous and encompassing, and often self-contradictory. Any naval action could probably be plausibly explained by some doctrinal "principle." Some causal factors other than doctrine must dictate which part of doctrine is operative in leading to any given naval behavior.

McGwire emphasizes Soviet naval construction as both cause and effect of Soviet naval behavior. For example, he writes (SND: 176):

> All interpretations of Soviet policy must involve some element of conjecture . . . analysts of Soviet naval policy are unusually fortunate in having available a range of specialized and reasonably concrete data from which to derive, and against which to test their hypotheses. By its very nature, the significant evidence comprises large discrete items (warships) . . . The number and characteristics . . . of ships reflect the mix of operational requirements . . . Operational activity reflects established tactical concepts and assessments of threats . . . Patterns of deployment reflect decisions on the employment of available capability . . . all this evidence can be combined . . . [to] provide a reasonably clear outline of underlying naval policy and its evolution.

This risks being magical thinking to the degree that it underestimates the flexibility of naval vessels and overestimates the connections between decisions and behaviors. That is, particular naval missions do not entirely fix all parameters of naval construction (engineering and economic limits also play their roles), nor is the mission flexibility of a Soviet ship totally fixed by its technical parameters (i.e., availability of bases can greatly extend the mission radius of ships). Soviet ships respond to design limits and requirements that are not simply naval, and Soviet ships may undertake (not necessarily successfully) missions for which they are ill-fitted. Soviet ships and ship construction need not simply reflect their immediate naval causes or effects.

McGwire does not explicitly imply otherwise, in fact, he repeatedly states that he considers and combines various lines of evidence (what the ships were built and used for, what the ships must attempt to

do) in his estimates of what the Soviets plan to do with their naval forces. Nevertheless, a heavy emphasis on ship and system features may predispose the hardware analyst more than others to expect that the causes of Soviet ship construction will resemble the ships, and that the effects of Soviet construction will resemble the ships' capabilities. The hardware analyst may tend to underestimate the degree to which Soviet ship construction responds to nonnaval pressures, or the degree to which Soviet ships may undertake missions or have effects for which they were not designed.

Capabilities and intentions. McGwire himself describes a similar problem (SND, prologue): that the close examination of the evolution and character of Soviet capabilities may lead the analyst to suggest that Soviet intentions follow from (or resemble) capabilities.' The consequence is that the analyst may calculate the worst that the Soviets could do and reason that this is also what they will do. An equally misguided tendency is to estimate the limits of Soviet capabilities and then assume the Soviet Navy would not undertake missions that exceed those apparent limits. An assumption that intentions simply resemble capabilities is the most dangerous manifestation of the representativeness heuristic in naval analysis but one which seems quite rare among the Soviet naval estimates sampled.

However, because the conclusion that intentions will resemble capabilities is so dangerous, naval analysts should be especially alert to this bias, particularly in more subtle shapes. For example, an analysis that Soviet ships are especially limited with respect to some mission (e.g., sea control, cf. Friedman, NPSP) should not be taken to automatically imply that the Soviet Navy would never attempt that mission. Prior to Pearl Harbor, for example, U.S. naval analysts reasoned that, because it was ill-suited for sea control, the Japanese Navy would not undertake massive offensive operations far from home waters. On the other hand, a particularly strong Soviet capability should not be taken alone as evidence that the Soviets planned or intended to use that capability. The Soviets' submarine strength provides them a strong capability to interfere with Western sea lines of communication, but it is questionable whether this mission has a very high priority in the Soviet Navy (see, e.g, STS for discussion of the pros and cons on this issue). Analysts should avoid letting capabilities alone suggest causal schemes for Soviet intentions, or assuming intentions must resemble capabilities.

Causal search biases. Two recent studies by Shaklee and Fischhoff (1977, 1979) suggest that people analyze a situation until they have identified a minimal set of sufficient causes. Other possible causes are thereupon ignored or dismissed. People seem to believe that multicausal situations are unlikely to occur. When they know or believe one cause is present, they believe other causes are absent. Potentially ambiguous multicausal events are treated as unambiguous single causal ones. Shaklee and Fischhoff label these tendencies the "principle of minimum causation."

Minimal causation

Complex problems have simple, easy-to-understand wrong answers.

Grossman's misquote of H. L. Mencken

People seem to feel quite comfortable with the first causal explanation that is congruent with their observations. There may be very little or no searching or testing beyond the first monocausal hypothesis that fits the data. Since the first plausible cause may not be the best explanation, Shaklee and Fischhoff (1977) observed, "the true cause may never be evaluated because the question is closed prematurely."

Shaklee and Fischhoff (1979) tested whether people tend to use a "parallel search" (i.e., examine information on all possible causes before making causal judgments), "serial search" (examine one possible cause after another), or a "truncated search" (examine one cause without considering other[illegible]). They found people tend to investigate only the first cause they believe to be acting, i.e., they truncate the search among possible causes after the first plausible one is found. Once an event could be plausibly explained, there was very little investigation of other possible causes.

Minimal causation is consistent with a belief that facts and events fit only one theory. Since people may select first the cause that is most available or most representative of effect, other causes less available or representative will tend to escape attention. Because the first plausible cause is not likely in general to be the best causal explanation, people will tend to be overconfident regarding the accuracy of their causal theories. That is, people will tend to believe they have settled on, not just the first, but the best explanation.

Causal hydraulics. The tendency to assume that events can be explained by a single cause and to seek the first plausible causal explanation leads to a simple hydraulic notion of causality, i.e., a belief that causes compensate for and compete with each other in producing for effects. That is, analysts may believe that if strong political factors are causing a Soviet behavior, strong military or naval factors cannot also be causing that behavior. Alternatively, an analyst might perceive a strong causal influence by the Soviet Navy on the actions of the U.S. Navy and not perceive very strong U.S. causal influences acting on the U.S. Navy at the same time.

For example, a popular technique in foreign policy analysis is to examine decisions as the product of bureaucratic competitions between agencies and interests in a nation (e.g., SNP, chs. 2-6). Often these bureaucratic actors are seen as necessarily competing: more for the Navy implies less for the Army; more guns means less butter. For example, Warner (SNP: 79) wrote of Soviet weapons acquisition:

> The process is fundamentally political. It is permeated with pulling and hauling over roles, missions, budgetary priorities, and many other institutional considerations.

A more sophisticated outlook acknowledges the consequences of rational engineering and policy-making: the Army and Navy may compete for respective shares but do so in a manner that leads to more for both. A nation may increase productivity so that there are more guns and more

butter, even though military and civilian sectors continue to compete. As game theorists have long ago noted, a zero-sum game (strict competition) admits only one motive, while nonzero sum games allow players to act with a variety of mixed motives.

Causal hydraulics imply that if an analyst perceives a factor to be present which is known (or strongly believed) to be sufficient cause for some action, then other factors are assumed by the analyst to be nonexistent or relatively unimportant. This leads to the fallacious belief that, if A can be shown to have caused B, this somehow proves that C did not cause B. The presence of A is then taken as evidence for the absence of C. None of the naval analysts sampled seemed to adhere to this fallacious belief. To the contrary, Weinland (SND: 297), for example, notes that

> We are forced to infer antecedents from their observable consequences -- remembering all the while that any action can, and most actions probably do, have more than one antecedent.

Its operation in the larger context of Soviet analysis is not unusual, however. For example, some who argue the Soviet Union is motivated by a desire to obtain military superiority seem to believe this implies that the Soviets could not also be motivated by a desire for detente with the United States.

Parsimony, misguided parsimony, and indiscriminant pluralism.

> If it's worth doing, it's worth overdoing.
>
> Roger's Blessing
>
> Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system.
>
> W. v. O. Quine, 1953

Naval analysts occasionally suggest that one of their number has an overly simple causal model of Soviet naval behavior. For example, Weinland (SND: 294) wrote the following of McGwire's analysis of Soviet forward deployment:

> Most of what he says is right. Some of what he says, on the other hand, is so right that it is wrong. Soviet naval policy and practice are in fact predominantly reactive in character and defensive in orientation. In McGwire's argument, however, they are only that -- the imperatives of strategic defense are depicted as all-powerful and totally pervasive, and the Soviets are clearly denied freedom of initiative. Strategic defense is the reason for, and explanation of, all that they do.
>
> This is determinism [sic] par excellence; and single-factor explanations of this nature -- especially explanations of

> the behavior of organizations as large and multifaceted as the Soviet Navy -- require a considerable amount of direct empirical evidence in their support before they can be accepted. That kind of evidence simply does not exist in this case.

Weinland's critique suggests a common assumption in naval analysis and political science, namely that the rule of Occam's Razor does not apply. That is, the notion of parsimony per se, is not highly valued in explanations, and instead efforts are expended to account for everything, rather than to account for most things. Weinland seemingly argues that all Soviet actions must be explained. He writes, for example (p. 295)

> . . . the evidence does support the interpretation of Soviet naval policy and practice as predominantly reactive in character, and defensive in orientation. Some of their actions, however, require other explanations . . .

He challenges McGwire's parsimonious strategic defense hypothesis [which may be reductionistic but is not deterministic] because it fails to cover each Soviet naval move. A requirement to explain all rather than most is absolutely deterministic: it implies there exists a cause for every Soviet action and leaves no room for chance, luck, random factors, or causes yet undetermined.

Weinland makes it clear that he does not expect every Soviet move to be fully explained, but he seems to argue that this should be the ultimate objective. He also notes that the causes of Soviet naval moves are not always Soviet, that Western and international factors play a role in causing Soviet policy and practice.

The widespread notion that all Soviet naval actions should be explained, and that different causes, or combinations of causes, could be found for every Soviet event, might lead to a more sophisticated form of "minimal causation." That is, a plausible cause is applied to each different event, until that cause no longer seems plausible, at which point a new, plausible cause is selected, and so forth, until every event is explained.

The effort to explain every Soviet event leads to overfitting the data; i.e., the ability to fit an explanation to all events under examination. With a sufficiently large number of explanatory variables, any set of events can be explained to whatever degree of precision is desired. The consequence however, is "shrinkage," the inability to accurately predict future events. A parsimonious explanation may not be a perfect fit, but is more likely than all-encompassing explanations to hold true in future cases. Fischer (1970) terms the multiplication of causal components "indiscriminate pluralism" and notes that, without a specific means of weighting causes, no clear interpretations of effects are possible. Neither, we would add, is accurate prediction possible.

Calling a cause a cause. Fischer (1970) has also noted the tendency to use linguistic subterfuges to introduce multiple causes without labeling them as such. The terms "antecedents," "factors," and "rules" are some common cover words for causes. It is not clear why analysts avoid the term "causes" but readily cleave to "rules" or

"factors." What does seem clear is that, as used by analysts, "rules" or "factors" differ little from causes, and the use of the former terms tends only to complicate issues. For example, Dismukes and McConnell (1979: 278) have the following to say about their "rules" of Soviet naval diplomacy:

> The rules are . . . explicitly not meant to provide detailed predictions of superpower action in future Third World crises. By identifying regularities in past behavior, and hence reducing uncertainty about future behavior, the rules can aid the policy-maker in diagnosing the situation and making contingent forecasts . . . the rules provide a model of behavior against which the superpowers' actual behavior can be assessed.

How the rules can help make contingent forecasts without making detailed predictions is not clear, nor is it clear how a model of future behavior differs from a causal model. It seems possible to speak of "causes" without foresaking one's belief in complex cause-effect relationships.

Fundamental attribution error.

> It must be remembered that, among all changes, the nature of man remains much the same; the personal equation, though uncertain in quantity and quality in the particular instance, is sure always to be found.
>
> Alfred T. Mahan

People tend to estimate the causes of an actor's behavior as predominately or exclusively motivational and dispositional; i.e., as due to the actor. On the other hand, they tend to underestimate the impact of the environmental or situational pressures as causes of behavior. Psychologists have labeled these tendencies the "fundamental attribution error" (Ross, 1977).

One explanation of this bias is due to perceptual salience, observers tend to focus on the actor rather than on aspects of the situation. Such "perceptual focusing" tends to produce causal attributions; whatever or whoever we focus our attention on becomes more apt to be perceived as causing the events we see (Duval and Hensley, 1976; Storms, 1973; Taylor and Fiske, 1975). This implies that observers, who are focusing on the actor should perceive the actor (or rather the actor's dispositions and motives) as causing the actor's behavior, whereas the actor, who is focusing on the situation and environment, should perceive environmental pressures as causing his behavior. This is, in fact, what psychological studies repeatedly find: actors attribute their own behavior to the environment while observers attribute the same behavior to the actor (Jones and Nisbett, 1971).

The implication of the fundamental attribution error is that estimators may too readily infer broad personal dispositions and expect consistency and predictability in behavior across a wide variety of situations and contexts. There is a tendency to draw hasty conclusions about dispositions while overlooking relevant environmental forces and

constraints. Jervis (1968) suggests several ways this tendency may lead to misperceptions of intentions: estimators may see the behavior of others as more centralized, disciplined, and coordinated than it is; and may tend to perceive the position of a state's Foreign Office [or Navy] as the position of the state.

Naval analysts who argue that Soviet naval peacetime and crisis behavior is largely determined by a long-term Soviet foreign policy plan may be victims of the fundamental attribution error. Similarly, those analysts who see Soviet naval evolution as the exclusive handiwork of Admiral Gorshkov are probably overestimating that already formidable commander.

Jervis (1968) also notes that when states interact, one government may overestimate the degree to which some desired behavior by the second state is due to the influence of the first, and will overestimate the degree to which undesired behavior by the other state is due to internal forces. That is, estimators may generally see other states as behaving negatively for dispositional reasons and behaving positively because of influence exerted by the estimators' own government. In this latter event the estimator overemphasizes his own government's influence in bringing about the positive behavior. These tendencies have been noted in psychological research. Taylor and Koivumaki (1976) found that people are perceived as the cause of positive, desired outcomes while situational factors are regarded as causing negative outcomes. They labeled this the "positivity effect." Snyder, Stephan, and Rosenfield (1976) found an egotistic tendency to make attributions that put oneself in the best possible light, attributing good outcomes to one's own skills while bad outcomes are attributed externally.

This tendency has been repeatedly exploited by the Soviets, according to McGwire (e.g., SND: 3), through what he terms the "Western amplifier" effect (p. 4). McGwire argues that Western observers have overestimated the extent to which the Soviet Navy has deterred the U.S. Navy during superpower crises in the Middle East. In effect, he accuses the observers of falling victim to the "deterrence trap" described above, i.e., of believing that the Soviet naval presence caused a nonevent (i.e., U.S. nonintervention), and ignoring the situational factors that also influenced U.S. behavior.

From the Western perspective, attributing the course of Middle Eastern crises to the Soviets (rather than to U.S. and Middle Eastern dispositions) may be one means of justifying the negative aspects of these crises (e.g., the oil embargo and subsequent energy and financial crises). The egotistic tendency would lead Western observers to look for external causes for these negative outcomes for the West and the Soviets' presence provides a convenient explanation. In turn, the Soviets have exploited these Western perceptions to advertise their influence over the U.S. Navy, and their ability to protect Third World clients. Several naval analysts seem to give too little attention to U.S. forces as a counter to Soviet intervention in crises and too much to Soviet forces as a counter to U.S. intervention. After all, it was the Soviet Union that explicitly threatened to intervene in the 1973 crisis and the United States that forestalled their intervention, both directly, through a world-wide military alert, and indirectly, by forcing Israel to moderate its actions.

The fundamental attribution error can also be noted in efforts by

naval analysts to explain actions which are largely just naval as serving Soviet foreign policy ends. For example, analysts have elaborated political links between a Soviet naval exercise and the "Cod War." They have explained Soviet port visits in the Indian Ocean as almost wholly diplomatic (despite the limited diplomatic character of Soviet actions in that region). They have attributed Soviet naval activities to influence-building despite numerous severe setbacks in Soviet foreign affairs (e.g., Indonesia, Egypt, Somalia) which reflect a largely negative diplomatic balance sheet, and a need for foreign bases that goes beyond influence building.

Assessing dispositional causes. Given a tendency to overestimate dispositional causes and to underestimate situational causes, especially when unliked actors have performed "bad" deeds, what steps might observers take to make more realistic estimates? Two skills are necessary at the outset: the ability to detect covariations and to estimate their strengths. Unless the observer can estimate covariations fairly accurately (that is, weigh all four cells in the present-absent table, not just the ++ cell), the other steps outlined below are of little value. In fact, if based on faulty estimates of covariations, the steps below may be seriously misleading.

The initial requirement of causal evaluation is to determine that the suspected cause reliably covaries with its effect. This assessment can and should be done retrospectively, as well as for data which is obtained in the future. That is, all relevant past events can be reexamined to determine if the suspected cause was reliably related to events. The caveats noted above for covariation assessment, e.g., the possibility of illusory correlation perception, should lead the analyst to adopt a data-driven rather than theory-driven approach. Secondly, the difficulty in perceiving negative covariations should lead the analyst to take pains to explicitly hypothesize inhibitory causes and search through the data for these, as well as for facilitory causes.

Given evidence of reliable covariation, the analyst can then investigate three further dimensions to determine whether the relationship seems dispositional and intentional, rather than environmental and situational. These dimensions (suggested by Kelley, 1967, based on J. S. Mill's method of difference) are consensus, consistency, and distinctiveness.

Consensus refers to the degree to which other actors demonstrate the same causal relationship. Suppose the suspected causal relationship is "Soviet interest in Third World influence" (cause) led to "forward deployment of the Soviet Navy" (effect). If other nations also undertake widescale naval deployments, the action (or effect) is not unique, i.e., it has high consensus. On the other hand, if only the Soviet Navy undertakes far-flung deployments, their behavior is unique and has low consensus.

Consistency refers to context, time, and modality of response. Has the Soviet Navy consistently deployed forward over time, or did forward deployment develop suddenly? Does the Soviet Navy consistently maintain a forward deployed force, or does the force expand or contract with different situations? Are Soviet forward deployments consistently composed of the same units or types of ships, or does the modality of response vary?

Distinctiveness refers to the target of the response. Is the Soviet forward deployment limited to particular target areas within the Third World or to the entire Third World? Are Soviet responses in certain regions different than in others? Highly distinctive responses are ones that occur only for a small set of possible targets. Low distinctive responses are ones that occur across all targets.

Soviet forward deployment can most confidently be attributed to Soviet dispositions and intentions if it is low in consensus (other nations do not undertake far-flung naval operations), if it is highly consistent (the Soviets consistently maintain a high level of forward deployment over time and deploy forces of consistent composition), and if it has low distinctiveness (the Soviet Navy deploys throughout the Third World). To the degree that the Soviet response is high in consensus or distinctiveness, or low in consistency, the analyst should begin to look beyond the Soviets' dispositions and intentions, and consider environmental or situational forces. In particular, specific information on the distinctive features that lead to Soviet responses and the causes of inconsistencies (i.e., changes) in the timing or modality of Soviet responses provides the analyst with predictive power.

It is instructive to apply Kelly's covariation method to McConnell's classification of Soviet Third World diplomacy of force (Table 4.3.). The first three case categories are actions that most naval states, not just the USSR, would undertake in the Third World. The first two of these cases would only be undertaken by the Soviets in certain contexts and with distinctive targets, e.g., the Soviets do not always support the domestic government of a Third World country, it depends on the context and the country. The last four cases all evidence the same pattern, low consensus (other naval nations would not tend to do these things) which suggest Soviet disposition and intention, low consistency which implies the Soviets take these actions infrequently or only in limited contexts, and high distinctiveness which means these actions are taken only on behalf of certain distinctive countries.

The only case for which Soviet action has high consistency and low distinctiveness is case category 3: defense of citizens and property. The Soviets are likely to perform this action in most contexts and regardless of the character of the offending nation. However, the consensus is also high, most other naval nations would also undertake to defend citizens and property.

In none of these cases do we observe the "perfect disposition pattern" i.e., a pattern that would suggest the Soviets are acting only on the basis of their own intentions and dispositions. Nor do we observe a purely situational pattern, i.e., evidence that the Soviets only react to specific contexts and targets and do so as all other naval nations do. Instead we see a pattern that is situational relative to context and targets, but relatively unique to the Soviets, especially with respect to Third World international security. Categorizations like McConnell's point to the critical features of context and targets that predispose Soviet actions, and thus perform a highly heuristic function for both cause-effect analysis and prediction. The application of Kelley's covariation categories may serve to refine the relative weight and importance of these critical features in the analyst's model, especially as individual cases in a category are assessed.

Table 4.3. Analysis of McConnell's Classification of Cases in Soviet Third-World Diplomacy of Force in Terms of Consensus, Consistency, and Distinctiveness.

Case Divisions and Categories	Consensus		Consistency		Distinctiveness	
	High	Low	High	Low	High	Low
I. Security on the High Seas						
1. Demonstration of intent to protect USSR assets, or the assets of clients, at sea.	X			X	X	
II. Third World Domestic Security						
2. Demonstration of support for the domestic authority of an established government.	X			X	X	
3. Demonstration against an established government in defense of Soviet citizens and property.	X		X			X
4. Military support to a domestic faction during an interregnum when the U.S. is inhibited from counterinvolvement.		X		X	X	
III. Third World International Security						
5. Demonstration of intent to protect a client in a confrontation with a state that, held in odium by the international community, does not enjoy U.S. patronage.		X		X	X	
6. Demonstration of support for a client threatened (or that might be threatened) by a Western great power or in actual conflict with such a power.		X		X	X	
7. Demonstration or actual intervention against a U.S. client that is defeating a Soviet client.		X		X	X	
A. Perfect Dispositional Pattern		X	X			X
B. Perfect Situational Pattern	X			X	X	

Causal warnings. In closing this section on assessing causes and effects, it is worth noting that Booth (SNI, ch. 24) offers his fellow naval analysts several pieces of advice on cause-effect assessment "to prevent an influence relationship being seen in too simple a fashion (p. 470)." His warnings closely parallel some of the observations made above. Booth's points are listed on the left in Table 4.4 with psychological mechanisms on the right. The fact that Booth takes pains to summarize these various problems in causal analysis in detail, and offers clear warnings to other analysts, suggests that causal biases are not minor problems in naval analysis. Booth himself notes "the frequent difficulty of identifying the precise relationship between naval causes and political effects." To some degree these difficulties originate in the problems all people have in assessing causes and effects.

Prediction

Predicting and forecasting are central tasks in intelligence analysis. However, naval and military intelligence analysts often have been charged with being unable to accurately predict either the capabilities or intentions of adversaries. For example, Albert Wohlstetter (1974a, b, c; 1975a, b) demonstrated how U.S. intelligence analysts persistently underestimated several measures of Soviet strategic forces in the mid-1960s and early 1970s. Several scholars have noted the consistent tendency in the late 1950s and early 1960s for U.S. intelligence analysts to overestimate these same measures (e.g., Dick, 1972; Gray, 1972, Licklider, 1970). Lee (1977) described how the U.S. CIA underestimated Soviet defense expenditures by a factor of two and ultimately publicly revised its estimates upward. Whether the CIA is presently over- or underestimating Soviet defense expenditures remains an open and heatedly debated issue which puts into question current CIA forecasts of Soviet defense efforts (cf. Cockle, 1978; Holzman, 1980). That is, a methodology that misestimates past spending cannot be relied upon to estimate future spending.

Not only are intelligence analysts charged with failing to predict capabilities accurately, they often are accused of inability to estimate future intentions. Many writers have commented on the seemingly frequent failures of intelligence to forecast surprise attacks, crises, changes in policies, etc. (some recent papers are Ben-zvi, 1976; Betts, 1978, 1980; Chan, 1979; Gazit, 1980; Handel, 1980; Shlaim, 1976; for a review of literature and an analysis of intelligence failures to predict intentions, see Stech, 1979). Some writers (e.g., Betts, 1978) argue that failure to predict intentions accurately is inevitable.

General and specific causes of failure. Many of the critics of intelligence failures emphasize the political nature of this task. Some note that intelligence analysts are subjected to pressures by their own government to come up with the "right" predictions; i.e., those that serve the decision-makers' purposes. Other critics point to the dilemmas intelligence analysts face as they attempt to act out organizational and political roles and also produce objective estimates. Still other critics point to nonpolitical factors, such as the complexity of the international political environment; the significant problem of separating warning signals and intention indicators from random background noise or deception

Table 4.4. Booth's Causal Warnings to Naval Analysts and Related Biases*

Booth's Warnings	Related Biases
1. We must not be hypnotized by ships at sea, but instead see them in relation to a country's overall influence-building tactics.	1. Salience and vividness.
2. We must not mistake what a particular government claims or hopes to be its influence and what its influence actually is; nor must we confuse the effort to achieve influence with its actual achievement. ...(We) are likely to know far more...about the policy aims and tactics...than (about) the effects of those aims and tactics...	2. Representativeness bias. Deterrence trap. Misguided parsimony.
3. Whether or not naval tactics increase one country's influence with another, for organizational reasons navies are likely to claim influence potential, while decision makers will be predisposed to make some political use of naval forces in order to try to maximize what is an increasingly costly investment.	3. Implicit linking of foreseeability and intention. Mixing capabilities, intentions, and causality.
4. We must be careful to distinguish the possible influence of A's acts on B with behavior by B that might be the result of a coincidence of interests. The pertinent question is: "Would B have behaved in that way in any case?" This is a particularly important warning because the evidence for influence is often only circumstantial.	4. Biases of covariation assessment and hypothesis testing. Magical thinking. Fundamental attribution error. Mistaking chance for causation.
5. We must keep in mind that influence is at least a two-way process...we are interested in more than the influence of A on B. The multidimensionality of the influence process means that we must speculate about the effects of A's acts on not only B's rulers, but on all significant groups in B's political life, on third parties (C, D, E, F, and so on), and on political groups within A. To the extent that A's naval acts affect the expectations of all these groups, it has influence-building potential, in a multifaceted way.	5. Causal search biases, ciminal causation, and causal hydraulics.

*From Booth (SNI: 470).

action; organizational problems; or personality problems among intelligence chiefs.

The argument that the problems of intelligence prediction are unique to the characteristics of the subject, i.e., predicting capabilities and intentions, would gain strength if other forecasting and predicting tasks were performed accurately. That is, if, e.g., economists could accurately predict the growth of the U.S. economy, we would be more willing to attribute failures in the estimation of Soviet defenses as being due to some unique aspects of the latter problem. Similarly, if forecasters could accurately estimate the energy needs of the future, or the demand for transportation, we would be more willing to grant that intelligence estimates of future intentions pose unique problems because of their political or national security nature.

It is not the case, however, that forecasts and predictions of nonpolitical, nonnational security matters are accurate, (see e.g., Ascher, 1978). If anything, intelligence forecasts and predictions are about as accurate as forecasts in such fields as population, economics, energy, transportation, and technology; i.e., not very accurate at all. This suggests that the problems are with the tasks of forecasting and predicting in general, and are not specific problems of intelligence work. This judgment is consistent with the overall approach here, i.e., that people are not particularly good at processing data, integrating it into covariation or causal models, or making predictions. This does not mean that there are not unique problems in intelligence prediction and forecasting. There may be features in intelligence work that make prediction harder than in other fields. But if the ability to predict and forecast in general is weak (and it is) there would seem to be little payoff in addressing unique features and ignoring the general features that keep people from making accurate predictions. And since intelligence analysts' failures are attributed by critics to a variety of biases, weaknesses, vulnerabilities and temptations, it is only fair to compare these failures to those that seem to characterize virtually all forms of systematic prediction and forecasting. In short, nobody forecasts or predicts anything very well, intelligence analysts are apparently no exception, and the problems of forecasting and prediction are probably general as well as specific to a field.

Illusions of control. People have a strong need to master and control their environment. One expression of this need is the attempt to predict future events. In dealing with chance events people often are willing to predict outcomes with great confidence if they have a perception of control, e.g., an experience of early successes, or control over the date, or the opportunity to choose alternatives (Langer, 1975). Given such opportunities people attribute successes and accurate predictions to their own efforts. Hogarth and Makridakis (1979) note the "uncanny similarity between . . . forecasting and experiments concerning the illusion of control." Forecasters are likely to perceive the future as more predictable than it is simply because of their efforts to predict it.

Illusion of order. Related to the effects of the illusion of control that may result from the effort of predicting is the tendency to see patterns where none exist, an "illusion of order." The need to

structure the environment, to make sense out of it, forces people to organize perceptions, in keeping with "Gestalt" principles of symmetry, closure, proximity, good continuation, and common fate (Hogarth, 1975). People abhor randomness and persist in imposing patterns. Hogarth (1975: 273) quotes Simon and Sumner (1968) on this tendency:

> People appear to have strong propensities . . . to discover patterns in temporal sequences presented by the environment and to use these evidences of pattern for prediction . . . The urge to find patterns extends even to phenomena where one may well doubt whether pattern exists.

The need to find patterns is so strong that it extends ironically to those who generate random patterns when parts of a series of random numbers may not, themselves, appear sufficiently random (Lopes, 1980):

> Fisher and Yates found that some of the random numbers they produced failed certain tests of randomness. What did they do? The obvious -- they fiddled with the numbers until they were random enough to pass the tests. Kendall and Babington Smith had a similar problem involving 10,000 of the numbers they generated, which they solved by the simple expedient of suppressing the offending numbers.

One reason people are relatively insensitive to randomness and uncertainty is that they rarely consider the hypothesis that the pattern or process they perceive is probabilistic rather than deterministic (Brehmer, 1980). In studies of probabilistic inference tasks people seem to make these assumptions in the following order: (1) there is a rule, (2) it is deterministic, (3) it depends on the case (rather than on, say, the sequence), (4) the rule is functional (rather than, say, conceptual), (5) the function is positive linear. Brehmer wrote (p. 231)

> When these rules . . . fail, the subjects tend to assume that there is no rule at all, rather than to seriously consider the possibility that the rule may be probabilistic in character.

This insensitivity to the probabilistic character of the events they are attempting to predict seems to underlie many of the problems of forecasters and estimators.

Problems with long-range forecasts.

> Assumption is the mother of all screw-ups.
>
> Wechern's Law

Long-range forecasts (two years or more into the future) are rarely accurate. In an extensive and well-documented appraisal, Ascher (1978) compared predictions to outcomes in five forecasting areas: population, economics, energy, transportation, and technology. In general, prediction errors are large, ranging from a few percentage points

to several hundred. He detected systematic biases as well as unpredictable biases. Commonly, all forecasts in a given period show the same bias. There was no way of knowing beforehand which approach or forecaster would be most accurate. Differences between the accuracy of different methods (e.g., simple or sophisticated) was minimal. Ascher did find some evidence of a learning effect, the predictions tended to be less inaccurate in fields with the longest history of forecasting experience. In general, however, the more distant the forecast target date the less accurate the forecast accuracy. The one central difficulty among forecasts was in providing good core assumptions regarding the phenomenon being predicted:

> The core assumptions . . . the forecaster's basic outlook on the context within which the specific forecasted trend develops, are the major determinants of forecast accuracy . . . When the core assumptions are valid, the choice of methodology is either secondary or obvious. When the core assumptions fail to capture the reality of the future context, other factors such as methodology generally make little difference (Ascher, 1978: 199).

Ascher's conclusions about the critical role of key assumptions about the future context bear a striking resemblance to the conclusions of cognitive psychologists assessing the central elements of problem-solving expertise. For example, Larkin, McDermott, Simon, and Simon (1980: 1342) write that:

> In every domain . . . knowledge has been found to be an essential prerequisite to expert skill . . . large numbers of patterns serve as an index to guide the expert . . . to relevant parts of the knowledge . . . This knowledge includes sets of rich schemata that can guide a problem's interpretation and solution and add crucial pieces of information. This capacity to use pattern-indexed schemata is probably a large part of what we call . . . intuition.

In other words, the patterns the expert problem-solver perceives guides the problem solution, just as the forecaster's core assumptions about the future context guide the forecast. If the wrong patterns for the problem are perceived, or the wrong assumptions about the future context are made, experts tend to make errors, despite their knowledge or methods.

Medium-term forecasting. In many forecasting fields the analysis of patterns in past trends is best accomplished by statistical analysis. To the degree that medium-term (three months to two years) future forecasts are based on trend projections, a simple quantitative model should provide fairly accurate estimates. In fact, for a variety of quantitative predictions (stock market and economic forecasting, mutual fund performance, future sales and earnings estimates, currency exchange ratios, oil prices, future product demand) Hogarth and Makridakis (1979) found, in a review of the literature, that simple quantitative models were more accurate than the judgments of forecasters.

Short-term forecasts.

An ounce of history is worth a pound of logic.

Oliver W. Holmes

Similarly, the present states of many variables are predictive of the short-term (less than three months) future. Again, Hogarth and Makridakis found that

> . . . rather simple, mechanistic methods such as those used in time series forecasting can therefore often make accurate short-term forecasts and even out-perform more theoretically elegant and elaborate approaches used in econometric forecasting . . . quantitative models outperform judgmental forecasts . . . simpler models are often at least as accurate as sophisticated ones . . . simple decision rules can often be as effective as elaborate forecasting and planning procedures.

To what extent are naval analysts guilty of making sophisticated or complex judgments about the short- or medium-term future when simple quantitative projections would probably serve as well or better? Since this study undertook no explicit comparisons of predictions and outcomes, we cannot relate quantitative evidence of inappropriate sophistication, however, it was repeatedly encountered. Rather than simply extrapolate trends, naval analysts tend to develop elaborate models of process to predict future events.

A not atypical example is shown in Figure 4.5, which was displayed in an article on a methodology for threat projections (Ivanoff and Murphy, NPSP: 149). On the left is an index for antiship missile momentum which is simply a measure of the payload weight (W_{wh}) times the missiles maximum speed (V_m). The solid points represent successive generations of Soviet missiles. The open circles are the estimates for future Soviet missiles given by the authors' methodology. The triangular point is the U.S. Harpoon antiship missile. The analysts forecast a sharp exponential acceleration in the value of momentum -- a jump unjustified by the historical data. They explain this jump in capability as resulting from (1) the absence of past improvements and (2) the availability to the Soviets of the necessary propulsion, guidance, and materials technologies that would allow such a jump.

Assumption (1) is similar to the gamblers' fallacy, i.e., the belief that, e.g., the roulette wheel which has not come up even for ten consecutive spins is "due" for an even number. The analysts seem to take the fact that the Soviets have not made technical improvements in a system in many years, even improvements of which they are fully capable, as implying they are "due" for a change. The Soviets may have compelling nontechnological reasons to leave an adequate design alone -- e.g., perhaps it leads to missiles that are very easy to service, cheap, reliable, compatible with other systems, etc. The ability to do it better does not imply a need or desire to do it better.

The Soviets did not make simultaneous exponential breakthroughs in technologies in the mid-1970s. The technologies needed to make the

Figure 4.5. An Example of Oversophisticated Forecasting*

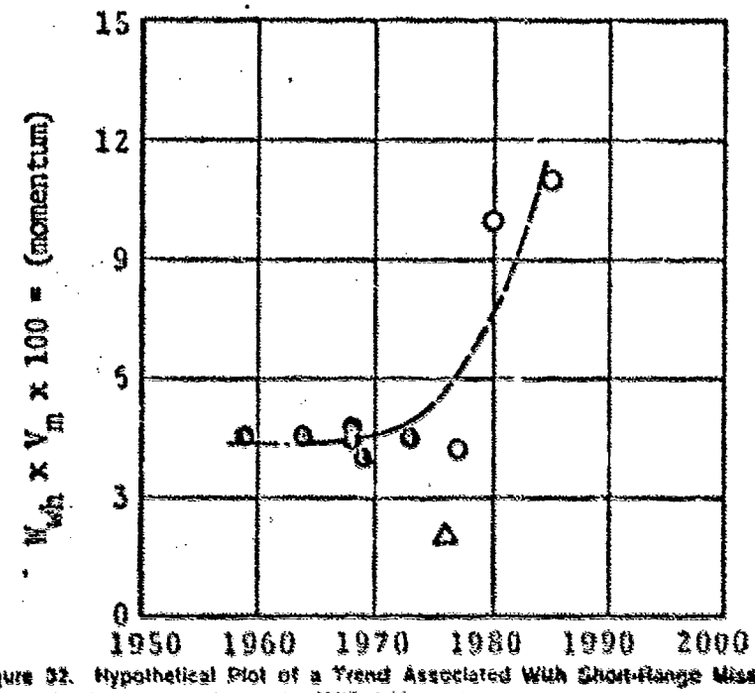


Figure 32. Hypothetical Plot of a Trend Associated With Short-Range Missile Composite Assessment Parameter MCP-2 Momentum.

*Reproduced from Ivanoff and Murphy, NPSR: 149.

estimated jump in momentum have been evolving throughout the historical period of Soviet antiship missile development. Successive generations of Soviet missiles have not reflected (in their momentum) any such evolution, yet the analysts estimate that suddenly they will. The fact that the Soviets have consistently produced missiles that are less than state-of-the-art is ignored and the heretofore unnoted practice of building state-of-the-art weapons is accepted as the trend for the future. This is worst-case estimating: the Soviets are estimated to be planning to build the worst they possibly can, although they seem not to have done this in the past (i.e., they often follow the maxim that the "better is enemy of the good.")

Clinical versus actuarial judgment. Forecasters and naval analysts are not the only judges whose estimates are more sophisticated than but often inferior to simple quantitative models of the data. Psychologists have found that human judges in general are inferior to such mathematical or actuarial formulas derived from the diagnostic input data (see Dawes and Corrigan, 1974; Goldberg, 1968, 1970, Meehl, 1955, Slovic and Lichtenstein, 1971 for reviews of this literature). Even mathematical models of the human judges themselves (rather than of the input data) outperform the judges because the models are more reliable and consistent than the humans.

Human judges are less accurate predictors than the mathematical models whether the humans have more information than is fed into the formulas, or the same data. Even if the judges are informed of the weights given data in the actuarial formula, or the specific predictions of the formula, the judges do worse.

Not only are human judges less accurate than the optimal mathematical regression models of the data; they are less accurate than almost any regression equation. Even if the model weights input data arbitrarily, as long as the weights are nonzero, positive, and linear, the models outpredict human judges (Dawes and Corrigan, 1974). Human judges apply invalid weights to data and weight data unreliably. The models outpredict humans because of improved reliability alone.

In many respects this research with human judges actually gives the humans their best advantage. The judges are typically skilled professionals, making predictions in fields for which they have special knowledge and much experience. The judges generally receive feedback on their predictive accuracy. They have ample time to study the data and perform calculations. The data are often presented in well-organized and categorized form.

Predictions and representativeness. One reason why analysts may [illegible] overly sophisticated predictions that are inferior to simple [illegible]relations of quantitative projections is the representativeness heuristic. That is, people tend to ignore base-rate data, e.g., prior probabilities, population proportions, data on averages or central tendencies. On the other hand, people overemphasize data on the specific target case, the particular event, object, or phenomenon about which the prediction is to be made. Nisbett and Ross (1980: 142) write:

> People are thus insufficiently concerned with the relative frequencies of the various outcome possibilities and are

overly concerned with the degree to which features of the target are similar to features of some particular outcome.

In the case of the capabilities predictions made by Ivanoff and Murphy (NPSP, see above), the anlysts have overemphasized the state-of-the-art case and underestimated the base-rate of the Soviets' past track record. Unless the analysts are entirely confident that future Soviet missiles will resemble the state-of-the-art case, they should regress their predictions toward the average of past Soviet performances. To the degree that the case data are believed to be less than perfect predictors the estimated missile characteristics should be regressed toward the average of past performance.

Regression effects. Closely related to the neglect of base-rate data is the tendency to assume that predictor variables are perfectly related to predicted variables. That is, an extreme score on a predictor variable is often taken as predicting an equally extreme score on the predicted variable. This is true, however, only if the two variables are perfectly correlated, i.e., the predictor is perfect. When the predictor variable is not a perfect index of the predicted variable, the estimator should regress the prediction from the predicted value toward the mean value of the entire data set. To the degree that the correlation between predictor and predicted variables decreases (i.e., the predictor becomes less diagnostic) the prediction should be shifted closer to the mean.

A hypothetical example of nonregressive prediction might be a naval analyst who estimates that the high levels of Soviet out-of-area operations in the 1973-1974 period indicates future high levels. Having shown that they were capable of such extensive forward deployment, and having gained diplomatic advantages from them the Soviets might be perceived as likely to continue this level of operations. In this case the analyst is using an extreme case (the high levels of deployment, prompted in part by a Middle East crisis) as a predictor. Since this was an extreme case, unless it is a perfect predictor (unlikely since the same factors, e.g., crisis, would not persist), a more regressive prediction would be in order.

Regressive predictions through dilution. Psychologists have found that people sometimes make regressive predictions, but for the wrong reasons. That is, when people were given data on predictor variables and information that was totally nondiagnostic (i.e., known to be unrelated to the prediction task) the predictions were less extreme (i.e., regressed toward the mean). In other words, information that was worthless for prediction served to dilute the effects of extreme information that was highly diagnostic (Nisbett and Ross, 1980: 154-155). The regression of the prediction toward the mean and away from the extreme value of the predictor variable was normatively appropriate, but occurred for inappropriate reasons. Rather than integrating the highly diagnostic, extreme information with nondiagnostic data, the estimator should integrate the extreme value with base-rate data, regressing the extreme value toward the mean value of the predictor variables. While the dilution bias may favorably offset nonregressive case-specific predictions, it could also occur at inappropriate times, e.g., overdiluting highly predictive indicators. A more appropriate means of

adjusting predictions is described below under "sensitivity testing."

Diagnostic and nondiagnostic data. Psychologists have found that people will make heavy use of case-specific information even when those data are held to have low predictive validity and to be unhelpful in prediction, at the same time that base-rate data are ignored. People will thus rely on specific information about the target that they recognize as relatively invalid and nondiagnostic while ignoring data averages which are highly diagnostic and should be rejected only if very strong diagnostic data on the target case are available (Kahneman and Tversky, 1973).

Even when data on the target case are perceived by the judge to be absolutely worthless, judges persist in neglecting available data on base rates. Instead the judges proceed as if no data were available. Only when there is no information at all on the target case did judges make use of the base-rate data (Kahneman and Tversky, 1973; Lyon and Slovic, 1976).

When base-rate data are used. Several psychologists have challenged Kahneman and Tversky's (1973) conclusion that base-rate data are totally ignored when case-specific data are available, and a series of studies have shown that people do attend base-rate information in certain circumstances. In particular, base-rate data are used when they have a clear causal relationship to the base being predicted (Ajzen, 1977; Bar-Hillel, 1977; Tversky and Kahneman, 1977), when case data are absent (Kahneman and Tversky, 1973), when the process generating the base data are clearly understood (Howell and Burnett, 1978), when the base-rate data are highly concrete (Manis, et al., 1980). Even in those cases in which base-rate data is employed in prediction, people failed to weight it as heavily as would be normatively appropriate, and would often use the base-rate data without being aware of doing so (Manis, et al., 1980).

As Nisbett and Ross (1980) point out, people are quite willing to make broad generalizations and predictions on the basis of a single, vivid case study, but will refuse to make predictions on the basis of data averages that may reflect large numbers of cases. Such an outlook on data synthesis and prediction is scientifically unjustified. The belief that a single case, sampled from a larger population, is representative of the entire population is an extreme instance of what Tversky and Kahneman (1971) labeled "the law of small numbers": a belief that accurate inferences and predictions about the population can be based on very small samples of data. Their research demonstrates that even scientists have a very inadequate understanding of the relations between sampling and statistical inference and prediction.

Sensitivity testing induced base-rate use. People can be induced to use base-rate information even when it is not causal, concrete, produced by a clearly understood process, and when case data are also present. This increased use results when people are asked to conduct subjective sensitivity tests, i.e., to determine how their predictions would be changed if the base-rates had very different values. When people considered several base rates they tended to make more use of base-rate data and their predictions shifted in the normatively appropriate direction (Fishhoff, Slovic, and Lichtenstein, 1979). Similarly, when

people considered various levels of validity between the predictor variables and the outcome, they weighted data more appropriately, i.e., giving less weight to less valid predictors. None of the naval analysts reviewed described any use of sensitivity testing.

Predictions, scenarios, and compound probabilities. A scenario consists of a series of events linked together in narrative form. According to the mathematics of probability, the likelihood of a multievent scenario happening is the multiplicative product of the probabilities of the individual links. The more event links in the scenario, the lower the probability of the entire scenario's occurrence. The likelihood of the scenario's least likely link sets the upper limit on the probability of the entire narrative. Compound events cannot be more likely than the least probable of the simpler events that constitute the compound.

People typically do not evaluate compound events or scenarios in this way. They typically judge the probability of a multilink scenario on the basis of the average likelihood of all its links (Schum and Pfeiffer, 1973; Slovic, Fischhoff, and Lichtenstein, 1976). Strong links appear to compensate for weak links and scenarios are constructed with perceived probabilities that increase as they become longer, more detailed, and less normatively probable. Individual weak links in a scenario may be perceived as having low probability, but, if they can be combined to tell a good story, the weak links are buried in masses of coherent details and the scenario is given much more credibility than it deserves.

Ross and his colleagues (Ross, Lepper, Strack, and Steinmetz, 1977) have found that merely identifying potential antecedents to explain an event increases that events' subjective likelihood. Even when people knew they were formulating hyothetical explanations, they believed events were more likely simply because they had produced explanations for why the events would or could occur. In other words, when analysts generate a plausible account suggesting how a particular event might be predicted from past data, they may inappropriately make the inferential leap from possibility to probability. Merely searching for possible links between antecedents and specific consequences may produce increased subjective probabilities for the predicted events. Similiarly, merely imagining the occurrence of an event can increase the person's expectation that the event will occur (Carroll, 1978).

The clearest naval example in our sample of an elaborate scenario influencing an analyst's estimates of likelihood is Ra'anan's account of Soviet decision-making in the Middle East 1969-1973 (SNP, ch. 11). Ra'anan takes pains to fill in unknown gaps with "a logical reconstruction of the course of events, in which causes and effects are linked coherently and motivations are explained lucidly" (p. 183). The number of gaps thus filled is quite large, and the individual probabilities of many of his links are quite low. Ra'anan judges his scenario to have a "speculative flavor" (P. 210), however, it is more likely that his reconstruction is quite improbable (cf. Kerr, 1975).

Theory-driven overpredictions. Theories are an excellent aid for making predictions. By specifying "if . . . then . . ." relationships theories predict when certain events can be expected to happen. Predictions can be made faster and more confidently when the analyst has a

theory for the event domain than if he or she does not. The presence of a theory may lead, however, to more extreme predictions and more analyst confidence in predictions than is warranted by the data (Fiske and Kinder, 1978; Taylor and Crocker, 1979). Thinking about an event in terms of a theory seems to lead to overestimation of the probability of the event occurring (Ross, et al., 1977). Taylor and Crocker (1979) suggest that imagining events in terms of a theory evokes temporal sequences between antecedents and consequences and clarifies the links in the causal chain connecting them. This clarified view leads to a higher estimation that the theorized chain of events will actually transpire. Hence, both theories and scenarios may lead to excessive analyst confidence in predictions.

Illusion of validity. Psychologists have repeatedly demonstrated the fallability of subjective prediction, the lack of predictive and forecasting accuracy by experts in a variety of fields, and the greater accuracy of simple statistical prediction models compared to expert predictions. Despite this evidence, experts and laymen persist in believing in the superior accuracy of their forecasts and predictions (Einhorn and Hogarth, 1978). Experts and laymen are overconfident in their predictive judgments, a phenomenon Kahneman and Tversky (1973) labeled "the illusion of validity":

> . . . people are prone to experience much confidence in highly fallible judgments . . . Like other perceptual and judgmental errors, the illusion of validity persists even when its illusory character is recognized.

People are often most confident of their predictions (e.g., when predictor cues are highly redundant or extreme) when, in fact, the data they use generate the least accuracy (Kahneman and Tversky, 1973).

The reasons why people persist in this illusion of validity and are often overconfident in fallible predictions are discussed in greater detail below under "Forming, maintaining, and changing theories."

Surprises and sudden events. Several analysts of forecasting and predicting have reached the same conclusions about the difficulty of accurately predicting distant events, namely that surprises, or unexpected sudden events, are more common and more important than forecasters anticipate. Forecasters and predictors seem to fail to understand the full extent of uncertainty. Hogarth and Makridakis (1979) quote the economist Paul Samuelson on this issue.

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

Ramage, a meteorologist, makes a similar point (1980: 268):

> Despite great increases in observation density, communication speed and computing power, [weather] forecasts are not much better now than a generation ago. The reason seems to lie in unpredictable turbulence "bursts" which contaminate forecasts beyond two or three days.

Ramage notes that the negative impact of such sudden events is not confined to weather but afflicts predictions in other natural and social sciences as well.

Ramage argues that sudden events are characterized by unpredictability and environmentally constrained randomness; i.e , they are events from the extreme tails of frequency distributions, and cannot be adequately incorporated by studies of past events.

Ascher (1978: 210) reaches a similar conclusion about forecasts and their core assumptions. Forecast models can only reflect a finite number of relationships, a limited number of relationship-modifying factors, and must include some invariant relations. This limits the structural flexibility of the model. Ascher notes, however, that:

> There is reason to believe that the turnover in "structures" defined at any level is increasing in this rapidly changing and changeable world. . .The likelihood of unanticipated structural changes in the long run (i.e., ten years or more) is high, so the usefulness of progress in forecast methods relying on any level of fixed structure is minimized for long-term forecasts.

To cope with this high probability of unexpected events, Ascher calls for "specialized surprise-sensitive forecasting." By this he means methods that would be "highly sensitive to potential surprise outcomes, even if they are not the most likely outcomes." The future problems and crises that can result from what is unlikely to happen are severe enough to warrant "lookout institutions," designed to anticipate and illuminate unlikely problem areas before they develop beyond control.

Such institutions are not unknown in military and naval intelligence, and can be found under the titles of "indicators," "warning," "current intelligence," "watch offices" and other labels. The main problem with these efforts is formulating clear patterns to watch for that will indicate events that have never happened, or happen rarely. Since the analysts are highly unlikely to anticipate the "right" pattern exactly, most (or all) warnings will have to be based on a very rough fit between the indicator patterns (or templates) and the pattern of unfolding events. At best, the warning office may be able to issue only uncertain alerts.

Ascher offers several recommendations for surprise-sensitive forecasting which compare favorably with recent recommendations on how intelligence analysts should deal with surprise (e.g., Gazit, 1980). First, deviations from long-term patterns should be examined as possible new trends rather than as merely ephemeral deviations. Since most deviations will be only ephemeral and not new trends, the surprise analysts will register many false alarms and will seem to waste a great deal of time chasing the will-o-the-wisp. This problem may be compensated for by their ability to register new trends, although not all new trends will be important. Being the first to note them may offset their many false alarms.

Second, surprise forecasters should be allowed, wrote Ascher, to make estimates free of plausibility checks. Their projections should not

be required to conform to anyone's intuitive notions of what is plausible. This does not mean that the forecasting methods should not be plausible, only that forecasters and their customers should consider seriously rather improbable or even outlandish projections. If such projections suggest outcomes that would be extremely dangerous, it will be worth while to monitor their possible development.

Third, surprise-sensitive forecasting cannot resort to consensus amalgams of the forecasts of others. Averaging of opinions, or reiterative judgments aimed at arriving at consensus will, by definition, eventually exclude the improbable, implausible, and unexpected events that surprise forecasting is aimed at predicting. The use of Delphi techniques, for example, is fine if the aim is to determine the middle-of-the road estimate, but it cannot possibly be relied upon to detect surprises.

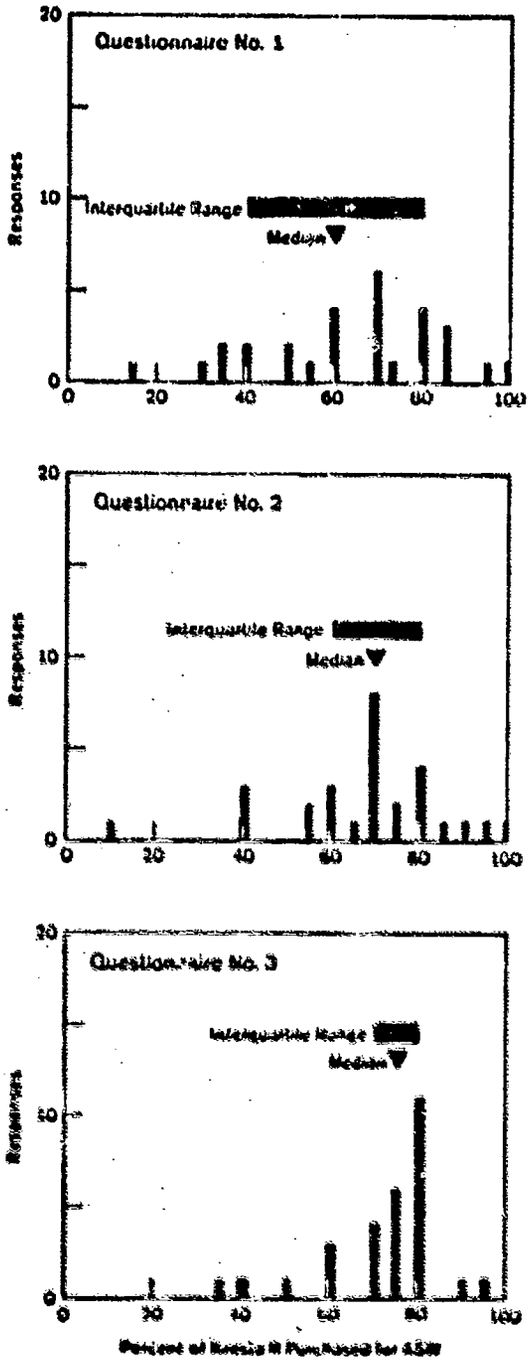
Because Delphi techniques are commonly used in forecasting, it is worth exploring this last point with an example from naval analysis. Thorpe (NPSP, ch. 8) used Delphi questionnaires to determine the consensus of experts on the missions of various Soviet combatants. Figure 4.6 reproduces the responses for one such question: the percent of the Kresta II's mission that is ASW. Note that the median response distinctively shifts to the right and that the overall range and interquartile range shrink dramatically. While some two or three respondents on the first questionnaire thought the percentage might be as low as 18 percent or as high as 100 percent, by the third round, no experts thought the percentage lower than 30 percent or higher than 95 percent. The extreme cases are explicitly eliminated by the Delphi technique. It is just these extreme, improbable cases that are of most importance in surprise forecasting.

It is beyond the scope of this study to completely analyze the problem of surprise prediction. For both forecasting and intelligence this effort is highly uncertain and speculative (cf. Ascher, 1978; Gazit, 1980; Handel, 1980). The problems with surprise prediction and intention estimation go far beyond the cognitive and methodological limitations of analysts. [A review of the literature on the estimation of intentions and the prediction of surprises by intelligence analysts is found in Stech, 1979. A general framework for analyzing surprise is outlined there which is based on sociological studies of disasters and of scientific discoveries.] An important element of surprise and surprise detection is the theory held by the analyst.

Forming, Maintaining, and Changing Theories

The analyst's theories play a fundamental role in the preceding six estimation tasks. Theories imply or specify which data are important and which are not, and lead the analyst to perceive some data and not others. Theories suggest how data should be weighted and conceptualized. Covariations may be theory-driven, i.e., implied by a theoretical relationship. Theories include "if . . . then . . ." statements which relate causes to effects, and imply causal hypotheses. Theories and schemas are thus strong influences on causal reasoning. Finally, theories produce predictions and influence forecasts by shaping the core assumptions used by the estimator. In effect, the process of estimation, which we have divided into seven stages, actually begins and ends with theory, and each stage influences the estimator's theories and, in turn,

Figure 4.6. Responses to Delphi Questionnaires on the ASW Mission of the KRESTA II Cruiser.*



*From Thorpe (NTSP, ch. 8: 160).

is influenced by these theories. An estimate is, in many respects, nothing more than a very specific theory.

In this section we examine evidence that people tend to persist in holding theories and beliefs long after overwhelming evidence discrediting them has been received. This general tendency to conservatively adhere to theories arises from several problems of data processing, integration, and inference. Theories and beliefs persist to such a degree that even after the evidence that produced them is totally discredited, the beliefs remain active. People seem to realize that they need schematic mental constructs like theories and beliefs to organize their mental world and to cope with their environment, but they seem to depend on a given theory long after they should have rejected it for another. People rarely employ more than one theoretical vantage, and are highly inefficient both in confirming and in disproving theories.

Theory conservatism.

> An easily-understood, workable falsehood is more useful than a complex, incomprehensible truth.
>
> Thumb's Second Postulate

A theory is valid and accurate if it makes verifiable predictions that generally turn out true. Adhering to a bad theory implies that accuracy and validity are sacrificed. Theories, however, serve other purposes than making accurate predictions. First, they organize knowledge and streamline investigations by providing useful categories, relationships, and weighting criteria that systematize data. Even though the system produced by the theory is invalid, i.e., unable to accurately predict, it may be better for organizing than the next best alternative, or no system. Second, accuracy is rarely obtained without effort, and an existing inaccurate theory may be more sensible, from a cost-benefit viewpoint, than undertaking the expense of developing a new theory of greater accuracy. Many decisions and estimates based on invalid theories are nonetheless reliably accurate, and even if the theory leads to many errors, these may carry very low costs. A bad theory, in other words, may yield high benefits and low costs, and seem quite good as a result. Revising a theory, or replacing it, is likely to entail considerable effort, with few assurances of a payoff in more beneficial accuracy or less costly errors. Even thinking about a new theory may be costly (see Shugan, 1980). People will rightly settle for a "satisficing" theory (one that yields results that are good enough to satisfy) rather than seeking for an optimal theory.

Scientific theories.

> Science is true. Don't be misled by the facts.
>
> Finagle's Creed

> Progress does not consist in replacing a theory that is wrong with one that is right. It consists in replacing a theory that is wrong with one that is more subtly wrong.
>
> Hawkin's Theory of Progress

Scientists are no exception to the tendency to persevere in believing a theory despite contradictory data. There are good reasons for this scientific conservatism. Data are often poorly generated, falsely interpreted, or otherwise invalid. Scientists rightly view new, inconsistent data with more skepticism than they do old trusted theories (Polanyi, 1958).

Mitroff (1974) interviewed 40 NASA scientists and found a surprising number were highly committed to confirming their own theoretical positions and saw this commitment as necessary and desirable. They argued that without it, many good, new, but undeveloped ideas would die as a result of premature falsification. Those scientists who held such views tended to be those who were rated by peers as especially successful and prominent.

Scientists often believe they should not drop a theory on the basis of a few, stray questionable facts, but only in the face of a large, consistent, interpretable body of data. In effect, a theory is kept until a competitor theory develops to the point that it can replace the first theory. Scientists try to protect themselves from over-reacting to new facts that may turn out to be in error. Francis Crick, discoverer of the structure of DNA, observed (Crick, 1979):

> Misleading data, false ideas, problems of personal interrelationships occur in much if not all scientific work . . . [when the double helix was discovered] I think we realized almost immediately that we had stumbled onto something important . . . I recall going home and telling my wife Odie that we seemed to have made a big discovery. Years later she told me that she hadn't believed a word of it. "You were always coming home and saying things like that," she said, "so naturally I thought nothing of it."

How then does new data ever become accepted as fact? The evaluation of new data in science is a social endeavor in which epistemological purity matters little, and anticipating objections and raising the costs of disagreement count for much. Scientists carry out experiments to "test" new data by broadening and refining the contexts in which the facts are true. Facts are accepted as such because of the specific conditions under which they were created, but in becoming facts they are stripped of reference to context and deemed equally true for all situations (Bearman, 1979). Scientific facts are socially invented by the community efforts of scientists. The seeming discovery of facts is a social creation; it is an illusion that scientific facts are found without the active involvement of social processes that interrelate scientists and scientific thought.

All cognition and perception of a scientific finding occur in a particular thought style which is developed among scientists, what Fleck termed a "thoughtful collective" (1979). Many of the biases of theory change and maintenance that affect laymen seem also to affect scientists because scientists are also dependent on a social basis of knowledge, and everything the scientist needs to know cannot be known scientifically.

Thus, while conservatism protects theories from being rejected by erroneous data, it also leads some scientists to hold theories long after other scientists have accepted data which support a better theory (Kuhn, 1970; Mahoney, 1976; 1977).

These empirical and social findings on the beliefs and practices of scientists contrast with widely accepted views on the appropriate strategies for testing theories and hypotheses. These normative views of scientific inference specify disconfirmation and tests of multiple alternative hypotheses as the major strategies (Platt, 1964; Popper, 1962). Popper's entire philosophy of science is built on the concept of disconfirmation. Platt has proposed that scientists construct experiments and tests to disconfirm successive generations of alternative hypotheses. He labels this method "strong inference." Not only do laymen not follow this normative advice on theory testing, neither do many scientists.

Theories and evidence. In the section above on "assessing covariations" we noted how people overestimate the strength of theory-driven correlations, to the degree that correlation is perceived where none exists. Theoretical illusory covariations were almost impervious to contrary evidence, e.g., data that were strongly and negatively correlated only slightly reduced beliefs that the data covaried positively. Psychologists have found that covariation theories are not alone in being relatively immune to contradictory evidence. Studies of theory formation and changes support four conclusions (Nisbett and Ross, 1980; Ross and Anderson, 1980; Ross and Lepper, 1980): (1) Preexisting theories, when exposed to probative evidence (whether supporting or opposing the theory), tend to be held correct to a greater degree than the evidence warrants. (2) When people form a theory on the basis of new evidence, the new theory resists disconfirmation by subsequent evidence. (3) If evidence that forms the basis for a theory is demonstrated to be false, people nevertheless will continue to believe the theory is still true. (4) People are able to find or create evidence to support any theory they believe.

Maintaining existing theories. People with a pre-existing theory or belief treat new evidence in a highly biased manner. Evidence which confirms their theory is accepted, the method that produced the evidence is deemed sound, and belief in the theory is reinforced by the evidence. Evidence which contradicts the theory is not accepted, the methodology which generated the evidence is strongly challenged, and despite the negative evidence, belief in the theory is increased (Lord, Ross, and Lepper, 1979). In other words, belief in a theory may be greater after reading negative evidence than after reading no evidence. Different standards are applied to research methods when they yield evidence that opposes an accepted theory than when their evidence confirms the theory. Lord, et al., (1979) also found that mixed evidence, which supports opposing theories equally, does not reduce confidence in theories but instead reinforces the confidence of those holding either view. Experimental results in agreement with the accepted theory were rated "more convincing" and the studies "better conducted" than results and studies which contradicted the theory, regardless of the methods of the studies.

One of the best instances of mixed evidence tending to confirm

differing theories can be found in the interpretations by naval analysts of Admiral Gorshkov's articles and book (see chapter 3, case 4). Gorshkov's many opinions can and have been taken to support a wide variety of views on the Soviet Navy's future. Rather than weakening an analyst's beliefs in his theory of Soviet Navy peacetime developments, the Gorshkov papers seem only to strengthen them. Analysts seem quite adept at interpreting Gorshkov as opposing the theories the analyst opposes and supporting the theories the analyst supports.

False consensus bias. People often judge the need for explanation by reference to themselves; if someone else acts much as we do, we see little need for a special theory to explain their behavior. In general, people are prone to assume that their own behavior or beliefs are quite common and needing little special explanation. Furthermore, when others act or believe as we do, there is no urgent need to explain their behavior. When other's behaviors or beliefs are different from our own, however, there is a strong motivation to consider various theories to explain the differences (Ross, Greene, and House, 1977). Since people overestimate the commonness of their own beliefs and behaviors, they tend to be surprised by the different behaviors and beliefs of others, and view these as more significant and uncommon than they are. People also underestimate the significance of beliefs or behaviors similar to their own, tend to overestimate how common they are, and assume they occurred for the same reasons.

Moore, Flanigan, and Helsel (SNI: ch. 7) suggest that in the period 1956-1966, naval analysts misinterpreted the Soviets' motivations for constructing missile submarines and misread Soviet missile submarine operations as being similar to Western motives and operations. Since the Soviet and Western submarines seemed similar, analysts assumed that they were intended for similar missions. Moore, et al., argue that the Soviet submarines had quite a different mission of anti-ship attack, and did not share the strategic deterrent role of Western missile submarines. They wrote of the episode (p. 160):

> Apparent mismatches between technical parameters and operational employment, as well as seemingly inexplicable Soviet pronouncements, often cause Western analysts little concern, as long as there is possible a "mirror image" or "the-way-we-do-it" interpretation of the data.

Theory formation. The fact that theories are so useful in organizing and understanding streams of new data seems to lead people to apply theories prematurely. Since theories are relatively insensitive to disconfirming evidence and highly sensitive to confirming evidence, theories formed early may gather strength too quickly and become insufficiently sensitive to subsequent negative data. People become prematurely committed to a theory and unable to revise it appropriately as further information is obtained.

People test a hypothesis from a new theory by trying to confirm it, i.e., looking for evidence consistent with the hypothesis and the theory, rather than by looking for evidence that would refute the hypothesis. This strategy allows people to determine which hypotheses work, but not which are valid. By overlooking negative information they

are unable to efficiently reject hypotheses and theories. There are ecological reasons why people have this tendency (which are related below), but this does not lessen the fact that confirmation is a weak testing strategy.

An experiment by Wason (1960) demonstrated this tendency. Wason asked people to guess the rule generating a series of numbers. The first numerical example given was "2 . . . 4 . . . 6." People guessing the rule could generate any series of numbers they liked and could ask if the series fit the rule or not. Most people generated examples consistent with the most common guess, that the rule was "ascending consecutive even numbers." Few people generated inconsistent or disconfirming examples, and few guessed the actual rule: "any ascending series." People tended to confirm the rule they thought to be correct, and did not attempt to refute incorrect hypotheses.

A series of studies by Mynatt, Doherty and Tweney (1977, 1978; Doherty, Mynatt, Tweney, and Schiavo, 1979; Tweney, Doherty, Worner, Pilske, and Mynatt, 1980) have revealed (a) a strong tendency to use confirming strategies and not use disconfirmation, (b) an ability to use explicit falsifying evidence when available to correctly reject wrong hypotheses in simple problems, (c) difficulty in using disconfirmation or multiple hypothesis testing strategies in complex problems, and (d) tendencies to seek information of no diagnostic value. In analyzing a simple "artifical universe" people consistently tended to confirm hypotheses, but if explicit disconfirmatory evidence was obtained it was used to reject false hypotheses. In assessing a more complex "artificial universe" people had great difficulty employing a disconfirmation strategy or disconfirmation evidence. Partially correct hypotheses were often completely rejected because of disconfirming data. A mixed strategy of confirmation, disconfirmation, replication, and experimental variation seemed to yield the best results. The testing of multiple alternative hypotheses seemed to impose a debilitating cognitive strain while the strategy of considering two related hypotheses at a time led to much improved performance.

Tweney, et al., (1980) note that disconfirmation is likely to be an efficient strategy only after several "good" working hypotheses have been determined by means of confirmation. Disconfirmation also requires that the disconfirmed case be analyzed as fully as possible. People tend to treat disconfirmations as negative information, i.e., they largely ignore it. Similarly, in testing multiple hypotheses, people may use information on the one "real" hypothesis, and ignore information on the other "dummy" hypotheses. The strategy of multiple hypothesis testing and disconfirmation may be epistemologically efficient, but cognitively very difficult to perform.

In a series of studies on the ability of people to test hypotheses about others, Snyder and colleagues found (Snyder, 1978; Snyder and Cantor, 1978; Snyder and Gangestad, 1979; Snyder and Swann, 1978a, b) that people systematically formulate and carry out confirmatory strategies and gather evidence that tends to support the hypothesis under examination. Hypothesis-disconfirming strategies are not formed, and people make little use of disconfirming evidence These tendencies led to a readiness to accept whatever hypothesis was under test.

There seem to be several cognitive processes that combine to

produce this confirmation bias in theory testing. People tend to recognize the relevance of confirming evidence more readily than that of disconfirming evidence, and therefore can more readily search for and detect confirming data (Wason and Johnson-Laird, 1965). In searching memory for relevant evidence, people are more likely to recall confirming than disconfirming evidence (Snyder and Cantor, 1979).

These tendencies are probably closely related to the tendency to overemphasize the frequency of "positive hits" in the four-fold, present-absent table in assessing covarations (see above). The entries in the ++ cell are much more likely to influence such assessments than the entries in the other cells. The ++ entries are, of course, confirmations.

Given the strength of the confirmation bias it is not surprising to find that in our sample of estimates by naval analysts only two essays made use of an explicit multiple hypothesis and disconfirmation approach. Many papers comment briefly on alternative hypotheses and then quickly discard them, usually on the basis of one or two items of contrary evidence. The hypothesis that is accepted may also face contrary evidence, but this is usually "explained," rationalized, or otherwise reinterpreted by the analyst so that it does little or no damage to the hypothesis. In contrast, Kelly (SNP, ch. 15) and Blechman and Levinson (SNI, ch. 22) explicitly test a body of evidence against a variety of competing hypotheses. A paper by Hardt (SNP, ch. 8) takes a traditional but more explicit approach, i.e., one hypothesis is considered but the evidence pro and con is described and evaluated. In contrast, Kelly lists four hypotheses to explain a particular Soviet action and lists the evidence for and against each one. This tends to underline the degree to which one piece of evidence may confirm several hypotheses and also disconfirm several others. Blechman and Levinson hypothesize six different reasons for a Soviet action and organize evidence for each, rejecting most as the primary explanation for the Soviet action, and leaving one as a strong candidate, and two as possible, partial motivations.

This is not to say that other naval analysts do not consider multiple hypotheses or assess hypotheses equally against the accumulated positive and negative evidence. To some degree, all analysts do some multiple hypothesis disconfirmation. What we are pointing out, however, are the strong tendencies not to do this and not to do it explicitly. It is far too easy to disconfirm hypotheses one does not believe in, and to confirm hypotheses one already accepts, especially if one relies on informal or ad hoc confirmation-disconfirmation procedures. Analysts would be far better off explicitly stating the various possible hypotheses and listing the pro and con arguments, as Kelly does, than in leaving such comparisons largely implicit and loosely related, as is more typically done. Analysts would also be better able to specify the degree of support a set of data give a hypothesis, and the degree to which several hypotheses remain viable given some data set, by clearly specifying the hypothesis-data relationships. This would also allow the analyst to better assess the level of confidence he or she should have in any one hypothesis. A failure to explicitly list all points of evidence inconsistent with hypotheses seems to be one primary cause of overconfidence in judgments (see below).

Perseverance of discredited theories. Quite often theories are formed on the basis of data which is later discredited. That is, later evidence is obtained which clearly demonstrates that the earlier data were incorrect. In this case, the theory based on the discredited data should be significantly modified, if not abandoned altogether. Psychologists find, however, that such modification is typically very slight, and such a theory survives discrediting evidence largely intact.

Ross and Lepper have investigated what happens to beliefs when their entire evidenciary basis is totally discredited (Ross, Lepper, and Hubbard, 1975; Lepper, Ross, and Lau, 1979; Anderson, Lepper, and Ross, 1979). They find that beliefs and theories persist after the basis for the beliefs is shown to be valueless. For example, Anderson, Lepper, and Ross (1979) requested people to explain either a relationship between a prognostic variable (risk preference on a paper and pencil test) and success or failure as a firefighter. One group was given a case study that suggested a positive relationship between risk preference and success and a second group received a case study suggesting a negative relation. After the people gave their explanations of the relation, they learned that the case study was totally fictitious and that others had been exposed to the opposite relation. Despite the discrediting of the basis of their explanations, peoples' beliefs in the accuracy of their explanations remained high, and they insisted that, notwithstanding the discrediting, it was obvious that the particular relationship they had explained was the correct one. Several expressed surprise that people in the other group had been so gullible as to accept the opposite relations as possible.

It is quite difficult to assess the degree to which the theories of naval analysts are resistant to the discrediting of earlier evidence. Information on the Soviet Navy is constantly being modified, weapons systems thought initially to be anti-ship later come to be recognized as anti-submarine, estimates of weapons' ranges expand and contract as additional data are obtained, estimates of Soviet tactics and doctrines are modified as Soviet exercises and deployments are studied, patterns of Soviet deployments and construction reveal unexpected plans or motives. Whether these changes in the data base produce appropriate changes in analysts' estimates is impossible to adequately address in a general study such as this. Analysts almost never describe exactly how their estimates change when they learn that a previously accepted "fact" was actually in error.

Analysts do, however, update their estimates. For example, Hudson (SND: 388) concluded (in a 1973 essay) that Soviet naval doctrine "even approximated U.S. doctrine" and the Soviet fleet "would be developed along the same lines as the American." A few years later Hudson (1976: 90) explicitly reversed his earlier estimates, concluding that Soviet doctrine "reflects a distinct view of naval utilization, differing in important respects from current Western thinking." To a degree Hudson argues that Soviet thought on these issues had shifted, but he also seems to suggest that his own thinking had changed, but it is not clear exactly why.

Perhaps the best example in our sample of explicit updating of estimates is McGwire's (SNS, ch. 25) essay on the evolution of Soviet naval policy: 1960-74. McGwire expands his estimate, that Soviet forward deployment was due to strategic defensive needs, to include evidence that the Soviet Navy was playing a greater role in inhibiting the

U.S. Navy power projection in peacetime (cf. p. 520). McCGwire makes clear that his estimate had changed in response to new evidence and a better understanding of ongoing trends (p.505). He also attempts to specify the degree to which his previous estimate remained intact, putting the two hypotheses of strategic defense and forward peacetime power projection into a relative perspective. It is also clear that other naval analysts would not agree that McCGwire has sufficiently adjusted his estimates in light of these new data (see chapter 3, cases 5, 8, and 9). Of course, in responding to his critics' charges that he had not modified his old theories enough, given the new evidence, McCGwire could respond that his critics have gone too far in accepting early returns as supporting their new theories, and have failed to consider adequately that part of the new evidence that also fits the older theories.

Overconfidence. Accumulating research on judgment, decision-making, and probability estimation shows a substantial lack of ability of both experts and nonexperts. However, people have great confidence in their fallible judgment. People believe their theories, judgments, and estimates are correct far more than they actually are; people are more confident than they are accurate (Lichtenstein, Fischhoff, and Phillips, 1976; Fischhoff, Slovic and Lichtenstein, 1977). Typically, for example, when people state their subjective probability of being correct is .70, the relative frequency with which they are correct is only 60 percent; when they are 90 percent certain, they are only right 75 percent of the time, and so on.

Fischhoff, et al., (1977) suggest several reasons why confidence exceeds accuracy. As people estimate an answer, they are forced to reason from the known to the unknown. As we have seen repeatedly in this chapter, people are insufficiently critical of their inference processes. They may fail to adjust their confidence levels to reflect weak assumptions, or problems with their own reasoning processes. People use inferential strategies almost automatically and unconsciously and give little thought to the logical steps required. We have noted how people perform many of these steps improperly or incompletely, while assuming they have done them correctly. Since the validity of the inference process is assumed to be virtually perfect, the product of the process is assumed to be correct.

People also believe their memories are exact copies (although faded or incomplete) rather than reconstructions of experience. Rarely are the biases of memory storage, reconstruction, and retrieval adequately considered by the individual when discounting confidence. The weakness of such memorial processes as eyewitness testimony are well-documented (e.g., Buckhout, 1974; Loftus, 1974) but eyewitnesses themselves rarely question their capabilities to report what they saw.

People may begin solving an estimation problem by referring to their theoretical knowledge and adjusting their estimates from a theoretical baseline by an amount that reflects any new information. Tversky and Kahneman (1974) have found that such adjustment tends to be too conservative; people fail to adjust their estimates from the anchoring point of theory sufficiently to reflect new data. The range of apparent possible answers is thus likely to be narrower than it actually is, and people may be overconfident that their theory-anchored inference process considered an adequately wide band of possibilities.

Complex inferential problems may be broken up and processed serially, a piece at a time. In doing so, people may ignore the uncertainty inherent in their solutions for early parts of the problem. This will reduce the cognitive strain in dealing with later parts. Rather than setting their confidence limits in accordance with the least accurate or certain part, people may average confidence levels for the various parts, or adopt the highest confidence level.

Lichtenstein and Fischhoff (1977) examined the confidence of experts and nonexperts in terms of calibration and resolution. The perfectly calibrated estimator assigns confidence probabilities that accurately reflect the proportions of correct answers. For example, when the perfectly calibrated estimator gives a confidence rating of .75, three out of four times the estimator is correct. Resolution reflects the ability of the estimator to discriminate those items he or she is more likely to correctly answer from those he is less likely to get correct. Experts were less overconfident than nonexperts, but the best experts showed some underconfidence; i.e., experts were better calibrated than nonexperts but not perfectly calibrated. Experts were not better at resolution than nonexperts, and seemed largely insensitive to how much they did or did not know. As Lichtenstein and Fischhoff noted, the reason experts were better calibrated than nonexperts, despite being no better at resolution, is simply that experts got more items correct, and their overall accuracy more closely matched their overall confidence level. Lichtenstein and Fischhoff (1980) found some improvement in calibration with feedback training on related tasks, but very little evidence of generalization to different tasks.

Another promising approach to improving calibration of confidence judgments is sensitivity testing. Koriat, Lichtenstein, and Fischhoff (1980) required people to list the reasons for and against each of the responses they chose. The people then estimated the probability they were correct. This procedure forces people to consider reasons why their answers might be incorrect. This sensitivity test produced marked improvements in the calibration of confidence judgments. In a subsequent study Koriat, et al., found that only the listing of contradicting reasons (i.e., reasons why an answer might be wrong) improved the calibration of confidence. These results suggest that confidence judgments tend to reflect the amount and strength of evidence in favor of the selected answer, and tend to neglect contradictory and disconfirming evidence. In view of the difficulty people have in perceiving and retrieving disconfirming evidence when testing hypotheses (see above), it is not surprising that disconfirming evidence has little impact on confidence judgments unless people are specifically required to consider it.

Einhorn and Hogarth (1978), in their review of overconfidence and the "illusion of validity" of predictions (see above), also conclude that the difficulty people have in using disconfirming information to test hypotheses prevents people from obtaining the positive and negative feedback from experience that would reduce overconfidence. They note two other factors that increase the difficulty of learning from experience to make more accurate judgments: lack of awareness of environmental effects on outcomes (e.g., regression effects, base-rates), and the use of unaided memory for coding, storing, and retrieving outcome information.

Data-driven and theory-driven overconfidence. People tend to be

overconfident in the accuracy of their answers when answering knowledge questions, estimating numerical values, or judging probabilities. Dawes (1980) has found that people are not overconfident, however, when making perceptual judgments but were when making intellectual judgments. He found that psychologists' recollections of the eye colors of colleagues did not reflect any appreciable overconfidence, while the answers of the same psychologists on a test of psychology and general knowledge items reflected considerable overconfidence, i.e., average assessed confidence was .80 while the percent correct was .70 for the knowledge items, but .67 and .75 respectively for the eye color perceptions. Perceptual tasks produced the least average confidence and the greatest average accuracy, while the intellectual tasks produced the highest average confidence and the lowest average accuracy. Dawes' study is a replication of the Lichtenstein and Fischhoff (1977) finding that resolution is generally poor, and people do not distinguish those types of items for which they are more accurate (e.g., perceptual tasks) from those on which they are less accurate (intellectual judgments).

Dawes' results also seem consistent with Einhorn and Hogarth's (1978) analysis of conditions that lead to overconfidence. That is, people probably have greater experience with perceptual illusions and evidence that disconfirms their own perceptions than they have with cognitive inferential illusions. People are probably more accustomed to checking the validity of their perceptions (e.g., by changing the vantage point, using a different sense modality, increasing the base-rate by obtaining a better look or a second opinion, etc.) than they are at checking the validity of their logical inferences. Secondly, people are probably much more familiar with the effects of environmental factors on perceptions than they are with the effects of environmental factors on inferences and predictions. People readily understand how their perceptions may be distorted by environmental factors (e.g., haze, ophthalmic weaknesses) but seem to have no appreciation of the effects of such environmental factors as regression, base-rates, treatment and placebo effects, and probabilistic independence and dependence on intellectual judgments. Finally, it seems likely that human memory is at least as adequate for coding, storing and retrieving data on perceptions (e.g., "when I recognize a face is it usually someone I have met before?"), as it is for information on intellectual judgment outcomes (e.g., "when I recognize a face do I usually associate the correct name with it?").

Learning from experience. The review by Einhorn and Hogarth (1978) implies people will tend not to learn from experience that their theories and hypotheses are wrong. People will thus continue to have confidence in weak or fallacious theories because of the difficulty they have in developing evidence from experience to test their theories. Einhorn and Hogarth (1978) and Brehmer (1980) suggest that the experiences people have with their theories facilitate the operation of biases which prevent learning: the tendency to use confirmatory evidence, disregard of negative information, and assumptions about causality. Brehmer argues that people tend to confirm hypotheses, for example, because it is rarely clear what dimensions of a problem are important, nor what boundaries distinguish one concept from another (i.e., hypotheses are "fuzzy set" concepts). Learning that a concept fits the data may be the only useful

increment in information the environment allows, while an attempt to disprove a hypothesis may fail because of the difficulty in specifying exactly what the hypothesis allows or does not allow. Negative information is far more useful in laboratory experiments, where the boundaries and dimensions of concepts and hypotheses are clear-cut. In the "real world" of fuzzy concepts, being told the Soviets do not practice sea control may not be as useful as being told they do practice sea denial. Learning that a Soviet action is not representative of sea control will be less interesting than the hypothesis that an action "fits" the sea denial concept.

This does not mean that naval analysts should avoid using negative evidence or disconfirming strategies for testing theories. It does mean that real world hypotheses must be carefully conceptualized for such methods to be useful. There are means by which fuzzy concepts can be distinguished more clearly (chapter 5). Given the difficulty of drawing distinctions and the absence of disconfirming strategies analysts should be aware of the inherent weakness of confirmatory strategies, and attempt to supplement them when possible.

Brehmer also notes that people must supply the hypothetical realtionships between concepts and that hypotheses generation tendencies are often inappropriate for judgment tasks. That is, people assume that the rules relating concepts follow the pattern:

> . . . there is a rule, rather than that there is no rule, that this rule is deterministic, rather than probabilistic, that the values to be predicted from the cue values do in fact depend on the cue values, . . . that the rule is functional .. that the rule is a positive linear function. (Brehmer, 1980: 231).

This pattern fits peoples' experiences with psychophysical stimuli; sensation usually is a positive function of the stimuli (in fact, a log function, e.g., Fechner's law). It is not a pattern that is broadly applicable to intellectual judgment tasks. This pattern is also consistent with a highly causal view of experience, which may be more applicable to psychophysics than to probabilistic prediction tasks. The deterministic pattern may be overlearned from extensive psychophysical experience, while real world experiences tend not to allow inferential strategies for probabilistic judgments to be efficiently learned, and they later are thus underlearned. Peoples' accurate confidence in their psychophysical strategies seems to be inappropriately extended to their inferential strategies. In summary, peoples' faith in what they have learned from experience about their theories may be misplaced, and people will erroneously conclude that because a theory fits some data the theory is correct.

Not all analysts view intention estimation as a matter of judgment and information integration, but rather as a form of historical narration. The last section of this chapter summarizes some problems with narration logic.

NARRATIVE LOGIC

> The only thing one learns from history is that nobody ever learns anything from history.
>
> **Hegal**

The chief practical use of history is to deliver us from plausible historical analogies.

James Bryce

In an earlier chapter we used Scheibe's (1979) distinction between sagacity and acumen to separate two different approaches to intention estimation. Sagacity refers to prediction based on the detection and exploitation of patterns and prior probabilities in past experience. These regularities are utilized to infer the likelihood of future behavior. In contrast acumen is based on empathy with the actor whose behavior one is attempting to predict. Empathy is an ability to take the other's role, sensing the other's perceptions and impressions, and experiencing the other's thought process. Whether acumen is a skill (i.e., something that can be taught), intuition (something whose source is ineffable), or experience, is not clear (Minear, 1980, offers an interesting perspective on these issues, see also Ornstein, 1977).

Pipes, an historian, (in Godson, 1980: 180) has described how acumen applies to intelligence prediction and estimation:

> . . . fundamentally, when you are dealing with political analysis, there is no substitute for methods which have come from scholarship . . . a deep knowledge of the country, of its general culture and its political culture in particular, from what the Germans call Fingerspitzengefuhl, the feeling "on the tips of your fingers" for a given culture where you know that some things are more probable and others less so.

Pipes recommends (p. 175) that the political analyst "be given very thorough general training in the historical method." The problems of political intelligence, he wrote (p. 174) "are rational only to a certain predictable point and are often emotional, psychological, cultural, and therefore of a kind that no amount of science can ever fix or predict."

Sarbin, a psychologist, agrees (1980) with Pipes that the phenomenon of greatest interest to intelligence analysts, future intentions, are inherently unpredictable from a scientific viewpoint (i.e., sagacity) because they are fundamentally unique, but may be predictable through acumen. Sarbin sees the intelligence analyst's task as unfolding in an environment of contextualism rather than science's environment of causality. He wrote (p. 116)

> The root metaphor of contextualism is the historic event in all its complexities . . . its home is . . . history and literature.

The task of the intelligence analyst, according to Sarbin (p. 117), is to organize a chaotic mixture of events into a comprehensible plot, following a story line, so that, like a literary critic, the analyst can "fathom the intentions of the author [of the events], to 'understand' to decipher the

meaning of the creative work " The product of this intelligence process is a narrative that meets tests of coherence. Sarbin does not reject prediction based on sagacity, he just argues that it is not enough, that the prediction of an adversary's future behavior requires acumen because that behavior is often unique, and tied to no past patterns. He wrote (p. 137)

> To predict the unique case, the known actions of a target person or organization serves as the framework upon which a plot is constructed . . . acumen follows from locating the noticed events in a plot . . .

Historical Perspectives

> Historians begin by looking backward. They often end by thinking backward.
>
> Nietzsche

Pipes and Sarbin would have their history in just one way, a flow of unique and unpredictable events, and it is true that some historians treat evidence from this perspective. Other historians do not:

> The rationalist sees evidence as exemplifying deep uniformities; the intuitionist sees evidence as testimony to uniqueness (Kuzminski, 1973).

In effect, whether an historian elects to be an intuitionist rather than a rationalist may depend on the degree to which he or she views most important processes in history as being unpredictable sudden events rather than as more orderly processes which are occasionally punctuated by sudden events.

This split between a noncausal view of existence and a causal outlook is one of the oldest issues in philosophy, and need concern us no further. Because some historians and intelligence analysts adhere to the intuitionist view, and practice narration rather than causal prediction the purpose of this section is to quickly summarize some of the problems with narrative logic. Not being an historian I rely heavily on the analysis of historians. It is clear, however, that taking the intuitionist position on prediction does nothing to eliminate the possibility of cognitive biases weakening the process of estimation. The analyst who attempts to predict with acumen, rather than sagacity, must still make sense out of his or her history. In doing so biases and fallacies are likely to occur. Many are the same as those affecting the sagacious analyst.

Fallacies of history.

> Even God cannot change the past.
>
> Agathon, Nicomachean Ethics

It has been said that although God cannot alter the past, historians can; it is perhaps because they can be useful to Him in this respect that He tolerates their existence.

Samuel Butler, Erewhon

Fischer (1970) lists and describes over one hundred fallacies commited by historians. Of these two groups are especially relevant for intelligence estimation, fallacies of inquiry and fallacies of explanation (see Table 4.5; for a list of Fischer's fallacies, see Appendix). Some of Fischer's fallacies are refinements of each other (e.g., 39 and 75, the numbers refer to the fallacies listed in the Appendix), while others are directly related to cognitive biases of the types described above. For example, the fallacy of identity (61) is closely related to the representativeness heuristic. The availability heuristic is similar to the fallacy of prodigious facts (24). In general, Fischer's category of inquiry relates to the categories used above on perceiving, weighing, and categorizing data; while his category of explanation relates to assessing covariations, assessing causes and effects, prediction, and theory. Several of Fischer's fallacies appear in naval analysis.

Inquiry.

History doesn't repeat itself -- historians merely repeat each other.

First rule of history.

Fischer suggests that facts may be falsely verified by pseudo-proofs; falsely precise and specific statements of no real evidentiary value (fallacy 11). This is occasionally a problem for analysts of Soviet naval doctrine. Any given statement by a Soviet official is questionable evidence that a particular doctrine does or does not exist -- the Soviets are rarely so specific or clear, especially regarding changes in doctrine. Further, a given statement at one time by a Soviet official cannot validly be used as evidence on Soviet views at any earlier time, although doctrinal analysts persist in making just this inference (e.g., Hudson, SND). The need for Soviet spokesmen to demonstrate the prescience of Marxist-Leninism obliges them to hindsightfully revise their earlier views (occasionally several times), and the Soviets are never shy about rewriting history to suit their needs. As we have noted several times earlier, Admiral Gorshkov's statements in 1967 on what the Soviet Union was planning for its Navy in 1954 are largely specious, although they are sometimes accepted at face value by analysts. Similarly, Soviet naval officers' retrospective claims that they opposed the party's views on nonnuclear wars, or surface ship construction, or balance (cf. Hudson, SND: 280) should not be taken at face value as reflecting what actually transpired. Since the Soviet authors and the naval analysts are resorting to the same source for guidance in these matters, i.e., hindsight, it is not surprising that analysts accept the Soviet claims that this is how things were back then. Indeed, given outcomes, things might have been as the Soviets say, but their latter day statements are hardly the best evidence.

Table 4.5. Types of Historical Fallacies*

I. Inquiry

- A. Question - Framing
- B. Factual Verification
- C. Factual Significance

II. Explanation

- A. Generalization
- B. Narration
- C. Causality
- D. Motivation
- E. Composition
- F. False Analogy

*From Fischer (1970). For a list of fallacies, see Appendix.

Several of the cognitive and conceptual biases of analysts contribute to the problems Fischer noted regarding factual significance. Analysts often seem to select "representative" facts in the form of case studies to demonstrate their point. Having abstract information about the Soviet Navy's behavior in the representative case, the analyst implies certain things about future Soviet behaviors. To make these inferences more precise the analyst may use several cases and organize them into a classification or categorization scheme (e.g., Dismukes and McConnell's 1979 "rules of the game"). The trick in this is knowing which cases are most representative for prediction (i.e., representative of causes), and not to select cases solely because they are representative of effects. The latter cases are useful for post hoc explanation, but may offer little information for predictive purposes.

On the other hand, the analyst may be an intuitionist rather than a rationalist, and select cases that are representative of the essences (the 'inner core of reality') of Soviet naval behavior (fallacy 23). This narrative approach amounts to a refinement of the analyst's theory or schema of what those essences actually are. As we noted above, people are quite adept at finding evidence or altering evidence to confirm their theories, and largely incapable of disproving their own theories without explicit assistance.

Some naval analysts advocate seeking an explanation for every Soviet naval behavior, which implies that every Soviet naval action has some explanatory or predictive significance (fallacy 63), an unlikely situation. A related tendency is to strive to explain or predict what is strange or unusual about the Soviet Navy (compared to Western navies), but to leave similarities unexplained. This assumes that the Soviet Navy does things that Western navies do for similar reasons (fallacies 71 and 79 and the egocentric attribution bias) and that unique Soviet naval behavior has some special significance (fallacy 24).

For example, naval analysts spend much effort assessing why the Soviets have consistently built more submarines than anyone else. Few of these assessments make much (if anything) of these prosaic facts (1) submarines are small, shallow draft, and the only capital ships that can be built at inland factories and moved in pieces along Soviet rivers and railroads, (2) they have very small crews, (3) they were often used as surface vessels (e.g., for shore bombardment) in World War II, (4) of all ships they are the most tactically compatible with a highly centralized system of command and control. The facts typically assessed as important for explaining Soviet submarines have to do with naval traditions, strategies, economics, technologies, philosophies and doctrines, and naval effectiveness (cf. Moore, Flanigan, and Helsel, SNI, ch. 7).

Explanation. Many of the fallacies of generalization noted by Fischer are directly related to cognitive biases. For example, insufficient sampling (32) is the same problem as the "law of small numbers" pointed out by Tversky and Kahneman (1971): a belief that valid generalizations can be made from extremely small samples, or even from a single case (fallacies 33: generalization from "the lonely fact," and 43: generalization from "the overwhelming exception").

Several naval analysts seem to be willing to generalize from a "lonely fact" or an "overwhelming exception." For example, Petersen and Durch (in Dismukes and McConnell, 1979: 144-152) assess Soviet naval

involvement in the Angolan crisis as reflecting "the state-of-the-art" in Soviet coercive diplomacy. They view (p. 148) Soviet deployments as attempts to "signal in advance [Moscow's] opposition to any U.S. naval deployments toward the Angolan theater:

> The evidence suggests a concerted Soviet effort to discourage [U.S.] deployments, actively, via the surface combatants deployed at Gibraltar and, perhaps, the Juliett [submarine], and passively, by curtailing the operations . . . nearest Angola . . . To the extent that these Soviet actions were taken in anticipation of possible U.S. moves, they were unprecedented in Soviet naval diplomacy in the Third World (p. 150).

In evaluating the evidence for this lonely fact and overwhelming exception, Petersen and Durch repeatedly qualify or hedge their judgments "the evidence suggests . . ., the possibility . . ., to the extent that . . ." In generalizing, however, these possibilities become facts about Soviet "state-of-the-art":

> In signaling its opposition to possible U.S. deployments, however, the USSR maneuvered its navy with considerable restraint.

In his analysis of this incident, McConnell concludes (in Dismukes and McConnell, 1979: 257-259) that the U.S. Navy

> . . . appears to have been the target for a Soviet attentional show of force . . . [the Soviets] apparently made a naval show of force, either to demonstrate an interest in protecting USSR-Angolan lines of communication or in countering U.S. naval threats against Cuban involvement on the mainland. Angola was the boldest initiative yet taken by the USSR . . . Surely there is a trend in this direction. . .

In their conclusion, Dismukes and McConnell (1979: 289) give this episode the status of a full-fledged fact:

> In the Angolan crisis in early 1976 [the USSR] went even further [than merely anticipating what the U.S. Navy would do], deploying surface warfare forces into the central Atlantic in anticipation of a U.S. naval counter-deployment that never materialized . . . in a theater where one can with confidence identify [Soviet] motives as predominately (if not exclusively) politico-military in nature.

It is worth noting that the Soviet units which were known to be in the South Atlantic to counter the possible deployment of a U.S. carrier task group were a newer ASW cruiser, a SAM destroyer, and an older (J-class) cruise missile submarine (cf. Dismukes and McConnell, 1979: 144-157), not the typical Soviet anti-carrier warfare (ACW) task force (the Soviets typically remove their newer ASW ships when an ACW task force is

assembled). The only meaningful activity noted in anticipation of a U.S. deployment was the sailing of the ASW cruiser in the direction of the crisis.

Herzog (NPSP: 41) goes even further in generalizing from this incident [note how the two Soviet surface combatants have become "its fleet"]:

> The Soviet Union finished its apprenticeship in the use of naval power to advance what Admiral Gorshkov calls "state interests" when the presence of its fleet off Angola helped to ensure the success of the Soviet backed faction . . With the success of the Soviet-backed forces, and the lack of naval action on the part of the United States, a complete reversal of roles was witnessed . . . by 1976 the United States had given up its place of naval dominance and had assumed the role of spectator . . . The Soviet Navy . . . [is] able to bring naval power to bear in most any part of the globe.

The main point is that a clearly atypical and anomolous episode is generalized into a trend for the future of Soviet and U.S. naval influence.

An instance of the fallacy of generalizing from false extrapolations (39) was reviewed above (see "Prediction - Short-term Forecasts" and Figure 4.6). The problem of false interpolations (fallacy 40) seems to occur when naval analysts attempt to demonstrate the accuracy of Admiral Gorshkov's 1967 speech on what the Soviets planned for their navy in 1954. The only way analysts seem able to get from one end point to the other is to ignore much of what happened in between (see Chapter 3, Cases 1 and 5).

Fischer's (1970: 125) description of the "double-reversing generalization" (fallacy 42) could be applied to most estimates and naval analysts will recognize their own (or at least each other's) hedging as sometimes approaching Fischer's extreme limit:

> . . . a maze of mutual qualifications or a cunning balance of casuistic contradictions, or a trackless wilderness of pettifogging detail, or a slippery ooze of substantive (as well as semantical) shilly-shally.

In Chapter 2 we described the general problem of lack of specificity in predictions and estimates.

Similarly, analysts will recognize their estimates as often succumbing to the temptation to make insidious generalizations (fallacy 41); i.e., sneaking in generalizations without any specific numbers via the loose use of terms like 'few, many, some, normal, common, often.' A common tactic is to deny any intention to make specific predictions, or to estimate intentions, and then to do so with insidious generalizations. For example, Chapter 15 of Securing the Seas begins (p. 411):

> We . . . claim no special wisdom that allows us to foretell the future intentions of the Soviets or even to grasp fully the motives implicit in some of their current activities.

It then relates a long list of predictions and explanations of future Soviet naval behavior, e.g. (insidious qualifiers underlined):

> The Soviets seem to be on a steady course, with ample resources allocated to maintain at least a modern 775-ship navy . . . (p. 415).
>
> . . . the emergence of a more confident and expansive Kremlin leadership, suggest(s) that Soviet naval elements could now be used in more offensive roles . . . Soviet naval forces can now operate in some force across all the world's oceans, under a variety of strategies to threaten the West (p. 416-417).
>
> . . . the Soviets may elect to resist U.S. involvement in some "local war" . . . [which] could ultimately produce a confrontation between naval elements of East and West . . . There is considerable indication that the Soviets feel they have entered a new era in which they are less constrained from challenging the actions of the West (p. 416).

Of the fallacies of narration, Soviet analysts are perhaps especially vulnerable to presentism (fallacy 45) since the tendency to use the consequences to assess the antecedents is a constant factor in Soviet historiography. Soviet authors must keep one eye firmly on the present as they explain the past. Soviet analysts are likely to accept more of these hindsightful Soviet views than is justified simply because so much Soviet material is based on this approach.

It is virtually inevitable that naval analysts practice some tunnel history (fallacy 47); i.e., explaining and narrating naval events from a naval perspective. Naval analysts seem well aware of the danger that tunnel historians may bump into each other and commence, buried out of the light, a profitless debate waged with fallacious "essences" (fallacy 23). The analysts in our sample continually attempted to overcome this problem of perspective by bringing economic, military, political, technological, even organizational perspectives and experts into the naval analysis study groups. Naval analysts seemingly realize they cannot altogether escape a tunnel perspective. McGwire and McDonnell introduced the latest volume (SNI: xxxvi) of the "Soviet naval studies group" proceedings with a call for more synthesis of Soviet naval policy studies with research on Soviet foreign, military, economic, and domestic policies. Efforts at synthesis have been part of all three proceedings. Booth (SNP: 371) termed current analysis of Soviet naval power as an instrument of peacetime policy "curiously underintellectualized" compared to other strategic studies. [The context makes clear that Booth did not intend this as a compliment.] McGwire repeatedly noted the need "to extend the range and depth" of analysis of Soviet naval development (SNP: x; also SND: ix; SNI: xxxvi).

Like the fallacy of essences, tunnel history is related to the representativeness heuristic: samples are believed to represent the characteristics of their population, the essence of naval history is naval events.

A few naval analysts in our sample seem to commit the fallacy of archetypes (51) in seeing the Soviet Navy as a reinactment of the Russian Navy, and Soviet efforts (e.g., the drive for warm water ports) as identical in character and motivations to Peter the Great's efforts (cf. Graham, NPSP: 275-277, 287, 298).

The static fallacy (53) conceives of a dynamic problem as a static one, as in the unfolding of an unchanging plan, or the emergence of some predetermined entity. As related to behavior the static fallacy is closely related to the "fundamental attribution error" (see above): the tendency to overestimate the dispositional, motivational, and intentional bases of behaviors and to underestimate the situational and environmental influences. Naval analysts are not unaware of these biases, especially in assessing the degree to which Soviet naval events reflect a plan. For example, Dismukes and McConnell (1979: 283) write:

> In the case of capabilities for coercive naval diplomacy . . . there has been a debate over how much may have been foreseen and provided for . . . and how much was simply ad hoc reaction to circumstances. We suspect much more was projected than is usually credited; and this is said without any of the sinister connotations . . . of the "master plan" or the "grand design" . . . the Soviet Union [does] have development strategies for long-term goals . . . this is not to say that the Soviets were absolutely committed to anything by the 1954 decision or that there were no false starts, or shifts in accenting the general framework, or much rethinking at critical junctures that required novel departures.

In other words, the Soviets have had and still have a plan for coercive naval diplomacy, but a flexible rather than static one.

Many of the fallacies of historical causality have parallel cognitive causal biases. For example, the tendency to believe big events have big causes (fallacies 57, 61) are similar to the profound motive fallacy (Rothbart and Fulero, 1978, see "Assessing causes and effects" above). The reductive fallacy (60) corresponds to the tendency toward minimal causal reasoning and acceptance of the first cause as best cause (see "minimal causation" above). We have already noted the problem of "indiscriminate pluralism" (63), or proliferation of causes, and the overfitting of explanations to data.

Many of the cognitive problems of testing theories probably contribute to the maintenance of fallacious motivational explanations. Once a motivation hypothesis is accepted it will be very difficult to dislodge in view of the confirmation bias and the fundamental attribution error. Similarly, the use of analogies (fallacies 84-91) corresponds to the use of theories and schemas for prediction, and many of the biases of theory and data characterization seem to underlie the fallacious use of analogy in historical narration.

CHAPTER 5. LITTLE-USED METHODS AND EXPERIMENTAL TECHNIQUES FOR BETTER ESTIMATES

This chapter attempts to relate little-used analytic methods and experimental techniques to the estimative problem areas described in Chapter 4. Some of these techniques are used now by naval analysts (but apparently infrequently). Others have been applied to other 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: i):

> 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 as 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 lists.

PERCEIVING DATA

The three main problems noted in Chapter 4 regarding the perception of data were (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.

Events and Nonevents

Naval analysts must render a continuous, undifferentiated stream of information on the Soviet Navy into discrete, discriminable, describable entities. Perception is discrete rather than continuous. The Soviet Navy, its major units, and actors are 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 (Newtson, 1976). 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 action into

grosser units. People resort to shorter units of action 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 then normally.

One consequence of this difference in event perception is that different coding categories are applied to the cases of expected and unexpected events. Soviet actions during a crises 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 (Kahneman, 1973).

Several studies by Newtson (1973, Frey and Newtson, 1973) 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 a few indicators (Janis and Mann, 1977) by analysts or estimate consumers, (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 State 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 naval analysis is striking. While several naval analysts made 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), the events were not elaborately coded and frequency analyses were 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).

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.

Friedheim and John (SNP, ch. 18) 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 John 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. She found, for example, that the 'Soviet Navy' was mentioned relatively often (338 times), ranking fourth behind 'Soviet armed forces' (949), 'Soviet Army' (599), and 'Soviet defense capability' (385). The fifth and sixth issues, 'Soviet troops' and 'Soviet security,' were mentioned much less frequently (173 times each). Each time the speaker mentioned a topic, coders assessed the speaker's evaluative attitude from the context on a scale of good (+3) to bad (-3). The Politburo speakers were less favorable toward the Soviet Navy than seven other defense topics (Soviet troops, Soviet Air Force, Soviet Army, Strategic Rocket Forces, Soviet military cadres, PVO, Soviet armed forces, in that order). Each reference was also coded for "potency" to reflect whether the speaker's reference to the topic reflected activity (+3) or impotency (-3). In this case Politburo references to the Soviet Navy again put that service behind seven other defense topics (Strategic Rocket Forces, PVO, DOSSAF, potential Soviet defense capability, Soviet military cadres, Soviet troops, Soviet Army), and equal to 'Soviet armed forces' and 'Soviet defense capability' in "potency."

It is unlikely that, given the effects of the availability heuristic, an analyst reading this number of Politburo speeches could accurately recollect the relative frequencies of such references. Furthermore, coding such dimensions as "evaluation" or "potency" may provide useful information on elite views which could not be accurately ascertained by intuitive judgment alone.

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 analysts should give more attention to statistics on central tendencies (means, medians, and modes) and dispersion (variance) in samples. Naval analysts rarely report such information and seem not to use these simple descriptive statistics. Even the moving average (a smoothing statistic), commonplace in trend analysis and forecasting, is absent in our sample of naval analysis estimates.

One technique used by naval 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 these features are selected, Dawes recommends that they be systematically tested in a model of predictive judgment to verify the adequacy of the expert-selected variables to 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 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.

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 these 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 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 perceptions which they cannot readily dissect without aid. Introspection may be inadequate to abstract these perceptual 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 describes 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 rrom their statements by analysts and evaluated with scaling or clustering techniques.

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

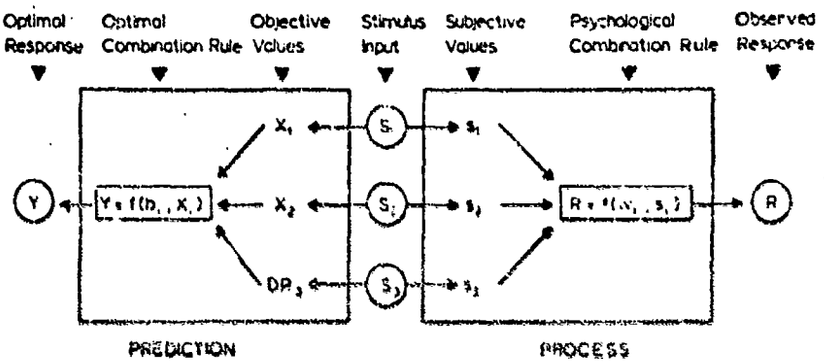
Figure 5.1 shows schematically how this can be done. A controlled set of data stimuli (S) are presented to the analyst and the analyst's estimative response is observed (R). The right side of the diagram suggests the analyst's estimation process. The input stimuli and the output response for this process 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, 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 captured objective weights to determine the self-insights of the analysts into their judgment and information integration processes.

Figure 5.1. General Estimation Diagram Illustrating Prediction Analysis (on the left) and Process Analysis (on the right). Stimuli (S_i) are common to both.*



*From Shanteau and Phelps, 1977: 258.

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 estimating 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. Fennessey (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 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 than 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 overreacting to a single piece of information, and to use the cumulative information contained in base-rate data.

Redundancy and Variance

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.

CHARACTERIZING DATA

In Chapter 3 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

In Chapter 4 we argued that 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 because 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, administration, 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 the example used in Chapter 4 to demonstrate that a naval analyst's conclusion was based on biased sampling (see Chapter 4, "Characterizing Data-Biased Sampling"), we 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, McConnell (SNI 612) wrote:

> MccGwire and Erickson . . . count on their subjective impressions of "tone" and "thrust" to tell them that Gorshkov is an advocate. I respect the conviction behind this approach, and perhaps 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, noted in Chapter 4, 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

In Chapter 4 we 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 these 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 (Sattath 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 Weiss, 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 judgements 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 onto 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' categorizations 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 semantical 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.

Factor Analysis in Intelligence

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. 5) 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 John (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 John were able to compare the views of the United States, the USSR, and Admiral Gorshkov on several Law of the Sea issues. Friedheim and John were also able to predict how often nations would favorably mention various themes by regressing characteristics of the countries (e.g., geographical characteristics, economic interests, 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 (SMI, 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.

ASSESSING COVARIATIONS

In Chapter 4 we 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 5.1, 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 5.1, 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 predictions. 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 underpredicts past criterion values, the analyst can make 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 behaviors. The actuarial method is the optimal means of assessing 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 5.1 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 human unreliability. 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 predictions made will be improved (Camerer, 1980). Furthermore, 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 bootstrapping models, it is important to reiterate a point made throughout Chapter 4, 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 variable.

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 judgments 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.

CAUSE AND EFFECT ASSESSMENT

The problems analysts may encounter in generating cause and effect explanations of intentions were reviewed in Chapter 4. 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.

Chapter 4 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 for 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 biases 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 cause (and the supporting evidence), and the rejected 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 in Chapter 4, "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 coalitions 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 relations, 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 overly simplistic perceptions. We use this term to reflect such problems as "minimal causation" (the tendency to search for the first plausible cause, see 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.

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 suspect relationships. 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 causes 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 termed "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 effects" 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" the naval events they seek to explain with causal trees and attach 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 than 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., McConnell'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 various the ways analysts are able to label or code an event, the more various 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 (see Chapter 4 "Prediction"), 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 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 no-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 mechanical 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 problem. While efforts to overcome impediments to causal search tree construction 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 brauches 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 symptoms, 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 an heuristic, 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

In Chapter 4 we reviewed several biases in the process of causal hypothesis generation, i.e., the tendency to generate some types of causal explanations and not others. We also noted in Chapter 4 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 described in Chapter 4 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 were reviewed in Chapter 4 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 in this 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 Huntington's 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, (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, technology, 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.

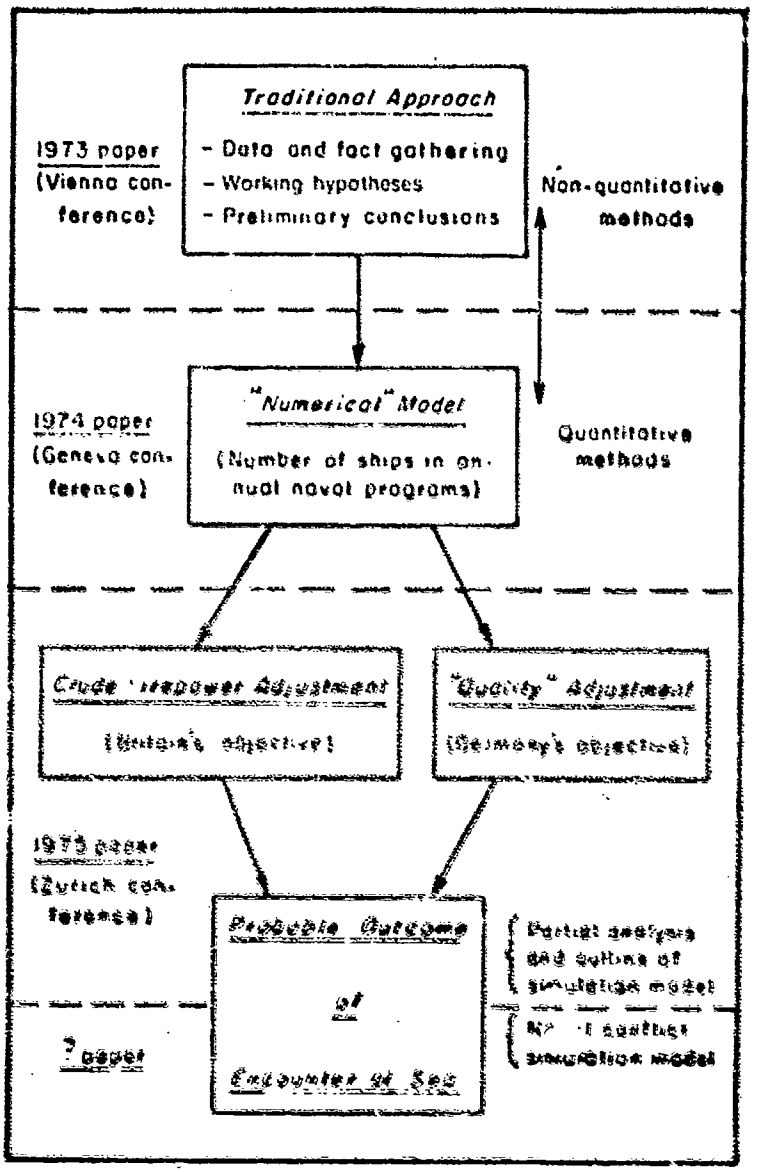
Explicit models of arms competition may yield some useful predictive indicators as well as systematize the analyst's reasoning on causal factors. For example, Wallace (1979) found that the product of the smoothed rate of arms increases for pairs of disputatious nations predicted whether war would follow the arms competition. Rapidly accelerated arms races escalated to war in 23 out of 25 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 5.2) 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

Figure 5.2. Lambelet's Approach to Qualitative and Quantitative Aspects of the Anglo-German Dreadnought Race.*



*From Lambelet (1976: 50)

actions and as being forced (or stumbling accidently) into good actions.

The methods outlined above in this section for systematizing causal hypotheses should assist the analyst to avoid these social cause biases. Another technique was outlined in Chapter 4: Kelley'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.

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. In Chapter 2 we noted that 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 analysts 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, noted in Chapter 4, 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 use 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 for 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' predictions. 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 noted in Chapte 4 were 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 her 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 crack 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 of these 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 events; 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 (see Chapter 4 "Prediction") 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 future 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 diff rent 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 [illegible] of the paradigm, or, no interpretable solution seems possible, or the anomaly is interpreted as a 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 those elements of the present (anomalies) 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 needed to anticipate a surprising future.

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 Meredith,
> The Ordeal of Richard Feverel

In Chapter 4 we reviewed 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 evidentiary 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 significance 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 an 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 Steeh, 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 scientific method. Once the analyst resolves to be scientific and to use 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 we outlined in 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 the other theory-revision methods noted in Chapter 4, and 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 how 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 [illegible] 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 a 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.

Theory and Naval Analysis: An Endnote

For many centuries scurvy was the greatest killer of seamen, worse than war, accidents, storms, or all other causes of death together. Mosteller (1981) recently summarized the course of the medical theorizing and research on this terrible naval scourage. The first clues as to scurvy's causes came when James Lancaster sailed with four ships to India for the British East India Company. Lancaster dosed the crew of one of the four ships with three teaspoons of lemon juice each day. Few of the men on this ship died, but more than a third of the men died on the other three ships.

James Lind, a physician, learned of Lancaster's experience and conducted another experiment on the ship Salisbury. He gave sailors with scurvy one of the following six treatments: vinegar, sea water, cider, vitriol elixir, oranges and lemons, or nutmeg. Those patients who received the citrus recovered, the rest did not.

The British Navy began using citrus juice on a regular basis and scurvy was wiped out in the fleet. The British Board of Trade followed the Navy and scurvy was eliminated from the merchant service.

Mosteller's tale of theory being put to practical use for naval service is an admirable case of naval analytic theorizing, save in one feature. Lancaster's discovery was made in 1601, Lind's experiment was conducted in 1747, the British Navy adopted lemon juice in 1795, and the Board of Trade did the same in 1865. As Mosteller (p. 886) noted, even today, 400 years after scurvy was identified as a naval plague, the U.S. Navy is concerned with the low levels of vitamin C in the plasma of submariners after patrols.

Theory is a powerful tool for naval analysis, if difficult at times to apply.

BIBLIOGRAPHY

AIZEN, I. (1977) "Intuitive Theories of Events and the Effects of Base-Rate Information on Prediction." Journal of Personality and Social Psychology, 35, 303-314.

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

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

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

ASHER, H. B. (1976) Causal Modeling. Beverly Hills, CA: Sage

BAR-HILLEL, M. (1975) "What Features Make Samples Seem Representative?" Journal of Experimental Psychology: Human Perception and Performance in press.

-----(1978) "The Base-Rate Fallacy in Probability Judgments." Acta Psychologica, in press.

BATHURST, R. S. (1979) Understanding the Soviet Navy. A Handbook. Newport, RI: Naval War College Press.

BEARMAN, D. (1979) "Review of Laboratory Life: The Social Construction of Scientific Facts." Science, 206, 500.

BEN-ZVI, A. (1976-77) "Misperceiving the Role of Perception: A Critique " The Jerusalem Journal of International Relations, 2, 2, 74-93

BETTS, R. K. (1980) "Intelligence for Policymaking." The Washington Quarterly, 3, 118-129

-----(1978) "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable." World Politics, 31, 1, 61-89.

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

BLECHMAN, B. (1973) The Changing Soviet Navy. Washington, D.C.: The Brookings Institution.

BOOTH, K. (1974) The Military Instrument in Soviet Foreign Policy 1917-1972. London: Royal United Services Institute for Defense Studies.

BRANSFORD, J. D. and J. J. FRANKS (1971) "Abstraction of Linguistic Ideas." Cognitive Psychology, 2, 331-350.

BRANSFORD, J. D., J. R. BARCLAY, and J. J. FRANKS (1972) "Sentence Memory: A Constructive Versus Interpretive Approach." Cognitive Psychology, 3 192-209.

BREHMER, B. (1980) "In One Word: Not From Experience." Acta Psychologica, 45, 223-241.

BUCKHOUT, R. (1974) "Eyewitness Testimony." Scientific American, 231, 23-31.

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

CANTOR, N. and W. MISCHEL (1979) "Prototypes in Person Perception." In L. Berkowitz (ed.) Advances in Experimental Social Psychology, vol. 12 New York: Academic.

CARROLL, J. S. (1978) "The Effect of Imagining an Event on Expectations for the Event." Journal of Experimental Social Psychology, 14, 88-96.

CHAN, S. (1979) "The Intelligence of Stupidity: Understanding Failures in Strategic Warning." American Political Science Review, 73, 171-180.
CHAPMAN, L. J. (1967) "Illusory Correlation in Observational Report." Journal of Verbal Learning and Verbal Behavior, 6, 151-155.
CHAPMAN, L. J. and J. P. CHAPMAN (1969) "Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs." Journal of Abnormal Psychology, 74, 271-280.
-----(1967) "Genesis of Popular but Erroneous Psychodiagnostic Observations." Journal of Abnormal Psychology, 72, 193-204.
COCKLE, P. (1978) "Analysing Soviet Defence Spending: the Debate in Perspective." Survival, 20 (5),209-219.
COHEN, R. (1980) "Rules of the Game in International Politics." International Studies Quarterly, 24, 129-150.
COMBS, B. and P. SLOVIC (1978) "Causes of Death: Biased Newspaper Coverage and Biased Judgments." Decision Research Report 78-8, Eugene, OR: Decision Research.
CRICK, F. (1979) "Crick Looks Back on DNA " Science, 206, 667.
DAWES, R. M. (1980) Confidence in Intellectual Judgments vs. Confidence in Perceptual Judgments," in E. D. Lautermann and H. Feger (eds.) Similarity and Choice. Bern: Hans Huber.
-----(1974) "The Role of the Expert in Constructing Predictive Systems." Proceedings of the IEEE Man and Cybernetics Conference, 522-525
-----(1971) "A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making " American Psychologist, 26, 180-186.
-----(1966) "Memory and Distortion of Meaningful Written Material." British Journal of Psychology, 51, 77-86.
DAWES, R. M. and B. CORRIGAN (1974) "Linear Models in Decision Making." Psychological Bulletin, 81, 2, 95-106.
DE SOTO, C. B. (1961) "The Predilection for Single Orderings." Journal of Abnormal and Social Psychology, 62, 16-23.
DICK, J. C. (1974) "The Strategic Arms Race, 1957-1961: Who Opened a Missile Gap?" Journal of Politics, 34, 1062-1110.
DISMUKES, B. and J. MCCONNELL (1979) Soviet naval Diplomacy. New York Pergamon.
DORNER, D. (1980) "On the Difficulties People Have in Dealing with Complexity " Simulation and Games, 11 (1), 87-106.
DOHERTY, M. E., C.R. MIRAIT, K. D. TWENEY, and M. D. SCHIAVO (1979) "Pseudodiagnosticity." Acta Psychologica, 43, 111-121.
DUVAL, S. and V. HENSLEY (1976) "Extensions of Objective Self-Awareness Theory: The Focus of Attention -- Causal Attribution Hypothesis " in J. H. Harvey, W. J. Ickes, and R. F. Kidd (eds) New Directions in Attribution Research. Hillsdale, NJ: Erlbaum.
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.
EINHORN, H. J. (1974) "Expert Judgement: Some Necessary Conditions and an Example." Journal of Applied Psychology, 59 (5), 562-571.
EINHORN, H. J. and R. M. HOGARTH (1981) "Behavioral Decision Theory Processes of Judgment and Choice." Annual Review of Psychology, 32, 1-60.
-----(1978) "Confidence in Judgment: Persistence of the Illusion of Validity." Psychological Review, 85 (5), 395-416.

EINHORN, H. J., D. N. KLEINMUNTZ, and B. KLEINMUNTZ (1979) "Linear Regression and Process Tracing Models of Judgment" Psychological Review, 86 (5), 465-485.
FAIRHALL, D. (1971) Russia Looks to the Sea: A Study in the Expansion of Soviet Maritime Power. London: Andre Deutsch.
FENNESSEY, J. (1977) "Improving Inference for Social Research and Social Policy: The Bayesian Paradigm" Social Science Research, 6, 309-327.
FISCHER, D. (1970) Historian's Fallacies. New York: Harper and Row.
FISCHHOFF, B. (1980a) "For Those Condemned to Study the Past: Reflections on Historical Judgment" in R. A. Shweder and D. W. Fiske (eds.) New Directions for Methodology of Behavioral Science. Fallible Judgment in Behavioral Research. San Francisco: Jossey-Bass.
-----(1980b) "Clinical Decision Analysis." Operations Research, 28 (1), 28-43.
-----(1977a) "Perceived Informativeness of Facts." Journal of Experimental Psychology: Human Perception and Performance, 3, 2, 349-358.
-----(1977b) "Decision Analysis: Clinical Art or Clinical Science?" Paper prepared for the Sixth Research Conference on Subjective Probability, Utility and Decision Making, Warsaw, Poland.
-----(1975a) "Hindsight & Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty." Journal of Experimental Psychology: Human Perception and Performance, 1, 3, 288-299.
-----(1975b) "Hindsight: Thinking Backward?" Psychology Today, 8, 71-76.
FISCHHOFF, B. and R. BEYTH (1975) "'I Knew It Would Happen': Remembered Probabilities of Once-Future Things" Organizational Behavior and Human Performance, 13, 1-16.
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.
FISHBEIN, M. and I. AJZEN (1975) Belief, Attitude Intention and Behavior. Reading, MA: Addison-Wesley.
FISKE, S. and D. R. KINDER (1978) "Schemas and Political Information Processing." Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario.
FLECK, L. (1979) Genesis and Development of A Scientific Fact. Chicago. University of Chicago Press.
FRANKS, J. J. and J. D. BRANSFORD (1971) "Abstraction of Visual Patterns." Journal of Experimental Psychology, 88, 409-413.
FREEMAN, J. R. and K. L. JOB (1979) "Scientific Forecasts in International Relations: Problems of Definition and Epistemology." International Studies Quarterly, 23 (1), 113-144.
FREY, J. and D. KAVIPSON (1973) "Differential Attribution in an Unequal Power Situation: Biased Inference of Biased Input?" Proceedings 81st Annual Convention of the American Psychological Association.
GANTZEL, K. J. (1973) "Armament Dynamics in the East-West Conflict: An Arms Race?" Papers, Peace Science Society, 20, 1-24.
GARTHOFF, R. L. (1978) "On Estimating and Imputing Intentions."

International Security, 2 (3), 22-32.
GAZIT, S. (1980) "Estimates and Fortune-Telling in Intelligence Work." International Security, 4, 36-56.
GEORGE, A. (1959) Propaganda Analysis: A Study of Inferences Made From Nazi Propaganda in World War II. Evanston, Ill.: Row, Peterson.
GERWIS, 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 1980's: 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.
-----(1968) "Simple Models or Simple Processes? Some Research on Clinical Judgments." American Psychologist, 23, 7, 483-496.
GRAY, C.S. (1972) "Gap Prediction and America's Defense: Arms Race Behavior in the Eisenhower Years." Orbis, Spring, 257-274.
HAMILTON, D. L. (1979) "A Cognitive-Attributional Analysis of Stereotyping" in L. Berkowitz (ed.) Advances in Experimental Social Psychology, vol. 12. New York: Academic.
HAMILTON, D. L. and R. D. FALLOT (1974) "Information Salience as a Weighting Factor in Impression Formation." Journal of Personality and Social Psychology, 30 (4), 444-448.
HAMILTON, D. L. and R. K. GIFFORD (1976) "Illusory Correlation in Interpersonal Perception: A Cognitive Basis of Stereotypic Judgments." Journal of Experimental Social Psychology, 12, 392-407.
HANDEL, M. I. (1980) "Surprise and Change in International Politics. International Security, 4, 57-85
HEIDER, F. (1958) The Psychology of Interpersonal Relations. New York Wiley.
HEISE, D. (1975) Causal Analysis. New York: John Wiley & Sons.
HENRICK, R. W. (1980) Soviet Naval Strategy. Annapolis, MD: Naval Institute Press.
HEUER, R. J. (ed.) (1978) Quantitative Approaches to Political Intelligence: The CIA Experience. Boulder, CO: Westview.
HILSMAN, R. (1956) Strategic Intelligence and National Defense. Glencoe, Ill.: Free Press.
HOGARTH, R. M. (1975) "Cognitive Processes and the Assessment of Subjective Probability Distributions" Journal of the American Statistical Association, 70, 271-294.
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, 503 529.
HOLZMAN, F. D. (1980) "Are the Soviets Really Outspending the U. S. on Defense?" International Security, 4 (4), 86-104.
HOWARD, R. (1980) "An Assessment of Decision Analysis" Operations Research, 28 (1), 4 27.
HOWELL, W. C. and S. A. BURNETT (1978) "Uncertainty Measurement: A Cognitive Taxonomy." Organizational Behavior and Human Performance 22, 45 68.
HUDSON, G. R. (1976) "Soviet Naval Doctrine and Soviet Politics: 1953-1975" World Politics, 29 (1), 90-113.

HUNTINGTON, S. P. (1958) "Arms Races: Prerequisites and Results." Public Policy, 8 41-86.
HUSBAND, W. B. (1979) "Soviet Perceptions of U.S. Positions of Strength Diplomacy " World Politics, 31 (4), 495-517.
JANIS, I. and L. MANN (1977) Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment. New York: Free Press.
JERVIS, R. (1968) "Hypotheses on Misperception," World Politics, 20 454-479.
JONES, C. D. (1977) "Soviet Hegemony in Eastern Europe: The Dynamics of Political Autonomy and Military Intervention." World Politics, 29 (2), 216-241.
-----(1975) "Just Wars and Limited Wars: Restraints on the Use of the Soviet Armed Forces." World Politics, 28 (1), 44-68.
JONES, E. E. and R. E. NISBETT (1971) The Actor and the Observer: Divergent Perceptions of the Causes of Behavior. Morristown, NJ. General Learning Press.
KAHNEMAN, D. (1973) Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall.
KAHNEMAN, D. and A. TVERSKY (1973) "On the Psychology of Prediction." Psychological Review, 80, 237-251.
-----(1972) "Subjective Probability: A Judgment of Representativeness." Cognitive Psychology, 3, 430-454.
KAPLAN, S. S. (1981) Diplomacy of Power: Soviet Armed Forces as a Political Instrument. Washington, DC: Brookings Institution.
KELLEY, H. H. (1967) "Attribution Theory in Social Psychology, In D. Levine (ed.) Nebraska Symposium on Motivation, 1967, Vol. 15. Lincoln, Neb.: University of Nebraska Press.
KLAN, N. H. (1975) "Soviet Influence in Egypt, 1967-73 " in A. Z. Rubinstein (ed.) Soviet and Chinese Influence in the Third World. NewYork: Praeger.
KIRK, E. J. (1980) Soviet Foreign Policy Perspectives in the 1970s: Group and Individual Perceptions. Bethesda, MD: Mathtech, Inc.
KLATZKY, R. L. (1975) Human Memory. San Francisco: Freeman.
KLOTZ, I. M. (1980) "The N-Ray Affair." Scientific American, 242, 168-175.
KONIAT, A., S. LICHTENSTEIN, and B. FISCHHOFF (1980) "Reasons for Confidence." Journal of Experimental Psychology: Human Learning and Memory, in press.
KRISCHER, J. P. (1980) "An Annotated Bibliography of Decision Analytic Applications to Health Care." Operations Research, 28, 97-113.
KUZMINSKI, A. (1975) "The paradox of historical knowledge " History and Theory, 12 (3), 269-289.
LANCHESTER, J. C. (1976) "A Complementary Analysis of the Anglo-German Dreadnought Race, 1903-1919." Papers Peace Science Society, 26 49-66.
-----(1975) "A Numerical Model of the Anglo-German Dreadnought Race." Papers, Peace Science Society, 24, 39-48.
-----(1973) The Anglo-German Dreadnought Race, 1903-1914." Papers Peace Science Society, 22., 1-45.
LANGER, E. J. (1975) "The Illusion of Control." Journal of Personality and Social Psychology, 32, 311-328.
LARKIN, J., J. MCDERMOTT, D. P. SIMON, and H. A. SIMON (1980) Expert and Novice Performance in Solving Physics Problems." Science, 208, 1335-1342.

LEE, W. T. (1977) The Estimation of Soviet Defense Expenditures, 1955-75. New York: Praeger.

LEPPER, M. R., L. ROSS, and R. LAU (1979) "Persistence of Inaccurate and Discredited Personal Impressions." Unpublished. Stanford University.

LICHTENSTEIN, S. and B. FISCHHOFF (1980) "Training for Calibration." Organizational Behavior and Human Performance, 26, in press.

-----(1977) "Do Those Who Know More Also Know More About How Much They Know?" Organizational Behavior and Human Performance, 20, 159-183.

LICHTENSTEIN, S. C., B. FISCHHOFF, and L. PHILLIPS (1976) "Calibration of Probabilities: The State of the Art," in H. Jungermann (ed.) Proceedings of the Fifth Research Conference on Subjective Probability, Utility and Decision Making, in press.

LICHTENSTEIN, S., P. SLOVIC, B. FISCHHOFF, M. LAYMAN, and B. COMBS (1978) "Perceived Frequency of Lethal Events." Decision Research Report 76-2. Eugene, OR: Decision Research.

LICKLIDER, R. E. (1970) "The Missile Gap Controversy." Political Science Quarterly, 85, 600-615.

LOFTUS, E. (1974) "The Incredible Eyewitness." Psychology Today, (December), 116-119.

LOPES, L. L. (1980) "Doing the Impossible: A Note on Induction and the Experience of Randomness." Unpublished. University of Wisconsin.

LORD, C., L. ROSS and M. R. LEPPER (1979) "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." Journal of Personality and Social Psychology, 37, 2098-2110.

LURIA, A. (1968) The Mind of a Mnemonist. New York: Basic.

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.

LYON, D. and P. SLOVIC (1976) "Dominance of Accuracy Information and Neglect of Base Rates in Probability Estimation." Acta Psychologica, 40, 287-298.

MAHONEY, M. J. (1977) "Psychology of the Scientist: An Analysis of Problem Solving Bias." Cognitive Therapy and Research, 1, 229-238.

-----(1976) Scientist as Subject The Psychological Imperative. Cambridge, MA: Ballinger.

MANIS, M., I. DOVALINA, N. E. AVIS, and S. CARDOZE (1980) "Base Rates Can Affect Individual Predictions." Journal of Personality and Social Psychology, 38, 231-248.

MARCH, J. (1978) "Bounded Rationality, Ambiguity, and the Engineering of Choice." Bell Journal of Economics, 9 (2), 587-608.

McCLINTOCK, C. G. (1981) "The Meta-Theoretical Bases of Social Psychological Theories," in C. G. McClintock and J. Maki (eds.) Social Psychological Theories. New York: Holt (in press).

McCONNELL, J. M. and B. DISMUKES (1979) "Soviet Diplomacy of Force in the Third World." Problems of Communism, January-February, 14-27.

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

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

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

McGWIRE, M., K. BOOTH, and J. McDONNELL (eds.) (1975) Soviet Naval Policy: Objectives and Constraints. New York: Praeger.
MEEHL, P. E. (1955) Clinical Versus Statistical Prediction: A Theoretical Analysis and A Review of the Evidence. Minneapolis: University of Minnesota Press.
MINEAR, R. H. (1980) "Cross-cultural Perception and World War II." International Studies Quarterly, 24 (4), 555-580.
MITROFF, I. I. (1974) The Subjective Side of Science. Amsterdam: Elsevier
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.
MURPHY, P. J. (ed.) (1978) Naval Power in Soviet Policy. Washington, DC: U.S. Government Printing Office.
MYNATT, C. R., M. E. DOHERTY, and R. D. TWENEY (1978) "Consequences of Confirmation and Disconfirmation in a Simulated Research Environment." Quarterly Journal of Experimental Psychology, 30, 395-406.
------(1977) "Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference." Quarterly Journal of Experimental Psychology, 29, 85-95.
NEWTSON, D. (1976) "Foundations of Attribution: The Perception of Ongoing Behavior." in J. H. Harvey, W. J. Ickes, R. F. Kidd (eds.) New Directions in Attribution Research, Vol. 1. Hillsdale, NJ: Lawrence Erlbaum.
-----(1973) "Attribution and the Unit of Perception of Ongoing Behavior." Journal of Personality and Social Psychology, 28, 28-38.
NICKEL, T. W. (1974) "The Attribution of Intention as a Critical Factor in the Relation Between Frustration and Aggression." Journal of Personality, 42, 482-492.
NISBETT, R. E. and N. BELLOWS (1977) "Verbal Reports About Causal Influences on Social Judgments: Private Access Versus Public Theories." Journal of Personality and Social Psychology, 35 (9), 613-624.
NISBETT, R. E., E. BORGIDA, R.CRANDALL and H. REED (1976) "Popular Induction Information is Not Necessarily Informative," in J. S. Carroll and J.W Payne (eds.) Cognition and Social Behavior. Hillsdale, NJ: Erlbaum.
NISBETT, R. and L. ROSS (1980) Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice-Hall.
NISBETT, R. E. and T. D. WILSON (1977a) "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review, 84 (3), 231-259.
-----(1977b) "The Halo Effect: Evidence for Unconscious Alteration of Judgments." Journal of Personality and Social Psychology, 35 (4), 250-256.
NITZE, P. H., L. SULLIVAN, et al. (eds.) (Atlantic Council Working Group) (1979) Securing the Seas: The Soviet Naval Challenge and Western Alliance Options. Boulder, CO: Westview.
NORMAN, D. (1969) Memory and Attention. New York: Wiley.
O'LEARY, M. K. and W. D. COPLIN (1975) Quantitative Techniques in Foreign Policy Analysis and Forecasting. New York: Praeger.
ORNSTEIN, R. E. (1977) The Psychology of Consciousness. New York Harcourt, Brace, Jovanovich.
OSKAMP, S. (1965) "Overconfidence in Case-Study Judgments." Journal of

Consulting Psychology, 29 (3), 261-265.
PIPES, R. (1980) "Recruitment, Training and Incentives for Better Analysis," in R. Godson (ed.) Intelligence Requirements for the 1980's: Analysis and Estimates. New York: National Strategy Information Center.
------(1977) "Why the Soviet Union Thinks It Could Fight and Win a Nuclear War." Commentary, 64 (1), 21-34.
PLATT, J. R. (1964) "Strong Inference." Science, 146, 347-53
POLANYI, M. (1958) Personal Knowledge: Towards a Post-Critical Philosophy. New York: Harper & Row.
POLMAR, N. (1972) Soviet Naval Power: Challenge for the 1970s. New York: National Strategy Information Center.
POPPER, K. R. (1962) Conjectures and Refutations. New York: Basic Books.
POSNER, M. I. (1973) Cognition. An Introduction. Glenview, Ill.: Scott, Foresman.
------(1969) "Abstraction and the Process of Recognition," in G. H. Bower & J. T. Spence (eds.) The Psychology of Learning Motivation, Vol. 3. New York: Academic.
RAMAGE, C. S. (1980) "Sudden Events." Futures, August, 268-274.
RATTINGER, H. (1975) "Armaments, Detente, and Bureaucracy." Journal of Conflict Resolution, 19, 571-595.
ROCK, I. and P. ENGLESTEIN (1959) "A Study of Memory for Visual Form." American Journal of Psychology, 72, 221-229.
ROSCH, E. and C. B. MERVIS (1975) "Family Resemblances: Studies in the Internal Structure of Categories." Cognitive Psychology, 7, 573-603.
ROSENBERG, S. and R. JONES (1972) "A Method for Investigating and Representing a Person's Implicit Theory of Personality: Theodore Dreiser's View of People." Journal of Personality and Social Psychology, 22, 372-386.
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.
ROSS, L. (1977) "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in L. Berkowitz (ed.) Advances in Experimental Social Psychology, vol. 10. New York: Academic.
ROSS, L., T. M. AMABILE, and J. L. STEINMETZ (1977) "Social Roles, Social Control, and Biases in Social-Perception Process." Journal of Personality and Social Psychology, 35, 485-494.
ROSS, L. and C. ANDERSON (1980) "Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments," in A. Tversky, D. Kahneman and P. Slovic (eds.) Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.
ROSS, L., D. GREENE, and P. HOUSE (1977) "The False Consensus Phenomenon: An Attributional Bias in Self-Perception and Social Perception Processes." Journal of Experimental Social Psychology, 13, 279-301
ROSS, L. and M.R. LEPPER (1980) "The Perseverance of Beliefs: Empirical and Normative Considerations," in R.A. Shweder and D. Fiske (eds.) New Directions for Methodology of Behavioral Science: Fallible Judgment in Behavioral Research. San Francisco: Jossey-Bass.
ROSS, L., M. R. LEPPER, and M. HUBBARD (1975) "Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." Journal of Personality and Social Psychology, 32, 880-892.
ROSS, L., M. R. LEPPER, F. STRACK and J. L. STEINMETZ (1977) "Social

Explanation and Social Expectation: The Effects of Real and Hypothetical Explanations Upon Subjective Likelihood." Journal of Personality and Social Psychology, 35, 817-829.

ROTHBART, M. (1980) "Memory Processes and Social Beliefs," in D. Hamilton (ed.) Cognitive Processes in Stereotyping and Intergroup Perception Hillsdale, NJ: Lawrence Erlbaum Associates.

ROTHBART, M. and S. FULERO (1978) "Attributions of Causality for Important Events." Unpublished. University of Oregon.

ROTHBART, M.,S. FULERO, C. JENSEN, J. HOWARD and P. BIRRELL (1978) "From Individual to Group Impressions: Availability Heuristics in Stereotype Formation." Journal of Experimental Social Psychology, 14, 237-255.

RUBINSTEIN, A. Z. (1980) "The Evolution of Soviet Strategy in the Middle East." Orbis, 24 (2), 323-337.

-----(ed.) (1975) Soviet and Chinese Influence in the Third World. New York: Praeger.

SAATY, T. L. (1968) Mathematical Models of Arms Control and Disarmament: Application of Mathematical Structures in Politics. New York: John Wiley & Sons.

SAFRAN, N. (1978) "Trial by Ordeal: The Yom Kippur War, October 1973." International Security, 2 (2), 133-170.

SARBIN, T. (1980) "On the Psychological Analysis of Counterdeception," in D. Daniel and K. Herbig (eds.) Multidisciplinary Perspectives on Military Deception. Monterey, CA: Naval Post Graduate School.

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

SCHUM, D. and P.E. PFEIFFER (1973) "Observer Reliability and Human Inference." IEEE Transactions on Reliability, R-22 (3), 170-176.

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

SHAKLEE, H. and B. FISCHHOFF (1979) "Strategies of Information Search in Causal Analysis." Decision Research Report 79-1. Eugene, OR: Decision Research.

-----(1977) "Discounting in Multicausal Attribution: The Principle of Minimal Causation." Decision Research Report 77-11. Eugene, OR: Decision Research.

SHIFFRIN, R. M. and R. C ATHINSON (1969) "Storage and Retrieval Processes in Long-Term Memory." Psychological Review, 76, 1, 179-193.

SHLAIM, A. (1976) "Failures in National Intelligence Estimates: The Case of the Yom Kippur War." World Politics, 28, 3, 348-380.

SHUGAN, S. (1980) "The Cost of Thinking." Journal of Consumer Research, 7, 99-111.

SHWEDER, R. A. (1977) "Likeness and Likelihood in Everyday Thought: Magical Thinking in Judgments About Personality." Current Anthropology, 18, 637-658.

SIMON, H. A. and R. K. SUMNER (1968) "Patterns in Music," in B. Kleinmuntz (ed.) Formal Representation of Human Judgment. New York: John Wiley & Sons.

SLOVIC, P. and B. FISCHHOFF (1977) "On the Psychology of Experimental Surprises." Journal of Experimental Psychology: Human Perception and Performance, 3 (4), 544-551.

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

SLOVIC, P., D. FLEISSNER, and W. S. BAUMAN (1972) "Analyzing the Use of Information in Investment Decision Making." Journal of Business, 45 (2), 283-301.

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.

SMEDSLUND, J. (1963) "The Concept of Correlation in Adults." Scandinavian Journal of Psychology, 4, 165-173.

SMITH, E. K. and F. D. MILLER (1978) "Limits on the Perception of Cognitive Processes: Reply to Nisbett and Wilson." Psychological Review, 85, 355-362.

SMOLANSKY, O. (1977) "Soviet Policy Toward Aircraft Carriers," in M. McGwire and J. McDonnell (eds.) Soviet Naval Influence: Domestic and Foreign Dimensions. New York: Praeger.

SNYDER, M. (1978) "Seek, And Ye Shall Find: Testing Hypotheses About Other People," in E. T. Higgins, C. P. Herman and M. P. Zanna (eds.) Social Cognition: The Ontario Symposium on Personality and Social Psychology. Hillsdale, NJ: Lawrence Erlbaum.

SNYDER, M. and N. CANTOR (1979) "Testing Hypotheses About Other People: The Use of Historical Knowledge." Journal of Experimental Social Psychology, 15, 330-342.

SNYDER, M. and S. GANGESTAD (1979) "Hypothesis-Testing Processes," in J. H. Harvey, W. J. Ickes, and R. F. Kidd (eds.) New Directions in Attribution Research, vol. 3. Hillsdale, NJ: Lawrence Erlbaum.

SNYDER, M. and W. B. SWANN, JR. (1978a) "Hypotheses-Testing Processes in Social Interaction." Journal of Personality and Social Psychology, 36 (11), 1202-1212.

-----(1979b) "Behavioral Confirmation in Social Interaction: From Social Perception to Social Reality." Journal of Experimental Social Psychology, 14, 148-162.

SNYDER, M. and S. W. URANOWITZ (1978) "Reconstructing the Past: Some Cognitive Consequences of Person Perception." Journal of Personality and Social Psychology, 36, 941-950.

SNYDER, M. L., W.G. STEPHAN, and P. ROSENFIELD (1976) "Egotism and Attribution." Journal of Personality and Social Psychology, 33, 435-441.

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.

STORMS, M. D. (1973) "Videotape and the Attribution Process: Reversing Actors' and Observers' Points of View." Journal of Personality and Social Psychology, 27, 2, 165-175.

SVENSON, O. (1979) "Process Descriptions of Decision Making." Organisational Behavior and Human Performance, 23, 86-112.

TASHDJIAN, E. (1977) "Time Horizon: The Moving Boundary." Behavioral Science, 22, 41-48.

TAYLOR, S. E. and J. CROCKER (1979) "Schematic Bases of Social Information Processing," in E. T. Higgins, P. Herman, and M. P. Zanna (eds.) The Ontario Symposium on Personality and Social Psychology, vol. 1.

Hillsdale, NJ: Lawrence Erlbaum.
TAYLOR, S. E. and S. T. FISKE (1978) "Salience, Attention, and Attribution: Top of the Head Phenomena " in L. Berkowitz (ed.) Advances in Experimental Social Psychology, vol. 11. New York: Academic.
-----(1975) 'Point of View and Perceptions of Causality." Journal of Personality and Social Psychology, 32, 3, 439-445.
TAYLOR, S. E. and J. H. KOIVUMAKI (1976) "The Perception of Self and Others: Acquaintanceship, Affect, and Actor-Observer Differences." Journal of Personality and Social Psychology, 33, 4, 403-408.
TESSER, A. (1978) "Self-generated Attitude Change," in L. Berkowitz (ed.) Advances in Experimental Social Psychology, vol. 11. New York: Academic.
TSUJIMOTO, R. N. (1978) "Memory Bias Toward Normative and Novel Trait Prototypes." Journal of Personality and Social Psychology, 36, 1391-1401.
TVERSKY, A. (1977) "Features of Similarity." Psychological Review, 84, 327-352.
TVERSKY, A. and D. KAHNEMAN (1981) "The Framing of Decisions and the Psychology of Choice." Science, 211, 453-458.
-----(1977) "Causal Schemata in Judgments Under Uncertainty," in M. Fishbein (ed.) Progress in Social Psychology. Hillsdale, NJ: Erlbaum.
-----(1974) "Judgment Under Uncertainty: Heuristics and Biases." Science, 185, 1124-1131.
-----(1973) "Availability: A Heuristic for Judging Frequency and Probability." Cognitive Psychology, 5, 207-232.
-----(1971) "The Belief in the 'Law of Small Numbers.'" Psychological Bulletin, 76, 105-110.
TWENEY, R. D., M. E. DOHERTY, W. J. WORNER, D. B. PLISKE, C. R. MYNATT, K. A. GROSS, and D. L. ARKKELIN (1980) "Strategies of Rule Discovery in an Inference Task." Quarterly Journal of Experimental Psychology, 32, 109-123.
van der KROEF, J. M. (1975) "Soviet and Chinese Influence in Indonesia," in A. Z. Rubinstein (ed.) Soviet and Chinese Influence in the Third World. New York: Praeger.
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.
WALSTER, E. (1966) "Assignment of Responsibility for an Accident." Journal of Personality and Social Psychology, 3, 73-79.
WARD, W. D. and H. M. JENKINS (1965) "The Display of Information and the Judgment of Contingency." Canadian Journal of Psychology, 19, 231-241.
WARREN, R. S. (1972) "Stimulus Encoding and Memory." Journal of Experimental Psychology, 94, 90-100.
WASON, P. C. (1960) "On the Failure to Eliminate Hypotheses in a Conceptual Task." Quarterly Journal of Experimental Psychology, 12, 129-140.
WASON, P. C. and P. N. JOHNSON-LAIRD (1963) Psychology of Reasoning: Structure and Content. London: Batsford.
WASSERMAN, S. (1980) "The Failure of Intelligence Prediction." Political Studies, 8 (2), 156-169.
WINKELMAN, W. A. (1974) How to Solve Problems: Elements of a Theory of Problems and Problem Solving. San Francisco: W. H. Freeman.
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?" Foreign 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.
WOHLSTETTER, R. (1979) "The Pleasures of Self-Deception." Washington Quarterly, 2 (4), 54-63.
WOLFE, T. W. (1972) "Soviet Naval Interaction with the United States and its Influence on Soviet Naval Development." RAND Report P-4913. Santa Monica: RAND.
WRIGHT, P. and P. RIP (1980) "Retrospective Reports on the Causes of Decisions: Retrieved Memories, Public Theories, or Impression Management?" Unpublished. University of Chicago, Center for Decision Research.
ZIMAN, J. (1979) Reliable Knowledge: An Exploration of the Grounds of Belief in Science. New York: Cambridge University Press.

APPENDIX

HISTORICAL FALLACIES*

I. Inquiry

A. Question - Framing

1. Baconian Fallacy - pure induction and synthesis, no preconceptions

2. Many Questions - (a) requiring two answers and allowing only one
(b) begging the question
(c) making a false premise
(d) demanding a simple answer for a complex question

3. False Dichotomies - assuming either/or and ruling out and/both; assuming mutual exclusion and collective exhaustion

4. Metaphysical Questions - demanding empirical answers for metaphysical questions

5. Fictional Questions - "what if" speculative questions

6. Semantical Questions - confusion of names with events

7. Declarative Questions - questions which state their own answers

8. Counterquestions - posing a question as an answer

9. Tautological Questions - all possible answers are true, by definition

10. Self-contradictory Questions - all possible answers are false

B. Factual Verification

11. Pseudo-proof - falsely precise and specific statements of no real meaning

12. Irrelevant Proof - facts of no bearing on the question

13. Proof by Absence - no evidence for X taken as evidence for not-X

*Summary of Fischer (1970).

14. Presumptive Proof - shifting burden of proof onto others

15. Circular Proof - assuming what was to be demonstrated

16. Mass Opinion Proof - amassing expert opinions on issues, rather than proof

17. Possible Proof - showing that X is possible, rather than that X is true

18. Hypostatized Proof - accepting a theory of X as proof of X itself

19. Appositive Proof - proving A by contrast to B, where B is in error

20. Misplaced Literalism - assuming everything said is literally meant

21. Misplaced Precision - accepting more accuracy than is possible

C. Factual Significance

22. Holism - a sense of the whole must guide selection of details

23. Essences - the "inner core of reality" must be sought and gives certainty and completeness to facts

24. Prodigious Facts - sensational facts are mistaken for significant facts

25. Furtive Facts - secret facts are assumed to have more truth or significance

26. Moralistic Facts - morally edifying events assumed to be of greater significance

27. Pragmatic Facts - facts that are useful for a social cause assumed also to be more significant

28. Aesthetic Facts - elegant facts believed to be more probative

29. Quantitative Facts - facts which count best count most

30. Antinomian Facts - facts count most which count least. Commitment to uniqueness

31. Fortuitous Facts - information uncovered by accident or at the "right" time deemed to be of greater significance

II. Explanation

A. Generalization

32. Insufficient Sampling - conclusions based on too-small samples

33. Lonely Facts - generalization from a single case

34. Special Pleading - use of a double standard of evidence interpretation to deal with inconvenient data, e.g., rejecting "outliers"

35. Statistical Impressionism - impressionistic interpretations of precise numbers, or casting impressions into precise numbers

36. Statistical Nonsense - a generalization with no meaning for the case in the present context

37. Probability Fallacies - (a) assuming a distribution will be exactly repeated
(b) gambler's fallacy - assuming a random event is overdue to occur

38. Ecological Fallacy - measurement classes do [illegible] with the variable being measured

39. False Extrapolation - stretching a statistical series past the breaking point

40. False Interpolation - accuracy of interpolation assumed greater than the accuracy of its weakest end points

41. Insidious Generalizations - bootlegging in generalizations despite disclaimers via hedge words: few, many, some, normal, common, often etc., without using any numerical evidence

42. Double-reversing Generalization - a halting generalization, understatement of generality interpretive bet-hedging

43. The Overwhelming Exception - A generalization based on an a-typical case, or the generalization that ignores major

exceptions

B. Narration

44. Anachronisms - taking concepts out of their right time

45. Presentism (nunc pro tunc) - the outcomes explain the means; the significance of antecedents determined by their consequences

46. Antiquarianism - opposite of presentism, collection of dead facts, well preserved from the present

47. Tunnel History - sealing off a class of events from everything else that has happened

48. False Periodization - applying inappropriate temporal limits to a problem. False application of a time scheme from one realm to another

49. Telescoping - making a long story too short, reducing a trend to a transformation

50. Interminalism - making a short story too long, or a long story longer. A false extrapolation stretched past the breaking point

51. Archetypes - conceptualizing changes in terms of primordial archetypes. Assuming time is repeating itself exactly

52. Chronic Chronism - misplaced temporal literalism, over-rigid chronological sequence, telling everything in its temporal order

53. Static Fallacy - conceptualizing a dynamic problem as a static one. The notion of emergence or unfolding of a static plan

54. Presumptive Continuity and Change - assuming change or continuity require no explanations

55. Genetic Fallacy - mistaking a description of process for an analytical explanation

56. Didactic Fallacy - attempting to extract lessons of history and applying them to the present

C. Causality

57. Post Hoc, Propter Hoc - if B followed A, A caused B, big events have big results

58. Cum Hoc, Propter Hoc - assuming correlation equals causation

59. Pro Hoc, Propter Hoc - putting the effect before the causes

60. Reductionism - reducing complexity to simplicity, or diversity to uniformity. Confusion of necessary with sufficient causes

61. Identity - assumption that causes must resemble their effects, e.g., big events have big causes, economic events have economic causes, etc

62. Absolute priority - belief in an absolute first term in a series of causes

63. Indiscriminate Pluralism - multiple causes postulated, of indeterminate number and weights

64. Mechanistic Causes - assumption that causes are additive rather than interactive

65. Reasons as Causes - a reason for an event assumed to be the cause of the event

66. Responsibility as Causes - confuses agency with causality examines "who did it" rather than "how did it happen"

D. Motivation

67. Pathetic Fallacy - ascription of animate behavior to inanimate objects, generalizing to groups from individuals

68. Apathetic Fallacy - assuming mechanistic characteristics for animated actors. Denying rationality, treating ideologies as pathologies

69. Idealism - overemphasis on knowledge and

calculation in conduct, and underestimation of other motives

70. One-dimensioned Motives - single motive explanations, e g. power

71. Universal Motives - assumption that people are intellectually and psychologically identical

72. Mass Motives - ascribing motivations beyond those of individuals to groups

73. Man-mass Motives - taking an individual and enlarging them into a group

74. Historian's Fallacy - assuming that subjects of history have insights into their situations that historians have with hindsight
Assuming that people can know what is coming next

E. Composition

75. Composition Fallacy - generalizing a property of an individual to a group, false extrapolation

76. Division Fallacy - application of group qualities to all members

77. Differences - conceptualizing a group in terms of its special rather than its generic characteristics

78. Similarity - rendering a special judgment on a group for a characteristic which is not special to it

79. Ethnomorphism - mirror-imaging, conceptualizing another group in terms of character istics of one's own group

80. Ethnocentrism - exaggerating the role of one group in its interactions with other groups

81. Elitism - concept of a group in terms of its upper strata

82. Racism - confusion of genetics and culture

83. Cross-grouping - using terms of one group to con-

83. Cross grouping - using terms of one group to conceptualize another group

F. False Analogy

84. Insidious Analogy - unintended analogy embedded in language, e.g., metaphors

85. Perfect Analogy - claiming a complete resemblance for a partial resemblance

86. False Analogy - shifting of analogical terms from one analogue to another

87. Absurd Analogies - analogical inferences extended to nonrelated characteristics

88. False Multiple Analogy - bootlegging a second analogy into a main analogy to undercut or reinforce the basis for comparison

89. Holistic Analogy - analogy from some part to the whole

90. Proof by Analogy - treating an analogy as proof rather than as support

91. Prediction by Analogy - using analogy to predict future events

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-A098153	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATES OF PEACETIME SOVIET NAVAL INTENTIONS: AN ASSESSMENT OF METHODS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Frank J. Stech		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0727
9. PERFORMING ORGANIZATION NAME AND ADDRESS MATHTECH, Inc. 4630 Montgomery Ave. Bethesda, MD 20014		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, code 434 800 N. Quincy St. Arlington, VA 22217		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 209
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

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	Military Intelligence
Cognitive Processes	Military Psychology
Intelligence Estimates	National Estimates
Intention Estimation	Naval Intelligence

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

The process of intention estimation was examined to determine what is required, how it is done, the nature of the problems, and the possible solutions. Three general approaches to intention estimation were assessed: control, sagacity, and acumen. Intelligence analysts largely rely upon the latter two. The requirements for sagacity and acumen were described. Estimates of peactime Soviet naval intentions were evaluated

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BLOCK 19. Key Words (continued)

Political Psychology
Social Psychology
Strategic Surprise

BLOCK 20. Abstract (continued)

in terms of these requirements and also specificity, complexity, and logic. Conflicts between estimates on nine major issues were reviewed. To investigate prediction of naval intentions through sagacity, the estimation process was divided into seven related steps: perceiving data; assigning weights to data; characterizing data; assessing covariations; assessing causes and effects; prediction; and forming, maintaining, and changing theories. Each of these steps was evaluated in terms of cognitive and information processing characteristics. Estimates by naval analysts were examined to determine how these were affected by human limits of information processing and problem-solving. Intention estimation through acumen was evaluated by reference to fallacies of historical reasoning and narration. ~~A Many historical~~ fallacies were found to derive from cognitive and information processing biases. To offset these problems, several little-used and experimental techniques were outlined. Some of these techniques have been applied in intelligence analysis and offer the possibility of improving and reinforcing naval analysis. These techniques are analytic aids, not substitutes for analysts or analytic reasoning.

DISTRIBUTION LIST

Office of Naval Research

CMDR K. S. Hull (Code 434)
Mr. J. R. Simpson (Code 434)
Dr. Martin Tolcott (Code 455)
ONR Branch Office - Boston

Office, Chief of Naval Operations, OP96

Mr. John Pond

DoD, Office of Net Assessment

Dr. Andrew Marshall

DIA

Maj. Alan D. Dunham (DE-X)

CIA

Dr. Lloyd Jordan (ORD)
Mr. Gerald Dargis (OSR)

USAF Electronic Security Command

MG Doyle E. Larson

Center for Naval Analyses

Mr. Bradford Dismukes
Mr. James McConnell
Mr. Robert Weinland
Mr. Charles Petersen

DARPA

Dr. Judith Ayres Daly (CTO)

Director, Naval Research Laboratory (Code 2627)

Defense Documentation Center

National War College

CMDR Steven Kime
Dr. Ned Lebow

Naval Postgraduate School

Dr. Katherine Herbig
Mr. Richards J. Heuer, Jr.

The Brookings Institution

Dr. Richard Betts
Dr. Donald Daniel
Dr. Michael MccGwire

The Carnegie Endowment

Dr. Barry Blechman

Center for Advanced Study in the Behavioral Sciences, Stanford University

Dr. Robyn M. Dawes

Center for International Affairs, Harvard University

Dr. Michael Handel

The Mershon Center, Ohio State University

Princeton University

Dr. Edward E. Jones

Yale University

Dr. Garry D. Brewer

Decision Research

Dr. Baruch Fischhoff
Dr. Paul Slovic

International Information Systems

Dr. Stephen J. Andriole

Ketron

Dr. Robert Herrick

Science Applications, Inc.

Mr. Philip Romero

Embassy of Israel

General Ben Torat