

R-3689-CC

Studying First-Strike Stability with Knowledge- Based Models of Human Decisionmaking

Paul K. Davis

April 1989

Supported by the
Carnegie Corporation

Project on
AVOIDING NUCLEAR WAR:
MANAGING CONFLICT IN THE NUCLEAR AGE

RAND



RAND/UCLA
Center for the Study of
Soviet International Behavior

The research described in this report was prepared as a contribution to the project on Avoiding Nuclear War: Managing Conflict in the Nuclear Age, which is being conducted jointly by The RAND Corporation and the RAND/UCLA Center for the Study of Soviet International Behavior. The project is supported by a grant from the Carnegie Corporation of New York.

Library of Congress Cataloging in Publication Data

Davis, Paul K., 1943-

Studying first-strike stability with knowledge-based models of human decisionmaking.

"Supported by the Carnegie Corporation."

"April 1989."

"R-3689-CC."

Bibliography: p.

1. First strike (Nuclear strategy) 2. Nuclear crisis stability. 3. Nuclear warfare—Decision making. I. RAND Corporation. II. Carnegie Corporation of New York. III. Title.

U263D39 1989

355.02'17

89-4022

ISBN 0-8330-0953-2

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

Studying First-Strike Stability with Knowledge- Based Models of Human Decisionmaking

Paul K. Davis

RAND |

**RAND/UCLA
CENTER FOR THE STUDY OF
SOVIET INTERNATIONAL BEHAVIOR**

PREFACE

The RAND Corporation and the RAND/UCLA Center for the Study of Soviet International Behavior (CSSIB) have a joint project for the Carnegie Corporation entitled "Avoiding Nuclear War: Managing Conflict in the Nuclear Age." This project has two broad objectives: to understand better the process of escalation from peace through general nuclear war; and to identify and assess, while protecting vital national interests, unilateral and cooperative measures that might inhibit unintended escalation or improve prospects for reversing or controlling escalation once it has begun.

This report contributes to the larger project by describing efforts of the author and colleagues in the RAND Strategy Assessment Center to develop and use knowledge-based analytic models of national-command-level decisionmaking for better understanding and communicating issues of deterrence, escalation, and war termination. The intended audience includes researchers and government figures interested in crisis decisionmaking, related command and control problems, and a framework for thinking about first-strike stability that integrates both force-posture factors and behavioral factors.

SUMMARY

THESIS

This report argues that *efforts to understand and improve first-strike stability should be guided by a formal theory of human decisionmaking that accounts for behavioral factors* such as mindset, desperation, fatalism, perceptions, and fears. The theory should reflect the fact that human reasoning can be highly nonrational, with conclusions other than those the same decisionmakers would reach if they were reminded of other objectives and values, or if they were shown alternative ways to make tradeoffs. A theory of first-strike stability should incorporate force-posture factors such as the vulnerability of nuclear-weapon and command-control systems, but these factors are usually given too much relative prominence. The way they would probably affect decisionmaking in crisis is better understood within a behavioral theory than in the narrow context of mathematical exchange calculations.

A natural way to express and communicate such a formal theory is to develop knowledge-based simulation models with enough flexibility to represent a broad range of *alternative* human behaviors. Since such models cannot be empirically validated, their value lies primarily in making explicit: key issues and related variables, plausible reasoning patterns, and possible consequences of different reasoning chains. This is a traditional purpose of models, but it requires that the models be understandable and modifiable.

Previous RAND work has produced experimental models of national-command-level decisionmaking with what appears to be the necessary flexibility and understandability. This study is essentially a think piece that outlines an approach using the concepts and methods of that prior work to explore the issues of first-strike stability in some detail.

TOWARD A THEORY OF FIRST-STRIKE STABILITY

First-strike stability can be regarded as the likelihood that no one will conduct a nuclear first strike (here defined as a *large-scale* nu-

clear attack on the opponent's homeland) during some period of interest such as the next 100 years. First-strike stability therefore depends on the improbability of crises and the improbability that crises would result in first strikes. This study focuses on the latter criterion, which in turn depends on crisis decisionmaking by human beings.

In seeking to understand how a national leader (or other figures in control of large numbers of nuclear weapons) might decide to conduct a first strike, it proved useful to conduct research in three phases: (1) a divergent-thinking phase of scenario spinning, which involved writing short stories or vignettes leading plausibly to first-strike decisions (stories based on imaginative ideas from history, general human experience, fiction, strategic theory, political-military war gaming, and discussions with people involved in both nuclear planning and operations); (2) a research period reviewing the literature on cognitive sciences for known classes of decisionmaking error and to confirm speculations from the scenario work; and (3) an analysis period identifying and organizing likely determinants of first-strike decisions in the context of a decisionmaking model.

The results of this exploratory research can be summarized in several ways. *Figure S.1 shows a simplified "fault tree" indicating some very different reasoning paths that could lead to a first-strike decision.* The paths involve human conclusions such as those shown in the second row. Reading left to right, some of the most important are:

- We should *preempt*.
- Going first is our *only chance* for survival (however small).
- Going first would deny our opponent (or give us) the *upper hand* (an overwhelming military advantage that could be used to force surrender or acceptable war termination).
- Only by going first can we *assure the destruction of our enemy*.
- The final escalation is beginning so we should *strike now while we can*.

Most of these hypothesized intermediate conclusions involve desperation and fatalism rather than calculation or gamesmanship.

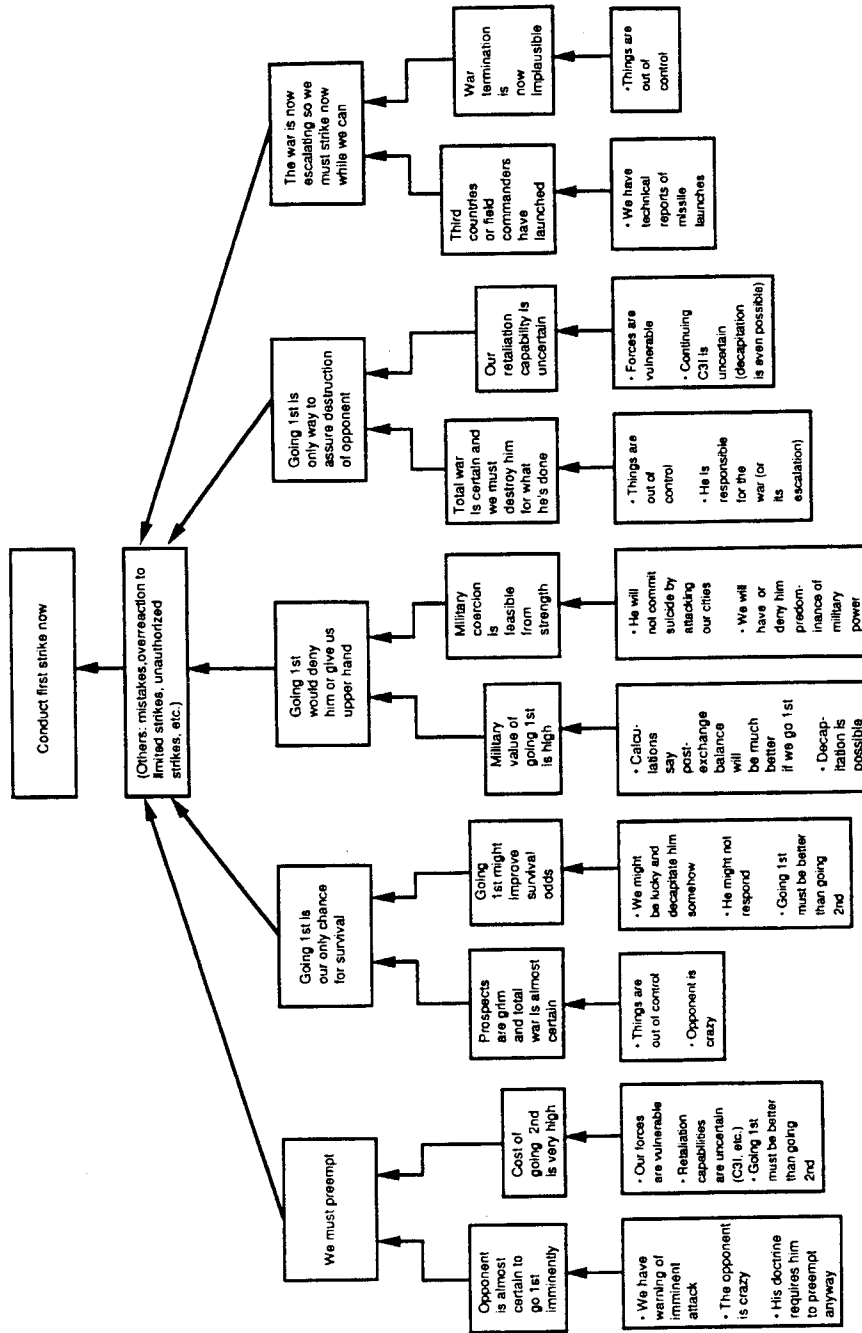


Fig. S.1—A partial fault tree of first-strike decisions

The decisionmaker feels compulsions more than he sees incentives. Especially significant are notions arising in various forms that amount to one of the following *dangerous ideas*:

- *Total war is inevitable and imminent.*
- *The enemy is almost certain to go first if we don't.*
- *The cost of going second (rather than first) is high in terms of (1) possible survival, (2) possible coercion of the enemy into war termination, or (3) assurance of destroying the enemy.*

In evaluating the cost of going second—an essentially qualitative concept informed by quantitative calculations but dependent on values and assumptions—humans in crisis would probably be erratic in considering best cases, expected cases, and worst cases. While peacetime analysts might focus on the alleged expected cases, decisionmakers in crisis would be likely to substantially weight worst cases (especially in rejecting a particular option) or best cases (especially when desperately “grasping at straws”).

There is a firm basis in the cognitive sciences for fearing that human decisionmakers might behave erratically and entertain such dangerous ideas even when they are not true. Many dangerous ideas could result straightforwardly (1) from classic types of nonrationality observable in history, such as the Japanese decision to attack Pearl Harbor, or (2) from laboratory experiments testing more general aspects of the human reasoning apparatus. This apparatus depends for its pattern-matching brilliance on heuristic simplifications such as thresholding (e.g., ignoring low-probability events), bolstering (e.g., building up the attractiveness of some options while tearing down the attractiveness of others to make the decision easier), and stereotyping (e.g., ascribing pure malevolence to one's opponent). It is also naturally difficult for humans to accept new information inconsistent with their current mindset—a feature of human thinking that could exaggerate perceptions of certainty and obstruct war-termination overtures or, at the other extreme, support termination-related strategic deception.

Mitigating factors include: the worldliness of political leaders, who are seldom overawed by technical military calculations; the inclination of many such leaders to seek out a range of advice; and staff pro-

cesses, which often balance information and impose a procedural structure counteracting rash judgments. Political leaders also tend to value enormously avoiding the destruction of their homeland or even a small part of it. All of this is grounds for optimism—on the average. However, the peacetime attributes of current political leaders would not necessarily prevail under the stress and desperate circumstances of nuclear crisis or war. Moreover, one or both sides might delegate some degree of decisionmaking to military commanders, who might see their job primarily in terms of accomplishing missions, such as assured destruction of the enemy, or who might be operating with lower quality information and advice. All things considered, *historical and psychological evidence on decisionmaking is sobering rather than encouraging when one considers the stakes*. Avoiding crises is a sound policy.

IMPROVING FIRST-STRIKE STABILITY

Desirability of Improvements

It is not self-evident that improving first-strike stability should be a principal U.S. objective. The usual argument against improvements emphasizes three points: (1) first-strike stability is already high; (2) there are reasons, such as maintaining strategic competitiveness and verification, for favoring some of the weapon systems (e.g., ICBMs in silos) disliked by those who favor improvement; and (3) there are logical tensions between improving first-strike stability and maintaining deterrence, especially extended deterrence: deterring attacks on our allies by threatening use of strategic nuclear weapons depends on the credibility of such attacks, which would presumably be reduced if first-strike stability were extremely high.

This study began with the view that these arguments were all serious and that the issue was by no means clear-cut, but *the arguments against improving first-strike stability appear unconvincing after analysis*—albeit analysis with some unavoidably subjective elements. Unless one believes that deterrence is on the verge of failing, the first point may be regarded as nonrational thresholding. Although the likelihood of nuclear first strikes in a given crisis is probably already small, the consequences of general nuclear war would be so great that we should seek to make it *extremely* small—seeking to avert such a

war not only in the present but also for the decades and centuries ahead.

The second point is simply misleading and is distressing to traditional defense-oriented strategic analysts as well as others. It can be readily demonstrated that *one can be for modern, survivable, impressively competitive, affordable, verifiable, and second-strike-effective weapon systems while also being for first-strike stability*. The confusion here is largely due to the vulgarization of the term “stability.” People tend to claim that systems they like are stabilizing and systems they don’t like are destabilizing, with no concrete sense of what the words mean.

The tension between deterrence and stability is much more complex. There are, however, many ways to improve first-strike stability that would almost certainly not reduce deterrence (e.g., continuity-of-government programs). Even those improvements that might have an effect would probably have little impact on the most important element of deterrence, establishing the following general (existential) mindset: the superpowers must avoid altogether crises that might possibly lead to nuclear war. It is true that some of an opponent’s measures to increase first-strike stability (e.g., deploying mobile ICBMs and dismantling silo-based ICBMs) might undercut using certain escalation threats to bring about a de-escalation once theater war has begun, but it is questionable whether extended deterrence would be credible in that event. Further, so long as neither side has anything approaching a disarming first-strike capability, and few advocates of extended deterrence seek or believe feasible such a capability, this element of extended deterrence in the course of war would depend far more strongly on the *will* to conduct even a limited nuclear strike on the opponent’s homeland than on the magnitude or detailed nature of that strike. However unconventional it might seem to analysts performing standard nuclear-counterforce calculations, the threat of striking vulnerable conventional-military or economic targets—such as army, navy, or air force bases or electrical power grids—has the same probable coercive effectiveness as the threat of a limited attack on ICBM silos. The tensions between extended deterrence and first-strike stability have been substantially exaggerated; however, tensions between nonextended deterrence and stability are discussed below.

Improvement Agenda

There are three principal mechanisms for improving first-strike stability:

- *Improve force-posture stability.*
- *Review and adjust nuclear policies, doctrine, and the way they are discussed.*
- *Improve the likely quality of crisis decisionmaking through efforts involving education, exercises, and staffing.*

There are opportunities to pursue all of these with both unilateral and either bilateral or multilateral efforts. Arms control can make all of them easier.

The first mechanism translates into the traditional recommendation of strategic analysts for survivable forces, survivable communication systems, and programs for continuity of government and military command. Consistent with the possible worst-case proclivities of decisionmakers in crisis, very high standards should be imposed in judging survivability.

The second mechanism involves strategic policy and doctrine. There *are* some significant tensions between deterrence—if based on hair-trigger assured-retaliation capability—and the desire to enhance first-strike stability, especially during periods when the nuclear powers see their opponents as serious and aggressive-prone military threats. The superpowers' current nuclear policies and doctrines appear to have been dictated largely by the perceived requirements of deterrence, with less emphasis on first-strike stability and with considerable recent emphasis on a hair-trigger response. It would seem that 1989 would be a good time for the sides to review these matters individually and jointly and make adjustments accordingly. In this regard *a few observations on the effects of policy and doctrine are especially relevant:*

- *Deterrence-motivated targeting policies and doctrines that emphasize prompt retaliation against political leadership, command-control systems, and nuclear forces necessarily contribute to first-strike instability.* This is because the odds of achieving targeting objectives would be higher in a first strike

(e.g., in a strike conducted before leaders and forces had been dispersed) than in a second strike, so the “calculated” cost of going second is bound to be high.

- *The combination of highly valuing counterforce and leadership targets, requiring that they be struck promptly, and depending on vulnerable silo-based ICBMs for the mission results in almost the worst possible scenario—the targets themselves are vulnerable only in a first strike.*
- *Launch-under-attack (prompt-retaliation) doctrine encourages an obsession with “using rather than losing” that could, in crisis, be transformed into perceived requirements for “pre-emption” ; this is because of doubts about the ability to launch under attack operationally, given warning and decision-time problems.*

Because, given the current force structures, difficult tensions play among the perceived requirements for deterrence, strategic competitiveness, and first-strike stability, it is not clear whether policies and doctrines will change. As a minimum, however, when “requirements” are levied on strategic forces and when their performances are being graded for net assessments, readiness, or other purposes, greatest emphasis should be given to the “requirement” (not currently achievable in general nuclear war) for national survival. Meeting this requirement is more important than meeting any other requirement; the constant reminder that damage limitation is not feasible even with a first strike should propagate the intuitive stabilizing judgment that the most significant cost of going second—the cost of national survival—is already moot. Many know this; others forget or are taught to think solely in alleged warfighting terms, although the warfighting in question would not assure national survival, much less more grandiose war aims.

The third mechanism to improve first-strike stability involves improving the likely quality of decisionmaking in crisis, including low-level crises well below war or nuclear war. Three general activities seem relevant: (1) general education of future political and military leaders, (2) specific education of high-level leaders through war gaming, background briefings, and readings, and (3) improved staff procedures for informing decisionmakers in crisis about relevant factors

and for assisting decisionmakers in avoiding recognized classes of bad reasoning.

The report discusses each of these to some degree, primarily in terms of establishing an agenda for further research. *Building both descriptive and normative decisionmaking models could be an important element in this agenda*, because models can be a powerful tool for communicating ideas and assisting individuals to discover important truths. Experience on other projects indicates that such models could be used as part of or as supplements to political-military gaming of crisis and war and could greatly extend insights. As in other domains, it is important to lay out variables and logic explicitly.

Such models could also help improve the quality of information presented to humans in crisis and war games and quite possibly in real-world nuclear crises. Insights gained from modeling and the use of models in war games and studies should have direct implications for informational displays, agenda setting in decision meetings, and procedural check lists designed to avoid predictable classes of bad reasoning. In particular, it should be possible to anticipate a significant fraction of the dangerous ideas and to prepare information displays that would defuse them as they arise. Some hypothetical circumstances result in dangerous ideas being correct (e.g., by virtue of highly unstable force structures creating objective incentives for a first strike); in these cases it should at least be possible to put the various considerations into perspective.

ACKNOWLEDGMENTS

This report benefited from the comments and suggestions of numerous individuals, but I would especially like to thank Alexander George of Stanford University and RAND colleagues William Harris, Arnold Kanter, Glenn Kent, and Robert McCalla, who provided detailed reviews. Discussions with Paul Bracken of Yale University, RAND consultant Joachim (Ed) Scholz, and Stephen Cimbala of Pennsylvania State University were also quite helpful. The remaining flaws, of course, are my own responsibility.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	xv
FIGURES	xix
TABLES	xxi
Section	
I. INTRODUCTION	1
Objectives	1
Definitions	1
Toward a Theory of Related Decisionmaking	4
Approach	11
II. REASONS FOR PLAUSIBLE FIRST STRIKES	13
A Survey of Miniscenarios	13
Observations	17
Heuristic Classification: The Fault Tree Concept	26
Key Variables	28
III. BACKGROUND CONSIDERATIONS	32
Possible Contradictions Between Stability and Deterrence	32
Built-in Shortcomings of Human Ability to Reason ..	37
Bounded Rationality	41
Effects of Declaratory Policy and Doctrine	43
Normative Issues: Desired Decisionmaker Values	44
Potential Effects of Strategic Defense	48
IV. REFLECTING ISSUES IN MODELS OF DECISIONMAKING	50
General Comments	50
Classes of Decision	51
Formal Representations	55

V. DISCUSSION AND CONCLUSIONS	62
Separable Issues	62
Dangerous Ideas	65
Next Steps	80
Appendix	
A. ON THE DESIRABILITY OF IMPROVING FIRST-STRIKE STABILITY	81
B. FORCE-POSTURE AND DOCTRINE ISSUES AFFECTING FIRST-STRIKE STABILITY	89
BIBLIOGRAPHY	93

FIGURES

S.1.	A partial fault tree of first-strike decisions	vii
1.	A process model of decision	8
2.	An illustrative partial hierarchy of variables	9
3.	Illustrative causes of desperation and compulsion	19
4.	Cascading implications of decision character	20
5.	Alternative ladders of objectives	24
6.	Illustrative changes of objectives as crises deepen	24
7.	Illustrative shift in perceptions about possible objectives	25
8.	A partial fault tree of first-strike decisions	27
9.	Key variables in a first-strike decision	31
10.	The Devil's Dilemma tradeoff	34
11.	Factors in mainstream decisionmaking under nondesperate circumstances	52
12.	Factors in mainstream decisionmaking under desperate circumstances	53
13.	Factors in Type-B decisionmaking under nondesperate circumstances	54
14.	Factors in Type-B decisionmaking under desperate circumstances	55
15.	An influence diagram of stability issues	64
16.	Structure used to help eliminate dangerous ideas	71
17.	A possible "military" perspective on the cost of going second rather than first	73
18.	A decisionmaker's possible view: no difference between first- and second-strike results	75
19.	A prudent decisionmaker's likely view	75
20.	Illustrative effects of uncertainty	76
A.1.	Postulated relationship between real and perceived likelihood of nuclear war	85
A.2.	Postulated extended deterrence versus threat	86
A.3.	Why the Devil's Dilemma tradeoff is not currently a problem	87

TABLES

1. Possible decisionmaking rules during ongoing war under nondesperate circumstances	56
2. Decision table for counterforce-oriented decisionmaking	59
3. A conservative's assessment	59
4. An optimist's assessment	60

I. INTRODUCTION

OBJECTIVES

This study has three objectives: (1) to broaden and clarify the concept of first-strike stability, (2) to identify the principal determinants of plausible first strikes, and (3) to discuss how first-strike stability can be measured, increased, and decreased. The study also comments on the tensions between improving first-strike stability and enforcing other national-security objectives such as deterrence and extended deterrence. Because the approach is new, the current work is something of a think piece, focusing more on ideas and frameworks than on details. Although extensive references are made to the literature, they are by no means comprehensive and are limited to materials actually consulted. The study concludes with suggestions about both additional research and concrete mechanisms for improving stability.

DEFINITIONS

In this study the term "first strike" means a *large-scale* nuclear attack on the opponent's homeland, an attack for the purpose of disarming and/or destroying the opponent, greatly limiting damage to one's own homeland, and/or significantly changing the counterforce balance (i.e., the balance of strategic nuclear forces). The definition does not, for example, consider small terrorist attacks or small demonstrative strikes in the context of a NATO/Warsaw Pact conflict to be first strikes.

In the literature, "first-strike stability" is often discussed but seldom defined. It is customarily associated with the absence of "incentives" for either superpower to conduct a nuclear first strike. In practice, the incentives have been measured technically with certain counterforce "exchange calculations" and additional assumptions. For example, if both sides' strategic forces consist of highly accurate and vulnerable MIRVed ICBMs, then certain technical calculations taken alone would indicate a considerable technical incentive for preemption in crisis—especially if each side assumed that its opponent would not launch on warning. The incentive might be described in terms of

the post-exchange ratio of residual nuclear weapons (a poor but common measure) or something more sophisticated, such as the sides' post-exchange capability for full-scale retaliatory attacks on other nuclear forces, general-purpose forces, command and control, and war-supporting industry.

In fact, these customary concepts and measures of stability address only one aspect of the problem, what might be called force-posture stability. They have been and continue to be quite useful in comparing stability-related aspects of alternative force structures, including force structures with different offense-defense relationships.¹ They can also be mathematically elegant, because they define the problem in terms of a few well-defined and quantifiable variables that can be manipulated with game theory calculations and simulations. At the same time, however, they ignore other factors that probably would, and should, be even more important to real-world decisionmakers faced with an actual nuclear crisis.

To improve discussion we need a broader definition of first-strike stability. At an abstract level *we can define first-strike stability over the time period T (e.g., the next hundred years) as $S(T)$, the probability that no first strikes will occur in that period.* Conceptually, we can imagine evaluating $S(T)$ by considering the range of possible world states, the probability of those states arising sometime during period T, and the probability of a first strike occurring in each of those states.² In this abstraction a world state is specified approximately by high-level variables, such as the current state of conflict among nations, the nature of the force postures, and the general nature of perceptions.

There are several worthy features of this definition, despite its abstraction:

¹See especially Kent and DeValk (1986), Wilkening and Watman (1986), and Kent, DeValk, and Thaler (1988). The first two include careful discussion of methodology and treat the interesting stability challenges associated with strategic defenses.

²More nearly precisely, but still notionally, $S(T) = 1 - \int P(x;T)F(x)dx$ where $P(x;T)dx$ is the probability that state x arises sometime during the period T, $F(x)$ is the probability that at least one of the sides finding itself in state x conducts a first strike at that time, and $\int \dots dx$ denotes an integration over all world states x (which should really be indicated as having multiple dimensions x_1, x_2, \dots, x_N). Since not all features of a real world state are defined by x (or x_1, x_2, \dots, x_N)—e.g., issues of health and the order of advice given to leaders—behavior in the state x is probabilistic.

- It reminds us that we are interested in avoiding first strikes not only in a particular crisis but altogether—both now and in the foreseeable future. Hence, we would prefer $S(T)$ to be large (1 or nearly 1) even for $T = \infty$. This means we should seek to assure that the likelihood of escalation in a given crisis is even smaller than we might otherwise assume adequate, because running a small risk many times may be to run a large risk overall.
- It highlights the problems of (1) reducing the likelihood of escalation in given world states and (2) avoiding world states in which escalation likelihoods are significant. Thus, avoiding crises is at least as important as managing those crises well. Also, avoiding certain force structures and force postures is important because they would be a destabilizing influence in many circumstances.

This study is largely concerned with the likelihood of first strikes in extreme crisis or conflict (the crux of what is often called *crisis stability*), but it also discusses the likelihood of lower-level crises arising or growing more extreme.

Consider this description of the conditions for first-strike stability in a specific situation:

First-strike stability in a given situation is high if neither national leaders nor any other people in control of nuclear weapons see incentives or feel compulsions to launch large-scale nuclear attacks against any opponent homelands.³

This apparently straightforward statement differs significantly from the more usual concept of first-strike stability based on incentives as measured by exchange calculations:

- It refers to “leaders” rather than “nations,” reminding us that degree of stability is largely determined by *human* considera-

³While first-strike stability in the given situation might be high, the situation might be unstable in other respects. For example, a leader might launch a very small nuclear attack (less than a “first strike”) or launch an attack on the opponent’s homeland with chemical or conventional weapons. The resulting situation might then be high first-strike instability.

tions⁴—including effects on perceived reality of biases, temperaments, reasoning style, and other situation-dependent behavioral factors—as well as by the quality of information systems.

- It recognizes that people other than national leaders could initiate a first strike.⁵
- It distinguishes between “seeing incentives,” which connotes allegedly rational calculations of positive gain, and “feeling compulsions.”
- It recognizes that more than two nations may be involved.

From this it becomes apparent that *any study of first-strike stability should reflect a theory of human decisionmaking*. Moreover, it becomes important to distinguish between how political leaders *might* behave in real-world nuclear crises, and how they perhaps *should* behave. That is, *it is of interest to develop both descriptive and normative decisionmaking theories*.

TOWARD A THEORY OF RELATED DECISIONMAKING

The Role of Models and Games

The traditional way of developing, assessing, and communicating complex ideas is by building a theory represented by *formal* models rather than by implicit and often unstructured mental models. Some models can be validated by comparisons with empirical information. The internal logic of other models can be validated even though empirical validation is less than decisive. In still other models, even internals cannot be validated because they depend on values, judgments, and images of how the world operates, as well as on measurable facts and rigorous deductive logic. The value of these models is in making variables, relationships, assumptions, values, and judg-

⁴This study does not consider cases in which computers control nuclear weapons, as in some hypothesized strategic defense systems requiring short reaction times and automated control.

⁵Nuclear weapons can, for example, be launched from U.S. SSBNs without depending on the receipt of “go codes” from the National Command Authorities (Brown, 1983), although there are stringent control measures (Cotter, 1987). In time of war there might be more extensive delegation of authority by both Soviet and U.S. leaders, and control of nuclear weapons generally could prove difficult (Bracken, 1983, 1987; Wohlstetter and Brody, 1986). SLBM weapons are not unique in this regard.

ments *explicit*. No external validation is possible, but the models can be very useful nonetheless.

Sometimes, much can be learned by studying the models themselves (e.g., the logic or the equations). In other instances, insight comes from seeing what the models predict under different assumptions. One important technique here involves computer simulations. Such computer models can generate “future histories” describing how the system of interest may unfold over time as the consequence of many interacting phenomena. Often, unaided human intuition is simply unable consistently to understand or predict such dynamics, even though—with the aid of such simulations—the likely or possible dynamics may be readily understood in retrospect. When the factors affecting system dynamics are highly uncertain, as is typically the case, then the simulation model can be used for sensitivity analysis.⁶

A somewhat complementary technique is human gaming. Of particular interest here is political-military gaming of top-level decision-making in crisis. Practitioners and participants of such gaming almost invariably report that the most important intellectual benefit is new insights—e.g., about how different activities interact, about how other people think, about how the “other side” probably thinks, about how commonly discussed strategies often prove either infeasible or dangerous, about how the “real issues” are often not the issues emphasized in analytic studies, and about how important and difficult it is to keep one’s eye on worthy objectives and values.⁷

Just as one might be concerned about using analytic models that cannot be externally validated, so also can one be concerned about whether the lessons learned from such games are mere artifacts. After all, the games depend on the idiosyncracies of the key players, the zeitgeist operative, and the designer’s objective. Schelling’s discussion of this is particularly good (Schelling, 1987, pp. 439-440):

What can one learn . . . [in games]? May not most of the things that one learns be game artifacts unrelated to life? Does the fact that

⁶Especially notable here is the pioneering work of Forrester (e.g., Forrester, 1969) at MIT. Always controversial, Forrester’s work has often been insightful and his methodology successful in bringing assumptions to the surface, where they have been debated vociferously on subjects as diverse as industrial production and the revitalization of cities.

⁷See, for example, Schelling (1987) and Bloomfield (1984) and other articles in the Winter 1984 issue of *Orbis*.

something happened in the game have predictive value for its happening in the world? Is not each game, especially ones that involve much human judgment and imagination and risk taking, a unique story that may never be repeated? . . . Is there not a danger that participants will be so carried away in this vicarious experience that they identify the game too much with the real thing and learn "lessons," perhaps overlearn them, that will prejudice their judgment in the future?

The answer is that games . . . are not different from real experience. Anyone who goes through a Bay of Pigs . . . , a Yom Kippur surprise attack, or a battle over the Falklands has had an enormous learning experience. For some it can amount almost to a rebirth. Each such crisis [however] is unique. Few people ever participate in enough of them to compare them or to get a sense of relative proportions. Whether one experienced the event personally or studied it as a historian, one must beware of generalizing The corresponding danger in games is probably no greater than in real experience—possibly less so because games can be replicated and varied experimentally.

The heuristic value should not be confused with the evidence value of games. Games . . . can produce ideas and insights, but [these] are not validated by having arisen in a realistic game. They are to be validated the way one would validate an insight derived, say, from reading fiction. The fact that something happened in a story doesn't mean that it would happen in a nuclear war, especially if the story was not about nuclear war in the first place. *But new ideas are so hard to come by that one should be ready to take them anywhere one can find them. Whether they are good ideas then depends on whether they stand up to the tests normally applied to new ideas* [emphasis added].

This advice is in close accord with that of historians, who believe that one can benefit greatly from studying history but who deplore the common tendency to learn the wrong lessons and apply them disastrously (Neustadt and May, 1986). In any case, there are good ways and bad ways to use history, games, one's own life experience—and unvalidated models—to try to improve the quality of reasoning in difficult circumstances, such as the crises in which first-strike decisions might be contemplated. Lacking an empirical foundation, the alternative is to stumble along relying upon randomly acquired intuition and informal arguments, which have caused so many problems

in the past and which are likely to cause problems in the future (see Sec. III).

Models of National-Command-Level Decisionmaking

In previous work the author and RAND colleagues have described a new methodology for developing models of national-command-level (NCL) decisionmaking that can be sensitive to both quantitative and qualitative factors and that can reflect many likely concerns of humans in crisis or conflict far better than the more usual mathematically oriented models dwelling on exchange calculations (although such calculations are still relevant and are included within the new methodology's models).⁸ First-generation versions of such models exist in the RAND Strategy Assessment System (RSAS), a large "knowledge-based" simulation for analytic war gaming (Davis, 1986a). *Some of the more important features of this approach to national-command-level modeling include:*

- *A process model of decision* that includes situation assessment; adjustment of assumptions about the opponent, third countries, and laws of war; projection of prospects and risks; objective setting; and selection of strategy (Fig. 1). When embedded in the full RSAS, the decision models can use a larger simulation of likely physical events and other actors' behaviors to "test" potential strategies and—as the feedback arrows in Fig. 1 indicate—can use the results to modify objectives and strategy before acting. The analogy here is to a decision-maker reaching a tentative decision, asking his staff to review it, and then adopting his second-choice strategy because the staff analysis convinces him that his first choice strategy would be either infeasible or disastrous.
- *Various hierarchies of variables*, which allow the model to make ultimate decisions on the basis of relatively few high-level variables, which in turn are determined by one or many layers of lower-level variables. In Fig. 2, for example, the per

⁸See Davis, Bankes, and Kahan (1986) for methodology building on earlier work on escalation theory (Davis and Stan, 1984). The motivation for applying the methodology to problems of deterrence, escalation control, and war termination is discussed in Davis (1987a,b).

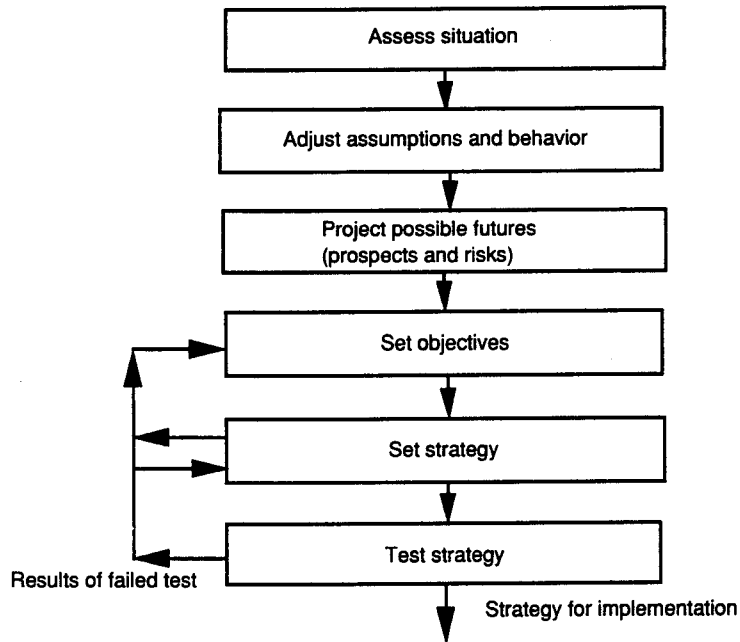


Fig. 1—A process model of decision

ceived likelihood of an opponent first strike is shown to depend on strategic and tactical context variables as well as on various types of warning. These, in turn, depend on more detailed variables, and so on. Real decisionmaking depends heavily on such hierarchies because it requires making tradeoffs, and making tradeoffs is difficult unless the number of top-level variables is small. Comprehensible models also depend on such hierarchical structuring for controlling complexity.

- *Alternative national-command-level (NCL) models* (referred to as Ivans and Sams), which can embody diverse assumptions about the temperament, value structures, and grand strategies of national leaders. Each Ivan or Sam has its own possible images of the opponent and has criteria for deciding which opponent model appears most appropriate.

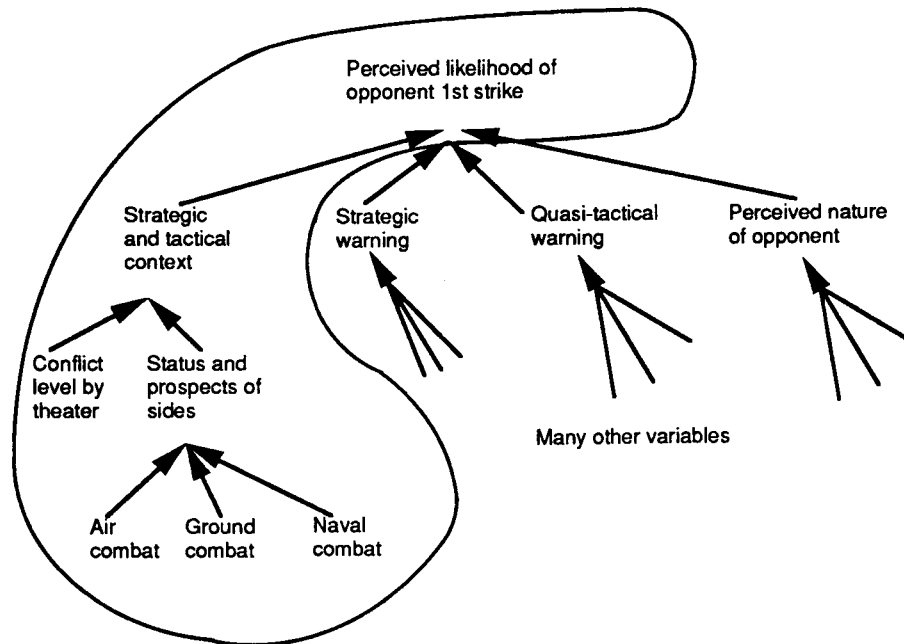


Fig. 2—An illustrative partial hierarchy of variables

- *Context-dependent behavior*, which means that even a given Ivan or Sam may exhibit drastically different behaviors under different circumstances. The emphasis is on developing relatively realistic complex models rather than models exhibiting pure-form behavior, such as stereotyped doctrinal warfighting. Realism must include the potential for degrees of irrationality.

The NCL models are said to be “rule based” because they are comprised of many statements of the form *If <some condition> Then <some conclusion or action>*. These If-Then rules are used for everything from assessment to decision. For example, a rule might be that *If Soviet strategic forces are dispersed, Then U.S. strategic warning is High*. Another rule might be that *If <condition 1 and condition 2 and condition 3 all prevail> Then conduct a first strike*. The rules may

refer to or incorporate mathematical algorithms as in *If, after a simulated first strike by the opponent, the ability to conduct an assured-retaliation attack is Low, Then institute launch-under-attack procedures*. Testing of the *If* clause might require a fairly complicated algorithmic exchange calculation. Collectively, the result of such rules is a “knowledge-based model,” since the rules themselves represent various classes of quantitative and qualitative human knowledge rather than mere recipes for numerical computations.

It is especially important to note that uncertainties about potential reasoning are fundamental rather than something that could be eliminated with further research. Moreover, the models cannot be validated in the sense that one thinks of validating models of physical systems. Hence, a basic element of the approach is to develop and experiment with the alternative models (Ivans and Sams) and even to treat individual rules as variables. The crucial rules are not laws of nature, but rather something similar to propositions or tentative judgments. For some purposes it is useful to make the rules stochastic.

In developing rules for a given model, we draw on: (1) logic; (2) reasoning observed in seminars, debates, essays, and memoirs; (3) the observed behavior of senior individuals in political-military war games intended to be realistic; (4) historical accounts of reasoning in crisis; (5) the considerable empirical and theoretical literature on decisionmaking; and, frankly, (6) our everyday experience with human reasoning. Some of the rules represent conscious reasoning; others attempt to reflect nonvolitional aspects of reasoning and behavior.

*The principal purpose of the models is to provide an integrative and flexible mechanism for making explicit (and clear): issues and related variables, alternative reasoning patterns, key assumptions, and potential consequences of decisions (rather than the elusive “truth”).*⁹ Much can be learned from building, and discussing, key model elements even if the computer programs are never completed and run (much as physicists learn from the structure of equations). If the computer models intended to treat seriously certain classes of issue are com-

⁹As emphasized elsewhere (e.g., Davis, 1986a), the models at issue should not be considered to be “expert systems,” even though the corresponding programs are implemented using related techniques. There are no experts in this domain, and the rulewriting burden is more an exercise in analysis and theory than a recounting of recipes known to work.

pleted, however, we can use them with an imaginative range of initial conditions to generate a plausible range of scenarios; in each scenario the decision models are faced with somewhat different challenges, and in some scenarios the models behave in ways one would not have predicted intuitively. The models sometimes exhibit unrealistic or even silly behavior but other times provide valuable insights. Distinguishing insights from model artifacts depends again on being able to understand the model reasoning and then making one's own judgments about what has the "ring of truth."¹⁰

This study is concerned less with developing and exercising complete models and related computer programs than with using the structures suggested by a general approach to organize thinking and change attitudes about what limits first-strike stability. In subsequent work we may indeed use the models to generate quasi-experimental "data" on plausible scenarios, a technique that we have found quite useful in prior (and as yet unpublished) work on launch-under-attack issues and the effectiveness of flexible-response strategy; in this study we are concerned more with basic concepts.

APPROACH

The remainder of the study proceeds as follows. Section II explores ways in which first strikes might occur by using scenario-spinning methods. Section III draws on the literature to discuss some of the more important underlying issues and to validate concepts proposed in Sec. II. Section IV discusses how all this can be represented in formal knowledge-based models and speculates about alternative classes of decisionmaking behavior. Finally, Sec. V summarizes

¹⁰A common question is whether such models could be used as real-time decision aids to suggest what decisions a policymaker should make in crisis. The answer is no. The models should instead be used analytically—in studies, war games, and other exercises—to provide systematic synthetic experience to help mold intuition and reasoning ability in both peacetime and crisis. To be sure, certain elements of the models (e.g., lists of key variables) could be very helpful checklist aids during crisis. Also, staffs developing materials (what are sometimes called decision aids but may be regarded here as briefing materials) could use such models along with other models and information systems—preferably with due respect paid to uncertainties and differences in opinion about facts and logic. The ultimate decisions, however, would and should be made in person-to-person discussions in which leaders exploit their knowledge and judgment to solve problems and make appraisals. In such deliberations "decision aids" are typically a distraction if permitted at all.

sources of first-strike instability and suggests prescriptive measures to improve it. An important concept in that discussion is the many *dangerous ideas* that national leaders should not entertain seriously during crisis. This has implications for doctrine, education, exercises, and decision aids. Section V also discusses briefly some possible directions for follow-up research and action.

II. REASONS FOR PLAUSIBLE FIRST STRIKES

A SURVEY OF MINISCENARIOS

To understand first-strike issues and reflect them properly in decisionmaking models, we need to understand when first strikes might occur and what reasons would lie behind the hypothetical decisions to enact them. In this section we use a scenario-spinning approach to bring out a range of possibilities. As in all scenario-spinning exercises, the reader will find some of the nine scenarios more realistic than others. The purpose, however, is not to assign subjective probabilities to various scenarios, but to go beyond what seems “likely” and consider deliberately the larger range of the “plausible.”¹

1. Preemption

A national leader is convinced during an ongoing conflict that the opponent is about to launch a first strike. Although he had never previously imagined that he would even consider a first strike, and had always believed that a first strike would be suicidal for both sides, he now finds himself thinking that his nation’s *only* chance for survival—however small—is a first strike that might somehow lessen damage or even avert an attack in ways that cannot really be calculated. It seems obvious to him, despite prior thinking on the matter, that if the outcome of nuclear war is not complete destruction, then going first must surely be better than going second. In desperation, he decides to preempt. After heated discussion, his decision is to launch a full-scale countermilitary attack. The attack includes targets in urban-industrial areas because cities contain communication nodes and important leaders and because of concerns that war might continue for a period of days or weeks: sparing important targets in

¹Scenarios are often used to provide alternative future histories, as in the late Herman Kahn’s work at the Hudson Institute. Here, “scenario” means something more akin to a short story describing the unfolding of a particular situation. The story format makes the issues less abstract and points out interrelationships. This use of scenario spinning is an exercise in divergent thinking: one starts with a particular notion of how a first strike might occur, begins sketching out a scenario, and then finds that other issues and possibilities suggest themselves.

cities might permit the enemy to collect, repair, and support his remaining forces for follow-on attacks.

2. A Fatalistic Nonpreemptive First Strike

A national leader is convinced in crisis that the ultimate war is imminent or has already begun at a low level. He is similarly convinced that the ultimate war will end with a full-scale nuclear exchange against the homelands. There have been some intelligence breakthroughs, and currently his nation's forces are able to locate and destroy a substantial fraction of the other nation's alerted nuclear forces, including those the opponent believes are secure reserves. The opponent is not yet on full nuclear alert and is not yet well postured for launching under attack. The leader decides on a surprise first strike, which he rationalizes in part as preemption, although we shall not consider it preemption in this report. Because he hopes to achieve the surrender of the enemy without a retaliatory strike, he limits his attack to nuclear-threat targets and related military command and control.

3. A Calculated First Strike from Strength

A national leader in the midst of a superpower conventional war is living in a world very different from today's, one in which both sides have substantial but vulnerable space-based strategic defenses. If the opponent should conduct a first strike, the leader's retaliation might be fairly ineffective because of these defenses—except perhaps for a suicidal retaliation focused on cities. The leader considers a first strike against the opponent's defenses, but concludes that the opponent will probably try to launch his strategic forces soon after detecting the attack and that it is therefore wise to attack not only the defenses but the homeland forces as well. In reaching this decision he is strongly affected by the belief that his own defenses will limit damage from a ragged retaliatory attack. He limits his attack to the opponent's nuclear forces.²

²Probably the earliest discussion of this type of defense-related instability is by RAND colleagues James Thomson and Russell Shaver. See Thomson (1987), Wilkening and Watman (1986), and Kent and DeValk (1986), all of whom note that defenses can be either stabilizing or destabilizing depending on circumstances.

4. Well-Intentioned but Inappropriate Unauthorized Use

The commander of a fleet ballistic missile submarine (SSBN) receives word that his homeland is under attack, after which communications fail. He is being trailed by enemy SSNs and is highly concerned about his own survivability. He decides, with the acquiescence of his key officers and crew³—all of whom are subject to the same pressures and perceptions, thereby undercutting some procedural safeguards—to launch his weapons immediately, even though that violates procedures. His weapons have been pretargeted against a range of installations, many of which are in urban-industrial areas. It so happens that the commander's information is faulty, or at least incomplete: his homeland had been the subject of an intendedly "demonstrative" strike causing only very limited damage, and the failure of communications was accidental. When national authorities learn that he has launched his SLBMs, they conclude that the die is cast and order a full-fledged nuclear strike with other forces.

5. Anticipatory Preemption

A national leader learns that one of his field units has just launched an unauthorized nuclear attack on enemy forces in a theater conflict that had previously been conventional.⁴ He concludes that at least one of the nations with which he is at war will probably counter momentarily with a nuclear attack against his homeland, including at least some urban targets. He further expects that this will cause a more widespread attack by all nuclear forces and that it is therefore important to act decisively while he still maintains some modicum of control. Once again, a key argument is that it must surely be better to go first than second. His strike is unconstrained.⁵

³Cotter (1987) describes procedures on U.S. SSBNs.

⁴As documented in Cotter (1987) and emphasized in a recent novel (Aaron, 1987), NATO's nuclear artillery shells and atomic demolitions do not have the permissive action link (PAL) devices that most people believe protect against unauthorized use. It seems likely that some Soviet weapons are similarly protected more by procedures than by physical devices.

⁵Another reason for anticipatory preemption could be the belief that the opponent's leadership is in disarray, unable to authorize war termination, and possibly not in full control of all nuclear weapons.

6. Misinterpretation of Coercive Signals

In the course of a conventional war, one nation launches a “demonstrative” strike against the other’s homeland to reestablish deterrence. However, the nature of the attack (number of weapons, duration, targets) is confusing to the recipient, in part because of technical malfunctions and poor processing of attack-warning information. The opponent interprets the attack apocalyptically and launches a full-scale countermilitary first strike (which he thinks of as a second strike, i.e., an attack against both nuclear and conventional forces).

7. Disproportionate Reaction

As in the previous scenario, a nation is the recipient of a “demonstrative” nuclear attack that kills a significant number of people. The nation’s outraged leader reacts by launching a full-scale “retaliatory” strike (a first strike by the definition used here). Rational *discussion* plays virtually no role in the decision, although the leader is assuredly sane.⁶

8. Disproportionate and Personalized Revenge

A national leader finds that his conventional invasion has turned into an unmitigated disaster. His forces are losing, his allies are realigning themselves, and there are reasons to believe his own life may be in danger from those who would consider overthrowing him. He feels enraged, betrayed, and potentially destroyed. His mind unsound—by the standards of others if not his own—he conducts a first strike out of a sense of generalized revenge. His attack includes cities.⁷

⁶This case may seem implausible to strategic analysts or game theorists, but it is highly consistent with the visceral attitudes of many people, including policymakers. To omit this scenario would be to overrationalize our analysis.

⁷One reviewer of this study objected to this and some of the other scenarios, arguing that it was unlikely that a national leader’s orders would even be obeyed if they were irrational. That sanguine view may well be correct, but irrationality is a matter of degree, madness can be masked, and the principal instinct of military organizations is to follow orders.

9. Assured Destruction of the Enemy

A national leader has already authorized a launch under attack (LUA) posture⁸ and is himself located in an airborne mobile command post. It seems likely that the opponent will strike at any time because the ongoing war has become extraordinarily destructive and war termination appears implausible. The leader is determined that the opponent is to be utterly destroyed if his own homeland is attacked; he has specified a full LUA response, including attacks on urban-industrial targets. The situation is worsening steadily, and communication problems continue to arise from time to time—exacerbated by the enemy's detonation of some nuclear antisatellite systems. Nuclear effects on distant satellites have been larger than anticipated. Military commanders are now concerned about the assured capability to receive tactical warning, transmit it to political authorities for decision, communicate the decision, and execute the LUA. Concerns about electromagnetic pulse effects are whispered about, although the official view is that National Command Authorities (NCA) communications will operate. The decision is made to go first, to assure that the strike will indeed take place (see also Steinbruner, 1984). The leader has no illusions about winning the war or even surviving, but is merely revengefully determined to destroy the enemy as surely as the enemy is likely to destroy his own homeland.

OBSERVATIONS

The scenarios are intendedly interesting in themselves, but their purpose is to help us recognize possibilities and variables so that we can develop a theory of first-strike stability. If one considers all of the scenarios to be utterly implausible, then they don't help much; but if one sees in them phenomena that "ring true," then observations from the mental experiments producing scenarios become something like data in affecting how we think.⁹

⁸The Department of Defense now uses the term prompt retaliatory launch (PRL) in preference to launch under attack.

⁹Political-military war gaming with senior people having a sense for policy issues can be a powerful influence in determining what "rings true" when reviewing scenarios. In some circumstances, such games can be remarkably prescient, although many caveats should be noted (Bloomfield, 1984, pp. 787 ff.).

One observation is appropriate at the outset: in a crisis the natural language of discussion and decision would have little to do with maximizing mathematical utility functions, as traditional decision theorists seem to believe.

Decisionmakers would be doing well indeed if they could achieve a modicum of rationality while sorting out what information to consider and believe.

Let us now consider some of the lessons that might be learned from the scenarios.¹⁰

The Role of Desperation

Although analytic discussions usually work in terms of first-strike *incentives* (a comparatively “positive” value), more commonly we see in the scenarios behaviors born in *desperation* (a comparatively negative value) arising from a multiplicity of factors (Fig. 3). For example, the leaders postulated in both Scenarios 1 and 2 feel *compelled* to go first—i.e., they feel they must do something to avert even greater catastrophes. Their purpose, then, is less to achieve something positive than to mitigate disaster, however slightly.¹¹ In Scenario 1 the result is *preemption*, defined here as a first strike conducted with the belief that a first strike by the opponent is imminent.¹²

As Fig. 4 suggests, desperation can have cascading and self-reinforcing effects on the entire decision process, and on the events themselves. Clearly, any effort to describe decisionmaking analytically

¹⁰In some cases, scenarios were generated by the author’s a priori concern about particular issues; in other cases, however, issues suggested themselves in the course of conducting the *gedanken* experiments.

¹¹Those looking for them will see the traditional first-strike incentives of strategic analysts in Scenarios 1 and 2 (“going first must surely be better than going second,” and “a good exchange ratio is likely,” respectively), but in neither case is the decisionmaker interested in calculational details or crude measures such as the ratio of nuclear weapons.

¹²Going first to avoid a *possible* first strike tomorrow or next week does not count as preemption even if, as in Scenario 2, the decisionmaker uses the term “preemption” as part of his rationalization.

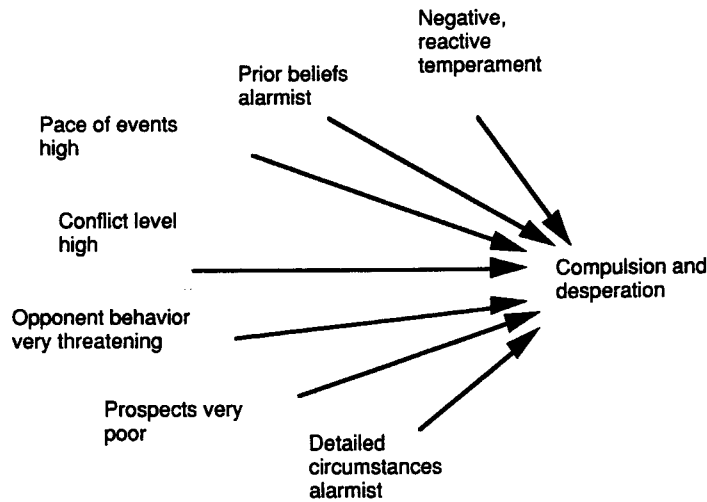


Fig. 3—Illustrative causes of desperation and compulsion

should capture these effects¹³ and serious efforts to improve first-strike stability should seek to mitigate them.¹⁴ Section V discusses some mechanisms for doing so.

The Issue of Rationality

As suggested by Fig. 4, desperation (and other factors) can lead to irrational decisions. Rationality, however, is notoriously difficult to

¹³It is fairly straightforward to do so using the methods discussed in Davis et al. (1986)—at least if the purpose is to *illustrate* influences and plausible behaviors. To be more predictive is virtually impossible, since the reasoning path followed by decisionmakers can be sensitive to details such as who talked to whom last, the decisionmaker's mental and physical health, the sequence of events, and the quality of staff. The author's reading of the history of the Bay-of-Pigs debacle, for example, suggests that President Kennedy's decision to withhold air cover was almost a random event, arising from a late discussion with Senator William Fulbright—not in the context of National Security Council (NSC) deliberations, but during an airplane flight in which the two men talked alone. The arbitrariness of decisionmaking is discussed in "garbage-can theories" (e.g., March and Weissinger-Baylon, 1986, pp. 313 ff.) and in accounts of how false analogies to historical events important to decisionmakers have proven mischievous (e.g., Neustadt and May, 1986, Chap 3).

¹⁴See also Foster and Brewer (1976), which discusses analogous issues in the context of war termination problems.

define. *This study considers a decision to be rational if it is reasonably consistent with objectives—given the information available.* Thus, a decision can be terribly wrong because of poor information (including judgmental “information” about the likely consequences of action) but still rational (this is the case in Scenario 4). A decision can be rather irrational by this definition for any of several reasons, notably:

- The decisionmaker is stark raving mad, unable to think coherently, and behaving erratically (a clear-cut case of irrationality, relatively less important than the following).
- The decisionmaker overlooks some of his own objectives and values and acts in a different way than he would if he were reminded of them. (An example is the leader in Scenario 7, who overreacts to a demonstrative nuclear strike and probably guarantees the destruction of his own country; he probably would not act this way if the decision process were more prolonged and deliberate.)

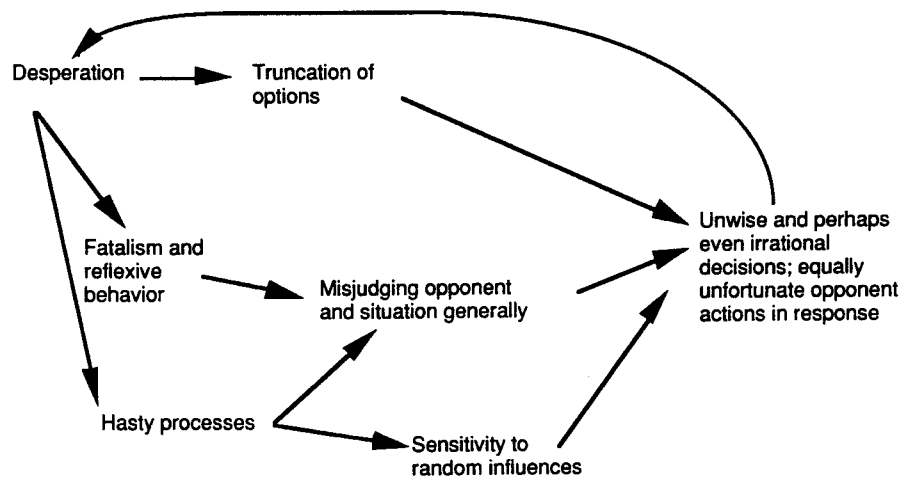


Fig. 4—Cascading implications of decision character

- The decisionmaker does the “calculation” wrong: he knows his objectives and values, but errs badly in estimating what decision would be best (by his own values). This may be regarded as a generalization of the previous problem.¹⁵

In this approach *rationality is a matter of degree and perspective*. Most of Hitler’s abhorrent decisions would be considered rational in this framework because they probably were appropriate to his own values, such as they were. Iran’s Khomeini would also be regarded as potentially rational.

There is more to these issues than pedantic definitions, because even “irrational people” like Hitler and Khomeini can be deterred: the often-heard claim that one can’t do anything about the possibility of irrational behavior is simply wrong. To deter or prevent such behavior may, however, require our understanding of the value structures and reasoning processes of humans we regard as loathsome and “crazy” in the vernacular sense. Most irrational people reason—in their own manner—and to assume they do not is to miss the chance to influence them.

The Importance of Force Structure

Most of the scenarios highlight the role of perceptions and flawed reasoning. However, in at least some scenarios (e.g., Scenarios 2 and 3) objective force structure matters substantially and first strikes may therefore have some benefits. Also, force structure has an implicit impact because intuition and common sense are affected by the background realities (or perceptions about those realities). Although we have lived for decades in a world in which first strikes appear extremely undesirable, different worlds have existed previously and could exist in the future.

¹⁵As discussed in the next section, there are additional complications regarding what constitutes a “correct” tradeoff across values. We should not assume that maximizing the expected value of a utility function is “right.”

The Importance of Special Factors

In contrast with peacetime decisions, nuclear crisis decisions could depend sensitively on such special factors as human-source information (the bug in the Politburo's conference room), supersecret detection and tracking methods (e.g., against SSBNs), faith in one's own tactics, or (on the other side of the coin) a side's knowledge of its own secret vulnerabilities (e.g., knowledge that an entire leg of the Triad was currently unreliable because of a recently discovered fusing problem). *We should not assume that conclusions emerging from peacetime policy-analytic studies are the conclusions that would pertain if a crisis arose—even if the ink on such studies were still wet.*¹⁶

The Importance of Command Level and Operational Milieu

What a decisionmaker would consider rational would clearly depend upon his command level and, less obviously perhaps, on what he was "doing." Consider the perspective of the SSBN crew in Scenario 4, in which "doing their job" corresponds to assured retaliation (and not, for example, war termination or damage limitation). Or consider a national leader who has authorized a launch-under-attack posture while boarding an airborne command post. Would such a leader have a broad view in weighing alternatives, or would the setting of the hair trigger now affect his priorities? Would "the" objective now be assured retaliation—so much so that "preemption" might be considered? In principle, one could also imagine that top-level military leaders exposed to hours of countermilitary exchange calculations might see issues in terms of force ratios and ability to meet targeting requirements.¹⁷ Or, as a last case, consider a Soviet commander of the Strategic Rocket Forces. Would it be surprising if the "use it or lose it" cliché actually described his thinking? Such issues could be sig-

¹⁶Special factors mattered in the battle of Midway (U.S. knowledge of Japanese codes), the Six-Day War (the Israeli preemptive air strikes on Egyptian airbases that exploited the minimum-altitude features of Egyptian radars), and World War II (German circumvention of the Maginot line).

¹⁷The phrase "in principle" is used here because senior military figures are often much less myopic about such matters than strategic analysts—in part, perhaps, because they are old enough to remember when the "requirements" were much smaller but the consequences of execution were still deemed to be cataclysmic.

nificant in contemplating and defining options for conditional delegation.

Ladders of Objectives

Although the word “objective” does not often appear in the scenarios, decisionmakers’ objectives can be inferred. Some of them are drastically different from those we associate with usual national security objectives. It is interesting to infer from the scenarios several notional ladders of objectives (e.g., those in Fig. 5). In these ladders higher rungs entail objectives requiring increasingly greater destruction of the opponent’s homeland. Which ladder applies depends on the particular decisionmaker involved, certain basic aspects of the force postures (e.g., is a disarming first strike feasible?), and level of action (e.g., submarine commander versus NCA). The leftmost ladder appears to be the most relevant to today’s world and today’s decisionmakers, but a decisionmaker who simply did not believe that meaningful damage limitation were possible might follow the center ladder. The rightmost ladder corresponds to a leader willing to entertain seriously coercive use of strategic nuclear weapons. If the world were different and first strikes were feasible, then the rightmost ladder might well apply. Other ladders are possible, and, as with Kahn’s escalation ladders, the metaphor is not supposed to imply that all rungs of the ladder must be climbed, or that climbing is at a constant rate. As the leftmost ladder in Fig. 6 suggests, the change of objectives over time might be quite different for different decisionmakers (A, B, C, and D). Moreover, leaders might hop from considering options on one ladder to considering options on another. Figure 7 illustrates the case of a decisionmaker who starts and ends believing in only crude alternatives (center ladder), but flirts along the way with damage limitation options and even a coercive first-strike option. The point here is that decisionmakers under stress would probably be torn not just among objectives within a given conceptual framework but among alternative frameworks.

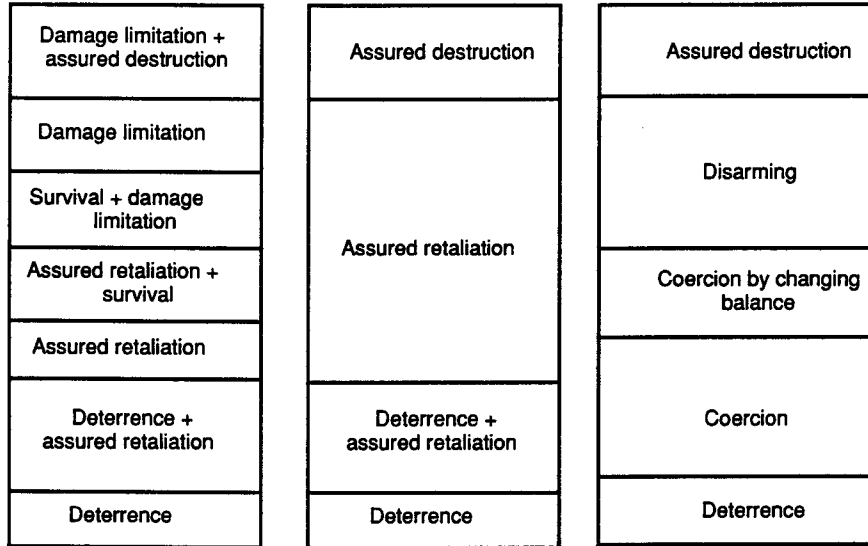
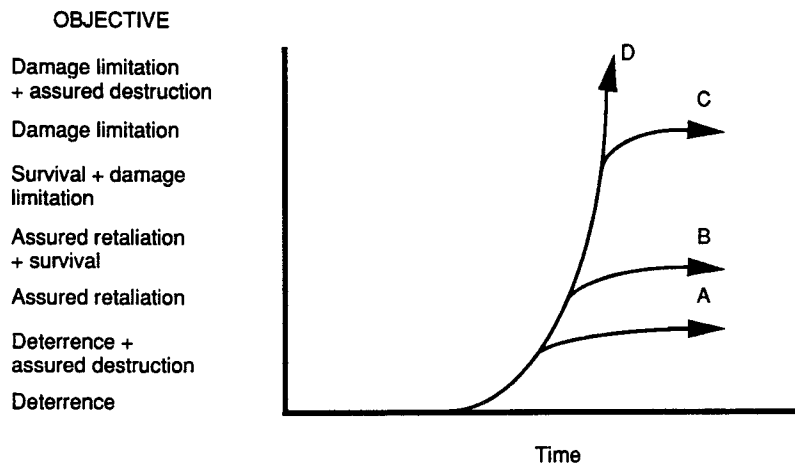


Fig. 5—Alternative ladders of objectives



Note: A,B,C, and D are alternative decisionmakers or circumstances.

Fig. 6—Illustrative changes of objectives as crises deepen

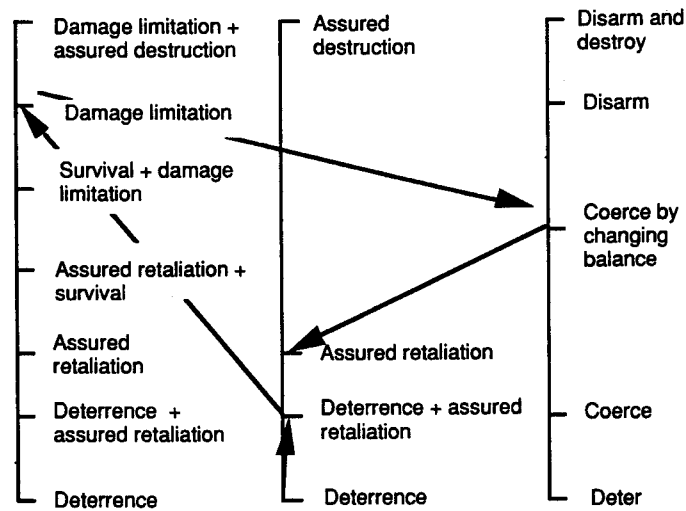


Fig. 7—Illustrative shift in perceptions about possible objectives

Other Observations

A few other observations stimulated by the scenarios are worth making at this point:

- In many circumstances that promote a first strike, there would be strong pressures to attack some targets in urban-industrial areas (e.g., submarine ports, bomber bases, the Moscow antiballistic missile (ABM) system, communication nodes, and control-related installations).
- In the most desperate circumstances, decisionmakers might well grasp at straws, focusing on the most hopeful of various unlikely ways to salvage or mitigate the situation (e.g., by a preemptive first strike that "might possibly" have more effect than one could confidently predict).
- At the same time, decisionmakers might dwell on risks when considering their ability to ride out attacks, risks that had previously been considered small and that perhaps would be small. This tendency might be exacerbated if prediction after prediction had failed on other matters. (How, after all, had the world found itself on the nuclear brink in the first place!)

Almost surely, there would have been system failures, communication failures, and obvious mistakes of judgment undercutting any confidence in predictions and analysis.¹⁸

HEURISTIC CLASSIFICATION: THE FAULT TREE CONCEPT

*An especially useful way to conceptualize the potential causes of a first strike is to construct "fault trees."*¹⁹ Let us consider a first strike to be a disaster for the world and lay out visually some paths by which this disaster might occur. Figure 8 shows a simplified fault tree that draws inspiration from some of the scenarios.²⁰ The words represent in capsule form what the decisionmaker may be thinking at every point (correctly or not, and as the result of rational or irrational processes).²¹ Figure 8 does not include decision paths dominated by extreme types of irrationality, but it gives a fairly insightful picture of some of the reasoning postulated in the other scenarios. For example, looking at the leftmost side of the tree, we see a path involving a decision to preempt. How might such a decision be reached? According to the figure, the decision would be the result of concluding that the opponent was very likely to be going first and, explicitly or implicitly, concluding that the cost of going second rather than first would be high—i.e., that going first would have value. These conclusions, in turn, might be based on a conclusion that only by going first would there be any chance, however small, of national survival (see discussion in Scenario 1). *Reviewing the scenarios by using fault-trees iden-*

¹⁸The likelihood of such problems *early* in crisis also raises questions about the plausibility of national leaders conducting limited nuclear strikes based on estimates that damage and perceptions of that damage could be strictly controlled.

¹⁹Other classification schemes are useful for other purposes. For example, we could classify the various miniscenarios in terms of: whether the first strike was authorized (and whether, if authorized, the decision was made by the NCA or someone to whom conditional launch authority had been delegated); whether the attack was considered by the decisionmaker to be preemptive; whether the decisionmaker was in desperate or quasi-desperate circumstances; whether he was rational (broadly construed); whether the wisdom of the decision was information-limited; and whether the wisdom of the decision was process-limited.

²⁰"Decapitation" in Fig. 8 really means "decapitation that effectively disarms or greatly reduces effectiveness."

²¹This type of fault tree displays what might be regarded as possible "cognitive maps." See Axelrod (1976).

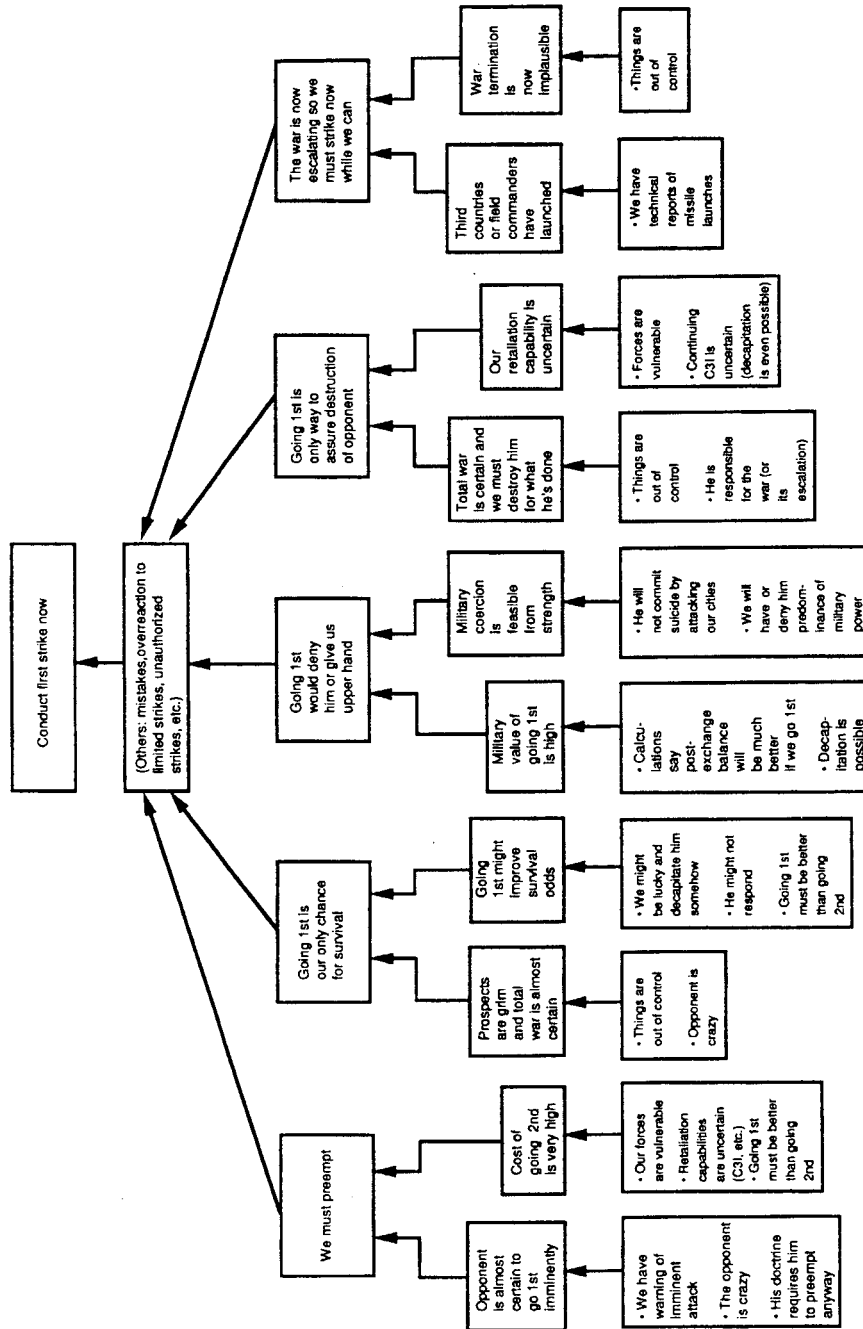


Fig. 8—A partial fault tree of first-strike decisions

tifies many “dangerous ideas” that we might hope decisionmakers would never hold. We shall summarize such dangerous ideas in Sec. V when we consider ways to improve first-strike stability.

KEY VARIABLES

As our final effort to milk the scenarios of insight, let us identify what appear to be the key variables affecting decisions to conduct a first strike. By “variables” we mean factors that might appear in a formal rule-based model of decisionmaking.

The first two variables that come to mind are nonvolitional and play no role in traditional discussions of first-strike stability; they would appear in a descriptive model of decisionmaking but would not be consciously recognized by the decisionmakers. *The most important nonvolitional variables are: (a) the particular decisionmaker; and (b) the decisionmaker’s “mindset.”* Any decision surely depends on which Ivan or Sam we are talking about, because Ivans and Sams have different temperaments, value structures, and grand strategies. The decision also depends on psychological mindset, by which we mean to distinguish, say, state of mind that might be called fatalistic-reflexive from a state of mind that might be called controlled-analytic (see Davis, Bankes, and Kahan, 1986).

The remaining variables that seem especially important would be recognized by the decisionmaker, explicitly or implicitly. Some of these volitional variables may at first glance appear to be the same as those in a traditional quantitative assessment of first-strike stability by strategic analysts. As we proceed, however, it should become clear that their meaning to decisionmakers is likely to be quite different from their meaning to analysts. Also, the way decisionmakers would evaluate the variables might well be highly nonanalytic. *The volitional variables most important to a first-strike decision appear to be:*

- *Context* (current situation and its origin, e.g., level of conflict in each theater, history of opponent behavior, gross status of forces and command, clarity of stakes and opponent intentions, and degree of both domestic political support and allied cooperation)

- *Prospects*, assuming current strategy or the best of strategies not involving a first strike
- *Likelihood of an opponent first strike* (a contributor also to “prospects”)
- *Survival value of a first strike* (i.e., survival value of going first rather than second, something determined primarily by damage-limitation capability)
- *Assured-destruction value of a first strike* (i.e., value for assuring the destruction of the opponent of going first rather than second)
- *Military value of a first strike* (i.e., value for achieving a substantial and usable military coercion capability and/or denying the opponent’s war aims and perhaps achieving one’s own war aims) of going first rather than second.

The last three variables (the values of a first strike) should be thought of as the result of calculations, however qualitative and however well or poorly made. In each, the issue is how much better off one would be by going first rather than second given that the opponent will try to retaliate. In some instances it is more appropriate to invert the terminology by referring to the “costs of going second”²² rather than the value of going first.

The way these volitional variables would be evaluated (whether intuitively or formally) would vary with the nonvolitional variables mentioned above: context, decisionmaker, and mindset. As a simplification consistent with the limited ability of people to consider multiple cases:^{23,24}

²²The definition here of the “cost of going second” is different from that of Wilkening and Watman (1986) and Kent and DeValk (1986). Here the cost is the *difference* in outcomes between going first and second.

²³In theory, the calculations should be based on integrals over all the possible cases using probability functions. In practice, just recognizing that one should perhaps think consciously about bounding cases would be a useful complication for many decisionmakers.

²⁴RAND colleagues Paul Hill and Robert McCalla have discussed in recent unpublished work some of the important traits of typical U.S. Presidents and the implications of those traits for decisionmaking about nuclear use. Hill and McCalla believe Presidents: would rely heavily on long-time advisers; would *not* be unduly impressed by purely military or technical arguments; would consult with and try to co-opt political opposition so as to share blame; would be especially concerned about survival of the political system defined by our Constitution; and would use nuclear weapons only if the onus for doing so could be squarely placed on the opponent. The

To each of the volitional and nonvolitional variables three possible "calculations" correspond, producing expected cases, best cases, and worst cases.

Here the "best-case" and "worst-case" calculations use, respectively, highly favorable and highly unfavorable assumptions for the side contemplating first strike.

Figure 9 provides an influence diagram of the problem, highlighting both the volitional and nonvolitional variables. As indicated schematically, mindset affects both the evaluation of the several first-strike values and the relative weighting of survival, military, and assured-destruction values. We shall return to these matters in Sec. IV and demonstrate how they can be treated in formal models.²⁵

Command-Control Arrangements

An important volitional variable affecting all of the above variables is the arrangement that has been made for enduring political and military command and control. At one extreme there might be no arrangement, with the national leader maintaining total personal control. At the other extreme a robust system might ensure expeditious devolution of both political and military authority in the event that some leaders are killed or otherwise unable to perform. Under some circumstances a leader might also conditionally delegate launch authority to a military commander (e.g., he might authorize prompt retaliatory launch). Arrangements increasing the certainty of retaliatory capability would reduce compulsions to go first by reducing the consequences of a decapitating attack on the political leadership—i.e., they would reduce the "military and retaliatory values of going first," to use the nomenclature of this study. Some such arrangements, however, would introduce significant risks of their own, including (1) the possibility that military authorities would not in the midst of

current report does not treat these points explicitly, but they are quite consistent with the report's approach and the character of the principal Sam and Ivan models that have been built so far using the methods described in Davis, Bankes, and Kahan (1986).

²⁵Figure 9 assumes a narrow definition of mindset in which it is not affected by prospects, value of first strike, and so on. In a more general sense, of course, mindset is affected by everything one feels and knows or thinks one knows.

war relinquish control of weapons once they had it and (2) the risk that military authorities might make faulty first-strike decisions on the basis of different information from that available to the political authorities (including human advice regarding opponent motives).²⁶

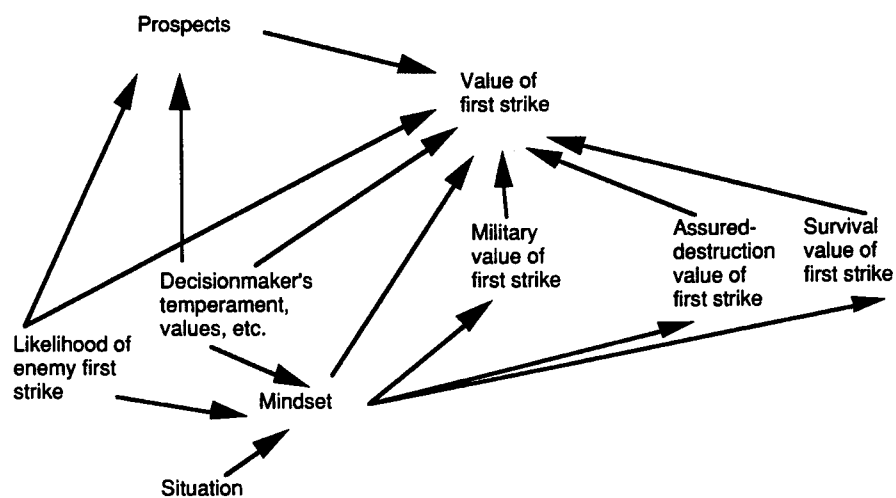


Fig. 9—Key variables in a first-strike decision
(using a narrow definition of mindset)

²⁶For a discussion of command and control issues, see Slocombe (1982); Bracken (1983, 1987); and Gottfried and Blair (1988). See also Davis, Stan, and Bennett (1983) for an early discussion of how such issues might be addressed with knowledge-based simulations.

III. BACKGROUND CONSIDERATIONS

Having used the scenario-spinning method to develop an intuitive sense for the variables affecting first-strike decisions, it is now appropriate to step back somewhat and discuss some of the underlying issues in more detail. What follows discusses in turn: possible reasons for not improving first-strike stability; what the empirical and theoretical literature tells us about relevant decisionmaking problems; the complications arising when one tries to define "good" decision in nuclear war; and some special issues raised by strategic defenses.

POSSIBLE CONTRADICTIONS BETWEEN STABILITY AND DETERRENCE

Deterrence and Extended Deterrence

First-strike stability is only one of a number of national objectives, and there are tensions among the objectives. In particular, there are tensions between first-strike stability and deterrence. To make the point, consider an example of pure logic. A two-sided world exists in which Side A has no forces, an aggressive Side B has many forces, and Side A therefore must do B's bidding. Neither has incentives or compulsions for a first strike, but the situation is undesirable for A because of B's coercive capability. Next, suppose that A contains a modicum of nuclear forces—not so many as B, but enough to threaten B's survival in the event that B were to attack A. We might then say that A had a minimum deterrent capability, which might or might not be credible. B's leaders think that A would use its nuclear weapons to attack B's cities if B attacked with either nuclear or conventional forces, even though that might mean A's own suicide; therefore the deterrent is effective in preventing such an attack or invasion. But it is effective precisely *because of first-strike instability*.

In reality, the tension between first-strike stability and deterrence is far less drastic than in this hypothetical case—because both sides currently have assured-retaliation capability that includes the capacity not only to devastate the opponent's cities, but instead or in addition to devastate the opponent's general-purpose forces and prevent

the opponent from achieving useful war aims. Although this could change, there is little reason to believe that it will change in the foreseeable future. As a result, those who see tension between deterrence and first-strike stability are usually concerned primarily with our ability to maintain *extended deterrence*, i.e., the ability to use the threat of nuclear strikes on the opponent's homeland to deter the opponent from highly provocative acts, such as invading close allies, or to help bring about an acceptable termination of war if deterrence fails initially.¹

*Specifically the United States and its NATO allies still depend on the concept of extended deterrence and the related strategy of flexible response.*² NATO's flexible-response strategy seems to deter Soviet aggression through having the capability for initial conventional defense, options for controlled nuclear escalation in Europe (perhaps extended to the Soviet Union), and a credible escalatory linkage to U.S. strategic nuclear forces. The concept of nuclear escalation to reestablish deterrence is explicit.

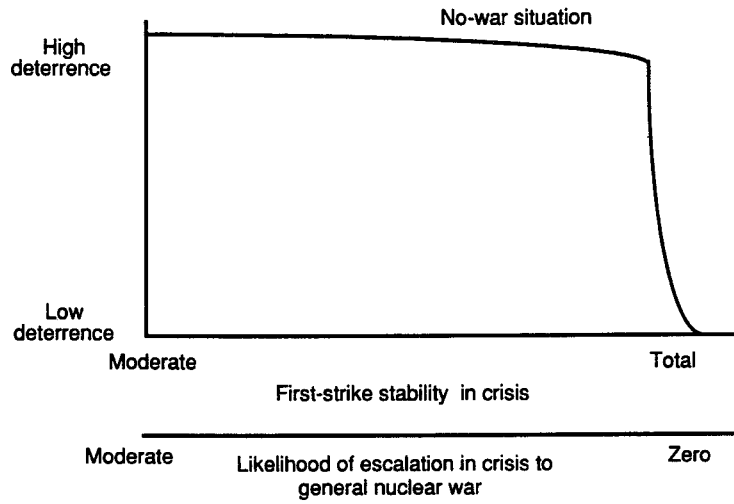
More generally, the United States has a worldwide strategic problem. In no area of likely conflict with the Soviet Union do the United States and her allies have an offensive conventional capability adequate to deter Soviet aggression outside NATO countries. And, if the Pact invaded NATO, the United States could look long and hard before finding a strategically decisive area in which to open a new front, although options for some degree of horizontal escalation have long been discussed. To some strategists, then, it appears that the United States should continue to maintain as much extended deterrence as possible.

The so-called Devil's Dilemma is this: if first-strike stability increased (in which case a U.S. first strike would be less plausible to the Soviet Union), the probability of conventional war in Europe and elsewhere would probably increase (Fig. 10a)—unless conventional deterrence were strongly improved, which many enthusiasts of strate-

¹The classic discussion of different types of deterrence and correspondingly different objectives for nuclear forces is that of Herman Kahn, who referred to what this study calls extended deterrence as Type-II deterrence (e.g., Kahn, 1960, p. 126). Another influential discussion of such matters is Nitze (1976), which is concerned more with averting coercive pressures than with deterring overt invasions or attacks. For a recent survey of issues and the strategic literature, see Cimbala (1988).

²For a definitive discussion of how flexible-response strategy arose and what it means, see Legge (1983), McNamara (1985), and Bundy et al. (1985).

a. Quality of extended deterrence (likelihood that war can be averted because sides fear escalation)



b. Quality of extended deterrence (likelihood that war can be averted because sides fear escalation or, if war has begun, that war can be terminated satisfactorily for the same reason)

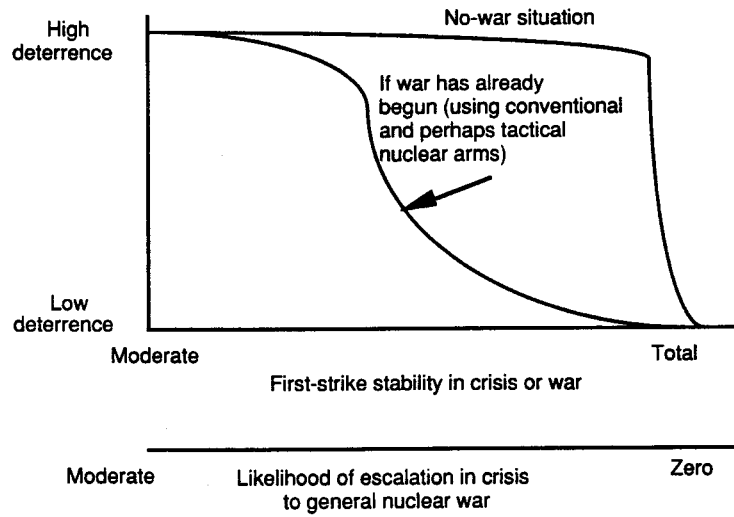


Fig. 10—The Devil's Dilemma tradeoff

gic arms regard as implausible. Study suggests that the NATO conventional balance is less adverse and more readily improvable than usually assumed (Davis, 1986b, 1988a), but uncertainties are legion and there is no substitute for the threat of nuclear war in turning such uncertainties into deterrence. As a number of commentators have noted, one thing we know from history is that conventional deterrence doesn't work.

In fact, the Devil's Dilemma is not so straightforward as it may appear (see also App. A). Extended deterrence has two aspects: general or "existential deterrence," which has long caused the superpowers to avoid any confrontations that might lead to war and possibly to nuclear war; and extended deterrence that might apply after war had already begun at a nonnuclear level (i.e., could the threat of strategic nuclear war bring about the resolution of conflict once begun?).³ Figure 10 suggests how different the tradeoffs might be for cases in which large-scale conventional war had or had not already begun. These figures obviously reflect the author's subjective judgment that:

- Only a modest amount of inherent first-strike instability may be sufficient to provide the extended deterrent function of deterring war generally, especially if the perceived instability is at least as great as the actual instability, and preferably much larger.
- By contrast, the amount of perceived instability necessary to reestablish deterrence once war begins—i.e., to bring about a satisfactory termination of war—might be much higher, and might require a level of actual instability that should be unacceptable because of the catastrophic consequences of full-scale nuclear war.

The question now arises as to whether one can distinguish among different classes of improvement measures. That, is can one find measures that would improve actual first-strike stability in crisis but that would not greatly reduce perceived first-strike stability in peacetime? Such measures would not decrease extended deterrence in situations short of an already ongoing war. The answer is that, in fact,

³This distinction has been made recently by Lawrence Freedman in unpublished work.

many improvement measures meet this criterion (e.g., improving the survivability of one's own forces and command-control system and educating leaders about mistakes of reason to avoid in crisis).

One-Sided First-Strike Stability

Consider another argument against first-strike stability: the ideal U.S. objective would be one-sided first-strike stability, in which Soviet leaders fear that NATO or the United States might escalate to nuclear use and possibly even a first strike, but the Soviet leaders neither see incentives nor feel compulsions to conduct a first strike themselves. This might apply, for example, if the United States had survivable and accurate strategic forces and C³I while the Soviets had vulnerable and somewhat unreliable strategic forces and C³I. As discussed in App. A, however, *there are two important implications of one-sided first-strike stability as an objective:*

- The result is a zero-sum game that the opposing superpower would have to play. That is, *the objective almost guarantees arms-race instability* as one or both sides seek unilateral advantage.
- Since the currency of discussion about first-strike stability tends to be exchange calculations, *the effect of the objective is to emphasize counterforce issues*, notably missile duels which have already been overemphasized to the detriment of public policy.

Appendix A discusses these and other arguments against improvement of first-strike stability. The arguments involve successfully competing in the realm of strategic arms, the desirability of certain weapon systems often labeled as destabilizing, and arms-control verification. The author's conclusion is that the arguments are overdrawn and that improving first-strike stability should be an important objective of the superpowers.

BUILT-IN SHORTCOMINGS OF HUMAN ABILITY TO REASON

Threshold Effects and Nonrational Reasoning

Section II speculated on the basis of scenarios that perfectly sane decisionmakers might behave in very nonrational ways. In fact, these speculations are confirmed by cognitive science.

The empirical literature is rich in evidence (e.g., Kahneman, Slovic, and Tverski, 1982) that humans systematically fail to reason in ways a decision theorist would regard as rational-analytic. In some circumstances, people consciously overlook possibilities they consider improbable—even though the consequences of overlooking them are extremely negative. In other circumstances people rule out options because of a perceived risk, even though those options might have a considerably higher expected return than the options they choose.⁴ In yet other circumstances, they pursue low-probability high-payoff options, even knowing that the expected payoff is poorer than with a more conservative choice. Several phenomena are involved here: (a) people are often unaware that their decision process does not maximize expected return; (b) even if they are informed of this flaw, they may cling to their choices because they don't really believe the alternative analysis; and (c) sometimes, even when they understand and believe the alternative analysis to some extent, they may hold to their choices in ways that suggest they are not trying to maximize expected value (e.g., they may be minimizing risk or maximizing the likelihood of a win above some threshold of value). This last phenomenon may or may not pass the test of being rational.⁵

⁴It can be argued that the extreme risk aversion the superpowers have demonstrated in the nuclear era is an example of this phenomenon. For example, it can be argued that President Kennedy would have been rational in taking a harder line during the Cuban Missile Crisis, since it was exceedingly unlikely that the Soviet Union would go to nuclear war over troubles a hemisphere away from them that did not endanger their more fundamental national interests. On the other hand . . .

⁵A related phenomenon is that experts are notorious for underestimating the probability of low-probability but highly adverse events. We have all learned to distrust such estimates, which often makes us more risk-averse than might seem superficially "rational." (Most people are quite happy, however, with policymakers who are extraordinarily risk-averse with respect to nuclear options.) At the same time, it should be mentioned that innovators regularly discount the low probability of success ascribed to their ideas by "experts." Thus, distrusting estimates applies to both ends of the spectrum.

We can see all of the above phenomena in our miniscenarios. For example, in Scenario 1 the decisionmaker finds himself focused on trying to maximize the likelihood of national survival; he is convinced (without careful analysis) that his “only chance” is to conduct a first strike, even though the odds of that strategy succeeding are small.

Another fascinating aspect of behavior is the selection of options that happen to be suggested at the time rather than of other options previously considered but currently out of mind. See for example, March and Weissinger-Baylon (1986), which includes March’s “garbage-can theory,” several accounts of real-world military operations in crisis, and a discussion of how staff processes mitigate some aspects of nonrational thinking.

Historical Evidence

If we turn to history, we can see some of the same problems in actual events. An example of threshold effects is the German decision to initiate the U-boat campaign against U.S. shipping in 1916. The Germans had correctly calculated that the United States could not have a very important impact on the war *in the next five months or so*, but to quote the German Secretary of State’s testimony after the war, “The possibility that the war could last two years . . . was not being considered seriously by anyone at that time [1916].” They were certain that England would instead sue for peace within months (see Ikle, 1971, p. 45).

One of the most remarkable historical sources on flawed decision-making is the record of deliberations by Japanese leaders prior to and immediately after the decision to authorize the surprise attack on Pearl Harbor (Ike, 1967). That record provides striking examples of how decisionmakers came to be locked within specific perspectives (in particular a view that war was inevitable because the United States and Great Britain were encircling Japan strategically); how in conflict with modern concepts of deterrence they decided to attack powers they knew to be much stronger than themselves, even though they acknowledged the possibility (or even likelihood) of a long war; how they were psychologically unable to discuss the likelihood of defeat or its probable consequences; and how fatalism overcame rational think-

ing.⁶ To illustrate fatalism, consider this conversation between Admiral Yamamoto Isoroku, who developed the idea of a surprise attack, and the Navy Chief of Staff, who was expressing hesitations. Isoroku reportedly remarked:

The only question that remains is the blessing of Heaven. If we have Heaven's blessing, there will be no doubt of success.⁷

This overt expression of (optimistic) fatalism perhaps explains Japan's remarkable formal discussion of war prospects. At the September 6, 1941, Imperial Conference, material was presented to answer the tough questions. From Ike (1967, p. 151):

What is the outlook in a war with Great Britain and the United States; particularly, how shall we end the war? (Answer:) A war. . . will be long, and will become a war of endurance. It is very difficult to predict the termination, and it would be well-nigh impossible to expect the surrender of the United States. However, we cannot exclude the possibility that the war may end because of a great change in American public opinion, which may result from such factors as the remarkable success of our military operations. . . . At any rate, we should be able to establish an invincible position: by building up a strategically advantageous position through the occupation . . . ; by creating an economy that will be self-sufficient . . . ; and by linking Asia and Europe . . . through our cooperation with Germany and Italy. Meanwhile, we may hope that we will be able to influence the trend of affairs and bring the war to an end.

Clearly, the Japanese were not making conservative cost-benefit calculations.

The Role of Perceptions

An especially serious facet of the reasoning problem is the role of perceptions. Ultimately, decisions depend entirely on perceptions,

⁶The introduction in Ike (1967) summarizes many of these phenomena, but some of the most relevant passages are on p. 152 (a document recording the inevitability argument), p. 201 (in which the Navy Chief of Staff explains why it was necessary to initiate the war then rather than later—despite the expressed views of others that war might not be inevitable and that risks were high), and p. 207 in which the Army Chief of Staff (Tsukada) summarizes the issues as he sees them—noting that “prospects if we go to war are not bright.”

⁷Ike (1967, p. xxvi).

which may or may not be in accord with reality. There is a considerable literature on how decisionmakers' perceptions can be badly flawed—not merely by virtue of bad luck or bad information, but by virtue of inborn mental heuristic processes that are crucial to our everyday performance but sometimes dangerous. Of particular importance here are such behavioral tendencies, well documented by historical incidents and to some extent demonstrable in controlled experiments, as: (1) forcing information into the mold of preexisting mental models; (2) assuming that opponents see the world in the same way as oneself and interpret one's own motives and preferences as intended; and (3) assuming that problems are the result of malevolent intentions and actions by the opponent.⁸ It is notable that additional information often does not have the impact that it should, especially when the root of the misperception is the images or attitudes of the decisionmakers rather than the recognized ambiguity or incompleteness of information (McCalla, 1987).⁹

Effects of Stress

All of the problems discussed in the above paragraphs can be exacerbated (or in some instances lessened) by crisis-induced stress. There is a considerable literature on this subject too, much of it reviewed in George (1986). Particularly relevant to this study is the following passage from that work (p. 538):

Another type of defensive avoidance, *bolstering*, may be resorted to when a decision is difficult to make, when it threatens great loss, or when it cannot be put off because external pressures or a strict deadline demands action. Under these circumstances the policymaker may make the task of choosing an option easier for himself by re-evaluating the courses of action before him, increasing in his mind the attractiveness of one option (which he will then accept) and doing the opposite for competing options (which he will then reject).

⁸Jervis (1968) is an excellent early example of this literature. See also Jervis (1976), Axelrod (1976), and McCalla (1987), the latter being a PhD dissertation with an extensive bibliography.

⁹See Kahan, Darilek, Graubard, and Brown (1983) for a more extensive survey of the literature relevant to the use of behavioral science in the study of war prevention. Lorell and Brown (1985) observe that we should not assume that nuclear war would be the result of misperceptions in the usual sense, or even that it would be inadvertent. It discusses alternative paths to war.

“Bolstering” is closely related to the grasping-at-straws phenomenon postulated in Sec. II. Later, in Sec. IV, we shall see how such phenomena can be reflected in formal rule-based models.

Although reproducing the conditions of interest in academic experiments is quite difficult (student subjects are not policymakers, the problems posed in experiments do not involve nuclear war, and so on), there is historical evidence on stress-related problems of top-level decisionmaking.¹⁰ George states that two important members of Kennedy’s inner circle were unable to cope during the Cuban Missile Crisis, becoming passive and unable to fulfill their responsibilities. The problems of Stalin during the early period after Hitler’s invasion and of Dayan during the early hours of the Yom Kippur War are two among the large number of available examples. Such problems are not restricted to individuals. As discussed in Janis’ work on “groupthink” (e.g., Janis and Mann, 1977), cohesive groups operating in crisis are subject to problems such as (1) global or undifferentiated thinking (simplistic views of the world and world actors), (2) a lost sense of proportion, and (3) a confusion of means with ends. While groups may pull together during crisis, their performance may suffer precisely because of the simplifications and reduced argumentation that is part of doing so.

BOUNDED RATIONALITY

Even if we put aside issues of misperception, stress, and man’s imperfect reasoning apparatus, high-level (“strategic”) decisions seldom have much to do with maximizing some utility function, whether explicitly or implicitly. While the techniques of decision analysis for formulating options, utility functions, subjective weights, and the like are undeniably useful in many circumstances, they are seldom useful to decisionmakers faced with *strategic* choices (“strategic” understood here in the most general sense of the word) and who are typically—not merely occasionally—faced with extraordinary and manifest uncertainties in many dimensions. Even if they accept the tenets of rational-analytic decisionmaking at some theoretical level, they may

¹⁰Not surprisingly, however, researchers interpret the historical evidence quite differently. See Blight (1987, p. 26) for a discussion of this with respect to the work of Janis and Lebow.

find it impossible to apply the techniques in practice because uncertainties dominate the problem and efforts to turn all the issues into probability distributions, utility functions, and the like obfuscate rather than assist.¹¹

Consider some of the uncertainties connected with entertaining the idea of a first strike:

- How likely is it that the opponent will go first?
- If he goes first, what will he want to attack? What can he find?
- If we go first instead, how likely is it that the opponent will launch under attack? What will his targets of retaliation be?
- How likely is it that the various weapon systems and command and control systems will even work? (This might include strategic defenses, both in space and on the ground.)
- Is there a realistic chance for de-escalation? Can we trust anything the opponent leader says, and is he even control?

Given such significant uncertainties, there would be little value in an elegant calculation of “expected value” or how to maximize it, even if the calculations could be performed.

It is true that game theory approaches can beautifully illuminate certain aspects of plausible nuclear crises (see, for example, Brams, 1985) and that decision analysis can more generally be used effectively, so long as the problem is simplified sufficiently and the analysis uses variables that are adequate surrogates for the larger issues at stake. The problem, however, is that the dilemmas are so multifaceted and the usual surrogates so extremely simplified as to obscure the real issues. So, for example, equating first-strike stability with the results of standard exchange calculations obfuscates most of the issues raised in Sec. II.¹²

¹¹See especially Simon (1980) for a highly readable account and Simon (1982) for technical papers on Simon’s Nobel-Prize-winning work on bounded rationality.

¹²In fairness, it should be noted that most discussions of first-strike stability using mathematical exchange calculations are ultimately concerned about the relative wisdom of alternative defense programs or arms control agreements; these calculations can be very revealing (e.g., see Kent and DeValk, 1986; and Wilkening and Watman, 1986). The author has used similar techniques many times over the years. His quarrel is really with those who claim that the calculations tell us much about likely decisionmaking behavior in a real-world crisis.

EFFECTS OF DECLARATORY POLICY AND DOCTRINE

The next problem is more insidious and involves interrelationships between declaratory and actual policies and strategies. To refer back to an earlier problem, the United States and its NATO allies depend significantly (whether wisely or not) on a deterrent strategy premised on the threat of nuclear escalation. For the strategy to succeed it must be credible. For it to be credible the Soviet leaders should presumably see the West practicing and internalizing military and political doctrines that plan for nuclear use under the circumstances at issue. But in the real world, where deception involving large organizations is very difficult, the West must actually *have* related doctrines, forces, and weapon systems for the Soviet Union to be seeing such things. And if it actually has them, they will affect *real* mindsets. Can one really expect that policy and doctrine (which is virtually defined as a set of precepts and procedures to be learned and internalized) will be completely cast aside if nuclear war should become a reality?¹³

The downside of deterrence that promises to wreak destruction has long been demonstrated by posing the following question in one or another variant (see especially the original discussion in Kahn, 1960, pp. 148 ff.):

Suppose one could construct an irreversible Doomsday Machine that would destroy the world upon detecting an attack on the homeland or upon detecting any of a number of other proscribed acts committed by the opponent (including creation of his own Doomsday Machine). Should the machine be activated?

¹³This discussion is deliberately provocative. It is more customary and politic to observe the warfighting doctrine of the Soviet Union and to then express concern that Soviet policymakers cannot be unaffected by that doctrine. While the author is well aware of Soviet military doctrine and believes it would indeed affect the decisionmaking of Soviet leaders to some degree, it seemed useful here to emphasize that the issues at stake are universal. Indeed, in the author's view, the Soviets have done a better job than the West in thinking through and then coming to grips with their nuclear strategy. For similar views by a noted Sovietologist, see Leites (1983). For discussion of Soviet behavior in crisis, generally cautious, see Adomeit (1982).

The difference is only in degree between the idealized Domsday Machine and a Western-style nuclear deterrent made credible through employment policies and doctrines that follow the precepts of declaratory strategy.¹⁴ The point here is that *while the ideal circumstance might have the opponent believe we have activated a Domsday Machine when in fact no such machine exists, the reality is that "keeping double books" (much less double doctrines and mindsets) is not easy,*¹⁵ at least for the United States, with our open system and cultural habits.

NORMATIVE ISSUES: DESIRED DECISIONMAKER VALUES

Let us next confront a troublesome issue undercutting whatever confidence one might otherwise have in the likelihood of so-called rational decisionmaking in times of the most extreme crisis. The issue raises fundamental questions about what good decisionmaking and aids to such decisionmaking might be. To illustrate the problem, consider the following:

What calculation would we want an American President to work through if the United States had already received a full-scale general nuclear attack by the Soviet Union, one that might effectively destroy the United States forever? (One could argue that the President would be rational and moral to

¹⁴For an interesting discussion on the dilemmas created by Western-style deterrence theory, see Builder and Graubard (1982), which reviews the relationship between assured-destruction concepts and international law, and in doing so places the conflict of values in sharp relief. The moral problems of assured-destruction deterrents have also been highlighted in recent years by a report by Catholic Bishops discussed briefly in an article by ex-CINC-SAC Dougherty in Carter et al. (1987). Roughly speaking, one can argue that the threat of assured destruction is moral (and legal), but the *execution* of such strikes may be another matter, even if they avoided targeting population per se.

¹⁵Policymakers have been trying to maintain double books for decades. For example, Robert McNamara has written that even in the event of a European war he would have recommended against nuclear first use by Presidents Kennedy and Johnson, despite the then-recent adoption of flexible-response strategy by NATO. Even Mr. McNamara cannot know with certainty, however, what he would have recommended in the full range of possible circumstances, especially if he had nuclear use in mind. Also, while McNamara believed Kennedy and Johnson "accepted his recommendation," it is not clear that agreement in the abstract would have translated into agreement in real crisis. Presidents have often found themselves doing things they had strongly wanted not to do. See McNamara (1985), as well as Bundy's introductory section in Bundy (1985).

forgo the retaliatory attack, because nothing positive would be accomplished by it.)

Many strategists have no problem with the concept that the President should “push the button” for assured retaliation. In the author’s experience, nonstrategists have even fewer problems with the concept once they accept the premise that the United States really has been attacked. Indeed, many people claim they would immediately order a full-scale retaliation even if the Soviet attack on the United States had been limited. They assert vociferously that the Soviets would behave similarly and that any notion of a limited exchange is nonsense. In their minds, they see assured retaliation as what they *would* truly do, not merely something they say they would do.

Would such a retaliation be rational? And, again, what decision would we as citizens want our President to make?

Those arguing that retaliation would be rational make one of two cases. The first is flippant but deeply felt: “Revenge is sweet!” It is also legitimate, they say, since revenge has been justified in teachings going back to the Old Testament. To the utility-theory practitioner, the implication here is that the value assigned to revenge is large compared with the value of sparing 100 million people who had nothing directly to do with the original attack.

The other response is more complex and goes something like this: “Oh, you don’t know—there *might* be value in retaliation. After all, the war would continue and somehow, in one way or another, there would be a ‘winner and a loser.’ Although both nations might be devastated, one or both nations might eventually recover; then the ultimate outcome might depend on the retaliatory strike. In any case, we should deny the enemy his war aims.”

These arguments would be more convincing if it were not for the fact that, while appealing to uncertainty and the need to think in terms of a protracted nuclear war, proponents of assured destruction are usually not interested in considering either. To the author’s ear, the explanation is also a rationalization for a dubious revenge.¹⁶

¹⁶As expressed by Wohlstetter and Brody (1986), “Nor is it usually clear what the victim of attack can accomplish by responding, except possibly a kind of revenge—a ‘dying sting.’ If mutual destruction is the nearly certain outcome, revenge has no

Is there a third explanation for why so many would applaud the ultimate retaliation if deterrence unequivocally failed? It may be that the answer lies deep within the human psyche, as determined by both evolution and culture, and in the impossibility of separating—at a human level—our declaratory policies and real values.

To explore this notion, consider first the unequivocal survival value of “standing tall” in the face of adversity. At the most mundane level, we all have learned that the schoolyard bully will usually back off if he sees he will have to fight hard—even if a calculation would indicate he will probably win. Or, to put it differently, people and nations have numerous opportunities to surrender preemptively or to stand their ground beyond where doing so is obviously wise. One could argue that those involved are performing a calculation and banking on deterrence; but if that were so one would see many quick surrenders once confrontations became conflicts (unless the participants included in their calculations the consequences for future generations of surrendering). There *are* examples of such surrender, for both individuals and nations, but there are many examples as well of the underdog, even the extreme underdog, continuing to the bitter end. This raises the question:

Why do free men sometimes fight to the death rather than accept slavery? If they surrendered, would there not be *some* possibility of eventual escape, and would not a “sound” calculation therefore imply surrender as the better course?¹⁷

The hypothesis offered here is that this behavior is not an accident but rather a manifestation of a deeply human characteristic needed in the development of personal liberty and national independence. It may be possible to interpret such behavior retrospectively and derive

function beyond itself. There would be no future attacker to deter and no future victim to do the deterring. Even as revenge, the response would be misdirected as bystanders with no responsibility for the attack avenged. One needs to consider more plausible objectives for the leaders of each side.” The objective of revengeful mass destruction is highly plausible, however sound this argument.

¹⁷To put it differently, is it rational to assign the outcome of surrender a value of negative infinity? Or, equivalently, is it rational to assume that the probability of surrender leading to an acceptable outcome would be zero?

implied utility values.¹⁸ But the more practical conclusion is that *a part of our humanity consciously rejects "rational-analytic thinking" in some circumstances and, instead of narrow calculations, prefers a simpler and purer form of reasoning based on ideals, simplifications, and thresholds.*¹⁹ *This behavior has provided survival advantages.* What does this tell us about what would constitute a good normative model of decisionmaking in nuclear crisis?²⁰ Would we want to undercut deterrence by teaching ourselves and our leaders to be completely rational in such instances?

It may be that—in the event—decisionmakers would restrain themselves from pushing the button because they would indeed consider full-scale retaliation to be irrational and immoral. It may be that the inability of most individuals to discuss this possibility is another example of the problem alluded to earlier: we can't keep double books. To survive in a difficult world we learn to play the deterrence game; but playing that game requires not flinching and, furthermore, not even thinking about flinching.²¹ As a practical matter, achieving credible deterrence usually takes precedence over worrying about what happens if deterrence fails.²²

¹⁸Using the example of surrender, one might argue that the person choosing to fight to the death must assign a very large negative value, $-V$, to the outcome of surrendering. However, so long as V is finite, there must be a probability P^* of surrendering turning out to be marginally acceptable (relative to death). If the person in question agrees that he would surrender if he thought P^* exceeded 0.1, then one could infer that he valued marginally acceptable life ten times more than death. This logic seems unnatural here, although similar reasoning can be powerful in other contexts.

¹⁹One such simplified reasoning is "the search for revenge," which is accepted as proper in many cultures and religions.

²⁰Western literature and tradition is stocked with heroes following noncalculating logic. The discussion is also reminiscent of Arthur Koestler's *Darkness at Noon* (The Macmillan Company, New York, 1941), which described the potential consequences of believing too strongly in one's ability to calculate consequences and make ends-means tradeoffs. The protagonist is persuaded ultimately to falsely confess his sins against the communist revolution and to accept execution.

²¹The unpleasant relationship between deterrence and the teenage game of "Chicken" was noted long ago by Herman Kahn.

²²The difficulties of war termination are discussed in Kecskemeti (1958) and Ikle (1971). The final pages of Ikle's book include a passionate and biting critique of our continued reliance on massive retaliation—e.g., "a scheme that would have been rejected as abhorrent in the Dark Ages by kings and the common people alike, appears to reflect the humane ideals of modern civilization." Also, "It is a tragic paradox of our age that the highly humane objective of preventing nuclear war is served by a military doctrine whose very purpose is to inflict genocide."

POTENTIAL EFFECTS OF STRATEGIC DEFENSE

Finally, let us consider briefly the potential significance of strategic defenses for first-strike stability. The starting point is Fig. 8 again, which, although by no means complete, suggests a number of ways in which strategic defenses could affect first-strike stability. Starting from the left, a particularly important possibility in crisis or conflict is that leaders will consider it necessary to "preempt" because of (1) the perceived near certainty that the opponent is about to conduct a first strike and (2) the belief that going second is far worse (the cost is far higher) than going first. If the side in question had survivable and robustly effective strategic defenses, however, its leaders would be less likely to believe the opponent were going to go first (to so believe might seem to require assuming irrationality). Perhaps they would also be less concerned about the consequences even if the warning of such a first strike were extremely strong.

This conclusion depends on the degree to which defenses (and defense-suppression systems) are survivable and effective. If both sides have vulnerable but otherwise effective strategic defenses, then both sides might be more inclined to worry about the enemy's first strike. Indeed, the instability could be substantially greater than any today because the side attacking first could change fundamentally the military balance while maintaining damage-limitation capability.²³

Moving rightward in Fig. 8, we see the case in which a first strike is conducted, not because of fear of imminent enemy attack, but because the situation is desperate (e.g., perhaps the war is going very badly in Europe's Central Region). It seems that the side in question is about to suffer "unacceptable" losses in power and influence. The decision to go first may depend in part on the view that general nu-

²³See Thomson (1987), Kent and DeValk (1986), and Wilkening and Watman (1986), who discuss both stable and unstable regions in the quantitative balance of mixed offensive and defensive systems. One crucial issue is whether the postulated defenses are effective against air-breathing systems. If air-breathing systems can continue to penetrate, then both sides can maintain assured-retaliation capabilities. Of course, one might then ask why anyone would even contemplate the high costs of a Strategic Defense Initiative (SDI) that still left one's country vulnerable to such attacks. Some proponents of SDI argue that fast-flying ballistic missiles have a special status that makes them more stabilizing (more warning time, presumably). That may conceivably be so now, although it is not evident, but with the emergence of low-flying cruise missiles and stealth aircraft, warning time for attacks by air breathers may be very small.

clear war involving the homelands is inevitable, but it may also depend on the unacceptability of the current war situation coupled with a view that going first would have real value. What value? Well, it might coerce the opponent into ceasing hostilities and perhaps moving back toward the status quo ante. If the side in question had even moderately effective strategic defenses and the other side did not, the potential coercive power of a first strike would increase.

Moving still farther to the right in Fig. 8, we see the case in which the first strike is motivated not so much by desperation as by calculation. This case is palpably absurd in today's world as it is seen by mainstream policymakers. However, that could change if strategic defenses were good enough. The result might be some value in first strikes on opponent defenses.²⁴ Moving farthest rightward, we see a first strike being motivated by things "getting out of hand" because of unintended actions by subordinates or third countries. Here strategic defenses could materially reduce the likely consequences of such unintended actions and therefore reduce the fatalistic incentives for a first strike.

The point of this discussion, then, has been to demonstrate that strategic defenses could affect first-strike stability in many different ways and in either direction; they would definitely introduce considerations absent from today's intuitive thinking.

²⁴Some would argue here that even limited strategic defenses could deter attacks based on military calculations by increasing greatly the size of the attack that would be needed to, say, achieve a decapitation and destroy vulnerable ICBMs. The calculations are valid, but they say more about how defenses could affect peacetime military grade cards used for balance calculations than about how nations would behave in crisis or conflict.

IV. REFLECTING ISSUES IN MODELS OF DECISIONMAKING

GENERAL COMMENTS

At this point let us discuss generally how we could reflect the various issues of the last two sections in formal rule-based models of decisionmaking. First, we must recognize that the issues would be treated according to where they arose in the decisionmaking process sketched out in Fig. 1. So, for example, the issues might be reflected in:

- *Situation assessment* (e.g., erroneous information or erroneous interpretation of information as the result of poor processing—as in ignoring information that doesn't fit the general pattern¹ or giving excessive weight to frightening information)
- *Adjustment of assumptions and behavioral pattern* (e.g., overreaction, causing the decisionmaker to characterize his opponent in the most irrational and threatening terms)²
- *Projection of prospects and risks* (e.g., misleading calculations about the likely and potential consequences of a first strike, based in part on erroneous information and in part on incorrect calculations, including calculations that ignore or underestimate risks)
- *Establishment of objectives* (e.g., not even contemplating certain objectives seriously, or overlooking important options) (George, 1980)

¹One of the defining characteristics of the alternative Ivans and Sams currently modeled in the RAND Strategy Assessment System (RSAS) is their degree of optimism or pessimism in evaluating information. There are always multiple indicators of how things are going, and people differ by whether they focus on good news or bad news. Arguably at least, an opportunist would be correlated with optimism, although an opportunist might revert to conservatism beyond some threshold. The tendency of real-world decisionmakers to impose their attitudes and prior beliefs on situation assessment is well discussed in Jervis (1976).

²See Jervis (1976) and George (1980) for examples of misassessments of one's opponent and the troubles this has caused historically. See also Nisbett and Ross (1980, Chaps. 7–8) for related psychological data and Neustadt and May (1986) for a discussion of how to use and not use historical examples.

- *Establishment of strategies* (e.g., limiting strategies to indiscriminate retaliation, or choosing a strategy (1) irrationally, (2) with inadequate or misleading information, or (3) without a reasoning process to improve the balance among contrary arguments)

Many of these issues are already reflected in first-generation national-command-level models incorporated in the RAND Strategy Assessment System, much as described in Davis, Bankes, and Kahan (1986) and Davis (1987a and b). The treatment of first-strike issues is by no means comprehensive, but the modeling principles have been demonstrated. What follows suggests how some of the particular issues raised in the preceding sections could be treated in formal models.

CLASSES OF DECISION

At the end of Sec. II we considered what appear to be key variables in the first-strike decision as inferred from the scenario-spinning exercise. The nonvolitional variables are the decisionmaker and the decisionmaker's mindset; the volitional variables are context, prospects, likelihood of opponent first strike, and various estimates of the value of going first rather than retaliating. To illustrate how these variables interplay, and to begin moving toward formal modeling, consider first Fig. 11. It indicates schematically how a current mainstream decisionmaker might approach a first-strike decision in a crisis that had not yet driven him to desperation. As noted in the box, a characteristic of this class of decision contexts is that neither side currently has disarming first-strike capability and both sides know it. In this and succeeding figures it is assumed implicitly that prospects under the current course of action are very poor (we assume that otherwise no serious thought would be given to first strikes).

We can read Fig. 11 as follows. First, we see that the decision on whether to conduct a first strike is determined by context and four other variables, each of which might depend on many layers of lower-level variables, only some of which are shown here. As suggested in Sec. II, some of the variables are more important than others (as indicated by the heaviness of the arrows). Here the likelihood of an oppo-

nent first strike is the most important issue, more general context is the second, and the capability to achieve assured destruction against the opponent is third. This decisionmaker is only slightly interested in the survival value and military value of a first strike—not because they might not be important in principle but because (1) he “knows” that meaningful national survival would be out of the question after a full-scale general nuclear war and (2) he has concluded that counting post-exchange weapons would therefore not be interesting either.

In the same way, Fig. 11 indicates that in evaluating the survival value and assured-destruction value of a first strike, the decisionmaker would not pay equal heed to the several bounding calculations. Instead, he would be inclined to use worst-case survival estimates, consistent with his basic view on the subject, and expected figures for assured-destruction capabilities in a first and second strike.

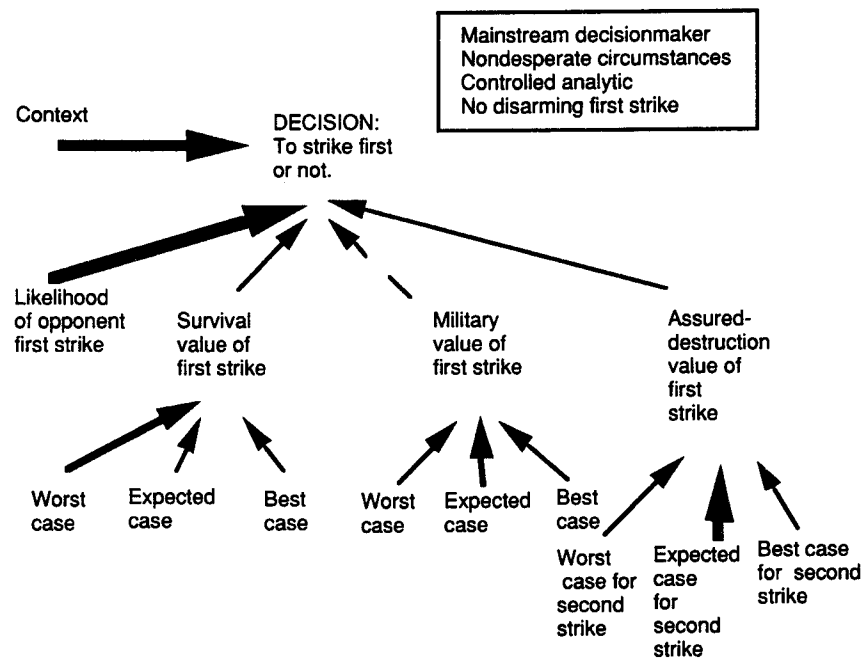


Fig. 11—Factors in mainstream decisionmaking under nondesperate circumstances

Figure 12 indicates how the same decisionmaker might think were he deep in nuclear crisis and desperate, so desperate as to have a mindset we might characterize as “fatalistic-reflexive.” This might be Scenario 1 of Sec. II, which ended in a preemptive first strike. Note that his evaluation criteria have now changed dramatically. He now is deeply interested in the possibility of survival—that is his most important concern even though his original view was that survival was a nonissue. Now he finds himself not only looking at the survival value of a first strike, but giving significant weight to the “best case” estimate thereof. This might be regarded as grasping at straws or bolstering or it might be regarded as simple pragmatism, depending on how certain it was that the opponent was indeed about to go first. At the same time, under these desperate circumstances, the deci-

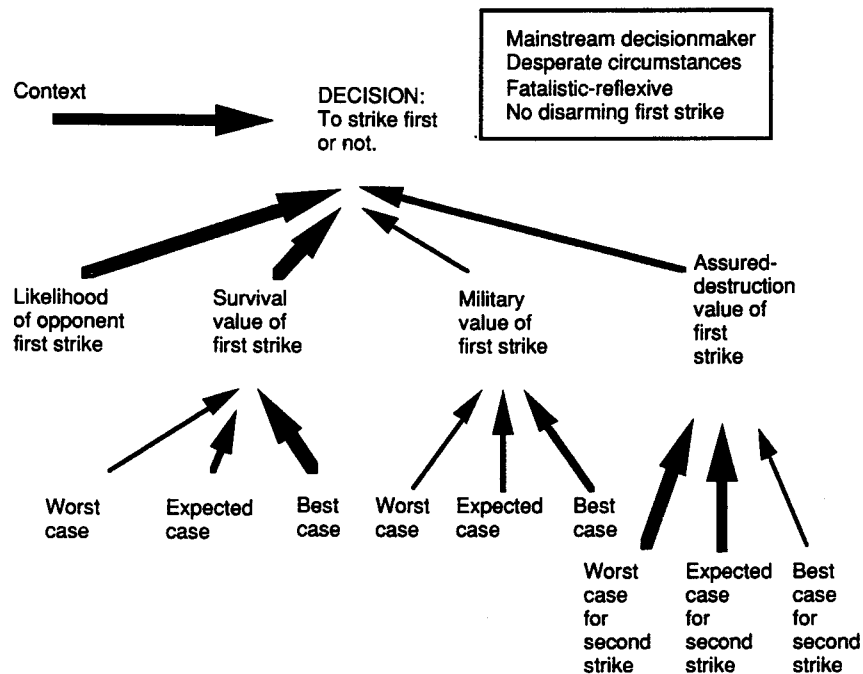


Fig. 12—Factors in mainstream decisionmaking under desperate circumstances

sionmaker might be inclined to be pessimistic about his ability to conduct an adequate assured-destruction retaliation—he might be very sensitive to ways in which the second strike would prove inadequate for his revenge.

Let us next consider, in Fig. 13, a decisionmaker with an importantly different incoming attitude—one that might be regarded as more counterforce-oriented. Further, let us assume that his emphasis has some basis—in his world both sides have significant first-strike capability, not a splendid first strike capability, but one permitting significant damage limitation as that might be judged by an analyst. Here we see that the decisionmaker is not much interested in assured destruction per se, but primarily in survival and sovereignty—i.e., in the first-strike value for survival and post-exchange military power. He is less sensitive to the likelihood of an opponent first strike. On the other hand, he is willing to consider preemption if the opponent is likely to go first at some point, even if not now. Because of his coun-

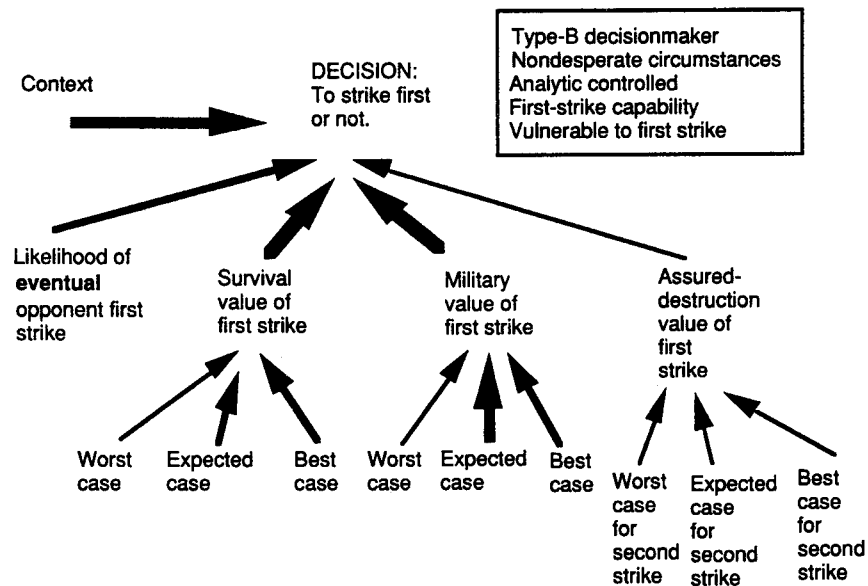


Fig. 13—Factors in Type-B decisionmaking under nondesperate circumstances

terforce orientation, he is predisposed to believe that going first must have value beyond what is readily calculated. Hence, he gives little weight to the "worst-case" estimates and more value to the expected-case and best-case values.

Finally, in Fig. 14, we depict the same decisionmaker in more desperate circumstances, but circumstances in which he maintains a relatively analytic and controlled mindset. There are many other cases, of course, but the purpose here is to illustrate how we can represent such cases schematically.

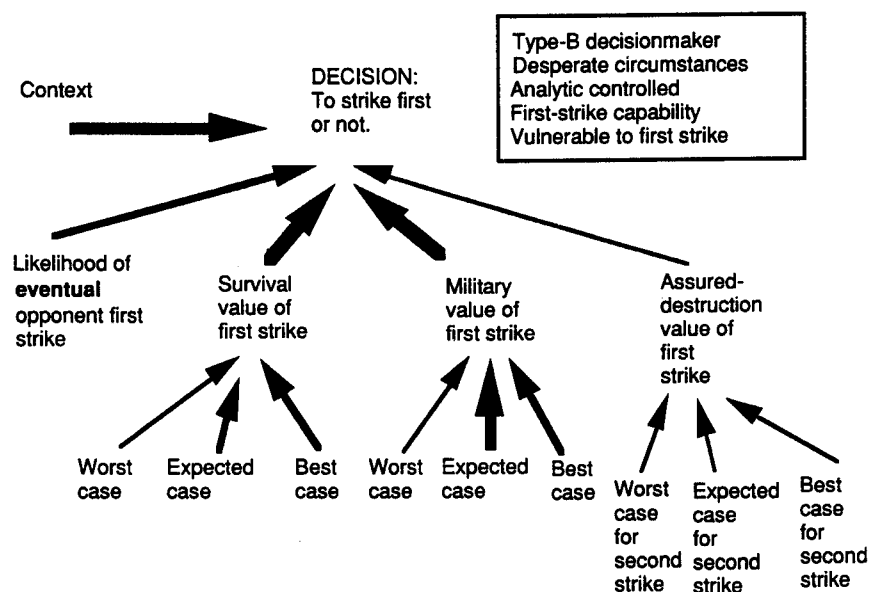


Fig. 14—Factors in Type-B decisionmaking under desperate circumstances

FORMAL REPRESENTATIONS

The schematics may be fine at a qualitative level, but the next question is whether the ideas involved can be formalized. Table 1 il-

Table 1

**POSSIBLE DECISIONMAKING RULES DURING ONGOING WAR UNDER
NONDESPERATE CIRCUMSTANCES**

(See also Fig. 11)

Survival Value of Going 1st	Likelihood of Opponent Going 1st	Retaliation Cost of Going 2nd	/ /First-strike /Decision
Very-Low	—	Very-Low	No
Very-Low	—	Low	No
Very-Low	—	Mixed	No
Very-Low	—	High	No
Very-Low	<Very-High	Very-High	No
Very-Low	Very-High	Very-High	LUA
Low	—	<=Mixed	No
Low	<High	>=High	No
Low	>=High	>=High	LUA
Mixed	<High	<High	No
Mixed	<High	>=High	LUA
Mixed	High	<High	No
Mixed	Very-High	<=Mixed	LUA
Mixed	Very-High	>=High	First strike
>=High	<=Mixed	—	LUA
>=High	High	<High	LUA
>=High	High	High	LUA
>=High	High	>=High	First strike
>=High	Very-High	—	First strike

NOTE: This and subsequent tables reflect "If-Then-Else" logic. The second line applies only if the conditions of the first line are not met, and so on.

illustrates how they might be so formalized for the case of the mainstream decisionmaker in nondesperate current-day circumstances.³ It shows a decision table much as it might appear in computer code using the RAND-ABEL[®] programming language developed originally for the RAND Strategy Assessment System (RSAS). We read the

³Evaluation of such variables as "retaliation-cost-of-going-second" can depend sensitively on command relationships and arrangements for continuity of both government and military control.

table as follows, using the first line below the header as our example: "If the survival value of going 1st is Very-low and the Likelihood of the opponent going first is anything at all and the retaliation cost of going 2nd is Very-Low, then the first-strike decision is No." By contrast, lower in the table are several boxed cases in which the decision is to conduct a first strike. In most cases the variables take on the values Very-Low, Low, Mixed, High, Very-High, so that " \leq Mixed" means, for example, "Very-Low or Low."

None of this means much until all of the variables and values are defined precisely. Let the following comments on definitions suffice:

- The likelihood of opponent first use is the likelihood as it might be described by the decisionmaker on the basis of all reports coming to him and all of his own information—except that the decisionmaker might add comments such as, "But I still don't believe it—I've met him, and talked with him, and I just don't believe" That potential for residual doubt is implicit here and is part of the character of mainstream policymakers.⁴
- The survival value of going first is understood as a judgment about how much better the quality of the post-exchange world would be if the side in question went first rather than second but the opponent did retaliate. Although in practice the value would be calculated using something like number of direct casualties, in principle the judgment would represent a conceptual integral over "possible outcome scenario of quality of life" times "the probability of that outcome scenario." This judgment would not be especially sensitive to the precise number of expected direct casualties, much less the calculation of damage to military bases. Instead, it would be a broad judgment taking into account the mainstream belief that a large-scale nuclear war could have unimaginable consequences such as widespread disease and suffering on the part of survivors. Indeed, mainstream policymakers today clearly believe that

⁴In existing national-command-level models, situation assessment, including tactical warning, is considered to be the output of the intelligence community. The final assessment of "likelihood of opponent first use," however, depends also on the decisionmaker's personal image of his opponent; the initial value of that image is an analyst-specified parameter, although the image is modeled to change as crisis and conflict proceed.

the survival value of going first is very low. In the words of Nikita Khrushchev, the living would envy the dead. *However, this judgment depends on the gross features of the military situation, if not on the details.* Thus, if the objective potential existed for damage limitation through effective strategic defenses or a disarming first strike, then that would affect the judgment.

- The “retaliation cost of going second” is deliberately a negatively defined version of the “assured-destruction value of going first.” The point here is that mainstream policymakers in non-desperate circumstances are unlikely to see much positive value in large-scale retaliation. The value is in the threat, not the execution. Indeed, carrying through on the threat may be regarded as immoral. On the other hand, knowing whether one truly possesses the assured capability for massive retaliation is of considerable concern. To a mainstream policymaker, the value of this variable is probably not very sensitive to calculational details such as the difference between destroying 70 percent or 80 percent of some category of targets. Thus, in this context, the “retaliation cost of going second” should be thought of as a measure of how good a world the opponent would be left with if one went second. Again, the concept amounts to a mental integral over “outcomes of quality of life” times “probability of outcome.”
- The quantitative assessment of variables such as the cost of going second can be accomplished using the same algorithms employed, for example, in Kent and DeValk (1986) (except for modest changes to reflect slight differences in definition). Thus, assessment would require exchange calculations using target bases and weapon-system characteristics such as CEP and reliability. If strategic defenses existed, assumptions would be necessary about such issues as whether they could preferentially defend some targets over others.

Table 2 illustrates how counterforce-oriented decisionmaking might be represented. Tables 3 and 4 illustrate how to translate the different notions of “weighting” into rules. Table 4, for example,

Table 2
DECISION TABLE FOR COUNTERFORCE-ORIENTED
DECISIONMAKING

Survival Value of Going 1st	Likelihood of Opponent Going 1st	Military Value of Going 1st	/ /First-strike /Decision
Very-Low	—	<=Mixed	No
Very-Low	<=Mixed	Very-High	No
Very-Low	>=High	>=High	First strike
Low	—	<=Mixed	No
Low	>=High	>=High	First strike
Low	<=Mixed	—	No
Mixed	<=Mixed	<=Mixed	No
Mixed	>=High	<=Mixed	No
>=Mixed	>=High	>=High	First Strike

Table 3
A CONSERVATIVE'S ASSESSMENT

Worst-case Evaluation	Expected- case Evaluation	Best-case Evaluation	/ /Net /Evaluation
Very-bad	Very-good	—	Mixed
Very-bad	Good	>=Good	Mixed
Very-bad	Mixed	>=Good	Mixed
Very-bad	Mixed	Mixed	Bad
Very-bad	<Mixed	—	Very-bad

would go along with Table 1 in describing mainstream nondesperate decisionmaking in today's world. Note from the first line that even if the expected-case evaluation of some variable like "the military value

Table 4
AN OPTIMIST'S ASSESSMENT

Worst-case evaluation	Expected- case evaluation	Best-case evaluation	/ /Net /evaluation
Very-bad	Very-good	Very-good	Good
Very-bad	Good	Very-good	Good
Very-bad	Good	Good	Mixed
Very-bad	Mixed	Very-good	Good
Very-bad	Mixed	Good	Mixed
Very-bad	Bad	Very-good	Mixed
Very-bad	Bad	Good	Bad
Very-bad	Bad	Mixed	Bad
Very-bad	—	—	Very-bad

of going first" is Very Good, a Very Bad worst-case evaluation would dominate the assessment.

In summary, it should now be evident that the kinds of phenomena discussed in earlier sections can be represented in formal analytic models. It is important to recognize, however, that such models are more valuable for their ability to represent and illustrate possible behaviors than for their ability to predict. After all, none of the behaviors posited in this report are in any way inevitable. Real-world behaviors would depend also on details of decision meetings, physical health of key figures, the pace of events, and perhaps the phase of the moon.

At the same time, the models can do more than merely repeat what has already been recognized: filling out structures like Tables 1 and 2 is an exercise in logic that extends the insights gained from scenarios and war games to other cases. Roughly speaking, each scenario corresponds to one line of a decision table. If we fill out those decision tables reasonably, using the same assumptions about decisionmaker temperament and so on that we find in the scenarios and war games, then we will see reasonable "predictions" about what that decisionmaker would do in a variety of circumstances.

The purpose of showing these decision tables is to demonstrate how down to earth the basic elements of a knowledge-based model of deci-

sionmaking can be. One need not be a computer programmer to understand the issues, review the logic, and suggest additional or alternative variables. The decisionmaking models as a whole are both large and complex (e.g., 15,000 lines of RAND-ABEL computer code describe a given Sam), primarily because of the many facets of situation assessment and the need to instruct the computer on the most mundane considerations. But the basic structure of operational national-command-level models already exists and is probably more than adequate as a framework within which to add more sophisticated and comprehensive rules dealing with the specific issue of first strikes.

When one operates the NCL models as part of the overall RAND Strategy Assessment System, an explanation log is automatically generated. This is essential for providing an integrated picture of the complex reasoning that might underlie a decision. The log will, for example, describe the basic results of each decisionmaking step as depicted in Fig. 1. One can specify the level of detail desired, seeing first a top-level view of the model's reasoning and then the bases for individual judgments.⁵

⁵For technically oriented readers interested in knowing more about what the models are like and how one can understand their reasoning in the midst of complexity, see Davis (1988b).

V. DISCUSSION AND CONCLUSIONS

SEPARABLE ISSUES

This final section turns to the challenge of improving first-strike stability. As mentioned in Sec. III and discussed in some detail in App. A, not everyone believes improving first-strike stability should be an important objective given the tensions between doing so and meeting other objectives. For reasons discussed in the appendix, however, *improving first-strike stability should be given greater emphasis than it has in recent years*; although in doing so we should be cautious not to adopt positions on weapon systems, targeting policy, and the like merely because proponents claim they would enhance first-strike stability. Many such claims are simply spurious.

In any case, the remainder of this study considers improvement measures. For a focal point let us return to the fault-tree in Fig. 8. In this depiction *many of the potential causes of first strikes can be thought of as specific dangerous ideas—suggesting that one way to improve stability would be to assure that such ideas either do not emerge or are not long entertained.*

Where do dangerous ideas come from? They come to some extent from objective realities (e.g., the facts regarding weapon systems, situation, and command-control) and to a large extent from information-gathering, information-filtering, and decisionmaking processes, all of which introduce perceptions that may differ significantly from reality. Most of this study deals with those processes. At the same time, it would be very misleading to encourage readers to believe that first-strike stability is *only* a matter of perception—i.e., something that can be dealt with adequately by appropriate “perception management.” There are several reasons for this:

- *In some cases, perceptions (and dangerous ideas) could be correct, in which case they may still be considered dangerous but may also be important and valid. Some theoretical damage-limiting cases of highly unstable force postures¹ depict ra-*

¹An unstable force posture is one that creates a “military incentive” to conduct a first strike—i.e., a force posture in which one or both sides would see advantages in going

tional decisionmakers who could decide to conduct a first strike exactly to limit damage (e.g., see Table 3 of Sec. IV).

- Moreover, even if the perceptions were incorrect, they could in periods of crisis weigh heavily on the decisionmakers' minds—much more so than rational analysis or the behavior of past leaders would suggest. For example, leaders might feel greater certainty about opponent intentions than would be justified by the information available. It would therefore be wise to hedge against nonrational decisionmaking by removing obvious sources of dangerous ideas, such as unstable force postures.
- Even if the decisionmakers would not be driven in crisis circumstances to conduct a first strike by virtue of force-posture instabilities, those same instabilities clearly cause a great deal of political and related arms-race instability as illustrated in Fig. 15 with an “influence diagram” in the style of the system dynamics school (e.g., Forrester, 1969).² Thus, they are undesirable in any case. Unfortunately, it often takes efforts from both sides to improve the situation.

In summary, to improve first-strike stability we should ideally want: (1) to assure that dangerous ideas are not true and (2) to assure that decisionmakers avoid dangerous ideas or, if they don't, are persuaded that such ideas are wrong or irrelevant.

Appendix B reviews current issues of force posture and doctrine affecting first-strike stability. The gist of that discussion is:

- The sides should emphasize survivability and should more generally avoid force-posture or command-control vulnerabilities that would make a disarming first-strike possible—or even postures making it possible for the side striking first to

first if they measured results strictly in terms of post-exchange military capabilities such as surviving connected weapons. In the limiting case, one side's first strike might disarm the other.

²One reads influence diagrams as follows. The diagrams consist of arrows and named items that have magnitude or levels (e.g., “Force-posture stability,” at the bottom, means the *degree* of force-posture stability). If a positively signed arrow connects two items, then an increase in the first tends to increase the second; if the sign is negative, then an increase in the first tends to decrease the second.

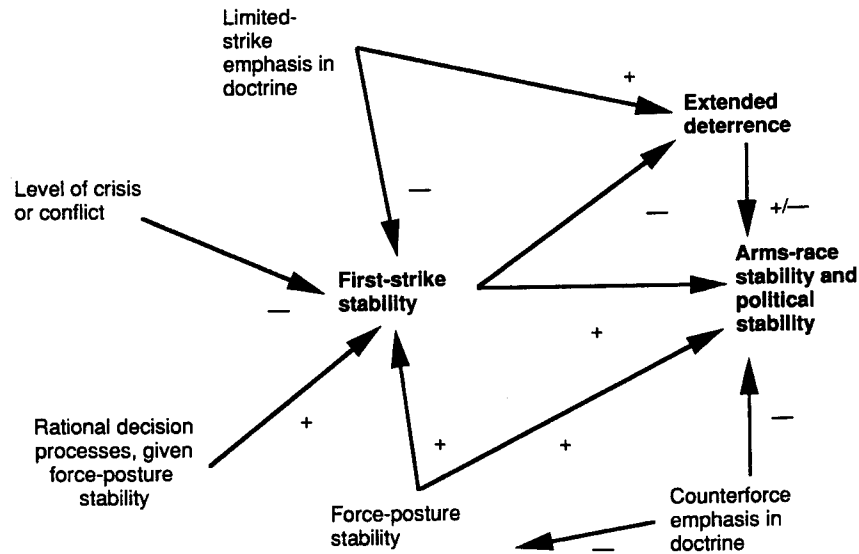


Fig. 15—An influence diagram of stability issues

improve substantially the perceived balance of power. They should also assure rigorous control of nuclear weapons under all circumstances, preferring to accept possible delays, inefficiencies, and reduced effectiveness in some contingencies to achieve increased confidence that “accidents” do not happen. Further, they should probably favor nuclear nonproliferation and urge nonsuperpowers with nuclear weapons to develop comparably high degrees of force-posture and command-control survivability and control. Finally, the sides should reconsider their strategic-nuclear doctrines, including targeting, to avoid emphasizing objectives achievable, if at all, only in first strikes.

DANGEROUS IDEAS

Dangerous Ideas to Have in Crisis

Having discussed the need to reduce sources of instability in forces, command-control, and doctrine, let us now focus more heavily on the perceptions and reasoning processes that are a main theme of this study. Let us begin by drawing on Secs. I–IV and Fig. 15 to list some of the dangerous ideas that might be reached during crisis, ideas that might encourage a first strike. These can be grouped as follows:

- General nuclear war is inevitable and perhaps imminent.
 - We have lost control of some of our forces and it is likely some of them will be launched (because of conditional delegations or the assumed propensity of military commanders disconnected from national authorities), so we should attack as decisively as possible.
 - Our opponent is irrational and total war is inevitable.
 - Our opponent has lost control of his forces, which means we must expect a nuclear attack at some point.
 - Once any nuclear weapons are used (or once any nuclear weapons are used against the homeland), we must assume total war is inevitable and launch a full-scale attack.
- Our opponent is almost certain to go first at any time.
 - His doctrine will require him to go first at some point.
 - He will surely go first because he expects us to and he will be required by doctrine to preempt.
- The cost of going second rather than first is very high.
 - We might be able to decapitate the opponent and thereby cause him to quit, retaliate ineffectively, or simply fail to respond out of paralysis.
 - We might be able to disarm the opponent or at least reduce his capabilities so drastically as to assure our own survival.

- We might be able to change the balance of forces so drastically as to coerce him into surrendering or ceasing his aggression and negotiating an acceptable outcome.
 - Our decapitation is possible and would be catastrophic in reducing the likelihood of an effective response.
 - The military cost of going second rather than first is high, even if we are assuredly able to retaliate.
 - By virtue of a first strike, we might actually survive an attack if it were less than fully effective. Indeed, our only chance to survive is by conducting a first strike and hoping it is more effective than our confidence level indicates.
- Along with a number of these dangerous ideas might be the notion that the opponent's response to a first strike would certainly be limited.
 - The opponent would respond to our counterforce attack with a retaliatory counterforce attack. That is, he would keep war as limited as possible. (This is a reasonable assessment, perhaps, of what a rational opponent would do, but not obviously a reasonable assessment of what would happen. At a minimum, there would be a substantial risk of full-scale retaliation.)

Avoiding Dangerous Ideas

If the above ideas were wrong because of stable force structure and command-control, then avoiding dangerous ideas would become avoiding misperceptions and flawed reasoning. The remainder of this study discusses some general concepts that might be pursued toward that end.

First, consider a simple list of improvement measures that come to mind readily—an unstructured list designed to provide a “feel” for the issues and challenges.

- Expose all military and political figures responsible for strategic issues to analyses demonstrating the extraordinary

uncertainties and risks that would attend any attempt to conduct a first strike, whether preemptive or not.

- At the same time, expose said figures to highly conservative and candid analyses assuring them that second-strike effectiveness would be high even in worst-case conditions (something not obviously true today, again underlining the significance of force-posture improvements).
- Have outside risk experts familiar with the flaws of standard treatments review the intendedly conservative uncertainty analyses.³ The experts should consider errors at both ends of the spectrum.
- Expose all military and political figures responsible for strategic issues to realistic gaming that incorporates phenomena such as conditional delegation of authority, continuity of government, and unauthorized launches (from both sides). Presumably, one lesson to emerge would be that temporary decapitation would not permanently prevent retaliation (nor would permanently decapitating the opponent necessarily defeat him). Such games should also contemplate how wars might end, exposing the likely desire of the opponent to avoid general nuclear war.
- Ensure that such games examine a wide range of scenarios or vignettes, including limited nuclear wars and wars in which one side or the other launches attacks that include some or many “value” targets such as population centers. A purpose here would be to force all participants to go beyond purely instinctive thinking about what actions and counteractions make sense.
- Purge misleading analyses that give erroneous impressions of the value (to the opponent or to us) of first strikes or preemption. These would include analyses focused strictly on the ratio of ICBM weapons to equivalent megatonnage (EMT), analyses that focus on the ability to achieve relatively arbitrary targeting “requirements” inadequate to disarm the opponent,

³One standard error is for analysts to conduct sophisticated calculations involving integration of distribution functions describing uncertainties of some key variables but ignoring other uncertainties altogether, especially uncertainties related to “squishy” subjects such as decisionmaker behavior. Another standard error is to conduct such analyses *assuming* that doctrinal procedures will be followed religiously.

analyses that implicitly assume vulnerabilities for strategic bombers far greater than those considered plausible by Strategic Air Command (SAC) or the Joint Chiefs of Staff (JCS) (analyses treating such vulnerabilities in excursions are quite another matter and are to be applauded), and analyses that assume dramatic C³I-related effects without attempting to examine likely events over a period of hours and days.

- Increase the visibility in analyses of probable direct and indirect civilian casualties, using best-estimate assessments of weapon targeting and heights of burst rather than assessments postulating a casualty-minimizing attack.
- Insist that educational games dealing with nuclear crises present information on such issues as the cost of going second and—assuming that we have made this true by virtue of proper force posturing—assure that those being educated understand that the “value of going first” and the “cost of going second” can easily be greatly overrated.
- Deal forthrightly with problems such as temporary disconnection of the NCA, the potential for big surprises in weapon effectiveness, and possible technical breakthroughs. If our forces have been properly configured, the result will be to greatly reduce the perceived need to act early in crisis.
- In both educational games and analysis, clarify reasoning errors caused by overlooking (or dwelling on) low-probability events—including the possibility that the opponent is eager to avoid general nuclear war despite a pattern of events seemingly to the contrary.
- Ensure that such educational games and analyses are strongly influenced by specialists familiar with the thinking of policymakers rather than with only military doctrine and exercises.
- Ensure that opposing political leaders communicate regularly on a person-to-person level to make it more difficult for them to accept alarmist characterizations of opponent intentions with regard to nuclear use.
- Encourage the Soviet government to adopt all the same measures proposed here; provide them with accurate technical analysis, which would assure that senior Soviet leaders, both

military and political, are at least as aware of first-strike follies as we are.

This unstructured list of possible improvement measures is useful as a starting point, but how would one *do* these things, and how would one judge which of them really make sense? We might think about the challenge this way: Who should be influenced, how could the influencing be done, and what generic classes of issue apply? Figure 16 summarizes possible targets of attention, techniques that might be brought to bear, analytic issues, and value issues. Generic objectives, tangible examples of which can be found in the previous list, would include:

- General education about nuclear-war issues.

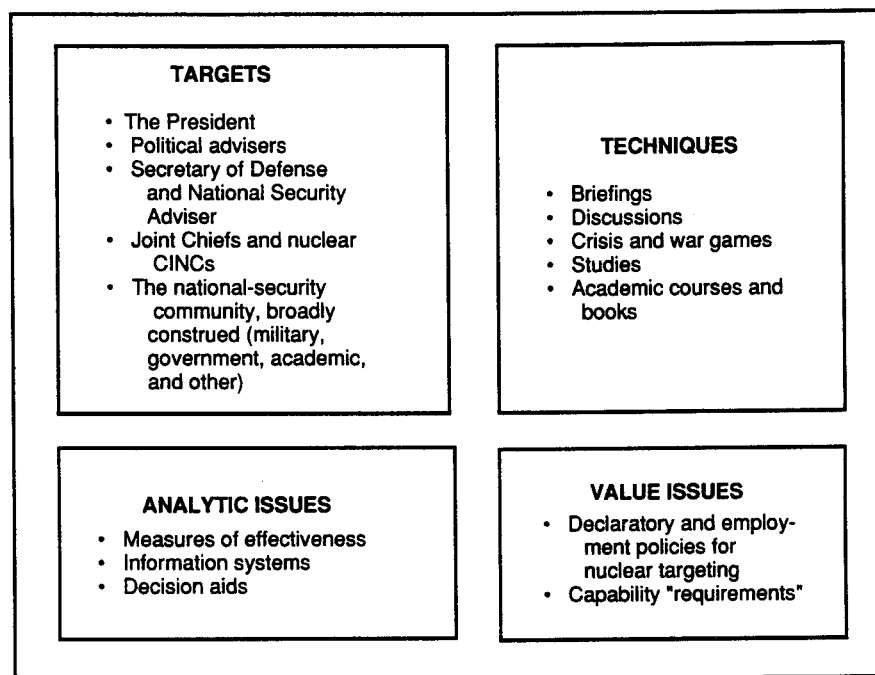


Fig. 16—Structure used to help eliminate dangerous ideas

- Specific education regarding decisionmaking under uncertainty in nuclear crisis—e.g., war games, simulations, briefings, and backgrounders designed to
 - Provide as good a base as possible of synthetic experience.
 - Provide exposure to the best thinking available on the nature and ideas of other world leaders in control of nuclear weapons.⁴ Encourage regular face-to-face discussions with those leaders.
 - Address difficult choices and dilemmas that might be encountered.
 - Address uncertainties that might be encountered (e.g., the worst-case effects of EMP on strategic command and control).
 - Clarify command arrangements, arrangements to assure continuity of government and military command, and the implications and adjustability of those arrangements.
 - Sensitize everyone to the importance of having some procedures to mitigate built-in aspects of individual and group decisionmaking that can lead to nonrationality and serious mistakes.

- Improved staff procedures to
 - Tailor information to the needs and styles of particular decisionmakers while emphasizing factors relevant to wartime decisions rather than peacetime balance assessments.
 - Identify nonintrusive (or at least comfortable) procedures for reviewing options in ways designed to avoid classic types of nonrational thinking (e.g., procedures as simple

⁴This might be especially important as an antidote for misconceptions created during the political process or developed from previous experience. Until recently, for example, many senior civilians and military officers appeared to believe that Soviet political leaders had a callous attitude about nuclear warfighting that in reality they probably never had, although Soviet military doctrine and decisions about force structure most certainly did reflect a warfighting perspective (and still do). See, for example, Davis and Stan (1984).

as providing a tailored checklist of questions for group discussion before final decision).

Targets

It is easy enough in a study to conclude (1) that the President and other likely members of the National Command Authorities should all be exposed appropriately to the kinds of issues raised here and (2) that it may be even more important that the Soviet General Secretary and his Defense Council are similarly benefited. Realistically, however, all these individuals have full agendas and their tolerance for discussion of Armageddon issues has typically been modest, at least in the United States. This suggests that efforts to discuss them had better be well designed and efficient. It also suggests that the greatest payoffs come from educating advisers and military and civilian officials and also the types of people who will eventually become top leaders, advisers, or important military and civilian officials. Some of this might be accomplished in war colleges and universities where senior officers and officials take specially designed advanced courses and seminars.

Techniques

The techniques of Fig. 16 may seem to need little explanation, but in fact there is much to discuss. Consider, for example, the issue of how the President and other top officials should think about behavior and decisionmaking in nuclear crisis. Does anyone believe a simple (or even a complex) briefing would suffice? The most important dangerous ideas are the ones least likely to be discussed in standard briefings (e.g., true worst cases or which nuclear weapons would be "right" for which circumstances). It is not clear how to address this challenge, but a possible approach would appear to be as follows:

- Have the Secretary of Defense and/or the National Security Adviser take personal responsibility for discussing the issues with others and then thinking them through, tailoring information systems and decision aids, and recommending particular policies and assumptions to be used in crisis. He would review these matters privately with the President and Vice President and provide them with appropriate written material as background. The President and Vice President might then participate in one or a very few war games designed to test the adequacy of this preparation, although they would not be requested to make, much less announce, their decisions.⁵
- Encourage the President to discuss some of the issues with his Soviet counterpart and/or selected leaders of nations with independent nuclear capabilities, or his most trusted national-security adviser might discuss the issues with his counterparts.
- Develop a "need-to-know-only" background book addressing all the known key issues in some technical detail. For each issue, record the judgments of the Secretary of Defense and/or the National Security Adviser.⁶

The reasoning here is that briefings and studies are almost always sanitized, especially with respect to normative issues, but it is precisely the normative issues that most need to be discussed. It is unlikely that any President or Soviet leader would discuss such matters deeply and candidly with any but his very closest advisers. Moreover, even those discussions would be of limited usefulness if no one presented serious strawman recommendations.

⁵Since Secretaries of Defense and National Security Advisers are almost as busy as Presidents, they might in turn call for assistance on individuals who had previously been in such posts. Ultimately, however, they would have to engage the issues themselves and develop or endorse the recommendations.

⁶There is a parallelism between what is being recommended here and the development, during the Carter administration, of "Contingency Guidance" for the Joint Chiefs of Staff and CINCs. The Under Secretary of Defense, Robert Komer, prepared a document for the Secretary to review and pass on to the military, a document that gave "real" policymaker judgments and assumptions to be used in operations planning for a variety of contingencies.

Analytic Issues and Value Judgments

Avoiding dangerous ideas must start with better staff work—with studies focused less on bean counts and politically generated measures of the strategic balance and more on the types of information that would be relevant in a real crisis. To illustrate some of what is involved, consider Fig. 17, which is in the spirit of many displays used in strategic analysis and war games.

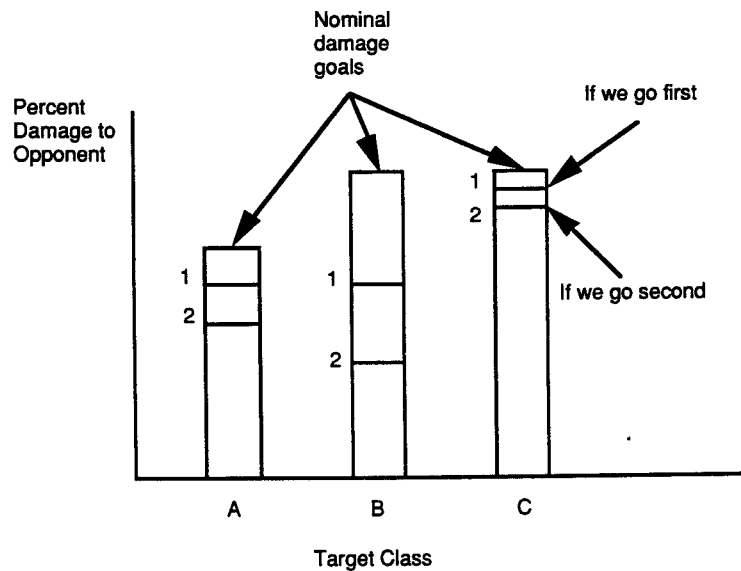


Fig. 17—A possible “military” perspective on the cost of going second rather than first

Three notional bar charts show the level of damage to the opponent in three target categories that would be expected from conducting a full-scale attack. The tops of the bars indicate nominal damage goals—the “requirements” established in peacetime for the military to use as goals in developing weapon systems, command-control systems, targeting, and doctrine. The levels for “1” and “2” illustrate what a national leader might be told to expect if he conducts the at-

tack as a first strike or as a second strike (a third case showing results for launch under attack might also be given). One implication of the figure would *seem* to be that the cost of going second rather than first is significant, as suggested by the 40 percent difference in damage levels for Target Class B (all the charts and numbers are, of course, purely illustrative, and not the result of calculations with actual forces and target lists).

Figure 17 is a classic example of an apparently value-free “military” measure of effectiveness. It is also a classic example of how information relevant for peacetime studies and purely military exercises might be altogether irrelevant to a prudent national leader’s decision.⁷ Figure 18 might be a bit more relevant: a display tailored to a particular decisionmaker. Here we see that the decisionmaker is less interested in how many Class A targets (nuclear-threat forces) could be destroyed than in how many would *remain* (see shading), since those forces could attack his homeland. On the other hand, in considering damage to Target Classes B and C, which might as an example represent economic and projection-force capabilities, he might judge that either the first-strike or the second-strike attack would cause massive damage—so much so that the effective difference in results would be zero.

Figure 19 goes farther in suggesting what a decisionmaker would probably be most concerned about, and properly so. It focuses less on damage to the opponent than on likely damage to his own country. More graphically than Fig. 18, it illustrates why the decisionmaker is not impressed by the “cost of going second,” as might be inferred from Fig. 17: even in the ideal case, the expected damage to his own homeland would be catastrophic.

The point here is that the standard measures of effectiveness used in studies and war games (measures similar to or inferior to those of Fig. 17) are largely irrelevant to wise decisions in times of actual crisis. A second point is that the “right” measures are not self-evident. They depend on the policymakers’ context, real value structure, and beliefs about the use of nuclear weapons, not on the values and beliefs

⁷As noted above, however, decisionmakers might be swayed more by purely military information than they “should be” or than would be expected from their basic character and peacetime behavior.

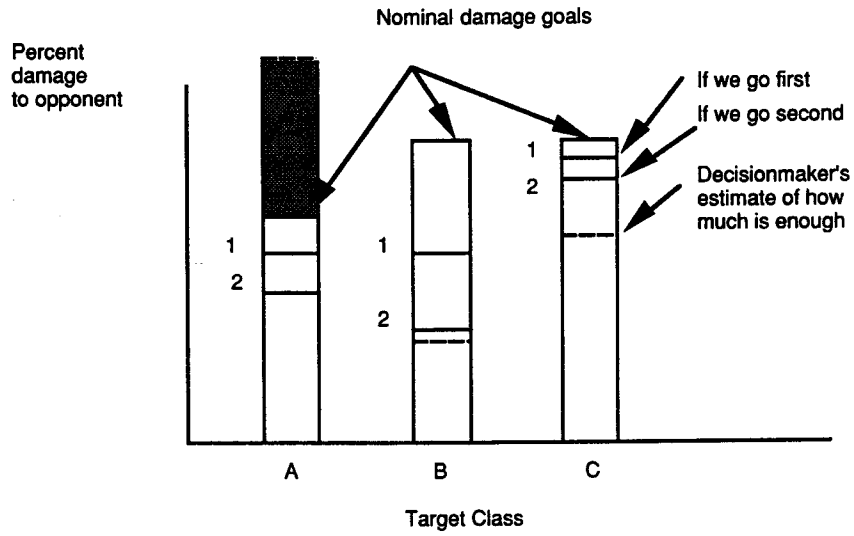


Fig. 18—A decisionmaker's possible view: no difference between first- and second-strike results

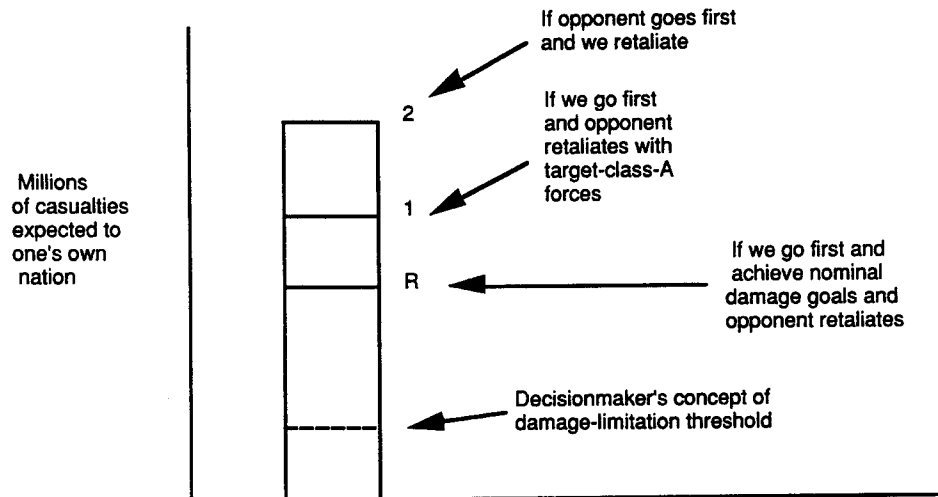


Fig. 19—A prudent decisionmaker's likely view

inferred from the same policymakers' preferred (deterrence- and competition-oriented) military requirements during peacetime. A serious review of appropriate analytic displays (measures of effectiveness, information displays, decision aids, or whatever one might wish to call them) would address explicitly all of the special perspectives identified in Sec. IV (e.g., Figs. 11 ff.). Note that Figs. 17 and 18 deal only with what Sec. IV defines as the military value of a first strike, and therefore implicitly with the "expected-case value." A serious treatment would not only be more detailed, but would also address the best-case and worst-case evaluations (e.g., Fig. 20) and the distinctions among survival value of first strike, military value, and assured-destruction value. Even *reminding* a decisionmaker of such distinctions (e.g., with adaptations of Figs. 12 through 14 or with versions of Tables 3 through 6) would be useful.

When might a first strike make sense, i.e., pass tests of rationality? From Fig. 8 and Table 3 one sees two principal issues: one's certainty

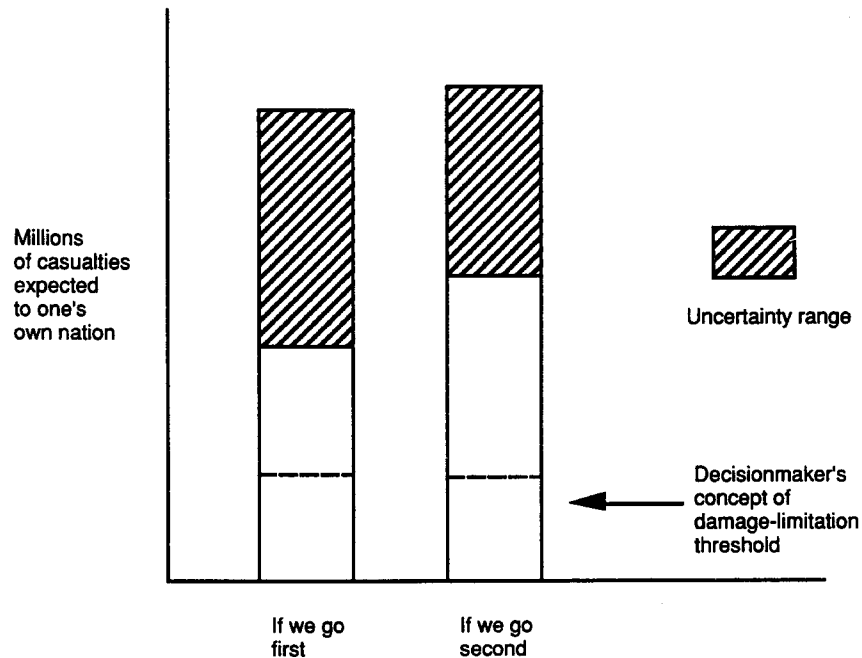


Fig. 20—Illustrative effects of uncertainty

value lies in going first instead of second. Assuming that command and control could be maintained or reestablished reasonably soon after an attack (hours, not weeks), then in today's world the prudent policymaker's assessment of the second issue would be that the value of going first, or the equivalent cost of going second, is small. At the same time, the cost of a nuclear exchange is extremely high. In a decision-analytic framework, the question becomes, "Which is larger: the value of going first multiplied by the likelihood that if one does not go first the opponent will, or the value of avoiding nuclear exchange altogether multiplied by the probability that if one does not go first neither will the opponent?" Put more intuitively, while there might be some modest value in going first, the consequences of any nuclear exchange would be catastrophic. Hence, going first would only make sense if one were absolutely certain that the opponent was about to do so. The standards for being "absolutely certain" should be extremely high, and discussions about certainty should include sample historical blunders.

On the Special Value of Models and Analytic War Games

Anyone accepting the preceding arguments should be sobered by the importance and challenge of providing good informational aids. Many key issues are seldom discussed straightforwardly, much less put into structures that could be used for group discussions (e.g., procedures for briefers and/or displays of both quantitative and qualitative issues). By and large, decision aids in other fields have not been notably successful; indeed, the decision-aids field itself is replete with failed examples.⁸

If there is a glimmer of hope, it is that years of experience at RAND, in government, and in a few academic institutions such as MIT indicate that military war games and political-military war games can teach and communicate effectively. Although these games seldom teach a particular lesson, careful game design can help participants discover and discuss important issues so that they emerge with

⁸A survey paper on decision aids and the decision-aid experience is being developed for the author by colleague Melinda Baccus.

greater insights than when they began.⁹ It is clear that some of the better games provide a synthetic experience permanently affecting the thinking of participants. The games themselves have varied from structured seminars to command-post exercises held by the military.

Unfortunately, human gaming has rarely been structured enough to penetrate many of the difficult issues raised in Sec. IV. The topics discovered and discussed, and even the lessons learned, are often very much the result of the particular individuals involved, recent events that happen to be on their minds, and other random factors. It has often been observed that a fair number of the lessons learned are probably wrong, but few studies have looked into this seriously. It would be interesting to review all influential war games dealing with launch-under-attack tactics and the related issues of targeting and planning assumptions and to examine whether the lessons learned hold up to scrutiny.

Exacerbating the limitations of gaming (and military exercises) has been the conscious reluctance of practitioners to employ anything remotely resembling decision aids or to provide any analytic structure. Whether a given team does well or poorly depends on whether its constituents are organized and systematic. There are four reasons for this. First, military officers are properly reluctant to force their views on policymakers. Second, many attempted decision aids fail (often because they are designed by technical and quantitatively oriented people who fail to understand or accommodate policymaker concerns or idiosyncratic styles). Third, a fair number of properly designed games have "succeeded" without such structures because the players discover interesting subjects and interact well and enjoyably if not efficiently. Indeed, one of the most interesting aspects of "free-form" games is how remarkably good humans are at problemsolving.¹⁰ However, while humans are "good," they are not as good as they could be with more tutoring on decisionmaking procedures and more back-

⁹See, for example, Schelling (1987), Jones (1985), Bloomfield (1984), and Masterson and Tritten (1987). Andrew Marshall has long been a proponent and sponsor of gaming activities within government.

¹⁰This depends to some extent on culture; Americans apparently work much better in groups of unfamiliar people than do Russians. Both anecdote and limited empirical experiments suggest that Russians have a great deal of difficulty dealing with free-form situations because the first order of business is to discover the hierarchy of power that "must" exist. The business literature is replete with examples of how differently Japanese and Americans work together in groups.

ground about facts and fallacies in specific subject areas. A fourth and final reason may be the visceral distrust senior people have, and properly so, of computerized decision aids. Computerized aids can be useful as background for analysts and staffers but should seldom if ever be allowed to intrude on serious face-to-face discussions.

On the Possible Role of Decisionmaking Models

And now it comes time to close the circle that began in Sec. I with the observation that first-strike stability should be investigated with formal models of human decisionmaking. As illustrated in Sec. IV, it should be possible to use model building to investigate with some degree of rigor the many issues and tradeoffs arising in human war games. The models can display common flaws of reasoning, alternative concepts (e.g., alternative cognitive maps), and logically consistent patterns of reasoning. They can improve communication and can serve as a decision aid or—to avoid this term that often conjures up negative emotional responses—a kind of information structuring mechanism. Especially effective are human games in which players are encouraged to *develop* the principal elements of key decision tables; in the author's experience such efforts are both interesting and useful. In essence, the task of laying out the decision logic pushes players into a rational mode, and having the structure of decision tables and explicit hierarchies of variables (however qualitative) as a guide reduces the common tendency for discussion to wander and hop chaotically. By contrast with many other efforts to impose rationality, encouraging players to develop these tables and variables broadens the scope of conversation. Because they are asked to identify the "real" variables, even if qualitative, players begin thinking about some of the issues raised in this study. It would seem, then, that efforts to study and improve first-strike stability could be significantly assisted by rule-based modeling of the sort described here. The author and colleagues have now had experience with numerous structured seminars which dealt with very different problems and in which participants sketched out such model primitives; the results have been very encouraging.¹¹ In considering, then, broad educational ef-

¹¹These include an unpublished study of launch-under-attack issues with colleagues Peter Stan, John Schrader, Laurinda Rohn, and John Setear, in which repeated

forts within the academic and research communities on first-strike stability, knowledge-based modeling of decisionmaking could be a powerful force for gaining greater insights and improving rationality. Moreover, important quantitative analyses of force-structure instabilities could be incorporated naturally and readily, thereby integrating worlds of study that have too long been separated.¹²

NEXT STEPS

This study has been an extended think piece—an attempt to lay out issues, challenges, and possible directions. Future work should focus on results. Doing so, however, will require time, effort, and the interaction of many individuals of diverse backgrounds and viewpoints. It is an irony that most of the people who would benefit from such efforts have agendas full of requests that are probably much less important in a cosmic sense, however urgent and pragmatic. The best hope for the approach discussed here, as a means for adding rigor to what have long been areas of “soft” analysis, is probably for it to be embraced by a large portion of the academic community concerned with national security issues.

seminar games led to very similar decision-table structures. The various participants could ultimately agree on the key variables even when they disagreed about what actions were desirable in a given, well-specified situation. Some of the key variables had been completely overlooked prior to the games.

¹²PhD dissertation research along these lines has recently begun in the RAND Graduate School.

Appendix A

ON THE DESIRABILITY OF IMPROVING FIRST-STRIKE STABILITY

Many seriously question the desirability of emphasizing first-strike stability issues because they are more concerned about other objectives, notably: (1) maintaining extended deterrence, (2) remaining an effective competitor in the perceptually and substantively important realm of strategic nuclear weapons, and (3) negotiating verifiable strategic-arms agreements. Let us consider each of these matters and how they relate to stability.

EXTENDED DETERRENCE

The problem here is the Devil's Dilemma tradeoff raised in Sec. III. Rather than having an "On the one hand . . . but on the other hand" discussion, let us consider some representative, strongly worded versions of the two polar extremes.

View One: The Case for Emphasizing Extended Deterrence over Stability

First-strike stability is not a *real* problem: stability is high today and has been high for a long time—perhaps 25 years or more. Policymakers in both countries realize that general nuclear war would be an unmitigated catastrophe to be avoided at all costs. The superpowers have been extremely cautious in crisis situations and will continue to be.

By contrast, there is reason to worry about conventional deterrence: the history of mankind is the history of war, the Soviet Union has major conventional advantages, and we can see plausible war scenarios as we look out over the next 20 years. Not only is it plausible that the Soviets would move against the oil supplies of Southwest Asia that are the heart blood of the West (including Japan), but it is increasingly plausible that uprisings will occur against the Soviets in Eastern Europe in the years ahead, uprisings that could lead to larger

conflict. Most important, were it not for the nuclear deterrent, Soviet ability to *intimidate* by virtue of USSR conventional superiority would be a major factor in its foreign policy. We can and must remain strong competitors in the realm of strategic weapons because that is the cornerstone of the overall balance of power.

While it is theoretically possible for NATO and other U.S. allies to protect themselves against Soviet aggression by building up conventional forces, any improvements in the balance are very unlikely to provide *robust* deterrence. The nuclear deterrent is the single most important reason for our longstanding peace. Since that in turn depends increasingly on the credibility of employing central strategic forces as part of a U.S. flexible response strategy—especially in the wake of the INF treaty—it is essential that NATO do nothing to further undercut that credibility. In summary, worrying about first-strike stability should not be high on our agendas. To the contrary, we should be making every effort to retain the capability to hold at risk a significant part of the Soviet nuclear force structure. Since attaining anything approaching a disarming first-strike capability is out of the question, such efforts will not materially decrease first-strike stability, but failure of such efforts would undercut extended deterrence. Those who don't realize this don't live in the real world.

View Two: The Case for Improving First-Strike Stability

The notion that first-strike stability is so high today that improving it doesn't matter is a classic example of people ignoring a low-probability possibility even though it involves catastrophic consequences. True, nuclear war does not seem likely today, but that situation could change as the result of changes in military balance, individuals in office, the general world situation, and a host of other factors. To cite one example, suppose that the Soviets broke out of the ABM treaty and deployed the ballistic-missile defenses they have been developing for the last decade or so. The world would then be different and first-strike stability would not be something to brush aside lightly. Deployment of an SDI system would probably be even more destabilizing and would lead to all kinds of worries and tensions, or even an attack on the other side's defenses.

Not only is first-strike stability a matter of fundamental and enduring concern, but the notion that a measure of first-strike instability can solve our other problems doesn't wash. Flexible-response doctrine and extended deterrence are increasingly exercises in self-deception: *the emperor has no clothes*, and if strategic nuclear analysts don't understand that extended deterrence isn't dependable, they should read what policymakers have said when they are no longer constrained by being within government. If NATO is threatened by the Soviet Union it will have to build up its conventional capability.

At the same time, although this may seem contradictory at first, a significant measure of extended deterrence will exist *independent* of doctrines, policies, and details of the military balance. That measure would not be undercut by increasing first-strike stability—except perhaps in the logical extreme, which is irrelevant. To put it differently, “existential” or general extended deterrence causing the superpowers to avoid crisis and conflict should remain strong so long as nuclear forces exist and are exercised seriously; but if superpower war should start, we should not expect extended deterrence to produce war termination on favorable terms.

When all is said and done, the people who argue for extended deterrence and against first-strike stability turn out to be nuclear strategists trying to keep their rice bowls filled, or nuclear strategists who still believe in strategic superiority, or, at the very least, nuclear strategists who still believe in the desirability of continued strategic-nuclear competition, which can be allegedly won by the United States.

The Author's View

There is no way to objectively resolve these two positions, but the author's subjective views are as follows:

- Unless one assumes away the possibility of future superpower crises or conflicts, the first sentence of View Two is correct: first-strike scenarios are plausible (i.e., at least some of the miniscenarios in Sec. II appear plausible). Thus, *dismissing first-strike stability as a nonproblem is inappropriate*.
- In support of this, consider Fig. A.1. It shows a postulated relationship between actual and perceived likelihoods of in-

tercontinental nuclear war.¹ If one accepts the general nature of the relationships, it is important to note that major oscillations in the perceived likelihood of war have occurred over the last 40 years and even within the last decade.² The potential for a new period of fears and instabilities then seems much higher than it does if we merely think in terms of today's relatively benign political environment.

- *The most important element of extended deterrence appears to be the willingness to raise stakes—even to the extent of attacking the other superpower's homeland—rather than the prospective size of an attack.* This is illustrated in Fig. A.2.³

¹The idea here is that if war is deemed extremely unlikely and almost unthinkable there will be any number of explicit and implicit barriers to its initiation (e.g., taboos against even potentially provocative challenges) except in the extreme: when nuclear use is believed almost inconceivable, and one nation then begins aggressive activities that lead to war and the concomitant raising of stakes. At the other end of the spectrum is the realm in which people believe nuclear war is likely and indeed almost inevitable. In such instances there may be considerable pressures to choose the time and place or to at least react quickly and decisively rather than agonize about uncertainty.

²President Kennedy stated during the Cuban Missile Crisis that he regarded the likelihood of nuclear war to be rather high, although less than 50 percent. It is less well known that during the early 1980s surveys indicated that a substantial fraction of American and Soviet young people had nightmares and more general anxieties about dying in a nuclear war. Americans visiting Moscow during that same period were sometimes asked by distraught citizens on the street why the United States wanted nuclear war. And during that same period the United States was willing to accept considerably more dependence on launch-under-attack tactics than would have been predicted in the late 1970s when the concept was still anathema. This was probably related to some government officials taking more seriously than they should have the claims of a "window of vulnerability" (a window that seems mysteriously to have closed, even though U.S. strategic-nuclear modernization during the 1980s has been unimpressive and has resulted in Peacekeeper ICBMs deployed in vulnerable silos). Yet another indicator that senior officials have taken the potential for nuclear war seriously is the attention given to and participation in nuclear command-post exercises during the early 1980s. Such exercises are highly desirable for deterrence, but an objective observer would surely note that the exercises would never have taken place but for the renewed anxiety of some U.S. leaders. It is likely that Soviet leaders have reacted with alarm to improved U.S. command-control procedures, much as U.S. leaders have reacted to the broad range of Soviet nuclear-war preparations throughout the 1970s and 1980s (see, for example, the emphasis given to Soviet leadership-survival activities in the 1988 edition of *Soviet Military Power*). Regardless of one's views on any of these activities, the likelihood of nuclear war was perceived to be significant by many people in the early-to-mid 1980s.

³Nuclear strategists arguing for a measure of first-strike capability to enhance extended deterrence are usually cagey about how much is enough. On the basis of game theories, discussions, and papers, however, Fig. A.2 is consistent with the *implicit* thinking of many such strategists. They would go further, however, and note that the essence of the deterrence is the willingness to inflict pain and that great pain could be

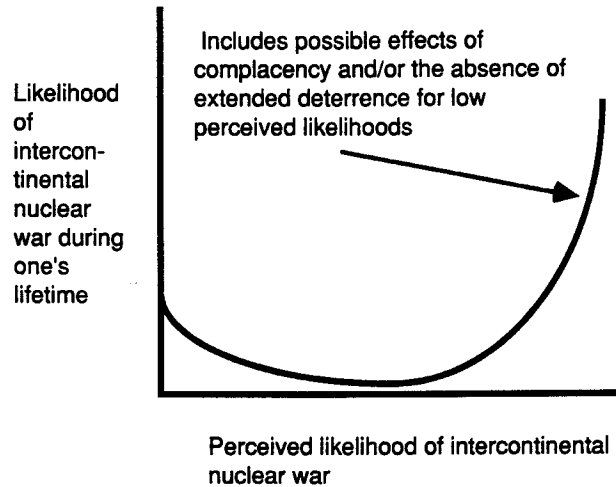


Fig. A.1—Postulated relationship between real and perceived likelihood of nuclear war

Virtually no serious proposal for improving first-strike stability would deny the superpowers the capability to launch for coercive reasons small- or medium-sized attacks against military targets if they chose to do so.

- If extended deterrence decreased, it could and probably would be resurrected in periods of desperate crisis. That is, if one side found itself gravely and imminently threatened, it might well rediscover and flaunt the nuclear threat, which in turn might well be effective almost instantaneously. By contrast, if first-strike stability worsened it would probably be because of factors that could not be remedied quickly (e.g., vulnerable forces, vulnerable command and control, and notions about how to measure the value of first strikes).

visited on the opponent with small numbers of nuclear weapons targeted, not on ICBMs in silos, but on general-purpose forces such as naval bases or economic targets such as power networks. Thus, the desire for extended deterrence through a credible but limited first-strike capability should not be equated with the requirement for substantial first-strike nuclear counterforce capability.

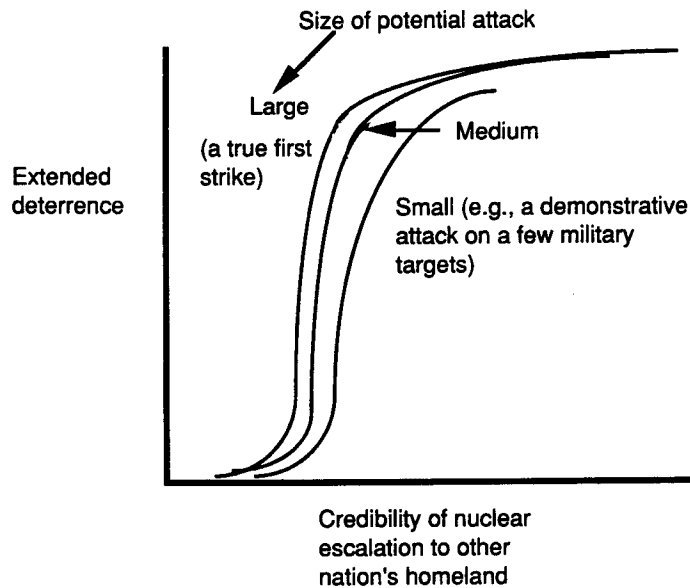


Fig. A.2—Postulated extended deterrence versus threat

On balance, then, first-strike stability should not be taken for granted, and plausible improvements in first-strike stability would not undercut extended deterrence that we possess significantly. As suggested by the notional points added in Fig. A.3, there is plenty of room to improve first-strike stability before extended deterrence starts to become less effective. This judgment, however, is subjective and each reader must decide for himself.

RED HERRINGS AND OTHER ISSUES

The other arguments against improvements in first-strike stability have their basis in confusing and misleading relationships. In particular, it does not follow that because one believes strongly in improving first-strike stability, one should judge individual weapon-systems by whether they are characterized by their proponents as "stabilizing." Nor should one be against systems because they are

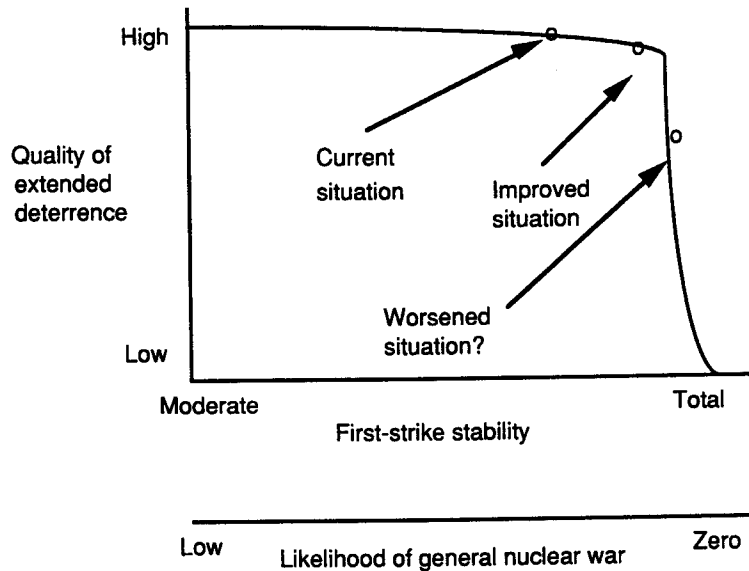


Fig. A.3—Why the Devil's Dilemma tradeoff is not currently a problem

characterized by their detractors as “destabilizing.” One of the more unfortunate aspects of the last decade’s low-quality strategic debate has been to rob the word “stabilizing” of meaning. As Harvard’s Albert Carnesale is fond of observing, it seems that the underlying criterion for using this word is usually, “Your systems are destabilizing and my systems are stabilizing.” So it is that the U.S. government has often deplored Soviet mobile ICBMs and the Soviet Union deplores the Trident II SLBM.

Another example of this confused meaning is the alleged stabilizing character of the single-reentry-vehicle (RV) Midgetman. The reality is that the Midgetman’s effect on first-strike stability would depend not on its size or number of weapons, but on its *survivability*. Are Poseidon SLBMs not stabilizing? Did not previous administrations conclude after extensive analysis that MX with multiple aimpoints would be stabilizing? The answer is that both systems are good for first-strike stability, despite involving large, MIRVed, but survivable

missiles. Some proponents believe that the single-RV Midgetman would establish a process leading to deMIRVing (putting the genie back in the bottle), which could indeed reduce military first-strike incentives, but I regard the genie-stuffing goal as unobtainable.

The point is that one can be for improving first-strike stability and still be in favor of strategic-force modernization, weapon systems with counterforce capability, and weapon systems that are much more cost-effective than the proposed Midgetman.

The last topic to be addressed here is verification. It so happens that some survivable (and therefore stabilizing) weapon systems are difficult to monitor under arms-control agreements. This is the basis for the on-again, off-again U.S. position that would prohibit Soviet mobile ICBMs. It should be noted, however, that many current and prospective strategic weapon systems are difficult to monitor without intrusive inspection measures. Further, in the wake of the Intermediate Nuclear Force (INF) treaty, intrusive inspection measures are now very much conceivable for strategic arms reduction talks (START). Actually, it can reasonably be argued that the national-security value of high-confidence verification has been exaggerated, but in any case verification problems are not an insurmountable obstacle in improving first-strike stability.

In conclusion, then, improving first-strike stability appears to be important; and the alleged tensions between doing so and achieving other objectives such as extended deterrence, verifiable arms control, and modernized strategic forces have been overdrawn.

Appendix B

FORCE-POSTURE AND DOCTRINE ISSUES AFFECTING FIRST-STRIKE STABILITY

This appendix identifies some of the many force-posture and doctrine issues affecting or allegedly affecting first-strike stability. The issues are posed here as questions, and some of them are difficult to answer. Let us first consider U.S. issues for decision.

- *ICBM Modernization or Alternatives.* Will the United States at long last find a way to deploy a new generation of survivable ICBMs or move away from dependence on ICBMs toward increasing dependence on survivably based and operated SSBNs and bombers? Or will it continue to deploy and highlight Peacekeeper ICBMs in silos (near caricatures of "first strike systems")? As a related matter, will the United States preferentially reduce silo-based ICBMs under START?
- *Vulnerability of Bombers.* Will the United States continue to depend heavily on bombers as we approach the next century, and if so how will it assure their survivability against threats such as submarine- or ship-launched cruise missiles?
- *Survivable C³I.* Will the United States continue to improve the survivability and connectivity of its strategic communications in worst-case scenarios without relying upon LUA by whatever euphemism? Will it continue to improve arrangements for assured continuity of both civilian and military command?
- *Strategic Defenses.* In pursuing research and development of strategic defenses, will the United States continue to emphasize the fundamental requirement that such defenses be survivable, so as not to invite first strikes (one of the famous "Nitze criteria")?
- *Targeting Doctrine and Capabilities to Enforce It.* How much emphasis will the United States give to these targeting objectives: (1) threatening the survival in nuclear war of

Soviet Party leadership and (2) holding at risk a significant portion of Soviet nuclear counterforce systems even as the Soviets emphasize dispersal and mobility? More generally, how much emphasis will the United States give to strategic programs designed to find relocatable strategic targets?¹ Emphasizing the targeting of relocatable strategic targets, whether leaders or nuclear-threat systems, creates an unintended incentive for first strikes.²

- *Targeting Values and Requirements.* Generally, how much weight will the United States continue to give the targeting of nuclear counterforce weapons? Is the purpose to “compete” and to create some degree of risk for those forces, or is the purpose to achieve damage limitation capabilities?
- *Targeting Existing Weapon Systems.* Will the United States begin to give its vulnerable nuclear forces, notably the Minuteman ICBM force, less and less important targets, or will it continue to rely heavily upon the Minuteman force to achieve important targeting goals?³
- *Requirements for Quick-response Hard-target Capabilities.* How much emphasis will the United States give to quick-response hard-target capability? How many targets “must” be strikable within ballistic-missile flight times? Will targeting “requirements” emphasize destroying nuclear counterforce weapon systems within a half hour or so (perhaps in the belief that prompt retaliatory attacks would reduce the likelihood

¹The issue is “how much?” because it is only prudent to pursue a vigorous program of R&D on such matters and to achieve some degree of capability, just as we have long pursued strategic ASW. If we failed to do so, we would be especially vulnerable if the Soviets made technological breakthroughs placing at risk a higher fraction of our strategic offensive forces. Moreover, it is unwise on first principles to grant an opponent sanctuaries when he grants us none (there is no question but that the Soviets would, in nuclear war, target our leaders and nuclear weapon systems). The question, then, is how far to pursue these matters in terms of dollars and doctrinal emphasis.

²The objectives and rationale for the countervailing strategy were discussed, for example, in the 1980 Annual Defense report by Harold Brown and in subsequent articles and books (e.g., Slocombe, 1982; Brown, 1983). The strategy was very much an effort to deny the Soviets any faith in their ability to win a nuclear war by their own criteria, without in the process encouraging a belief in nuclear warfighting within the United States. Creating first-strike incentives was not intended.

³It is important to note that calculations of the “cost of going second rather than first” depend on the *actual* targeting of weapon systems, their survivability, and the value of targets. If vulnerable weapons are the key mechanism for destroying highly valued targets, then the cost of going second will be considered high.

that Soviet forces could successfully launch on warning against second-strike U.S. forces—although against what targets is not clear), or will slow second-strike counterforce capabilities be deemed adequate?

- *Counterforce Doctrine.* More generally, how much emphasis will the United States give to counterforce targeting against nuclear-threat targets?

The Soviets have comparable issues. Indeed, it was Soviet emphasis on classic warfighting doctrine and warfighting forces that led to current U.S. strategic-nuclear doctrine and concerns about the window of vulnerability. Among the more important issues for the Soviets are:

- *ICBM Survivability.* To a much greater extent than we do, the Soviets depend on their ICBM force, and that force is becoming increasingly vulnerable as the United States deploys Peacekeeper and Trident II and continues to tune Minuteman accuracy. Will the Soviets actually make the long-discussed move to mobile ICBMs, which has gone quite slowly so far, or will they continue to rely upon silo-based ICBMs and, presumably, launch under attack?
- *Other-system Survivability.* Now that the Soviets have more substantial capabilities in their SSBN and bomber forces and their ICBM forces are becoming increasingly vulnerable, will they change their operational practices to increase SSBN and bomber survivability under peacetime or low-to-moderate crisis conditions?

It is appropriate to comment briefly here on the relative merits of one-sided and two-sided first-strike stability. As noted in Sec. III, in some respects the ideal circumstance for a given side is first-strike stability in which it can credibly threaten a first strike but the other side cannot (one-sided stability). For example, the first side may have invulnerable hard-target weapons and the other side may have only vulnerable ones; the second side can do little unless it believes it can find an Achilles Heel in the opponent's command and control.

One problem with a one-sided posture is that it is inherently unstable with respect to arms competition. The weaker side must obviously try to remedy the situation, which implies new systems and then the "requirement" for systems to defeat the new systems. The second problem is the highlighting of counterforce issues and the perpetuating of emphasis on nuclear counterforce—even though prospects for either side gaining anything like a disarming first strike are small. The resulting effect is the guarantee of large swings in anxieties and fears about the likelihood of nuclear war. As implied in Fig. A.1, I believe such swings are themselves dangerous.

As a practical matter, American decisionmakers needn't agonize over the issue. If we make our own strategic forces and command-control system survivable for even worst case, the Soviets will probably take care of themselves. And if they don't, "too bad for them." *The greater danger appears to be that U.S. concerns about two-sided first-strike stability (modest now but perhaps larger in the future) may discourage deployment of survivable forces, such as the Trident II, that are needed for other reasons (modernization, range, competitiveness, prompt second-strike counterforce, and second-strike counterforce more generally).*

Having itemized specific issues, let us now summarize some principles:

To improve first-strike stability, the sides should emphasize survivability. They should more generally avoid force or command-control vulnerabilities that would make a disarming first-strike possible. They should even avoid postures making it possible for the side striking first to improve substantially the balance of power. They should also assure rigorous control of nuclear weapons under all circumstances—preferring to accept possible delays, inefficiencies, and reduced effectiveness if necessary to increase confidence that "accidents" do not happen. Further, they should probably favor nuclear nonproliferation and urge nonsuperpowers with nuclear weapons to develop comparably high degrees of force-posture and command-control survivability.

BIBLIOGRAPHY

- Aaron, David, *State Scarlet*, Pocket Books, New York, New York, 1987.
- Adomeit, Hannes, *Soviet Risk-Taking and Crisis Behavior*, George Allen and Unwin, Ltd., London, 1982.
- Art, Robert J., and Kenneth N. Waltz (Eds.), *The Use of Force: International Politics and Foreign Policy, 2d edition*, University Press of America, Lanham, Maryland, 1983.
- Axelrod, Robert (Ed.), *Structure of Decision*, Princeton University Press, Princeton, New Jersey, 1976.
- Blight, James G., "Toward a Policy-Relevant Psychology of Avoiding Nuclear War: Lessons for Psychologists from the Cuban Missile Crisis," *American Psychologist*, Vol. 42, No. 1, 1987, pp. 12-29.
- Bloomfield, Lincoln, "Reflections on Gaming," *Orbis*, Winter 1984.
- Bloomfield, L. P., P. B. Bracken, G. D. Brewer, and L. H. Hoffman, "Political and Military Gaming," *Orbis*, Winter 1984.
- Bracken, Paul, *The Command and Control of Nuclear Forces*, Yale University Press, New Haven, Connecticut, 1983.
- Bracken, Paul, "Delegation of Nuclear Command Authority," in Carter et al. (Eds.), *Managing Nuclear Operations*, The Brookings Institution, Washington, D.C., 1987.
- Brams, Steven J., *Superpower Games: Applying Game Theory to Superpower Conflict*, Yale University Press, New Haven, Connecticut, 1985.
- Brodie, Bernard, *Strategy in the Missile Age*, Princeton University Press, Princeton, New Jersey, 1959.

Brown, Harold, *Thinking About National Security*, Westview Press, Boulder, Colorado, 1983.

Builder, C. H., *The Case for First-Strike Counterforce Capabilities*, The RAND Corporation, P-6179, July 1978 (available also in *Proceedings of the 41st meeting of the Military Operations Research Society*).

Builder, C. H., and M. H. Graubard, *The International Law of Armed Conflict: Implications for the Concept of Assured Destruction*, The RAND Corporation, R-2804-FF, January 1982.

Bundy, McGeorge, George Kennan, Robert McNamara, and Gerard Smith, "Nuclear Weapons and the Defense of Europe," in William P. Bundy (Ed.), *The Nuclear Controversy*, Meridian, New York, 1985.

Bundy, William P. (Ed.), *The Nuclear Controversy*, Meridian, New York, 1985.

Carnesale, Albert, and Charles Glaser, "ICBM Vulnerability: The Cures Are Worse than the Disease," *International Security*, Vol. 7, No. 1, Summer 1982.

Carter, Ashton, John Steinbruner, and Charles Zracket (Eds.), *Managing Nuclear Operations*, The Brookings Institution, Washington, D.C., 1987.

Cimbala, Stephen J., *Nuclear Strategizing: Deterrence and Reality*, Praeger, New York, 1988.

Cotter, Donald R., "Peacetime Operations: Safety and Security," in Carter et al. (Eds.), *Managing Nuclear Operations*, The Brookings Institution, Washington, D.C., 1987.

Davis, Paul, "A New Analytic Technique for the Study of Deterrence, Escalation Control, and War Termination," in Steven Cimbala (Ed.), *Artificial Intelligence and National Security*, Lexington Books, Lexington, Massachusetts, 1987a (available also from The RAND Corporation as P-7224, May 1986).

- Davis, Paul K., "Applying Artificial Intelligence Techniques to Strategic-Level Gaming and Simulation," in Maurice Elzas et al. (Eds.), *Simulation in the Artificial Intelligence Era*, North-Holland, Amsterdam, 1986a.
- Davis, Paul K., "Game-Structured Analysis as a Framework for Defense Planning," in Reiner Huber (Ed.), *Modeling and Analysis of Conventional Defense in Europe*, Plenum, New York, New York, 1986b.
- Davis, Paul K., "Knowledge-based Simulation for Studying Issues of Nuclear Strategy," in Allan M. Din (Ed.), *Arms and Artificial Intelligence*, Oxford University Press, New York, New York, 1987b.
- Davis, Paul K., *The Role of Uncertainty in Assessing the NATO-Pact Central-Region Balance*, The RAND Corporation, P-7427, April 1988a.
- Davis, Paul K., *Explanation Mechanisms for Knowledge-Based Models in the RAND Strategy Assessment System*, The RAND Corporation, N-2711-NA, August 1988b.
- Davis, Paul K., and P.J.E. Stan, *Concepts and Models of Escalation*, The RAND Corporation, R-3235, May 1984.
- Davis, Paul K., S. C. Bankes, and J. P. Kahan, *A New Methodology for Modeling National Command Level Decisionmaking in War Games and Simulations*, The RAND Corporation, R-3290-NA, July 1986.
- Davis, Paul K., P.J.E. Stan, and B. W. Bennett, *Automated War Gaming as a Technique for Exploring Strategic Command and Control Issues*, The RAND Corporation, N-2044-NA, November 1983.
- Forrester, Jay, *Urban Dynamics*, MIT Press, Cambridge, Massachusetts, 1969.
- Foster, James L., and Garry D. Brewer, "And the Clocks Were Striking Thirteen: The Termination of War," *Policy Sciences*, 7, 1976.

- George, Alexander L., *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice*, Westview Press, Boulder, Colorado, 1980.
- George, Alexander L., "Crisis Management: The Interaction of Political and Military Considerations," *Survival*, XXVI, No. 5, September/October 1984, pp. 223-234.
- George, Alexander L., "The Impact of Crisis-Induced Stress on Decision Making," in Frederic Solomon and Robert Q. Marston (Eds.), *The Medical Implications of Nuclear War*, Institute of Medicine, National Academy of Sciences, National Academy Press, Washington, D.C., 1986.
- Glaser, Charles, "Why Even Good Defenses May Be Bad," *International Security*, Vol. 9, No. 2, Fall 1984.
- Gottfried, Kurt, and Bruce G. Blair (Eds.), *Crisis Stability and Nuclear War*, Oxford University Press, New York, 1988.
- Hart, Douglas M., "Some Approaches to Crisis Management: The Military Dimension," *Survival*, XXVI, No. 5, September/October 1984, pp. 214-222.
- Ike, Nobutaka (Ed.), *Japan's Decision for War: Records of the 1941 Policy Conferences*, Stanford University Press, Stanford, California, 1967.
- Ikle, Fred Charles, *Every War Must End*, Columbia University Press, New York, New York, 1971.
- Janis, I. L., and L. Mann, *Decision Making*, Free Press, New York, New York, 1977.
- Jervis, Robert, *Perception and Misperception in International Politics*, Princeton University Press, Princeton, New Jersey, 1976.
- Jervis, Robert, "Hypotheses on Misperception," *World Politics*, Vol. 20, April 1968.
- Jervis, Robert, "Deterrence and Perception," *International Security*, Vol. 7, No. 3, Winter 1982/1983.

- Jones, W. M., *On Free-Form Gaming*, The RAND Corporation, N-2322-RC, August 1985.
- Kahan, J. P., R. E. Darilek, M. H. Graubard, N. C. Brown, A. A. Platt, and B. R. Williams, *Preventing Nuclear Conflict: What Can the Behavioral Sciences Contribute?* The RAND Corporation, N-2070-CC/FF/RC, December 1983.
- Kahn, Herman, *On Thermonuclear War*, Princeton University Press, Princeton, New Jersey, 1960.
- Kahn, Herman, *On Escalation: Metaphors and Scenarios*, Penguin Books, Baltimore, Maryland, 1968.
- Kahneman, Daniel, Paul Slovic, and Amos Tverski (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge Press, Cambridge, Massachusetts, 1982.
- Kecskemeti, Paul, *Strategic Surrender: The Politics of Victory and Defeat*, Stanford University Press, Stanford, California, 1958.
- Kent, G. A., and R. J. DeValk, *Strategic Defenses and the Transition to Assured Survival*, The RAND Corporation, R-3369-AF, October 1986.
- Kent, G. A., R. J. DeValk, and D. E. Thaler, *A Calculus of First-Strike Stability (A Criterion for Evaluating Strategic Forces)*, The RAND Corporation, N-2526-AF, June 1988.
- Landi, Dale M., Bruno W. Augenstein, Cullen M. Crain, William R. Harris, and Brian M. Jenkins, "Improving the Means for Intergovernmental Communications in Crisis," *Survival*, XXVI, No. 5, September/October 1984, pp. 200-213.
- Legge, J. M., *Theater Nuclear Weapons and the NATO Strategy of Flexible Response*, The RAND Corporation, R-2964-FF, April 1983.
- Leites, N., *Soviet Style in War*, The RAND Corporation, R-2615-NA, April 1982.

- Lorell, M. A., and N. C. Brown, *Avoiding Nuclear War: A RAND Research Approach and Agenda*, The RAND Corporation, N-2338-RC, September 1985.
- March, James G., and Roger Weissinger-Baylon, *Ambiguity in Decision: Organizational Perspectives on Military Decision Making*, Pitman Publishing Company, Marshfield, Massachusetts, 1986.
- Masterson, Rear Admiral K. S., Jr., and James J. Tritten, "New Concepts in Global War-gaming," *U.S. Naval Institute Proceedings*, July 1987, pp. 117-119.
- McCalla, Robert Bruce, *The Dynamics of Perception in U.S.-Soviet Crises*, Ph.D. Dissertation, Political Science Department, University of Michigan, 1987.
- McNamara, Robert S., "The Military Role of Nuclear Weapons: Perceptions and Misperceptions," in William P. Bundy (Ed.), *The Nuclear Controversy*, Meridian, New York, New York, 1985.
- Neustadt, Richard E., and Ernest R. May, *Thinking in Time: The Uses of History for Decision Makers*, The Free Press, New York, 1986.
- Nisbett, Richard, and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.
- Nitze, Paul H., "Assuring Strategic Stability in an Era of Détente," *Foreign Affairs*, Vol. 54, No. 2, 1976.
- Schelling, Thomas C., *The Strategy of Conflict*, Harvard University Press, Cambridge, Massachusetts, 1960.
- Schelling, Thomas C., "The Role of War Games and Exercises," in Carter et. al. (Eds.), *Managing Nuclear Operations*, The Brookings Institution, Washington, D.C., 1987.
- Schwabe, W., and L. M. Jamison, *A Rule-Based Policy-Level Model of Nonsuperpower Behavior in Strategic Conflicts*, The RAND Corporation, R-2962-DNA, December 1982.

- Shlapak, D. A., W. Schwabe, M. A. Lorell, and Y. Ben-Horin, *The Mark III Scenario Agent: A Rule-Based Model of Third-Country Behavior in Superpower Crises and Conflict*, The RAND Corporation, N-2363-NA, October 1985.
- Simon, Herbert A., *The Sciences of the Artificial*, 2nd ed., MIT Press, Cambridge, Massachusetts, 1981.
- Simon, Herbert A., *Models of Bounded Rationality: Economic Analysis and Public Policy*, Vol. 1, MIT Press, Cambridge, Massachusetts, 1982.
- Slocombe, Walter, "The Countervailing Strategy," *International Security*, Vol. 5, No. 4, Spring 1981.
- Slocombe, Walter, "Strategic Command, Control, Communications and Intelligence," in *Challenges for U.S. National Security-Nuclear Strategy Issues of the 1980s: Strategic Vulnerabilities; Command, Control, Communications, and Intelligence; Theater Nuclear Forces, A Third Report*, prepared by the staff of the Carnegie Panel on U.S. Security and the Future of Arms Control, Carnegie Endowment for International Peace, Washington, D.C., 1982, pp. 85-134.
- Smoke, Richard, *War: Controlling Escalation*, Harvard University Press, Cambridge, Massachusetts, 1977.
- Steinbruner, John D., "Launch Under Attack," *Scientific American*, Vol. 250, No. 1, January 1984.
- Survival*, Vol. 26, No. 5, September/October 1984, pp. 200-234.
- Thomson, James A., "Deterrence, Stability, and Strategic Defenses," in Fred Hoffman, Albert Wohlstetter, and David Yost (Eds.), *Swords and Shields: NATO, the USSR, and New Choices for Long Range Offense and Defense*, D.C. Heath and Company, Lexington, Massachusetts, 1987.
- Wilkening, D. A., and K. Watman, *Strategic Defenses and First-strike Stability*, The RAND Corporation, R-3412-FF/RC, November 1986.

Wohlstetter, Albert, and Richard Brody, "Continuing Control as a Requirement for Deterring," in Ashton B. Carter, John D. Steinbruner, and Charles A. Zraket (Eds.), *Managing Nuclear Operations*, The Brookings Institution, Washington, D.C., 1986.



RAND/R-3689-CC

