This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Purchase this document
Browse Books & Publications
Make a charitable contribution

For More Information

Visit RAND at www.rand.org
Explore RAND National Defense Research Institute
RAND Project AIR FORCE

View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.
This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.
The research described in this report was sponsored by the Office of the Secretary of Defense and conducted within RAND Project AIR FORCE and the International Security and Defense Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Department of the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community under Contract W74V8H-06-C-0002.

Library of Congress Cataloging-in-Publication Data

Long, Austin G.
   From Cold War to long war : lessons from six decades of Rand deterrence research
   / Austin Long.
   p. cm.
   Includes bibliographical references.
   I. Deterrence (Strategy) I. Rand Corporation. II. Title.

U162.6.L64 2008
355.02'17—dc22
2008027296

Cover photos courtesy of AP/Wide World Photos and iStock International.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

Cover design by Carol Earnest

© Copyright 2008 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2008 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org

To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
   Fax: (310) 451-6915; Email: order@rand.org
Preface

The concept of deterrence has been somewhat neglected in the nearly two decades since the end of the Cold War, particularly after the attacks of September 11, 2001. Yet deterrence will likely remain a major component of U.S. foreign and defense policy. The RAND Corporation was central to the development of modern deterrence theory and examining its more than six decades of research on the subject helps explain both why deterrence is so necessary for the United States and how to improve its practice with potential adversaries ranging from peer competitors to terrorist networks.

This research was sponsored by the Office of the Secretary of Defense and conducted within RAND Project AIR FORCE and the International Security and Defense Policy Center of the RAND National Defense Research Institute.

The RAND National Defense Research Institute

The RAND National Defense Research Institute is a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Department of the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community.

For more information on RAND’s International Security and Defense Policy Center, contact the Director, James Dobbins. He can be reached by email at James_Dobbins@rand.org; by phone at 703-413-1100 x5134; or by mail at the RAND Corporation, 1200 South

iii

**RAND Project AIR FORCE**

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the U.S. Air Force’s federally funded research and development center for studies and analyses. PAF provides the Air Force with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future aerospace forces. Research is conducted in four programs: Force Modernization and Employment; Manpower, Personnel, and Training; Resource Management; and Strategy and Doctrine. Additional information about PAF is available on our Web site: http://www.rand.org/paf/
Contents

Preface ........................................................................................................ iii
Summary ...................................................................................................... vii
Acknowledgments ...................................................................................... xi
Abbreviations .............................................................................................. xiii

CHAPTER ONE
Thinking (and Rethinking) the Unthinkable:
    RAND and Deterrence ........................................................................... 1

CHAPTER TWO
A Too-Distant Mirror? The Relevance of Prior Deterrence ................. 5

CHAPTER THREE
Department of Defense as Ministry of Fear:
    The Theory of Deterrence ..................................................................... 7

CHAPTER FOUR
Avoiding the Garrison State: Deterrence as a Strategy ....................... 17

CHAPTER FIVE
Deflecting the Sword of Damocles:
    Strategic Defense and Deterrence ....................................................... 23
The Least Miserable Options:
    Counterforce, Limited Nuclear Options, and Deterrence ............... 25
Firebreaks on the Battlefield: RAND and Tactical Nuclear Weapons .... 43
Modern Arms and Expensive Men:
  Conventional Deterrence After Vietnam................................. 46
  The Magic Bullets: Conventional Counterforce......................... 50

CHAPTER SIX
The Other Side of the Hill:
  Understanding the Adversary and Deterrence ......................... 53

CHAPTER SEVEN
Deterrence Then and Now .................................................... 59
The Long Shadow of the Garrison State:
  Deterrence in Future U.S. Strategy ................................. 59
Cold War Redux: Deterrence and the Peer Competitor.................... 62
Terrible Swift Sword:
  Preemptive Strategic Counterforce Versus Russia and China ....... 62
Bears on Land, Dragons at Sea:
  Conventional Deterrence Versus Russia and China ............... 65
Under Western Eyes: Russian and Chinese Operational Codes........ 68
  How Dire a Strait? Extended Deterrence, China, and Taiwan .... 68
Beyond the “Axis of Evil”: Deterrence and the Regional Power ...... 72
  Smothering the “Sea of Fire”:
    Deterrence and Counterforce on the Korean Peninsula .......... 77
The Men Without Fear? Nonstate Actors and Deterrence ............... 80

CHAPTER EIGHT
Conclusion ........................................................................ 85

APPENDIX
Annotated Bibliography ..................................................... 87
References ................................................................. 91
Deterrence is the art of producing in the mind of the enemy the fear to attack.

—Peter George, Dr. Strangelove

Deterrence was never a well-loved concept in the United States, but, following the end of the Cold War, it lost what little luster it held for most Americans. A strategy of nuclear deterrence in particular was an ugly policy, involving the threat of massive damage to entire societies. It was also an uncertain proposition, as affecting the enemy’s mind rather than its body is only partly susceptible to advance planning and calculation. Such movies as *Fail-Safe* and *Dr. Strangelove* document the combination of revulsion and dark humor that nuclear deterrence evokes. Conventional deterrence by punishment is little better regarded.

A decade after the end of the Cold War, many viewed the attacks of September 11, 2001, as confirming the end of deterrence. Despite overwhelming nuclear and conventional superiority, the United States suffered a major attack on its own soil for the first time since World War II. In response to this atrocity, the 2002 United States national-security strategy (White House Office, 2002, p. 15) proclaimed the

---

1 George (1988, p. 98). *Dr. Strangelove* remains a central element of popular culture for deterrence theorists in general and the RAND Corporation specifically. The title character references “the BLAND Corporation” (also on p. 98), and the character as portrayed in the movie by Peter Sellers includes elements of the personality and style of RAND analyst Herman Kahn. Kahn, Thomas Schelling, and others also consulted with the movie’s director, Stanley Kubrick.
irrelevance of deterrence to many important challenges from both state and nonstate actors.

Yet reports of the demise of deterrence were greatly exaggerated. The 2006 version of the national-security strategy returned deterrence to the lexicon of U.S. national defense. More importantly, it provides the framework from which this revival of deterrence springs. This framework is that “of a long struggle, similar to what our country faced in the early years of the Cold War.” Deterrence, the unpalatable but indispensable strategy of the old cold war, will be an equally indispensable part of the strategy of the new long war.

Since its inception six decades ago, the RAND Corporation has been one of the key institutional “homes” for the study of deterrence. Most if not all of the early deterrence theorists spent at least some time at RAND, where they conducted extensive research and applied theory to critical policy decisions. This book is an attempt to examine much of this research for lessons relevant to the current and future strategic environment. It is therefore part intellectual history and part policy recommendation, intended to encourage debate and discussion on how deterrence can best be incorporated into U.S. strategy.

This book serves the additional purpose of increasing the distribution of important RAND research. While much of RAND’s work on deterrence is well known and forms the backbone of the modern conception of deterrence, other important works have remained limited in distribution. Hopefully, making this past research more widely available will contribute to the quality of future debates on deterrence.

This book consists of seven main sections. The first presents a brief history of RAND’s role in the development of deterrence theory and policy. The second provides an assessment of the relevance of Cold War–era deterrence research to the challenges of the long war. The next section discusses the theoretical basis of deterrence and its components as well as some generic policy considerations that are derived from the theory. The fourth section describes why deterrence was the strategy the United States adopted for the Cold War and the benefits that accrued from this choice. The fifth section describes in more detail

---

various technical and doctrinal approaches to making deterrence effective that RAND studied. The next section describes RAND efforts to study the psychological and organizational elements of deterrence. The final section presents three contexts and scenarios related to the long war in which RAND deterrence research might be relevant.

In addition, an annotated bibliography of less well-known RAND deterrence research is presented. Some of the documents cited herein were part of RAND’s draft (D) series of publications, which were intended to promote discussion among researchers. Those publications were not reviewed and were never intended for external dissemination, yet provide interesting insights into the debate within RAND on these issues at the time. Not all D-series publications are available to the public.
A great many people provided valuable support to this project. James Dobbins provided commentary and guidance throughout. Reviewers Robert Jervis of Columbia University and David Gompert of RAND gave the author the benefit of their considerable expertise on deterrence. Paul K. Davis provided a number of critiques that strengthened the analysis and evidence presented. Carl Kaysen of the Massachusetts Institute of Technology (MIT), Andrew Marshall of the U.S. Department of Defense, and the late James Digby were all generous in discussing their experiences with both the theory and practice of deterrence. Marshall also provided extensive comments on an early draft. Colleagues at MIT Security Studies Program were also very helpful, particularly Brendan Green and Caitlin Talmadge. Very special thanks are due to Owen R. Coté Jr., associate director of the program, who was subjected to countless hours of discussion of this topic and bore it with good humor while providing invaluable help and comments on drafts. Finally, the author thanks the RAND Library staff for help with archival and classified materials. Any errors are the responsibility of the author.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>anti–ballistic missile</td>
</tr>
<tr>
<td>ACM</td>
<td>advanced cruise missile</td>
</tr>
<tr>
<td>ASW</td>
<td>antisubmarine warfare</td>
</tr>
<tr>
<td>AVF</td>
<td>all-volunteer force</td>
</tr>
<tr>
<td>C3I</td>
<td>command, control, communication, and intelligence</td>
</tr>
<tr>
<td>CCP</td>
<td>Chinese Communist Party</td>
</tr>
<tr>
<td>CEP</td>
<td>circular error probable</td>
</tr>
<tr>
<td>DoD</td>
<td>U.S. Department of Defense</td>
</tr>
<tr>
<td>FY</td>
<td>fiscal year</td>
</tr>
<tr>
<td>GDP</td>
<td>gross domestic product</td>
</tr>
<tr>
<td>GWD</td>
<td>generated with damage</td>
</tr>
<tr>
<td>GWOD</td>
<td>generated without damage</td>
</tr>
<tr>
<td>ICBM</td>
<td>intercontinental ballistic missile</td>
</tr>
<tr>
<td>IRBM</td>
<td>intermediate-range ballistic missile</td>
</tr>
<tr>
<td>LNO</td>
<td>limited nuclear option</td>
</tr>
<tr>
<td>LTTE</td>
<td>Liberation Tigers of Tamil Elam</td>
</tr>
</tbody>
</table>
MAO  major-attack option
MIRV  multiple independently targetable reentry vehicle
MIT  Massachusetts Institute of Technology
MLRS  multiple-launch rocket system
MRBM  medium-range ballistic missile
NTC  National Training Center
NUWEP  Nuclear Weapons Employment Policy
PGM  precision-guided munition
REFORGER  Return of Forces to Germany
RNO  regional nuclear option
SAC  Strategic Air Command
SADARM  sense and destroy armor
SAGE  Semi-Automatic Ground Environment
SAM  surface-to-air missile
SAO  selected-attack option
SIOP  Single Integrated Operating Plan
SLBM  submarine-launched ballistic missile
SSBN  nuclear-powered, nuclear ballistic missile–carrying submarine
SSGN  Trident-submarine conversion
TOW  tube-launched, optically tracked, wire command
UAV  unmanned aerial vehicle
As World War II ended, the commanding general of the U.S. Army Air Forces, General Henry H. “Hap” Arnold sought to ensure military access to the country’s intellectual elite. To that end, he and others convinced the Douglas Aircraft Company to create Project Research and Development (RAND) as a means to continue the forward-looking vision on technology that had led the United States to develop the atomic bomb. This arrangement lasted until 1948, when Project RAND separated from Douglas and became the independent, nonprofit RAND Corporation. RAND quickly grew into an interdisciplinary think tank concerned with the problems of the nascent cold war.¹

RAND, as the U.S. Air Force’s think tank, became the central location for the study of the terra incognita that was atomic warfare.² From the beginning, one of the principal concepts underpinning this study was that of deterrence. Bernard Brodie, though not yet at RAND, set this tone in the 1946 book *The Absolute Weapon*:

> Thus, the first and most vital step in any American security program for the age of the atomic bombs is to take measures to guarantee to ourselves in case of attack the possibility of retaliation in kind. . . . Thus far the chief purpose of our military establishment has been to win wars. From now on its chief purpose must be to avert them. (Brodie, 1946, p. 76)


² See Schelling (1978–1979) for comments on RAND’s central importance to the study of strategy in the Cold War.
In the 1950s, Brodie’s initial and simple description of deterrence in the atomic age quickly led to a welter of concerns. As RAND’s William Kaufmann noted at the time, “In principle, then, the requirements of deterrence are relatively simple. In practice, however, they turn out to be exceptionally complex, expensive, and difficult to obtain” (Kaufmann, 1958, p. 2). RAND analysts engaged all of these complexities, from optimal basing for bombers to understanding the nature of surprise attack to strategies for fighting nuclear war. One of RAND’s most famous analysts, Herman Kahn, literally wrote the book on the subject. His seminal *On Thermonuclear War* was not only a groundbreaking work; it was also a national bestseller (Kahn, 1961). RAND was also intimately involved in the development of disciplines that would later have bearing on deterrence, such as computing, economics, and war-gaming (see Mirowski, 2002).

The 1950s saw RAND devoting enormous effort to exploring all of these interrelated concepts, yet RAND staff remained only advisers with one client: the U.S. Air Force. The inauguration of President John F. Kennedy in 1961 marked a turning point. Many current and former RAND analysts were invited to join the administration in both the U.S. Department of Defense (DoD) and National Security Council. The policy advisers would become the policymakers. Many of those who did not join the administration were nonetheless involved in advising DoD. RAND’s ties to the government were thus broadened and deepened.

From 1961 forward, RAND’s research on deterrence, both theoretical and practical, had a profound influence on U.S. strategy. In the Kennedy and Johnson administrations, RAND researchers helped to implement “Flexible Response,” a program intended to expand U.S. capabilities to execute a wider array of operations than the choices of no response or full-scale nuclear war. This program sought to increase both deterrent capability (by making it more credible) and the ability to wage war at both the conventional and nuclear levels.

---

3 On this period, see Kaplan (1991) and Herken (1987).

4 See Wells (2001) and Ghamari-Tabrizi (2005) for additional insight into RAND at the time.
However, by the 1970s, the Soviet nuclear parity and the possibility of mutually assured destruction feared since the 1950s had clearly arrived. More than just Flexible Response was needed, so RAND strategists contributed to a program designed to generate limited nuclear options (LNOs). These options, combined with investments in command and control, were intended to further refine the ability of the United States to deter, and, if deterrence failed, to fight. RAND’s influence was, in some ways, higher than ever as the secretary of defense at this time, James Schlesinger, was the former director of strategic studies at RAND. Even after Schlesinger’s departure, RAND subsequently influenced the Carter administration’s “counter vailing strategy” (Kaufmann, now at the Massachusetts Institute of Technology (MIT), even recommended the term *counter vailing* to Secretary of Defense Harold Brown).

The 1970s also saw changes at RAND as the impact of the Vietnam War and the leaking of the Pentagon papers drove a wedge between RAND and academia. RAND sought to broaden its research outside of defense issues and became more cautious in its relations with academia. Daniel Ellsberg was, after all, no longer a RAND employee at the time he leaked the Pentagon papers; he was an associate of MIT’s Center for International Studies. Academia moved in the opposite direction, with classified research being prohibited at many universities and even defense issues generally shunned by many.5

Despite these changes, in the 1980s, RAND and its alumni continued to guide U.S. deterrent posture. Andrew Marshall, who had been at RAND almost since its inception, had become an important figure in the Pentagon after Schlesinger named him director of the new Office of Net Assessment in 1973. He continued in this position in the Reagan administration and provided guidance on further improving U.S. strategy for deterring nuclear war. RAND alumni, such as Fred Iklé and F. J. “Bing” West also served in senior DoD positions during the Reagan administration.6

5 See Wells (2001) on Ellsberg; on university responses, see Leslie (1993).
6 Iklé was undersecretary of defense for policy. West was assistant secretary of defense for international-security affairs.
In examining prior research on deterrence, the question of its relevance is almost immediately raised. The Cold War was a time of intense bipolar competition between enormously powerful rival states that were also ideological opposites. The long war appears murkier, involving possible peer or near-peer competitors, regional adversaries, and nonstate actors, such as insurgents, terrorists, tribal groups, criminals, and militias. At the same time, the United States towered over all other states in almost every measure of military capability. Given these radical differences, can cold-war deterrence tell us anything about deterrence in the post-9/11 era?

The answer is a qualified yes. Deterrence may have assumed a paramount place in the nuclear standoff that the Cold War eventually became, but it was by no means a new phenomenon. Thucydides, in his *History of the Peloponnesian War*, quoted Hermocrates of Syracuse:

Nobody is driven into war by ignorance, and no one who thinks that he will gain anything from it is deterred by fear. The truth is that the aggressor deems the advantage to be greater than the suffering; and the side [that] is attacked would sooner run any risk than suffer the smallest immediate loss . . . [W]hen there is mutual fear, men think twice before they make aggressions upon one another. (Thucydides, 1998, book IV, pp. 59–62).

Deterrence, though not nearly so refined a concept as later analysis would make it, thus clearly predates not only the nuclear age and cold war, but the modern era itself.
Because deterrence is a long-standing concept, the issue of applying cold-war research on deterrence to future challenges becomes one of separating the general phenomenon from the specific circumstances. This problem is one of historical analogy and the careful use of history. History for social scientists serves the purpose that experimentation serves in the physical sciences, yet it cannot be manipulated in the same way. This reasoning by historical analogy is a common social-science and policy-analysis technique.¹

The bipolar nature of U.S.-Soviet competition combined with the level of perceived threat on both sides is perhaps the most obvious of the specific circumstances. It seems likely (if a bit obvious) that lessons from the Cold War will be most applicable to situations that most resemble this competition. Determining resemblance then requires categorization of deterrence relationships to tell which most resemble the cold-war environment.

For purposes of this discussion, deterrence challenges in the current and future environments will be divided into three broad categories. These categories, in descending order of similarity to the Cold War, are peer or near-peer competitor, regional power, and significant nonstate actor. The peer or near-peer competitor is defined as a state or collection of states with the power to effectively challenge U.S. interests on a global scale (see Szayna et al., 2001). The regional power is defined as a state or collection of states with the capability to effectively harm U.S. interests in its own region (see Watman et al., 1995). A significant nonstate actor is defined as a group or collection of groups with the ability to effectively hurt the United States or its allies, either globally or regionally. These categories provide a framework for discussion of the policy relevance of lessons from the Cold War in the second half of this book. The next section discusses both the theoretical and the practical components of deterrence, beginning with a definition of the term.

¹ There is a considerable body of work on historical analogy and foreign policy. See, for example, Khong (1992), Ernest May (1973), and Neustadt and May (1986).
CHAPTER THREE
Department of Defense as Ministry of Fear: The Theory of Deterrence

A widely used definition of *deterrence* is the manipulation of an adversary’s estimation of the cost/benefit calculation of taking a given action. By reducing prospective benefits or increasing prospective costs (or both), one can convince the adversary to avoid taking the action. Yet this relatively bloodless definition can be simplified and made more visceral: Deterrence is the generation of fear.

RAND, home to engineers, scientists, and economists, used the more complex version of deterrence in much of the work discussed here. Yet it is clear that they never lost sight of the visceral definition. Such words as *terror* and *fear* are not uncommon in RAND’s work on deterrence. Albert Wohlstetter, one of the deans of nuclear strategy at RAND, produced the seminal article “The Delicate Balance of Terror” very early in the Cold War (Wohlstetter, 1958). Thomas Schelling’s Nobel Prize–winning work on game theory includes a chapter titled “The Reciprocal Fear of Surprise Attack,” which emphasizes the effect of nervousness and fear on deterrence (Schelling, 1960). In a list titled “Desirable Characteristics of a Deterrent,” Herman Kahn placed “frightening” first (Kahn, 1961, p. 146).

Given that *deterrence* shares its Latin root with *terror*, this is perhaps unsurprising and may seem merely pedantic. However, this duality in deterrence is more than semantic hair-splitting. While deterrence is clearly rooted in thought and calculation, it inherently contains an element of emotion as well. Fear exists in the mind of individuals, giving it qualities that are sometimes idiosyncratic or not easily observed from outside. Further, the emphasis on fear illustrates the
importance of the unknown in deterrence. Cost/benefit calculation relies on known inputs, while much of deterrence rests on uncertainty about those inputs. This duality is important to bear in mind when thinking about deterrence.

Further, the fear characteristic is particularly paramount in the arena of nuclear deterrence. Atomic and, later, nuclear weapons enabled the devastation experienced by Europe and Asia during World War II to be both magnified in scope and compressed in time. What had taken years could now be done in hours or even minutes. Brodie’s choice of title for what is perhaps the first real examination of the nuclear revolution, *The Absolute Weapon*, aptly captures the fact that the fear of even the possibility of nuclear devastation was so great that it made war much less likely. Few benefits could rival this potential cost.

Deterrence, as both the manipulation of cost/benefit calculation and the generation of fear, is a form of coercion (or, more colloquially, blackmail).¹ Like all forms of coercion and threat, it requires two elements: the credible capability to harm and the credible intent to carry out this harm. Kaufmann noted in 1958,

> deterrence consists of essentially two basic components: first, the expressed intention to defend a certain interest; secondly, the demonstrated capability actually to achieve the defense of the interest in question, or to inflict such a cost on the attacker that, even if he should be able to gain his end, it would not seem worth the effort to him. (Kaufmann, 1958, p. 2)

Kaufmann’s definition points out that an intention includes two parts: “an expressed intention” and a “certain interest.” The first part is a declaratory policy that makes clear what is to be deterred. While this may be simple in theory, ambiguous declaratory policies made by the

¹ Betts (1987, pp. 3–7) has perhaps the best discussion on deterrence as blackmail. Coercion differs from what Schelling refers to as *brute force*, in that it involves the threat of future harm rather than the actual inflicting of harm. Brute force, in contrast, simply inflicts harm to achieve results. If coercion is blackmail, then brute force is smash-and-grab robbery.
United States appear to have contributed to the initiations of both the Korean War and the 1991 Gulf War.2

At the same time, as Schelling noted, deterrence is often enhanced by not being entirely clear in declaratory threats, instead “leaving something to chance.”3 This again highlights the importance of the unknown in deterrence. If an opponent believes that taking an action that one wishes to deter will set in motion events that may escalate beyond the control of both parties, then uncertainty will make him less likely to take an action. This is particularly effective if the deterrent threat is matched by a counterthreat of equal or greater magnitude from the opponent. While it might be irrational (and therefore not credible) for one to intentionally use a deterrent threat knowing that it would invite one’s own destruction, it is arguably more credible for one to argue that the deterrent threat might be used unintentionally as a result of escalation. Of course, relying on one’s own response to provocation during a crisis becoming uncontrollable as way to deter would be deeply troubling to many, including some policymakers.

An additional point on intention and deterrence is that, as the upholding of the status quo, deterrence is generally held to be easier than changing the status quo by coercion (which is termed compel-lance). Experiments in a type of behavioral economics known as prospect theory provide some insight into why this is the case. Humans as a rule tend to be risk-acceptant when facing loss and risk-averse toward gain. As long as maintaining the status quo is not a clear path to loss, most people will be risk-averse in taking steps to upset it.4

The second part of intention, interest, is more ambiguous. As discussed next, there is seldom any way of directly measuring interest. The

---

2 In a speech in January 1950, Secretary of State Dean Acheson’s remarks appear to place South Korea outside of U.S. interests. A meeting in July 1990 between U.S. Ambassador to Iraq April Glaspie and Saddam Hussein appears to have downplayed U.S. interest in the Iraq-Kuwait dispute. In both cases, the available evidence suggests that neither Acheson nor Glaspie’s remarks were a “green light” for aggression but were ambiguous enough to reduce the efficacy of deterrence.

3 Schelling (1960, chapter 8) is titled “The Threat That Leaves Something to Chance.”

4 See Kahneman and Tversky (1979) and Farnham (1994). Of course, two parties may not even agree on what “the status quo” is, making deterrence more difficult.
most prominent question of interests for the Cold War was simple, yet hard to evaluate. It was in the interest of the United States to ensure that Western Europe was not dominated by a hostile power, but it was not clear to what extent that interest exceeded the interest in not losing millions of citizens. Other theorists picked up on this question.

Kaufmann’s definition also includes two distinct forms of deterrent capability. The first is “the defense of the interest in question,” which has been termed deterrence by defense or deterrence by denial (see Snyder, 1961). The logic is simple here in theory as well; what enemy would start a war if it knew for certain that the war would not achieve its aims (unless perhaps the war was expected to be inexpensive)? However, the nuclear revolution made deterrence by denial unpromising, particularly after the creation of intercontinental ballistic missiles (ICBMs). Only late in the Cold War would a form of deterrence by denial return.

Thus the second form of deterrence, the threat to inflict harm, dominated much of the development of cold-war thinking on deterrence. Of course, it is not just the ability to inflict any harm that is important for deterrence. The harm inflicted must be of sufficient magnitude to overcome potential gains from taking the action that is to be deterred.

Additionally, one aspect of deterrence that is sometimes downplayed is that of reassurance. Thomas Schelling noted the critical importance of reassurance to deterrence, as it means that the threatened harm will not be implemented if deterrence holds. Without credible reassurance, there is no incentive to comply with deterrent demands (Schelling, 1966, p. 74).

The importance of both the fear inherent in deterrence and the need for credible reassurance leads to one of the central insights of deterrence theory: the concept of the security dilemma. The security dilemma is a situation in which one party, by taking action that enhances its own security, makes another party feel less secure (see Herz, 1950, and Jervis, 1976). So country A’s seeking to build capabilities to deter action by country B frightens B and stimulates a similar response in B, which in turn makes another increase in A’s capabilities likely. This action-reaction process is termed the spiral model (see
Jervis, 1976, Chapter 3). Without credible reassurance, attempts to promote deterrence can backfire and make both parties less secure.

Credibility is thus the linchpin of deterrence, particularly the credibility of the threat (as without a credible threat, the credibility of reassurance is relatively unimportant). Reflecting on threat estimation in the Cold War, former RAND analyst Raymond Garthoff reiterated this point:

The question of Soviet intentions and attendant objectives was the fundamental element of threat assessment. Soviet military forces and capabilities to carry out Soviet leaders’ intentions necessarily constituted the second, but crucial element of that assessment. (Garthoff, 2003)

Yet the credibility of any threat can be hard to estimate in practice. The United States spends billions annually to estimate threats, yet its track record is mixed. Though some argue that this is indicative of incompetence in the intelligence community, this mixed track record is more a result of the incredible difficulty of the task. However, the efforts of analysts and scholars have produced some general propositions for estimating the credibility of both capability and intentions.

Ascertaining the credibility of capability is often easier than determining the credibility of intentions, if only because there are tangible elements, such as force structure, associated with capability. Generally, credible capability is composed of the following three elements: aggregate forces, proximity, and power-projection capability. Aggregate forces is an extremely broad concept that includes everything from static “bean counts” of number of people and pieces of equipment to more dynamic factors, such as training, leadership, and doctrine. Geographic proximity makes the use of force easier, so a nearby threat is more credible than a distant one. Power-projection capability is comprised of those

---

5 For an overview of many of these elements, see Millet, Murray, and Watman (1986) and Van Creveld (1982). For a more general model of modern conventional warfare, see Biddle (2004, Chapters 3, 4, and appendix).
forces that mitigate distance. It includes logistics, mobility and transportation assets, and long-range strike systems.\(^6\)

Of course, capability is not one-sided; it must be evaluated with respect to the other side’s capability. RAND’s Andrew Marshall noted in 1966,

> Estimating the military power of the United States, or any other country, can only be done relative to another country, or set of countries viewed as an alliance. . . . [M]ost attempts to explicitly measure military power are mere tabulations of forces of various sorts: the number of men under arms, the numbers of weapons of a given type, etc. This is itself an evasion of the problems of estimating military power, since it says nothing about the actual capabilities of the forces of one country to deal with another. (Marshall, 1966, p. 2)

This dynamic evaluation is the essence of what has come to be called net assessment\(^7\).

By performing a net assessment of those forces involved in a threat, an estimate of the credibility of capability can be determined (though even here, assessment can fail—few expected the rapid German victory of May 1940). Analysts using different methodologies may differ on the exact capability of a given threat, but they often agree on the general magnitude. For example, one can examine net assessments of the NATO–Warsaw Pact conventional balance in central Europe during the 1970s and 1980s. Though varying in degree, these assessments began to shift away from overwhelming pact superiority as the NATO countries, particularly the United States, invested heavily in improving conventional forces. Yet the pact remained a highly credible threat to all analysts.\(^8\)

---

\(^6\) See Posen (2003) and Thompson (1978) for discussion of some of the components of power projection.

\(^7\) For a good review of late–Cold War thinking on military assessment, see Friedberg (1987–1988).

\(^8\) See, inter alia, Canby (1973) and Mearsheimer, Posen, and Cohen (1989) for the evolving debate.
If credibility in the realm of capabilities is murky despite its tangibility, credibility in the intangible realm of intentions is much more opaque. This is particularly true when deterrence is extended beyond the borders of an individual nation-state. Schelling noted, “the difference between the national homeland and everything ‘abroad’ is the difference between threats that are inherently credible, if unspoken, and the threats that have to be made credible” (Schelling, 1966, p. 36). This question of level of interest led Schelling to distinguish between a warning and a threat. A warning sought to convey the deterrer’s true and inherent interest. A threat, in contrast, conveyed the deterrer’s commitment to a position that was not clearly in its true and inherent interest (Schelling, 1960, pp. 123–124).

The U.S. interest in protecting its population and territorial integrity was assumed to be fundamental. This was termed basic (or type I) deterrence and was (and is) considered a highly credible intention. Declaring an intention to retaliate for an attack on U.S. territory was no threat in Schelling’s formulation; it was a warning.

However, the United States wanted to extend its nuclear deterrence beyond its own borders. This would require one of Schelling’s threats, and the notion of threat beyond the homeland, referred to as extended (or type II) deterrence, would be perhaps the central concern of deterrence theory in the Cold War. How could the United States convince the Soviet Union that it would attack Moscow if Berlin were attacked, particularly once the Soviet Union could strike Washington in return? William Kaufmann charitably described this process as the “difficult and delicate problem of making intentions credible” (Kaufmann, 1956, p. 19). Bernard Brodie described threats of this nature as having an air of unreality about them (Brodie, 1958, p. 5). Paul Kecskemeti perhaps went furthest of all:

Blue [the deterrer], however, must assume unconditional commitments overriding the maximization principle. The strategy of deterrence cannot work unless such unconditional commitments are built into it. To fight back if directly attacked, for example, is an unconditional commitment of this kind; to honor alliance obligations is another . . . it is clear that decisions stemming from unconditional commitments are not rational. We shall say
that they represent a non-rational element in political conduct.
(Kecskemeti, 1960, pp. 14–19. Emphasis in original.)

Basic deterrence was nonrational but credible (though some would even question the credibility of basic deterrence). However, the same could not be said for extended deterrence.

Yet making these nonrational intentions credible was critical, so understanding the causes of credibility of intention became equally critical. How, then, would a government know a credible intent if it saw one? Reputation was posited very early as a source of credibility. Thomas Schelling declared, “what one does today in a crisis affects what one can be expected to do tomorrow” (Schelling, 1966, p. 93). This notion of reputation was intuitively appealing and appeared to be supported by historical evidence, most especially the appeasement crises of the 1930s that culminated in Munich.

The defense of reputation, however, is problematic. In the United States and Soviet Union alike, concern about the interdependence of commitments provided a large part of the rationale for costly interventions in Vietnam and Afghanistan. Yet these interventions did little to enhance extended deterrence and probably little to detract from it either (at least in the areas of highest concern, such as central Europe). The weakness of reputation was further explored in the game-theory work of Reinhard Selten, who argued that attempts to establish reputation were irrational from a cost/benefit perspective (though Selten himself felt that this said as much about game theory as about reputation) (Selten, 1978).

Further, reputation must be managed with one’s allies as well as one’s adversaries. The politics of alliance were as important to deterrence as any other element. If the Western Europeans did not believe in extended deterrence, it was almost irrelevant whether the Soviets did, as the Europeans might simply drift into the Soviet sphere of influence (a worry termed Finlandization). Further, all allies must share an

---

9 Interestingly, both the domino theory and the Brezhnev Doctrine were based in large part on the logic of reputation, yet many U.S. and Soviet officials and analysts felt great trepidation when confronted with intervention on those grounds. See Ford (1998, Chapters 1 and 2) and Westad (1996–1997).
understanding of what is to be done in a crisis and the forces that will be involved. NATO, for example, maintained a nuclear planning group that helped ensure that the allies had similar views on the requirements of deterrence.\textsuperscript{10}

If reputation is, at best, only somewhat helpful for the credibility of extended deterrence, are there other mechanisms that can be used to make these incredible intentions credible? Making the threat response automatic and hence disconnected from cost/benefit calculation is one possible method. Schelling and Kahn both discussed this concept, and it is parodied in \textit{Dr. Strangelove}, whose titular character notes, “because of the automated and irrevocable decision making process [that] rules out human meddling, the doomsday machine is terrifying. It’s simple to understand. And completely credible and convincing.”\textsuperscript{11} Yet Strangelove and RAND analysts alike conclude that this method is impractical.

Another strategy that Schelling discussed was embracing non-rationality and simply giving the impression that U.S. leadership was crazy. He, along with most other analysts, found this approach unpromising, though it is worth noting that recently declassified material suggests that, in 1969, by placing nuclear forces on high alert, Richard Nixon tried to convince the Soviet leadership that he was unstable (Burr and Kimball, 2003). Instead, leaders would “have to substitute brains and skill for obstinacy and insanity” (Schelling, 1966, p. 42).

It would, in many ways, fall to the RAND Corporation to provide “the brains and skill” of which Schelling wrote. Ultimately, making extended deterrence fully credible may simply be impossible. Certainly, it remained a problem throughout the Cold War, with European fears of abandonment never totally abating. Yet extended deterrence appears to have succeeded in keeping the peace for decades, weathering crises and shifts in the balance of power. The next two chapters discuss why deterrence mattered to the United States and then provide a discussion of RAND research into the factors that helped shape this remarkable success.

\textsuperscript{10} On NATO and nuclear forces for deterrence, see Legge (1983) and Thomson (1982).

CHAPTER FOUR
Avoiding the Garrison State:
Deterrence as a Strategy

Before discussing policies intended to make deterrence credible, a brief discussion of why deterrence was important in the first place is appropriate. This may seem obvious in retrospect, but a strategy based on deterrence was not a foregone conclusion. The postwar grand strategy of containment with its reliance on nuclear deterrence was not the only strategy available. The United States could have prepared much more fully for a protracted conventional war or done more to prepare for fighting a nuclear war (through active and passive defenses, for example) or both. That it did neither of these was not an accident. Rather, fear of becoming what sociologist Harold Lasswell termed a “garrison state” limited the drive for resource extraction that such an effort would have entailed (see Lasswell, 1941). U.S. grand strategy was not an optimal strategy for national security; rather, it was a “strategic synthesis” in response to both international threat and domestic pressures.¹

Several brief examples illustrate this point. First, despite its perceived importance to future war (conventional and, to a lesser extent, nuclear), universal military training was never adopted (Friedberg, 1992, pp. 125–128). Similarly, efforts to introduce massive civil-defense programs (advocated in many cases by RAND analysts) were never enacted, despite the significant reduction that civil defense offered in U.S. vulnerability to nuclear attack.² Any version of contain-

¹ This interpretation of the interaction of domestic and international pressure is drawn from Friedberg (2000). See also Hogan (1998). The term strategic synthesis, which Friedberg used, is from Milward (1977).

² For example, see Kahn (1958).
ment other than significant reliance on nuclear weapons would have required taking such major steps as these, which few were willing to countenance. In short, other strategies would have required a remaking of U.S. society beyond all recognition, essentially producing a United States that was a mirror image of the USSR.

As another alternative, the United States could have launched a preventive war on the Soviet Union before the Soviets developed a significant nuclear capability. This would have obviated the need for both containment and deterrence. It was even considered by some and advocated by a few in the United States. Preventive war would have eliminated the Soviet threat once and for all, though at high cost. As discussed more next, this strategy was ultimately rejected, as were subsequent plans to launch preventive war on China before it went nuclear in 1964.³

Deterrence based on nuclear weapons, in contrast to other strategies, promised to allow the United States to rely on sophisticated but relatively small forces in being. These forces would consist primarily of the nuclear forces of the Strategic Air Command (SAC) and a standing conventional force in Europe. These forces, while massive by U.S. historical standards, were sustainable without radical societal change. Resource extraction and social alterations, such as the peacetime draft, could be minimized, and the long-run health of the country ensured. Nuclear weapons, though a source of vulnerability for the United States, could be combined with a deterrence strategy to provide security at a reasonable cost. The increase in peacetime military forces and the fear of atomic annihilation that deterrence brought with it were a large price to pay, but less than the alternatives.⁴ To paraphrase Churchill, deterrence was the worst strategy, except for all the others.

This willingness to assume the risk of deterrence failure in order to maintain the preferred societal arrangements carried with it the seeds of an important strategic advantage. If the United States could live with some risk, it could extract fewer resources for national defense

---


⁴ For history of the political development of this compromise position, see Hogan (1998).
in the short run. This, in turn, meant more resources for investment in other areas, which would contribute to more economic growth and increased productivity. A stronger economy meant that more resources would be available for defense. This virtuous circle would allow equivalent (or greater) absolute levels of defense spending, even as the relative extraction of resources fell.

For example, post–World War II U.S. defense spending as a percentage of gross domestic product (GDP) peaked in fiscal year (FY) 1953 at about 14.2 percent. In absolute terms, this was equivalent to about $416 billion in FY 2000 dollars. This peak quickly dropped to around 10 percent of GDP and slowly declined for the remainder of the Cold War. Thirty years after the peak, in FY 1983, U.S. defense spending as a percentage of GDP was about 6.1 percent. Yet in absolute terms, it was about $331 billion in FY 2000 dollars. Had the United States wished to spend more in absolute terms in 1983 than it had in 1953, it could have done so for about 7.7 percent of GDP or slightly more than half the relative level of 1953.\(^5\) The “weakness” of the U.S. state and the desire to preserve U.S. society had allowed the logic of compound interest to work in the favor of U.S. defense.\(^6\)

This strategic advantage was important, as the Cold War would be—as Eisenhower and others observed very early—“a long pull.” As discussed earlier and despite Eisenhower’s relative optimism, many in the early stages of the Cold War doubted that it would be a long pull, leading some to advocate preventive war against the Soviets (see Trachtenberg, 1988–1989). However, as the contest stretched on, the nature of this long-term competition between the United States and the Soviet Union became a topic of increasing interest at RAND. If U.S. state-society arrangements produced a virtuous circle, could the

\(^5\) Numbers from the Under Secretary of Defense (Comptroller) (2006, Chapter 7).

\(^6\) However, even as the defense component of the consumption of GDP declined, the total consumption of U.S. GDP by the federal government grew slightly. For example, in FY 1953, the entire federal government consumed about 20.4 percent of U.S. GDP, most of which was on defense. In FY 1983, it consumed about 23.5 percent, even though defense spending had declined as a percentage of GDP. Much of this increased consumption was due to increased nondiscretionary spending, such as social security and Medicare and Medicaid. See OMB (2006). FY 1983, interestingly, was the peak of federal expenditures since FY 1948.
Soviet Union be maneuvered into a vicious circle of increasing military budgets and declining economic growth? The Soviets, lacking the antistatist elements of U.S. society, were inherently more willing and able to extract greater resources for defense. If the United States could find areas of competition in which it held a comparative advantage and exploit them, then it could perhaps force the Soviets to spend disproportionately in competing in that area. This would, in turn, leave the Soviets the choice either of foregoing other uses of defense spending or of extracting more resources from Soviet society and economy.7

Air defense provides a good example of this phenomenon. As Thomas Schelling observed in the late 1950s,

> If at little cost we can force him into a costly diversion of air defense resources it may look like a good idea. But, if at great cost we force to divert a small amount of resources, it does not look good. (Schelling, 1964, p. 211)8

But would the Soviets keep spending on air defense if it was disproportionately costly? It seemed possible that it would, because the Soviets made strategic air defense a separate armed service very early in the Cold War.9 As such, it had developed an institutional interest in and bureaucratic infrastructure dedicated to air defense. Soviet bureaucratic interests would ensure that high levels of Soviet air-defense spending would continue, rather than simply abandoning the bulk of air defense when it became a losing proposition (as the United States did).

---

7 See, inter alia, Becker (1981) and Ofer (1980). Note that estimates of Soviet economy and military spending were often controversial, even among RAND analysts and alumni. For example, Abraham Becker thought the CIA’s relatively conservative estimates of Soviet military expenditure and positive estimates of Soviet economic growth to be roughly accurate. Other RAND assessments, such as Wolf et al. (1983), noted that CIA estimates did not include significant costs that served as a drag on the Soviet economy. I thank Andrew Marshall, who sponsored The Costs of the Soviet Empire, for emphasizing this point.

8 This volume collected a set of RAND lectures and commentary from 1955 and 1959.

9 Known for most of the Cold War as Voyska Protivovozdushnoy Oborony Strany (Troops of National Air Defense) or simply PVO Strany. On the bureaucratic character and influence of PVO Strany, see Lepingwell (1988).
In short, the United States could provoke a vicious circle if it could find ways to threaten air defense at a lower marginal cost than the cost to defend against the threat. Andrew Marshall argued that the United States should exploit this possibility:

The massive Soviet Air Defense effort probably has cost the Soviets more than it is worth. . . . We can take advantage of this Soviet tendency to keep resources flowing where they are currently going (this is, of course, not solely a Soviet tendency). Specifically, as regards air defense, the extent that we really feel that they are overspending relative to other defense expenditures, probably we can keep that part of their defense budget above what it ought to be with minimal expenditures on our own side. Therefore, part of a strategy for the long-term competition would involve looking for areas where we would like to keep them spending resources, and finding U.S. actions that would keep them spending in those areas. (Marshall, 1971, p. 25)

The United States proceeded to do exactly this, though admittedly only in a small part, due to strategic calculation. Both cruise missiles and “stealth” technology created a challenge that Soviet air defense spent heavily attempting to counter.10 This was, in some sense, a form of “virtual attrition,” in which the response to a threat reduced overall enemy capabilities by spreading resources thin.

RAND research on long-term competition (also termed competitive strategies) expanded to include not just Soviet military expenditures, but also what came to be termed the costs of Soviet empire (Wolf et al., 1983). This included all the various subsidies to its satellites, as well as operations in support of the Brezhnev Doctrine, such as the war in Afghanistan. If the United States could raise these costs as well, then the drag on the Soviet economy would be even higher.

Yet all of these advantages hinged on the competition remaining a peacetime competition primarily based on the maintenance of the status quo through deterrence. In other words, containment would

---

10 On the possible expense of countering cruise missiles and stealth technology in the late Cold War, see Quinlivan (1989).
have to succeed. While the U.S. intention of supporting basic deterrence was quite credible, the intention of extending deterrence to Europe, as discussed previously, has inherently limited credibility. To shore up this weak credibility in the intention part of the threat, the United States would have to compensate in the capability portion of the threat. For at least the first two decades of the Cold War, it was widely (though not universally) conceded that the Soviet Union and Warsaw Pact had conventional superiority in Europe. As noted, the United States was not willing to extract the resources required to attain conventional parity, at least for the first few decades of the Cold War. U.S. capability to deter would thus rest principally on nuclear weapons, including the possibility of being the first to use nuclear weapons in a conflict. It would improve deterrence using nuclear weapons both by directly improving elements of capability and in better understanding the thinking of the enemy, to ensure that it developed capabilities that would be feared.

One key way in which the credibility of extended deterrence could be improved was to be able to strike the Soviet Union without sustaining a counterblow. This meant that extended deterrence rested on the bedrock of nuclear superiority (what Kahn termed a **credible first strike**) (Kahn, 1961, pp. 27–36). From 1945 to the early 1950s, this was clearly the case, as Soviet long-range nuclear assets were nonexistent or embryonic. As long as this was the case, the Soviet leadership faced the prospect of near-certain annihilation for any conventional aggression.

Yet this superiority was seen by many to be “a wasting asset,” and would soon vanish as the Soviets built up their nuclear arsenal. Two ways to handle this problem were readily apparent in theory but difficult in practice: strategic defense and a preemptive disarming or damage-limiting first strike. A third option was to rely on tactical nuclear weapons used on the battlefield. These technical and doctrinal aspects, as well as some that emerged later are the subject of the next chapter.
The first possible solution to maintaining a credible first-strike capability was to defend against Soviet retaliation. In the early 1950s, this would mean downing most if not all of the Soviet intercontinental bombers and medium bombers that could be sent on a one-way trip to the United States. RAND was already at work on air defense well before the Soviet Union had even broken the nuclear monopoly. One of the first RAND efforts ever, initiated in 1947 and completed in 1948, was titled *Active Defense of the United States Against Air Attack*. Though brief, it attempted to develop the theoretical and practical groundwork for air defense. A subsequent effort simply titled *Air Defense Study* was completed in 1951 and went into much greater detail on the problems and prospects for air defense (Barlow and Digby, 1951).

Other air-defense studies were under way in the early 1950s. The most important of these was Project Charles, conducted at the Massachusetts Institute of Technology (MIT) in 1950–1951. Project Charles’ conclusions were that air defense was possible on a continental scale if significant resources were dedicated to it (see MIT, 1951). From this recommendation would grow MIT’s Lincoln Laboratory and subsequently the MITRE Corporation, as well as the Defense Early Warning line of radar stations and the Semi-Automatic Ground Environment (SAGE) computerized system. RAND would be significantly involved in this process, providing many of the programmers for the
Whirlwind computer at the heart of SAGE as well providing numerous recommendations on the overall system.\(^1\)

However, the explosive pace of technological growth during this period soon cast doubts on the efficacy of air defense. Within two years of the initial studies, the Soviets tested their first hydrogen bomb, providing a multiple-orders-of-magnitude increase in the destructive power of weapons. At the same time, the possibility of developing ICBMs looked increasingly likely, as the work of RAND physicist Bruno Augenstein demonstrated.\(^2\) In the judgment of many at the time, the two combined would render air defense obsolete. RAND’s Edward Barlow arrived at this conclusion in 1953, arguing that, while air defense should still be pursued in the short run, missile defense would soon be needed (Barlow, 1953).

The problems of defense against piloted bombers paled in comparison to the challenge of ballistic-missile defense. RAND proposed possible solutions, but few felt them to be truly promising. The problem was that a defense that was effective against 90 percent of missiles fired against it was still a losing proposition for cities. If 200 missiles carrying multimegaton nuclear weapons were launched against such a system, the United States could still expect to lose 18 cities and millions of lives. If such a defense could be built cheaply, then perhaps it would be worthwhile, but it promised to cost far more than Project Charles’ already expensive air-defense system.

Further, ballistic-missile defense was on the losing side of competitive strategy. It was far cheaper to build a ballistic missile than defend against it. Decoys (termed penetration aids) could be added to ballistic missiles even more cheaply. Finally, the development of multiple warheads for each missile proved to be yet more cost-effective, making the saturation of defenses all but inevitable.\(^3\)

---

\(^1\) On this period, see Schaffel (1991) and Redmond and Smith (2000).

\(^2\) Augenstein did pioneering work on long-range missiles at RAND, beginning with *Long-Range Surface-to-Surface Ramjet Missiles: Preliminary Investigations and Results* (Augenstein, 1948). His most important work was *A Revised Development Program for Ballistic Missiles of Intercontinental Range* (Augenstein, 1954).

\(^3\) See Greenwood (1975) for the history and evolution of multiple-warhead technology.
Finally, anti-ballistic missile (ABM) systems would have to be based near U.S. cities. This proved to be a problem, as some cities objected to this. While nuclear weapons could remain comfortably distant at SAC bases and missile fields in the sparsely populated middle of the United States or on submarines, interceptors near cities provoked more of a response. When the plan to construct a Sentinel ABM system near Boston was announced in 1968, at the height of the Vietnam War, it provoked protests that eventually led to the suspension of construction at the site. The program was soon terminated.

However, these problems held primarily for cities, which were big, soft-area targets with populations that could protest. Missiles in hardened silos or aircraft in hardened shelters were a different proposition. Many at RAND, including Albert Wohlstetter, felt that missile defense of strategic forces would be worthwhile to ensure U.S. retaliatory capability against Soviet surprise attack. Unfortunately, a second strike secured by missile defense might be a good way to secure basic deterrence, but it did not seem a promising way to support a credible first-strike capability.

The Least Miserable Options: Counterforce, Limited Nuclear Options, and Deterrence

The second answer to the problem of damage limitation for a credible first strike was to obtain the ability to destroy Soviet long-range nuclear systems before they could be launched. Some at RAND envisioned this ability, which would become known as counterforce, very early. Andrew Marshall and Joseph Loftus, in particular, were to become advocates of counterforce by the early 1950s.

---

4 See Herken (1987, chapter 20) for discussion of Wohlstetter and the debate about the Safeguard ABM system intended to protect land-based missiles.

5 See Loftus (1955, 1959) as well as Herken (1987, pp. 79–81). RAND analyst Victor Hunt also helped develop the conceptual foundation for counterforce before his life was cut short by illness.
Before discussing the often-controversial topic of counterforce, three points need to be addressed. The first point is a brief digression on strategy and doctrine to clear up some confusion on deterrence and counterforce. Some scholars and analysts have framed nuclear strategy as a choice between a deterrent strategy and an offensive strategy. Counterforce, in this framework, is an offensive strategy. Yet this is a false distinction. Counterforce was not a strategy, in the sense of a full means-ends chain to security. An offensive strategy would have been a preventive war, which would have eliminated the threat via offensive action. This had been rejected in favor of containment and deterrence.

Instead the debate over counterforce is properly framed as one between stability and utility. Those who felt stability in the nuclear balance was most important were against first-strike counterforce, as it created “use-or-lose” incentives for one or both sides. In other words, getting off the first salvo would cripple the other side, while not getting off the first salvo meant that one’s own forces would be crippled. This put a high premium on launching first in a crisis and thus made crises unstable (thus the term of art crisis instability), as one side might feel that it had no choice but to fire first. Instead of first-strike counterforce, both sides should seek survivable second-strike forces, to ensure that no one had incentive to launch a catastrophic war.

In contrast, first-strike–counterforce advocates felt that they were proposing an enhancement of the utility of nuclear weapons for extended deterrence. In the days of U.S. nuclear monopoly and near-monopoly in the early 1950s, a doctrine of pure deterrence by punishment was sufficient, as the United States could launch it without fear of reprisal. In a crisis or war, a counterforce doctrine sought to shift the balance of nuclear forces back to that of the early 1950s through offensive action, so that extended deterrence by threat of punishment would again be effective. Counterforce did not represent a change in strategy,

---
6 This characterization draws heavily on Trachtenberg (1989), especially pp. 322–327.
7 For two prominent examples of those who felt counterforce to be of limited utility while decreasing stability, see Jervis (1984, pp. 70, 112–114) and Glaser (1990, Chapter 5). The two arguments are not identical, as Jervis conceded the possibility that first-strike counterforce could have utility. In contrast, Scott Sagan argued for second-strike “slow” counterforce; see Sagan (1989).
but rather a change in doctrine; it would be an *offensive* doctrine and force structure in support of a *deterrent* strategy.\(^8\) One of the downsides of this offensive doctrine, as discussed in more detail later, is that it tends to provoke or amplify the security dilemma.

The second point is to characterize first-strike counterforce based on its expected efficacy. A preemptive first strike that was expected to destroy virtually all of an opponent’s weapons capable of reaching the United States would be considered *disarming*. A strike expected to be less effective but nonetheless able to significantly shift the nuclear balance and reduce the enemy’s ability to strike back at the United States would be characterized as *damage limiting*. This distinction is important, as disarming counterforce became increasingly unlikely as the Cold War progressed, but damage limiting did not disappear (or so many thought; nuclear operations still remain thankfully untested).

The third point is to address objections that the U.S. government all but gave up on first-strike planning and counterforce options in the 1960s. The public statements of many U.S. government officials, including former secretaries of defense, over the past 40 years appear to support this position. If this is true, then preemptive counterforce appears to have been only a minor part of cold-war deterrence.

However, a variety of declassified documents show that, as is often the case, public rhetoric and classified planning are often different. Similarly, the development of U.S. force structure is at odds with these public statements. There are many reasons for such divergence, and to note it is not to accuse U.S. government officials of lying. Rather, it shows that public statements are for a variety of audiences (the U.S. public, allied governments, adversaries) and are made under a variety of understandable security restrictions. Further, senior officials and other participants in nuclear planning may have intended one outcome yet, for bureaucratic, political, and technical reasons, ended at another.

The story of counterforce is thus still a contested history. This ongoing debate is briefly but cogently captured in correspondence between current and former government officials Peter Flory and Keith

\(^8\) See Ravenal (1982) for a lengthy discussion of the links between extended deterrence and counterforce.
Payne on the one hand and Keir Lieber and Daryl Press on the other, published in the September–October 2006 issue of the journal Foreign Affairs (Flory et al., 2006). Flory, at the time assistant secretary of defense for international security policy, explicitly stated that the United States was not postured for a first strike. Payne amplified this point and argued that preemptive-counterforce thinking vanished in the 1960s. Lieber and Press argued that it blossomed in the 1970s and 1980s, pointing to declassified planning documents as well as force structure. This monograph, based on available evidence, takes the position that preemptive-counterforce options were a major part of U.S. efforts to deter the Soviet Union for the entire Cold War, while acknowledging that many disagree with this position.

Regardless of whether it would be pursued or even possible in the future, and perhaps reflecting the uncertainty that surrounds a type of warfare that has never been conducted, counterforce was a much-debated topic in the early 1950s. Counterforce as envisioned by Marshall and Loftus was controversial at RAND during this period; many thought it a good but impractical idea. The primary reason that counterforce was considered unworkable was the lack of intelligence on the location of Soviet nuclear systems. However, Marshall and Loftus had access to intelligence that most at RAND did not; Loftus had previously been director of target intelligence for the U.S. Air Force, and Marshall had ties to the intelligence community and was able to get the clearance necessary to see targeting intelligence.9 This intelligence included imagery of Soviet bases captured from the Japanese and Germans, which was supplemented by prisoner-of-war interrogations as well as dangerous penetrations of Soviet airspace by U.S. Navy and U.S. Air Force aircraft. Marshall tried to hint to others at RAND that intelligence might not be the problem they thought it was, but to no avail (with the exception of William Kaufmann, who was more receptive to the idea).10

9 For more discussion, see Andrew May (1998, pp. 340–343).
10 Herken (1987, pp. 81–83). Interviews with Andrew Marshall have provided additional insight into this period.
After a few years, others at RAND took up the idea of counterforce. James Digby headed a RAND study on counterforce beginning in 1954 (Digby, 1955). At the same time, an Air Force headquarters New Approaches Group was also examining the issue. Both groups made progress, but many in the Air Force and at RAND remained unconvinced. However, the major expansion of both U.S. nuclear stockpiles and the ability to produce more weapons in the 1950s made counterforce increasingly attractive simply because there were more than enough weapons for targets, if the targets could be found.

As discussed below, the Air Force was not against striking Soviet military assets per se; attacking Soviet nuclear weapons was already a high priority in SAC war plans in 1954. Rather, the Air Force did not believe in restraint during nuclear war, so Soviet military targets were just one part of a total and massive offensive that one military officer described as intended to reduce Russia to “a smoking, radiating ruin at the end of two hours.” Counterforce, particularly first-strike counterforce that sought to limit damage to the United States and its allies, was thus a central part of nuclear-war planning from early in the Cold War.

Of course, an expanded arsenal did little good if one did not know where to target it, so intelligence would remain central to counterforce. The emerging field of overhead reconnaissance, though intended primarily to document Soviet capabilities rather than to provide targeting intelligence, would ultimately mitigate the intelligence problem for counterforce against fixed targets. It was not accidental that two of those in the Air Force New Approaches Group, Brig. Gen. Bernard Schriever and Col. Richard Leghorn, were involved in overhead reconnaissance. Leghorn was an early advocate of what would become the U-2 spy plane, and Schriever would soon oversee both ballistic-missile development and the Corona spy-satellite program as head of the Air Force’s Western Development Division. The U-2 and Corona would

---


soon make counterforce targeting vastly more capable.\textsuperscript{13} Similarly, Digby believed that the counterforce had more utility than many at RAND thought. RAND during this period was also advocating the development of the reconnaissance satellites that would comprise Corona.

Intelligence was not the only important element of counterforce. A corollary to counterforce was the need to avoid targeting cities, at least initially. Disarming or damaging the enemy’s striking power without destroying its cities left those cities “hostage” to subsequent attacks while avoiding mass killing of civilians. The enemy would therefore have little incentive to use the surviving nuclear forces. However, many in the Air Force found this concept of withholding strikes to be a foolish dispersal of resources. In the late 1950s, Digby and Kaufmann would refine this idea of a “no-cities” doctrine (see Kaufmann, 1960). It was the no-cities approach that the Air Force initially did not find appealing about these early concepts of counterforce.

In addition to intelligence, the need to avoid killing large numbers of civilians with counterforce strikes meant that a premium would be placed on accuracy. Accuracy would mean that smaller weapons could be used to destroy Soviet military targets, which would in turn lead to smaller numbers of civilian casualties. This need for accuracy would grow if Soviet nuclear systems were protected or “hardened,” a step RAND was already recommending for U.S. nuclear systems (Wohlstetter and Hoffman, 1954).

By 1960, Andrew Marshall’s frustration over the issue of intelligence and counterforce at RAND had peaked. Writing at about the same time that the downing of Gary Powers revealed to the world the existence of the U-2 program, Marshall commented:

\begin{quote}
One cannot stand up in front of Air Force audiences, who know better, and make statements about how we cannot do this or cannot do that, or the Soviet Union has such and such an advantage over us, because we do not know where their strategic air bases are and expect that the audience will believe that one is an
\end{quote}

Deflecting the Sword of Damocles: Strategic Defense and Deterrence

expert on war. RAND has on a number of occasions appeared to be a set of babes in the woods in just this way. (Marshall, 1960, p. 43)

Perhaps fortunately, the exodus of many at RAND to the Kennedy administration only a few months later acted to finally ameliorate the problem of RAND access to intelligence.

This access led to a quick embrace by some of counterforce. In response to the tumultuous Vienna summit of 1961 and subsequent crisis over Berlin, two former RAND analysts began to sketch a counterforce first strike plan. Carl Kaysen and Henry Rowen, in consultation with others at RAND, laid out a surprise attack in which a small bomber force could penetrate at low altitude and destroy the relatively low number of long-range Soviet systems.14 Combined with recent advances in antisubmarine warfare (ASW) that made the handful of existing Soviet strategic submarines detectable, this plan stood a reasonable chance of achieving the destruction of most (if not all) of the Soviet ability to strike the United States.15 Further, it did so in a way that would minimize Soviet civilian deaths—though Kaysen conceded that the number of dead would be in the hundreds of thousands (Kaysen, 1961, Annex A, p. 3).

This plan was given some consideration, mostly as an intellectual exercise, and ultimately rejected. However, this was not the end of counterforce. Secretary of Defense Robert S. McNamara, briefed on the no-cities aspect of counterforce by William Kaufmann, initially embraced the concept. However, he began to withdraw his support as counterforce seemed to lead down the path to huge numbers of nuclear weapons (Herken, 1987, pp. 150–152, 168–176). Regardless of the con-

---

14 Kaysen had been at RAND briefly in the late 1940s before going to teach at Harvard. In the Kennedy administration, he was a special assistant to National Security Adviser McGeorge Bundy. Rowen left RAND to become deputy to Assistant Secretary of Defense for International Security Affairs Paul Nitze. A redacted version of their plan (calling for low-altitude penetration of the Soviet Union by a force of about 55 bombers) is available (Kaysen, 1961).

15 The Soviet strategic submarine fleet was small and of low quality in this period, while the United States sound surveillance system (SOSUS) had become operational in this period. See Press (2005, Chapter 3, Appendix A) and Coté (2003, Chapters 3 and 4).
cern about requirements, Kaysen and Rowen had made starkly clear the logic that linked extended deterrence and counterforce:

If each increase in the scale of our action is met by a corresponding and always dominating increase in the Soviet response, we will clearly be forced at some point to move from local to general action. . . . Soviet retaliation is inevitable; and most probably, it will be directed against our cities and those of our European allies. What is required in these circumstances is something quite different. We should be prepared to initiate general war by our own first strike, but one planned for this occasion, rather than planned to implement a strategy of massive retaliation. (Kaysen, 1961, pp. 2–3)

Subsequently, both the Air Force and civilians in the Office of the Secretary of Defense would be driven by this logic to embrace counterforce.

McNamara and the RAND alumni also sought to bring tools of system analysis to the broader questions of nuclear strategy. The Single Integrated Operating Plan (SIOP), first introduced under McNamara, sought to rationalize force employment across the nuclear triad and the services. By the mid-1960s, it included three main target sets: Soviet and Chinese nuclear forces (ALPHA); Soviet and Chinese other military targets (BRAVO); and urban-industrial targets (CHARLIE). The ALPHA target was clearly prominent even though McNamara himself was not sanguine about the prospects for counterforce. Further, the SIOP could be executed preemptively and included an option to strike only ALPHA targets. To be sure, few felt confident that a truly disarming strike was a possible, but damage limitation in the 1960s still seemed plausible.

Unfortunately, the early SIOP formulation for even the ALPHA-only target set would use thousands of weapons, many in the suppression of air defense, and would not limit collateral damage to civilians as much as might be wished. The Air Force, as noted, was never very accepting of limited nuclear war, considering it impractical from a command-and-control perspective, if nothing else. Until at least the
late 1970s, the Air Force assessment of command and control had considerable merit, a point underappreciated by many at the time.\textsuperscript{16}

RAND analysts in the 1960s discussed the counterforce imperative in some detail but were also concerned that the United States maintain a secure second-strike capability as well. The first-strike counterforce capability gave credibility to extended deterrence by limiting retaliatory damage to the United States, while a secure second strike reduced the incentive for the Soviets to launch their own disarming first strike. The advent of reasonably accurate and reliable ICBMs, such as Minuteman, could help provide the counterforce first strike, while the development of submarine-launched ballistic missiles (SLBMs), such as Polaris, made the secure second strike possible.

The United States was thus in a strong position in the early to mid-1960s; the Soviet response to the Cuban missile crisis showed as much. Yet some at RAND were already foreseeing the end of U.S. nuclear dominance. James Schlesinger, writing only a few months before the crisis over Cuba, noted,

During the next four or five years, because of nuclear dominance, the credibility of an American first strike remains high. A selective counterforce strike of soft Soviet military targets could cripple a sizable part of the Soviet military machine. The sparing of Soviet cities, in light of the increased strategic imbalance, would provide the Soviets with every incentive to avoid reprisals against NATO cities. At the same time, the possibility of a crippling American strike against [U.S.] military targets will in all probability dissuade the Soviets from very provocative moves against Western Europe. For the time being the argument is sufficiently persuasive to convince the Europeans. . . . What about the future? Unless the Soviets are less shrewd or less technically competent than we think them to be, eventually they will provide themselves with a relatively invulnerable counter-deterrent, which, permitting a much higher percentage of Soviet military might to survive, will increase the risks to the United States of a large-scale counterforce

\textsuperscript{16} Steinbruner (1978) was one of the earliest public appreciations of the problems of command and control in nuclear war.
strike. Will this be in 1965, 1968, or 1970? If massive nuclear exchanges are to be avoided—and both superpowers have a dramatic incentive to do so—the chief possibility for dissuading the Soviets from major provocations is to threaten carefully measured attacks against military targets or cities—deterrence on a tit-for-tat basis. (Schlesinger, 1962, pp. 8–9)

Schlesinger’s prediction of the end of nuclear dominance was prophetic; in 1965, the United States still retained some hope of first-strike counterforce, but, by 1970, it was gone.\(^\text{17}\)

Some dreaded this mutual vulnerability for a simple but widely debated reason known as the *stability-instability paradox*.\(^\text{18}\) This concept argues that, if both sides have a devastating, strategic, secure second-strike capability (stability) then, paradoxically, it may encourage risky and provocative behavior below the threshold that will trigger a strategic nuclear response. So the Soviets, once they developed an invulnerable counterdeterrent might feel that they could conventionally threaten western Europe with relative impunity. Given perceived Soviet conventional superiority in Europe, this made deterrence seem tenuous to some.

Not all analysts believed this, as some felt that Schelling’s “threat that leaves something to chance” would work to deter these lower-level provocations. Why would the Soviets run even a small chance that U.S. strategic nuclear forces would come in to play accidentally in exchange for some minor gains in Europe?\(^\text{19}\) Yet it remained a concern that could

---

\(^{17}\) In 1965, the Soviets had only 224 ICBMs and 195 intercontinental bombers, while the United States had 854 ICBMs (many of which had reasonable counterforce capability against existing Soviet systems) and 807 intercontinental bombers. By 1970, the Soviets had 1,220 ICBMs, to the United States’ 1,054 (virtually none of which was a multiple independently targetable reentry vehicle [MIRV]), while the number of bombers was unchanged for the Soviets and decreased to 501 for the United States. See Berman and Baker (1982, pp. 42–43).

\(^{18}\) First articulated in Snyder (1961).

\(^{19}\) It is also worth noting that the high value placed on Europe and the Soviets’ ability to devastate it should also have given the Soviets a very credible deterrent against the United States.
not simply be dismissed, particularly if one believed that the Soviets were bent on dominating Europe, if not the world.

Answering the challenge posed by the stability-instability paradox would then fall to the Nixon administration, the first administration to be faced with a robust, survivable Soviet nuclear threat. Even as it sought strategic arms limitations, many in the administration were seeking to promote counterforce. Schlesinger, as secretary of defense, would be among them. While his response to the problem of declining credibility in 1962 had been to cautiously explore the option of an all-European nuclear deterrent through some form of multilateral force, by 1973, he sought to promote improvements in both LNOs and in counterforce.

The LNO approach essentially called for the “deterrence on a tit-for-tat basis” referred to earlier. This was not attractive for two principal reasons. First, it was not clear at all how well nuclear war could be managed. Doing so would require not only highly robust command, control, and intelligence but also adversaries that were willing to adhere to controlled escalation under highly uncertain and dangerous conditions. Even if escalation could be successfully managed, it quickly became a contest of Schelling’s “obstinacy and insanity.” Schlesinger noted this in 1962 as well:

In a war of nerves, with limited encounters, which side will prove the stronger—especially when we have reached the city-swapping stage? How long will the American public accept a game played by these rules? Thus the final question appears: what does the decline of nuclear dominance do to the protection offered to Europe by a sophisticated deterrent [that] remains under American control? (Schlesinger, 1962, p. 10)

Despite his concern in 1962, Schlesinger was one of the major contributors to a RAND/Air Force study beginning in the mid-1960s on LNOs known as NU-OPTS. NU-OPTS sought to introduce more flexibility into the SIOP to make more-discriminate attacks possible. These discriminate attacks could serve a variety of functions, but the principal one was to harm the Soviets sufficiently to lead to war termination without provoking a catastrophic response. These limited
options often included a strong element of counterforce of the ALPHA target type, though they were almost (by definition) not capable of significantly limiting Soviet ability to damage the United States in retaliation.

Following the election of Richard Nixon, NU-OPTS came to be influential as its ideas spread. Some NU-OPTS participants took influential positions in the Nixon administration, most notably Schlesinger. When briefed on the SIOP, Nixon and his national security adviser, Henry Kissinger, were appalled at the rigidity of the options and began seeking to change nuclear targeting. This created an opportunity for proponents of NU-OPTS. The administration’s preoccupation with Vietnam limited change in the first term of the Nixon presidency, but a number of important reviews and studies were initiated that would subsequently affect nuclear planning (see Burr, 2005).

By 1972, with U.S. involvement in Vietnam drawing to a close, the issue of nuclear planning again became highly salient. Drawing on both the earlier studies and recent work done by an ad hoc group under the director of defense research and engineering, John Foster, the Nixon administration formulated a new nuclear-targeting plan. Formally embodied in national security decision memorandum (NSMD) 242 and the subsequent Nuclear Weapons Employment Policy (NUWEP) 74, this new set of guidance significantly changed the SIOP.

Instead of the five large preplanned options of the earlier SIOP, the new guidance was to base the SIOP around four types of options: major-attack options (MAOs); selected-attack options (SAOs); LNOs; and regional nuclear options (RNOs). NUWEP 74 called for four MAOs, two each against the Soviet Union/Warsaw Pact and the People’s Republic of China. The first MAO against each was a major counterforce attack combined with destruction of conventional military targets. The second was a more comprehensive attack that was essentially massive retaliation. There were 11 SAOs with a variety of

20 The findings of the Panel to Assess the Reliability, Safety, and Security of the United States Nuclear Stockpile (the Foster Panel) remain classified; a summary is available in a memorandum prepared by the National Security Council staff for Henry Kissinger. See Odeen (1972).
target sets, principally military, which will be discussed in more detail below. The MAOs and SAOs also had a number of prohibited target categories and optional “withholds” to make them even more tailor-able. LNOs and RNOs were discussed in less detail but were intended to be miniature SAOs with most targets withheld (Schlesinger, 1974).

Even as Schlesinger implemented these changes to permit for tailored and limited nuclear options, counterforce continued to be a prominent part of nuclear planning in two ways. The first was through what has been termed second-strike counterforce, the idea that the appropriate response to a Soviet limited counterforce attack was to retain the ability to strike back at Soviet withheld nuclear forces. This “assured retaliation” would convince the Soviets that attacking U.S. strategic nuclear forces was not worthwhile.

Second, damage-limiting (though not disarming) preemptive counterforce continued to be a major concern. Even though most analysts had concluded that highly significant damage limitation was no longer possible with the nuclear balance as it was in 1970, the need for a plausible first strike still drove the United States to plan for damage-limiting counterforce. NUWEP 74 made this continuing search for damage limitation clear. It stated that, if LNOs fail to control escalation and general war occurs, then damage limitation is second in priority only to damaging the enemy’s postwar recovery assets.

Further, NUWEP 74 made clear that it was expected that both MAOs and SAOs could be executed from a force posture known as generated without damage (GWOD). This clearly meant a preemptive first strike, as U.S. forces would be at their maximum readiness (i.e., not a day-to-day alert posture) yet would not have sustained damage from a Soviet attack. All of the GWOD MAOs emphasize the importance of damage-limiting counterforce (Schlesinger, 1974, pp. 4-6, A-1–A-2). Many of the SAOs also point to the importance of damage-limiting counterforce. SAO 1 (that is, the first listed SAO to be planned) has the objective “Neutralize the Soviet nuclear threat to the United States.” Five of the 10 remaining SAOs have at least some objective associated with destroying Soviet or Chinese nuclear forces (Schlesinger, 1974, pp. A-2–A-6; quotation from A-3). NUWEP 74 also emphasized the
need to at least disrupt Soviet nuclear forces in a preemptive attack even if the expectation that they would be destroyed would be low:

In a U.S. attack planned with fully generated undamaged forces on the Soviet nuclear threat to the United States and its allies, not less than one warhead should be applied to each ICBM site, each IRBM [intermediate-range ballistic missile] and MRBM [medium-range ballistic missile] site, each base for heavy, medium, and light bombers, and each base for missile-launching submarines, even if a high damage expectancy cannot be achieved or only short-term damage can be realized. (Schlesinger, 1974, p. A-7)

This guidance meant that a GWOD SAO 1 (in the acronym-heavy jargon of nuclear strategy) would be an attempt at major damage-limiting counterforce while avoiding cities as much as possible, even if the odds of very significant limitation were not good.

NUWEP 74 would be the guide to nuclear targeting until at least 1980, when it was superseded by a new NUWEP issued by Secretary of Defense Harold Brown. Most accounts of subsequent NUWEPs (and SIOPs) suggest that changes after NUWEP 74 were primarily incremental; for example, more SAOs appear to have been added. The SIOPs of this era did incorporate additional planning for second-strike counterforce by inserting a posture known as generated with damage (GWD), an alerted force that had been attacked and was now called on to retaliate.21

Even as new targeting policy was implemented, DoD was proceeding with the development of counterforce capabilities that would improve the ability to target the Soviets’ nuclear arsenal. This build up would gain speed during both the Carter and Reagan administrations, with such systems as the MX ICBM and Trident II D-5 SLBM. Combining extraordinary accuracy with high-yield MIRV warheads, both the D-5 and MX added very substantially to the hard-target kill potential of U.S. strategic forces and provided a potentially effective

---

21 See Kearl and Locke (1979). According to unclassified sources, the next NUWEP after 1974 was NUWEP 1980; see Ball and Toth (1990, pp. 67–68).
counterforce capability, assuming that the Soviets did not launch a significant portion of their own forces on warnings.\textsuperscript{22} 

One of the central problems of the prompt hard-target counterforce improvements that began under Schlesinger and the similar Soviet program (which culminated in such systems as the SS-18) was that they had the potential to exacerbate the security dilemma by creating incentives to strike first. First-strike incentives were further enhanced by the vulnerability of command, control, communication, and intelligence (C3I) systems. If C3I could be struck simultaneously with nuclear forces, the enemy might be unable to order its surviving forces to attack (at least quickly).\textsuperscript{23} This would potentially allow for retargeting on those assets that survived or, in the case of mobile targets, more time for them to be found and destroyed. Another related worry was that first-strike counterforce gave the other side an incentive to launch on a perceived warning of attack. This raised the chances for an accidental war wherein an attack was believed to be under way and, given the short timelines of ballistic-missile warfare, a snap judgment to launch was made, even though the warning was in error.

RAND analysts were not unconcerned with the first-strike instability problem that counterforce posed. They recognized that there was conflict between some of the objectives of U.S. policy. In a major analysis of strategic forces and first-strike stability, RAND analysts Glenn Kent and David Thaler noted,

\begin{quote}

The most important conflict [arises] between the objectives of enhancing first-strike stability, on one hand, and extending deterrence and limiting damage on the other; i.e. the more robust the Soviets believe first-strike stability to be, the less they might hesitate to precipitate a deep crisis by engaging in serious aggression, for example, in Western Europe. Balancing between first-strike stability and extended deterrence presents a problem in the planning of strategic forces. . . . Indeed, one might argue that an optimal amount of first-strike instability is possible: that is, enough
\end{quote}

\textsuperscript{22} On the D-5’s counterforce capability, see Coté (1991). On MX’s counterforce capability, see Soule (1978, Chapter 4).

\textsuperscript{23} A good assessment of the vulnerability of U.S. strategic communications is Blair (1985).
to deter the Soviets from generating a major crisis, say by invading Western Europe, but not enough to allow a major crisis to spiral out of control. Whether or not such an optimum actually exists, the concept provides the proper intellectual framework in which to think about the trade-off between first-strike stability and extended deterrence. (Kent and Thaler, 1989, p. 5)

Paul Davis termed this trade-off the devil’s dilemma, further noting that many thought the ideal objective of U.S policy (even if it was not possible in practice) would be “one sided first-strike stability,” i.e., the United States had both a secure second strike and a plausible damage-limiting first strike, and the Soviets would have neither.24 This asymmetric balance might lead to arms-race instability but, at least in the short run, would probably be advantageous to the United States and would also limit problems of crisis instability.

Trident D-5 was perhaps the ultimate, and asymmetric, solution to the stability-versus-utility debate. Based on the quiet Ohio-class ballistic-missile submarines, D-5 was nearly invulnerable, unlike the MX ICBM in silos (a proposed mobile basing system for MX was abandoned, in part for political reasons).25 D-5 gave the Soviets no incentive to shoot first, as they could not target it, so stability was preserved. At the same time, each D-5 could deliver eight W88 warheads (with a yield of 475 kilotons each) to within a few hundred feet of targets 4,000 nautical miles away (Coté, 1991). It denied effective first-strike counterforce to the Soviets at the same time as it ensured it for the United States.

It is important to note that D-5’s hard-target counterforce capability was not a mere inevitable incremental improvement over previous SLBMs, driven by the inexorable progress of technology. It took considerable effort, both politically and technically, to accomplish this feat, including introducing a stellar-inertial guidance system that many

---

24 Davis (1989, pp. 32–36, Appendix A). Davis was personally more concerned about first-strike stability than about extended deterrence and felt that much could be done to improve first-strike stability without undermining extended deterrence.

25 On the debates about basing the MX, both political and technical, see U.S. Congress (1981) and Edwards (1982).
long-time participants in the Navy’s Fleet Ballistic Missile program did not favor (see MacKenzie, 1993). Schlesinger was particularly important, as he spent considerable time advocating and then initiating the eight-year, $600 million Improved Accuracy Program needed to give D-5 hard-target counterforce capability (see MacKenzie and Spinardi, 1988, and Coté, 1996). The W88 warhead was similarly a major and expensive improvement and one many in the Navy did not want. Like the Improved Accuracy Program, it was intended to give D-5 incredibly good hard-target kill capability (see Francis, 1996, pp. 155–160).

In addition to offensive nuclear systems, intelligence remained a critical component of counterforce. The intelligence problem had been more or less solved for fixed targets, such as missile silos, but mobile targets remained important. One form of mobile target continued to be Soviet nuclear-powered, nuclear ballistic missile–carrying submarines (SSBNs). The Yankee class had been held at risk by U.S. ASW in the 1960s, but the 1970s saw the deployment of the Delta class of long-range SSBNs, soon followed by the quiet Typhoon class. The U.S. Navy, enjoying decades of acoustic advantage in the North Atlantic, began scrambling to regain the ability to engage these submarines. This was manifested in the maritime strategy to surge U.S. submarines into Soviet SSBN “bastions” near the Soviet Union, the decision to make attack submarines a larger portion of the Navy, and in the acquisition of improved Los Angeles–class attack submarines and the planning for the Seawolf-class attack submarine.26

Another target of increasing concern was the mobile ICBM, which the Soviets had begun developing in the early 1970s. Mobiles could not always be monitored continuously from space, so only intermittent sightings could be made if they were not in garrison. This drove plans for such programs as the Advanced Airborne Reconnaissance System, a highly classified, stealthy, unmanned aerial vehicle (UAV) intended to loiter at high altitude and conduct continuous surveillance of mobile missiles. This data would then be communicated to the forthcoming

26 See Coté (2003, Chapter 5). U.S. strategic ASW capabilities remain highly classified, though two reports on lengthy “trails” of Soviet submarines have been declassified.
B-2 bomber, which would target and destroy the mobiles (see Ehrhard, 2000, pp. 134–159).

Communication improvements also began in the 1970s. These improvements served multiple purposes. They helped ensure that retaliation would take place for any Soviet first strike but also helped make LNOs more feasible, as the communication system would degrade much less quickly. These improvements also made coordination for damage-limiting counterforce easier, particularly communicating with the SSBN fleet.²⁷

As the nuclear balance evolved, RAND continued to explore and advocate counterforce. Carl Builder was one of the strongest proponents of first-strike counterforce in the late 1970s. Builder acknowledged that the days of “splendid,” completely disarming counterforce were gone but that the United States could still try to obtain “objective” counterforce. He defined *objective counterforce* as the ability “to shift significantly and irrevocably the balance of strategic offensive forces” (Builder, 1979, p. 5).

Builder’s rationale for U.S first-strike counterforce sprang from conventional inferiority in Europe combined with the unattractiveness of either theater nuclear operations or LNOs. He reiterated and carried forward the arguments of Kaysen, Rowen, and Schlesinger:

> We seem to have forgotten, in our efforts to control strategic arms, that our conventional and nuclear capabilities have been and will remain linked. Our conventional capabilities have never been good enough to give away a credible, advantageous nuclear initiative. The only advantageous nuclear initiative left to us as a deterrent is the credible threat of striking first with an effective counterforce strike. (Builder, 1978, p. 18)

However, even as Builder argued the need for first-strike counterforce, shifts in the U.S. force posture and technological base were presenting new opportunities for defense. At the same time, others at RAND

---

²⁷ For description of developments intended to mitigate vulnerability and ongoing risks, see Carter (1987).
saw a different way to address the problem of credibility and nuclear weapons.

**Firebreaks on the Battlefield: RAND and Tactical Nuclear Weapons**

Though counterforce eventually gained many supporters at RAND (though even supporters were divided on how much and what kind was needed), some also questioned it. Bernard Brodie felt that counterforce would do little to address the central problem of credibility unless it was nearly perfect, a situation that seemed unlikely. Writing in 1965, he noted,

> The general consensus approving the no-cities targeting philosophy . . . is overlaid with a growing concern that counterforce targets may prove of steadily diminishing attractiveness. The reason is that even at best the residual damage-producing capabilities after an American counterforce strike are likely to remain huge. From being a good “damage-limiting” system, a strategy stressing counterforce targets may become simply the least bad system. (Brodie, 1965, pp. 31–32)

Brodie went on to note the consensus view that general war with the Soviet Union was most likely to emerge from a local conflict in Europe.

Where Brodie differed from many was in his belief in the utility of tactical (also termed battlefield) nuclear weapons (see Brodie, 1954). Smaller in power than strategic weapons, tactical nuclear weapons could be used to offset the Soviet numerical advantage in conventional forces as well as to reply to potential Soviet tactical nuclear attacks. This would create an additional step in the escalation ladder from local war to general war, strengthening Flexible Response.

Brodie was not alone in seeing utility in tactical nuclear weapons. The tactical nuclear concept was first extensively explored by Project Vista, a research project similar to Project Charles, conducted
by the California Institute of Technology in 1951. Vista, which was primarily an Army project, concluded that tactical nuclear weapons would be useful for the defense of Europe, an idea the Army embraced (albeit briefly) with the creation of the Pentomic Division, which was optimized for dispersed combat on the nuclear battlefield. The Army eventually developed very small nuclear weapons, such as the Davy Crockett recoilless rifle that could be used by a squad-sized section.

However, not everyone embraced the tactical nuclear concept. Some analysts, including RAND’s Alain Enthoven, felt that any use of nuclear weapons would cross a “firebreak” that would lead from a local tactical nuclear exchange to a general strategic one (Enthoven and Smith, 2005, pp. 127–129). Brodie’s counterargument was that an explicit firebreak could allow escalation below the level of that firebreak (i.e., unlimited conventional action below the tactical nuclear threshold). This worked to the advantage of the Soviets, who had conventional superiority. Further, it relied on Soviet acceptance of the firebreak, which the Soviets rejected in both public statements and in their military writing (Brodie, 1965, pp. 62–70).

Brodie noted that some Europeans had qualms about tactical nuclear weapons, not wanting to become the battleground between two superpowers that would remain untouched by the conflagration. He argued against this position by stating that tactical nuclear weapons were more credible, as they would not automatically entail Soviet strategic retaliation, which, he felt, counterforce would. He wrote,

> Where Europeans insist that we will surely be unwilling to hazard national extermination for them, it is useless to try to persuade them that we will indeed be ready to do so. We should rather emphasize the fact—which is to their decided advantage—that there is no need to initiate (and thus to undergo) strategic bombing in order to defend them effectively. The threat of an effective local defense—which is to say one that is serious enough to succeed in itself or to open the possibilities for large scale action—is

---


29 On the Pentomic Division, see Bacevich (1986) and Wilson (1998, Chapter 10).
a deterrent as good as or better than any threat of general war, especially since it is far less subject to being doubted. (Brodie, 1965, p. 5)

Brodie also commented that, in his experience and in contrast to attitudes about nuclear war, Europeans did not seem to harbor much rancor about the fact that, in both world wars, they had been devastated while the United States remained untouched. It was simply the nature of the geostrategic environment (Brodie, 1965, pp. 5–6).

In parallel to Brodie’s argument for tactical nuclear weapons was the work of RAND physicist Samuel Cohen. Cohen realized in the late 1950s that small nuclear weapons could be constructed that would deliver large amounts of prompt radiation in the form of high-energy neutrons (see Cohen, 1960). These weapons would be very low yield so they would not generate massive blast and heat effects, and if detonated sufficiently far above ground, they would produce little radioactive fallout. This combination of characteristics, Cohen felt, would make these weapons ideal in the tactical nuclear role, as they would have significantly less long-term environmental impact than other types of weapons. Cohen believed that it might even be possible to use these weapons on advancing Soviet armored units without inflicting major damage on West German towns or their inhabitants (see Cohen, 1975, 1978).

The weapons that Cohen envisioned would become known as enhanced radiation weapons or, more infamously, neutron bombs. They were widely decried by many Europeans, particularly those on the left, who felt that they were the ultimate “capitalist” weapon, sparing infrastructure while killing people. The neutron bomb also stoked some of the very fears that it was intended to quell that Europeans had about nuclear use. This paradox was striking, and it undermined Brodie’s argument that the Europeans would learn to live with tactical nuclear weapons. The entire point of the neutron bomb was to make such weapons more usable on the battlefield by making the effects more discrete and less environmentally damaging than previous nuclear weapons. However, many European leaders focused on the fact that the neutron bomb made nuclear use more likely rather than the fact that it made that use more discrete. Ultimately, these political considerations pre-
vented widespread adoption of these weapons (see Wasserman, 1983, and Auger, 1996). Yet even as the neutron bomb debate took place, conventional deterrence became more promising.

**Modern Arms and Expensive Men: Conventional Deterrence After Vietnam**

The year 1973 was, in many ways, the low water mark of the U.S. Army. The Army had not brought the Vietnam War to a successful conclusion yet had reduced readiness in Europe to fight the war. The draft was on its way out, to be replaced by an all-volunteer force (AVF). Discipline was shaky, morale low, and combat-effectiveness levels grim.30

In response to this crisis, the Army, with support from such civilians as Schlesinger, initiated a series of reforms that resulted in nothing short of revolution (see Bradford and Brown, 1973, and Nielsen, 2003). These reforms drew heavily on a belief that conventional battle in Europe was the Army’s primary (perhaps only) mission and that this battle would be, in many respects, similar to the 1973 Yom Kippur War. The central lessons of the Yom Kippur War appeared to be that modern combat was incredibly lethal, but that a smaller, well-equipped, and well-trained force could defeat larger armies even when beginning from a disadvantageous position. The U.S. Army, traditionally centered on deliberate mobilization, would have to be able to mobilize and win quickly.31

The changes to meet this requirement were extensive. Briefly summarizing, the Army sought to buy higher-quality equipment in somewhat lower quantity, while vastly improving training for units. On the acquisition side, this meant the so-called Big Five systems, including what would become the M1 Abrams tank and the M2 Bradley infantry fighting vehicle, as well as additional systems, such as the multiple-launch rocket system (MLRS). On the training side, the Army con-

---

30 For a series of personal vignettes on this time, see Kitfield (1995).

31 On this period in the Army, see Lock-Pullan (2003), Herbert (1988), and Romjue (1984).
structured the high-technology National Training Center (NTC) at Fort Irwin and emphasized improving the quality of individual recruits. Personnel, no longer free, had to be maximized in quality, even if this meant paying more per soldier. In short, the Army invested heavily in quality capital, both physical and human.

Additionally, the Army began attempts to increase coordination with the Air Force for the central European battle. Traditionally, Air Force and Army cooperation had been limited and contentious, with a few notable exceptions in wartime. However, the Air Force was going through its own post-Vietnam changes, and many in the service felt that improving relations with the Army was important. In addition to working to develop joint doctrine for fighting in central Europe (eventually termed *AirLand Battle*), the Air Force purchased its first modern dedicated close air-support aircraft, the A-10.

In combination, these reforms (once they came to fruition) meant that, for the first time, the United States could lessen its reliance on nuclear dominance. The Army/Air Force combination now shared the same characteristics that made nuclear weapons attractive for deterrence: a relatively small force in being that was both effective and could be maintained without major societal disruption. The key difference was that nuclear weapons were primarily a tool of deterrence by punishment, while the conventional forces were a tool of deterrence by denial. As with nuclear weapons, few if any felt that the deterrence offered by improved conventional forces were highly robust; nonetheless, the ability to rely significantly on conventional forces alone to deter was a significant improvement in U.S. deterrence.

Conventional forces were greatly aided in this transformation by recent advances in technology, from advanced armor and fire control

---

32 For an overview of the all-volunteer force, see Rostker (2006).

33 On this period in the Air Force, see Anderegg (2001), Rasmussen (1978), and Worden (2002, Chapter 8).

34 Arguably, the conventional balance had not been precarious in the early 1960s either, as Robert McNamara and his system analysts sought to demonstrate. See Enthoven and Smith (2005, pp. 132–158).

35 On conventional deterrence generally, see Mearsheimer (1985).
for the M1 to laser-based training systems at the NTC. Among the most important of these was the set of technologies that allowed the introduction of effective and relatively low-cost precision-guided munitions (PGMs). These systems, including laser-guided bombs and missiles (such as Maverick; tube-launched, optically tracked, wire command [TOW]; and, later, Hellfire) promised the efficient destruction of both fixed targets and the avalanche of armor the Red Army was expected to send against NATO. PGMs had already demonstrated their effectiveness against armor and mechanized units in both the Yom Kippur War and in the U.S. air operations that halted the North Vietnamese Easter Offensive in 1972.36

RAND was an early advocate of these systems, holding seminars on their importance in the early 1970s. Jim Digby was the project leader on many of these studies; by 1975, he had concluded,

On balance, however, the advent of PGMs results in a major advantage to NATO in defending against an armored thrust. . . . Most important, the critical nature of target acquisition for this generation of weapons gives an inherent advantage to the defender, who can find a good hiding place as he waits, over an approaching attacker who [must] expose himself. Communicating these prospects to the NATO allies can give an important psychological boost, especially if coupled with practical observations [from] the U.S. analysis of the October war in the Middle East. (Digby, 1975, p. xiii)

This view was subsequently supported by many in the Office of the Secretary of Defense, leading to further refinements in PGM capabilities against armor. Though some of these efforts were more successful than others, many were highly effective.

Digby’s analysis noted not only the technical importance of PGMs, but also the psychological importance. Part of NATO’s problem had always been the seeming hopelessness of conventional deterrence, which led to a self-fulfilling prophecy. Alain Enthoven and

---

36 On PGMs in the Yom Kippur War, see Safran (1977). On PGMs in the North Vietnamese invasion, see Anderegg (2001, Chapters 12 and 13).
others had attempted to prove otherwise, but various NATO members were still not convinced. No country would invest substantially in a hopeless cause, particularly when U.S. nuclear superiority was a cheap alternative. With no major investment, conventional deterrence stayed hopeless. PGMs made the cause less hopeless even as the end of nuclear superiority meant that the cheap alternative was much less tenable.

In addition to efforts to exploit and improve the already proven first generation of PGMs, more ambitious efforts were proposed. Beginning with the Assault Breaker concept in the mid-1970s, these efforts sought to destroy targets far behind the forward edge of battle. This required a host of technological developments, including the beginnings of the Joint Surveillance and Target Attack Radar System as well as autonomous guidance (“brilliant”) anti-armor submunitions. In pushing the envelope of technology, some of the proposed weapons failed to yield immediate benefits. However, these systems would provide the underpinning for current U.S. efforts at realizing a “revolution in military affairs” or “transformation.”

It is important, however, to be clear in what these technologies were intended to do. The intent was to provide a relatively low-cost way to offset the massive numerical advantage held by the Soviet and Warsaw Pact armies in armor and artillery, the areas that would make a rapid conventional victory possible. RAND’s studies were also careful to point out the need to supplement proven force structures with PGMs, rather than replace them entirely. Digby noted,

> there are some major deficiencies in the new weapons . . . and for a long time to come postures must be balanced between reliance on weapons like guns, with a high rate of fire and great effectiveness at lesser ranges, and reliance on PGMs like TOW, that are very effective at long ranges if all the circumstances permit their use. (Digby, 1975, p. vi)

The resulting balanced force structure would be one optimized for defense and counteroffense to destroy Soviet Tank and Shock armies without requiring a comparable force level. This clearly fit the pattern

[37 See Van Atta, Cook, et al. (2003) and Van Atta, Lippitz, et al. (2003).]
for long-term competition that Marshall had described earlier with regard to air defense. PGMs would offset Soviet investments in armor at acceptably low cost while forcing the Soviets to devise countermeasures or to develop similar technologies. From this perspective, PGMs proved to be an effective competitive strategy.

The Soviet concern about PGMs, as noted by comments in the Red Army’s newspaper by the chief of the General Staff, Marshal Ogarkov, led to significant attempts by the Soviets to develop their own PGM technologies (Ogarkov, 1984). The inability of the Soviet economy to effectively respond to the perceived need for these technologies made many in the Red Army unhappy. These officers were thus at least initially willing to support General Secretary Gorbachev’s push for restructuring, not out of an inherent desire for reform but from fear of losing this qualitative competition (see Becker, 1987, and Gottemoeller, 1989). PGM technology thus had strategic consequences for containment and deterrence well beyond the purely military.

### The Magic Bullets: Conventional Counterforce

Even as PGM technology made a competitive strategy for conventional deterrence possible, it opened up new possibilities in strategic nuclear deterrence. Though useful for many strategic strikes, the most important use for a highly accurate, long-range weapon might be to threaten Soviet strategic nuclear forces without the weapon needing to be nuclear itself. The possibility of conventional rather than nuclear counterforce was thus on the horizon.38

Carl Builder, already an advocate of counterforce, quickly realized the possibility of using conventional weapons for counterforce. Even as such programs as Assault Breaker applied PGM technology to the operational level of war, Builder proposed the possibility of using them at the strategic level. Writing in 1983, he commented,

---

38 For an overview of both U.S. and Soviet ideas about conventional counterforce in a variety of contexts, see Rosen (1987).
But the most challenging and provocative objectives for non-nuclear weapons are likely to be the military facilities and equipment that are often called counterforce and countermilitary targets. These targets include nuclear and conventional forces and the bases that support them. . . . It is these targets, numbering in the thousands, that pose the crucial test for the future of non-nuclear strategic weapons. For if they can be attacked with economically feasible nonnuclear forces, the prospect for these new weapons will not be limited to feats of technological virtuosity at the threshold of nuclear war. Rather, nonnuclear weapons will be destined to replace the bulk of the nuclear weapons arsenals and will become the pivotal military forces of the future. (Builder, 1983, p. 33)

Cruise missiles, the same technology that made a competitive strategy with Soviet air defense possible, would also soon make at least part of Builder’s vision possible. The family of cruise missiles developed in the late 1970s and early 1980s (Tomahawk, air-launched cruise missile, ground-launched cruise missile, advanced cruise missile [ACM]) would combine a relatively low-radar cross-section with long range and high precision. The Tomahawk, for example, was equipped with a terminal guidance system that made a previously unthinkable less than 10m circular error probable (CEP) possible. ACM, on the other hand, utilized new stealth materials and design techniques that made it extremely difficult to detect and shoot down (see Sweezey and Long, 2005).

Conventional ballistic missiles were not far behind in terms of the accuracy needed for conventional counterforce. The Pershing II IRBM was already being developed with an active radar terminal guidance system that would provide it with a CEP of 30–50m (see “MGM-31B Pershing II,” 2007). Though duplicating and then improving on this performance for an ICBM would be challenging, it did not seem out of reach.

Builder, writing a few years after his initial comments on conventional strategic weapons, presented two hypothetical ICBM systems with conventional payloads. One would precisely dispense fragmentation submunitions over a rectangular area. It would be capable of
rapidly destroying aircraft not in expensive shelters. The other would deliver PGMs equipped with a shaped charge warhead capable of piercing hardened silo doors.³⁹ Builder’s concepts were further elaborated in the report of the Commission on Integrated Long-Term Strategy, which was dominated by RAND alumni.⁴⁰

Of course, the intelligence requirement for conventional counterforce would be even more stringent than that required for nuclear counterforce. A slight miscalculation in target hardness or location is easily compensated for with the 475-kiloton W88 warhead carried by the Trident D-5; the same is not true of the 1,000-lb conventional warhead carried by a Tomahawk. Mobile targets are similarly more of a problem for conventional counterforce. One merely needs to get a nuclear warhead to hit in the vicinity of a soft mobile launcher, so target movement is less relevant (though not totally irrelevant if collateral damage from an area barrage is to be minimized).

However, the limited destructive power of conventional weapons for counterforce also reduced the collateral-damage effects as well. So Carl Kaysen’s disarming strike from 1961, instead of killing hundreds of thousands, would kill only tens or hundreds (almost all military personnel). This would reduce (in theory) the urge to retaliate by “city busting,” though not the “use-or-lose” incentives. A U.S. counterforce first strike would therefore be somewhat more plausible and thus a more effective deterrent, or so Builder’s logic held.

³⁹ Builder (1987). Builder, demonstrating that humor was not absent from RAND analysis, gave the first system in this paper the code-name Lucy, making reference to the Beatles’ song “Lucy in the Sky with Diamonds.”

The preceding chapter discussed the technical aspects of deterrence, with only modest reference to the psychological aspects. RAND analysis has unfairly been criticized as overemphasizing the rational nature of adversaries, particularly in crisis, and of underweighting the possibility of irrational action or the idiosyncrasies of decisionmakers. As evidence of this, critics point to Thomas Schelling’s oft-quoted remark that “you can sit in your armchair and try to predict how people will behave by asking how you would behave if you had your wits about you. You get, free of charge, a lot of vicarious, empirical behavior” (Archibald and Deutsch, 1966, p. 150). This is taken by these critics to mean that Schelling and other RAND analysts were insensitive to idiosyncrasies and the possibility of irrationality.

Yet RAND was also one of the foremost centers for the study of the nonmaterial components of deterrence. Both the cost/benefit calculus and the fear element of deterrence were studied at RAND, especially in the context of the Soviet Union. Analysts at RAND became some of the foremost “Kremlinologists,” seeking to understand how the various institutions and personalities that comprised the Soviet Union interacted and produced foreign and military policy. It is worth noting that Schelling was at least as sensitive to the possibility of unique characteristics of decisionmaking as his critics, as he noted immediately after the above quotation:

Where there are no strong personality and emotional determinants of behavior, people on the whole do the “right” thing. There must be enormous fields of human activity where this gets you
nowhere, but over some parts of economics this gets you a long way. Of course, you can make grievous mistakes if you think this is a hundred per cent, and you have to be alert for exception. (Archibald and Deutsch, 1966, p. 150).

RAND Kremlinologists thus sought to probe the Soviet Union for “exceptions” to common expectations. The two early, path-breaking RAND works in this field are undoubtedly Philip Selznick’s *The Organizational Weapon* and Nathan Leites’ *The Operational Code of the Politburo.* Both books sought to understand both the remarkable success of communist organization as well as its weaknesses. Selznick concentrated on the Communist Party as a “combat party” dedicated to seizing power through revolution. Leites, in contrast, focused on understanding how that party then handled the running of a country after it seized power.

Other RAND analysts examined what the Soviets thought of the concept of deterrence. If the Soviet concept of deterrence varied widely from that of the United States (or if the Soviets simply did not accept deterrence as a proposition), then the entire U.S. strategy rested on a weak reed. It might be that the Soviet leadership had goals (such as world revolution) that were as “nonrational” as the defense of territorial integrity was to the United States. This would make it nondeterrable in some important instances and would, in turn, argue for a U.S. preventive war rather than containment and deterrence. If war was inevitable, it should be launched when U.S. advantage was at its maximum.

The importance of this point cannot be overstated, and it was debated intensely in the early 1950s at RAND and at the highest levels of the U.S. government. Was the Soviet Union deterrable? At RAND, Bernard Brodie argued the affirmative, while mathematician John Williams argued the negative. The Eisenhower administration’s Project Solarium exercise replicated this debate in government, with some

---

1 Selznick (1952) and Leites (1950). Leites’s work was part of an even lengthier work; see Leites (1953).

2 Brodie (1953) and Williams (1953). For lengthier discussion of this debate, see Andrew May (1998, pp. 313–330).
in the Air Force supporting the case for preventive war. Ultimately, the argument was at least tentatively settled in the affirmative by President Eisenhower. The subsequent end of “splendid” disarming counterforce a decade later meant that a “splendid” preventive war was no longer a viable option and deterrence was the only course of action that would probably not result in Soviet retaliation on some level.

Even if the USSR was deterrable and the Soviet conception of deterrence was fairly similar to that of the United States, it could still vary in important ways. One of the most discussed examples was the tolerance of casualties. The USSR had been devastated by invasion in 1941, yet fought back to victory and postwar recovery even after taking millions of casualties. Few argued that the Soviets would therefore gleefully embrace another such outcome from nuclear war; the Great Patriotic War (as the Soviets called World War II) had been a time of near unimaginable sacrifice. However, it did seem plausible that, if it felt existentially threatened, the Soviets would tolerate the nearly instantaneous loss of millions of citizens. This single point, the “pain threshold” of the Soviets, was quite important. If it were quite low, then a minimal, survivable second strike would suffice for U.S. basic deterrence. If it were quite high, then the second strike would need to be quite large.

One of RAND’s early works on this subject sought to analyze how various U.S. and Soviet force postures and strategic moves would affect deterrence. Central to this analysis was an understanding of how the Soviets valued civilian and military losses. This analysis concluded that

Military losses are preferred to civilian losses if (a) the military losses do not preclude a satisfactory outcome of the war; (b) if the military losses do not prevent Russia’s maintaining a satisfactory postwar military posture toward allies (say, China) and neutrals; and (c) if the two foregoing conditions being satisfied, the civilian losses exceed a certain critical number, which we assume to be quite low. . . . Civilian losses, under such circumstances[,] are

---

3 See, for example, Lee (1953). More discussion of Solarium can be found in Trachtenberg (1988–1989).
more regrettable because they are less quickly and less easily recup-
erable and may have unpleasant internal political consequences
for the government and the party; they may, for example, increase
the political power of the military. (Goldhamer, Marshall, and
Leites, 1959, p. 33)

Other RAND research sought to make sense of the Soviet pro-
pensity for risk. If the Soviets were quite risk acceptant, they would be
more prone to make bold probes even from an inferior position. If they
were very risk averse, they would avoid provocation, even from a strong
position.

Many of the answers to these questions of interest suffered from
two interrelated problems. The first was a methodological and evi-
dentiary problem: The Soviet leadership was not going to openly and
honestly discuss most of these matters, so information was necessar-
ily incomplete. Even with complete information, mental calculations
of cost/benefit and the emotional response of fear are often not pro-
duced through a transparent, rational process. They are subject to the
limits of human cognition and emotion, resulting in a variety of pos-
sible biases.

The second problem was that most, if not all, responses are condi-
tioned by situational variables. This made generalization about Soviet
behavior even more difficult. It was thus plausible that, in some situa-
tions, the Soviets would be extraordinarily risk averse, while, in others,
they would be highly risk acceptant.

These factors meant that, even with some hindsight, interpreting
the behavior of the Soviets could be difficult. For example, viewed one
way, the Soviet decision to place missiles in Cuba was an incredibly
risky move provoked by desperation in the face of overwhelming U.S.
strategic superiority. Viewed another way, it was a low-cost attempt to
prevent the possible loss of an ally to U.S. invasion, an attempt that was
abandoned quite quickly in the face of confrontation.

One of RAND’s most methodologically sophisticated attempts
to deal with these problems and their relationship to deterrence was
initiated in the early 1970s. It was intended to evaluate a variety of U.S
force postures for their deterrent value against the Soviets. Rather than
postulating one model or operational code for Soviet behavior, analysts William Jones and Donald Emerson developed several plausible models, based on prior observations and experience. One model was heavily dependent on the particular constellation of Soviet leaders and their personality types; another was that of the “doctrinaire Marxist-Leninist”; another was similar to the bureaucratic-politics model of Morton Halperin and Graham Allison, with various institutions and interest groups jostling for resources and influence (see Jones, 1974, and Emerson, 1971).

These general models of the Soviet decisionmaking process could then be applied to a variety of scenarios to generate situational context. One scenario focused on the possibility of an inter-German confrontation that pulled in both superpowers; the interaction of U.S. responses with each of the Soviet decisionmaking models was then evaluated. In system-analysis fashion, each “model Soviet” had a complex worksheet of propensities and cost/benefit calculations. The most effective U.S. force structure for deterrence could then be developed based on these models.

In the case of NATO conventional forces, for example, it appeared that, across models, rapid mobility contributed greatly to deterrence. By making a rapid fait accompli, such as the seizure of a small German salient, seem unlikely if not impossible, even the most risk-acceptant Soviet model could be deterred without undue provocation (Emerson, 1971, p. 28). This analysis highlighted several force-structure considerations for deterrence. At the operational level, it showed the role of units, such as the corps-level armored-cavalry regiments, and the importance of new systems, such as PGM-armed attack helicopters, for deterrence. At the strategic level, it pointed to the importance of the Return of Forces to Germany (REFORGER) exercise, both as a deterrent signal and as a means of ensuring warfighting readiness.

Subsequent RAND analysis also sought to show how Soviet idiosyncrasies (termed Soviet strategic culture) could influence nuclear deterrence as well. Jack Snyder noted that the cooperative damage limitation implied by LNOs was even less accepted in the Soviet Union than the United States. While other RAND analysts doubted the U.S. resolve in the face of tit-for-tat exchanges, Snyder doubted that the Soviets
would be inclined to play along. The Soviets seemed to greatly prefer unilateral damage limitation, such as air defense or civil defense, to mutual restraint. This made strategic sense, because, as noted earlier, the Soviets were better able to mobilize and control the population for civil defense and better able to extract resources for military purposes, such as air defense (Jack Snyder, 1977). Some of this attempt by RAND and others fed into strategic-targeting doctrine in the 1970s. For example, the shift to targeting leadership and postwar recovery assets was in part due to the belief that these were things about which the Soviets cared more deeply than simple casualty counts.

RAND’s study of adversaries to be deterred also extended beyond the Soviet Union. RAND analyst Herbert Goldhamer wrote a lengthy personal account of his experience at the Korean War armistice negotiation that included observations on the effect of culture and attitude on perceptions and negotiations (Goldhamer, Marshall, and May, 1994). As the Sino-Soviet rift became apparent, RAND analysts began to explore the nature of the Chinese government as well. Particularly after the end of the main period of the Cultural Revolution, attention focused on civil-military relations within China as well as attempts to develop clearer profiles of Chinese leaders (see Robinson, 1970; Sung, 1975; and Whitson, 1973).

After the Cold War, the need to develop and model adversary behavior did not decline. RAND analysts noted that deterrence against Saddam Hussein in 1990 could have been improved by creating alternative models of Saddam’s decisionmaking and then using the models to inform decisionmaking. This recommendation was not well implemented and led to subsequent misestimation of Saddam’s calculus, discussed in more detail next (Davis and Arquilla, 1991).
CHAPTER SEVEN
Deterrence Then and Now

As noted previously, many future deterrence scenarios will look very different from the Cold War. Yet understanding the logic behind the United States’ adoption of deterrence in the Cold War and the theory and practices that underpinned it during the Cold War will be crucial in the future. Deterrence is an uncomfortable pillar on which to rest security, so it must be widely and well understood if policymakers are to rely on it when other options seem plausible.

This chapter will first discuss why deterrence will, in all likelihood, be a common part of U.S. grand strategy in the future. It will then turn to how the theory and practices of the Cold War can be applied to the three categories of possible adversary noted previously (peer/near-peer competitor, regional powers, and significant nonstate actors). Several examples will be provided from the recent and possible future international environments to illustrate the importance of these concepts.

The Long Shadow of the Garrison State: Deterrence in Future U.S. Strategy

At present, the U.S. defense burden is not egregiously heavy by cold-war standards. Including homeland security, non-DoD national-secu-

---

1 RAND continued to produce studies on deterrence after the end of the Cold War, including Watman et al. (1995); Johnson, Mueller, and Taft (2003); and Gompert, Watman, and Wilkening (1995).
Deterrence—From Cold War to Long War

Security programs, and a supplemental for Iraq and Afghanistan, the FY 2007 cost will be roughly 5 percent of GDP. However, the present level of expenditure and operations appears unsustainable for the long term. Spending on Iraq alone is running at close to 1 percent of GDP annually; numerous reports indicate that equipment is being damaged, destroyed, or worn out at a rate that will increase future maintenance and acquisition bills.

Further, the AVF, while maintaining coherence and meeting recruiting goals for now, is showing signs of strain. This is unsurprising, given that the post-Vietnam reforms (including AVF) were structured to make the military a better weapon for deterring and fighting short, conventional wars. While many patriotic men and women continue to volunteer, others who would previously have joined may be less likely to volunteer. Retention of well-trained personnel may also suffer.

A parallel can be drawn between the post-Vietnam U.S. military and the Israeli military on which it was consciously modeled. An unstoppable conventional force quickly occupies a country that it is then unable to govern. The occupation that follows dulls the reputation of a military previously thought invincible, drains valuable resources, and creates tensions in a society previously unified by the perception of external threat. This could be Iraq in 2003 as easily as Lebanon in 1982. The lesson that many Israelis eventually took away from Lebanon after more than a decade was to rely more on deterrence, both by punishment and denial.

Even as Iraq and Afghanistan strain the U.S. military, other potential threats remain on the horizon. The choice is therefore likely to be

---

2 The numbers are $439 billion for DoD, $21 billion for non-DoD, $35 billion for homeland security, and an assumed supplemental about equal to the FY 2006 supplemental. See OUSD(C) (undated) and OMB (undated). U.S. GDP is roughly $11.75 trillion.

3 Estimates of roughly $10 billion per month spent on Iraq yield an annual rate of $120 billion. See Weisman (2006). There is an ongoing debate about so-called reset costs for equipment maintenance and replacement, but it will not be small. See Belasco (2006).

4 The level of strain under which the AVF finds itself is also debated, but the consensus seems to be that it is significant. See the testimony provided by several witnesses on March 16, 2005, to the U.S. House of Representatives Committee on Armed Services Military Personnel Subcommittee as well as Hosek, Kavanagh, and Miller (2006).
between massive expansion of the U.S. military, with all the costs that would entail, or of relying on some combination of deterrence and reliance on proxies to deal with a substantial portion of these threats. For the reasons noted earlier, that massive expansion of the U.S. military proved fleeting during the Korean War, it seems unlikely to happen now or in the near future absent a catastrophic shock, such as a nuclear attack on the United States. This leaves deterrence as the only option for many future threats.

Even if expansion of the national-security apparatus were possible, it is not necessarily a good idea. The United States may well find itself in the position of “self-inflicted competitive strategy.” This is the situation Schelling described and on which Marshall elaborated, in which adversaries can counter—with less—every dollar spent on national security, leading to quickly declining marginal returns. While it is too early to really talk of “the costs of American empire,” continued expansion of national-security ends and means may change the situation. In some cases, the vast resources of the United States may make such self-inflicted competitive strategy worthwhile, but these situations must be carefully evaluated on a case-by-case basis.

National missile defense is an example of such a self-inflicted competitive strategy, costing far more to counter missiles than it does to develop and produce them. Nonetheless, it is possible to argue that a limited defense is worthwhile, given the potential consequences of a WMD attack. However, great care should be taken in evaluating how much should be spent on such endeavors.

Finally, as during the Cold War, the drive to expand the power of the national-security apparatus can be corrosive to U.S. society. It can lead to abridgement of freedoms, the alteration of the economy, and a host of other such unintended consequences. RAND terrorism analyst Brian Jenkins has eloquently noted that excessive concern about national security “can turn us into a herd cowering before imagined horrors, vulnerable to doomsayers and demagogues, ready to pawn liberty for security” (Jenkins, 2006, p. 153). The shadow of the garrison state did not disappear with the end of the Soviet Union, so deterrence will remain vital to U.S. strategy.
Cold War Redux: Deterrence and the Peer Competitor

The list of potential peer and near-peer competitors for the United States in the next 20 years is quite short. In descending order of likelihood, they appear to be China, Russia, the European Union, and India. Only the first two seem to be more than remote possibilities. Deterring these powerful states will closely resemble the Cold War, except that, in all likelihood, it will be much “colder” as well as easier.\(^5\) The following section applies past research to deterrence of these potential competitors and provides a short, illustrative scenario that will suggest future research needs.

Terrible Swift Sword: Preemptive Strategic Counterforce Versus Russia and China

In a pair of articles from 2006, academics Keir Lieber and Daryl Press argued persuasively that the U.S. nuclear modernization program begun in the 1970s has paid off handsomely. Combined with a decline in the readiness of Russian nuclear forces and early warning, the United States could launch a “bolt-from-the-blue” counterforce attack with a high likelihood of destroying all but a handful of Russian strategic forces. China’s much smaller nuclear arsenal is even more vulnerable to a U.S. first strike (Lieber and Press, 2006a, 2006b).

Lieber and Press, though maligned by some for even suggesting that the United States would ever consider a first strike, have only starkly described a situation that many already understood to be the case. RAND analysts came to a similar conclusion about U.S. strategic forces in 2003, noting,

> The force is larger than it needs to be if deterrence by threat of nuclear retaliation is the sole objective of U.S. nuclear strategy. Even a mildly expanded target base that included selected targets in emerging nuclear powers as well as chemical and biological weapons facilities in a larger set of countries would not necessarily require the sort of force that the United States plans to maintain. What the planned force appears best suited to provide beyond the

\(^5\) For a contrary view on the ease of deterring China, see Shulsky (2000).
needs of traditional deterrence is a *preemptive counterforce capability against Russia and China*. Otherwise, the numbers and the operating procedures simply do not add up. (Buchan et al., 2003, p. 92. Emphasis in original.)

Indeed, given the importance of preemptive counterforce in the history of U.S. strategic thought, it would be surprising if the United States did not have potent capability in this regard. While it is often seen as a taboo subject to broach, the logic of extended deterrence based on nuclear weapons discussed earlier has always rested on U.S. willingness to use nuclear weapons first. As noted, the threat to use them first is inherently more credible if it seems that one could do so without significant fear of catastrophic retaliation.

Lieber and Press found this U.S. nuclear primacy troubling, as it may mean that Russia and China will take risks that may make accidental war more likely. It may also mean that the United States will be more likely to run risks in confrontations with these states, secure in the knowledge that it once again possesses nearly splendid counterforce. The interaction of these dynamics may make crises extraordinarily dangerous in the future, possibly leading to a nuclear war that neither side truly intends (Lieber and Press, 2006a, pp. 31–33).

While Lieber and Press’s worries cannot be dismissed out of hand, they appear to be overstated. The United States, even when resting extended deterrence almost entirely on nuclear weapons, was always extremely circumspect about even obliquely threatening their use; this was no less the case during the 1950s when it still retained a near-monopoly on long-range nuclear weapons. In addition, at present and for the near term, U.S. conventional capabilities greatly reduce the need to rely on nuclear weapons for extended deterrence relative to the 1950s.

Further, Russia and China do not appear panicked by the current state of affairs. China has never sought to build an incredibly robust deterrent; U.S. forces have always had a counterforce capability against

---

6 Betts (1987) provided many examples of the reluctance to even directly threaten the use of nuclear weapons during the “nuclear golden age.”
Chinese forces. While there are signs that China’s force modernization may make this force less vulnerable, there has never been any sign of Chinese desperation about U.S. counterforce. Russia, while slowly modernizing its forces, also appears to be relatively unconcerned.

This relative lack of concern by both of these near-peer competitors is understandable, as they have not sought to extend deterrence to others in the same way that the United States has. Russian and Chinese nuclear forces exist almost exclusively to provide basic deterrence, which is inherently credible in intent. Given this highly credible intent to “trade Moscow or Beijing for Washington,” even a relatively small capability is very effective. What U.S. president would undertake an operation with even a 5 percent chance of resulting in the destruction of one or two major U.S. cities in any but the direst circumstances? A similar rationale underpinned the French and British nuclear-force structure in the Cold War; the survival of even a handful of nuclear weapons would give even the most hardened Soviet pause in launching a first strike. Here, the fear aspect of deterrence is clearly critical; even small uncertainties about relative gains from a first strike can deter in all but the bleakest scenarios involving highly credible threats (see Kahn, 1961, pp. 126–144; Jervis, 1984, p. 175, n. 47; and Kahneman and Tversky, 1979).

At the same time, neither Russia nor China has an incentive to launch first in most circumstances. The submarine portion of the U.S. deterrent alone is enough to inflict a devastating countervalue attack on either country even if its first strike succeeded in totally destroying both land-based legs of the nuclear triad. Given that neither Russia nor China is likely to be able to disarm a large portion of the land components of the triad in the foreseeable future, both have that much less incentive to fire first. Only in circumstances in which crises escalated to the point at which China or Russia felt that its basic deterrence was being undermined would this seem likely. This argues for limits on cer-

---

tain military options in crisis or confrontation, examples of which will be given in the scenario in the next section.

However, as in the Cold War, the credible threat of a damage-limiting (or even disarming) first strike by the United States helps makes extended deterrence more credible. This may be particularly important in East Asia, where U.S. guarantees to Japan reduce the likelihood of Japan going nuclear (see Hughes, 2007). The Chinese, at least for now, appear willing to accept a status quo in which their deterrent is vulnerable to a U.S. first strike in exchange for Japan remaining nonnuclear.

Additionally, as Lieber and Press noted, the deployment of a modestly capable missile defense has synergistic effects with preemptive counterforce capabilities. If only a handful of warheads survive and are launched nonsimultaneously, then the missile-defense system will be able to engage them singly or in small groups, thus making it vastly more effective. However, it seems unlikely to radically change the existing calculations of credibility. Even an 80 percent interception effectiveness against five warheads means that one is likely to get through. Given previous Russian and Chinese attitudes, this seems sufficient for basic deterrence.

**Bears on Land, Dragons at Sea: Conventional Deterrence Versus Russia and China**

Conventional deterrence against Russia and China will likely resemble that of the late cold-war period. However, like preemptive counterforce, it will be easier in all but a very small number of scenarios. China, for example, is not in position to threaten almost any U.S. allies with land forces. Of all U.S. allies, only Mongolia has a significant land border with China that is not extremely rugged terrain.

Russia, in contrast, does border several members of NATO. However, Russia’s quantitative edge in conventional forces during the Cold War has atrophied even as the United States and many of its allies have widened their qualitative edge. For example, Russia has more than 22,000 tanks; however, only 4,900 are modern (T-80/T-90), and it is unclear how well maintained any of the ex-Soviet armor is. In contrast, Germany alone has 2,200 tanks, all of which are modern (Leopard 1/
Leopard 2) and reasonably well maintained. Russia’s army, which is heavily dependent on conscription, totals only about 400,000 active soldiers; Germany’s army alone has 191,000 (more than half of whom are professionals rather than conscripts) in the active army. Combined with the eastern expansion of NATO and the resulting strategic depth, it appears likely that European forces alone provide significant conventional deterrence against Russia (Military Balance 2006–2007, 2006, pp. 74–76, 148–154).

The U.S. role in European deterrence against Russia will be in the area of intelligence and surveillance along with rapid reinforcement through naval and sea assets. In addition, it will provide some additional assets through the U.S. Army’s V Corps. It seems likely that the ability to successfully accomplish rapid yet limited operations will have a strong deterrent effect. This argues for maintaining a modest U.S. ground and air force presence in Europe, along with at least occasional major exercises with intent similar to that of REFORGER (which was last conducted in 1993).

One possible way to accomplish this is to conduct a “scale-model” REFORGER using something like an armored cavalry regiment. Even if done only every few years, such an exercise would ensure that any future crisis could be responded to smoothly. Moving forces to Germany rather than further east may also be less provocative, both in peacetime and in crisis. At the same time, such forces could also threaten the Soviet territory of Kaliningrad, which is wedged behind NATO states along the Baltic coast.

There are a few areas in which conventional deterrence will be tenuous. Ukraine and Georgia provide two examples of Western-leaning states that appear to be a bridge too far for conventional deterrence at present. This may be irresolvable, as further NATO expansion to these countries would be provocative.

In contrast to conventional deterrence of Russia, where Europe can provide a major component, much of the burden of conventional deterrence against China will rest on the United States. Japan still has only limited capability and continues to be bound by an extremely pacifist constitution. Though this is showing signs of change, for the near term, little can be expected. Other regional allies will have little
to contribute and may be extremely hesitant to commit to a conflict between two nuclear-armed great powers.

Fortunately, the United States, as noted, will be concerned primarily with deterring China in the maritime environment. The U.S. advantage over China in this area is substantial; even given Chinese force-modernization plans, this advantage will not erode substantially in “blue-water” operations for the near future. Only in contingencies in or near the Chinese littoral will Chinese force modernization be cause for significant concern. The most likely of these contingencies, confrontation over the Taiwan Straits, will be briefly discussed in the scenario to come.

It is important to note that the U.S. advantage in maritime power is likely to be enduring for reasons of simple geography. China, like the Soviet Union in the Cold War and Germany before World War I, is a land power with some pretensions to maritime power. Yet it must first and foremost ensure its territorial integrity against nearby rivals. In the case of China, these include India (though this border is extremely defensible) and Russia, and, to a lesser extent, Vietnam. This dilemma sets a lower bound on ground-force spending for the Chinese.

The United States, in contrast, has not faced a threat along its borders of any significance since the Civil War (bandits, such as Pancho Villa, notwithstanding). It is primarily a maritime power that has, over time, developed a strong army. It is no accident or mistranslation that the Chinese navy is officially called the People’s Liberation Army Navy. In contrast, the U.S. Department of the Navy contains its own army in the form of the U.S. Marine Corps. This enduring geostrategic advantage can be amplified by some thoughtful application of a competitive-strategy approach, examples of which are discussed next.

---

8 Blue water refers to open ocean.

9 For example, even those analysts most concerned by Chinese submarine developments concede that China will need to develop significantly more capability before seriously eroding U.S. capabilities. See Goldstein and Murray (2004, pp. 195–196).

10 On the ongoing importance to China of land-border defense, see Fravel (2007).
Under Western Eyes: Russian and Chinese Operational Codes

Significant efforts to understand Russian and Chinese decisionmaking structures and leadership attitudes are already under way. These efforts, both within the U.S government and in academia and think tanks, are generally understood, if sometimes underappreciated. In addition to simply recommending that these general efforts be maintained (if not expanded), one specific area that should be better explored is the area of civil-military relations.

Civil-military relations appear to have a serious impact on military professionalism and military effectiveness; recent works have argued that there is an important link between these factors. Improving both the theoretical and practical understanding of Russian and Chinese civil-military relations, professional education, and their respective officer corps not only would yield better understanding of how to deter these forces but also provides some insight into how effective these forces would be if deterrence were to fail. As discussed next, this research would have implications for states other than Russia and China as well.

How Dire a Strait? Extended Deterrence, China, and Taiwan

The potential for a U.S.-China clash in the Taiwan Strait is significant, though far from a foregone conclusion. The Chinese Communist Party (CCP), to shore up its domestic legitimacy as it moves away from communism and toward a market economy, has used the issue of Taiwan to burnish its nationalist credentials. Changes in the status quo, such as moves toward significantly greater Taiwanese independence, could be perceived as undermining the CCP’s legitimacy and therefore threatening the regime. This threat of loss could make the CCP relatively risk acceptant toward the use of force against Taiwan. The United States, while not formally allied with Taiwan, has all but pledged to ensure that this use of force does not take place. Ensuring deterrence in

---

11 See Brooks (2003) and Nielsen (2005) for a lengthier discussion of the need for such a research program.
the Taiwan Strait will require a careful balance of capabilities to deny Chinese success and reassurance that the United States will “rein in” Taiwanese moves toward independence. The following section briefly discusses a few scenarios for Chinese use of force and steps the United States could take to mitigate them.  

Chinese threats to use force fall into two categories: brute force (in the form of invasion) and coercion through physical and economic harm. The threat of successful invasion, according to several analyses, is relatively low, particularly if Taiwan is militarily supported by the United States (O’Hanlon, 2000; Shlapak, Orletsky, and Wilson, 2000). Amphibious and airborne assaults are difficult even when one has air and naval superiority; if the United States is involved, China will likely have neither. As other RAND analysts have noted, the United States can take certain actions to help ensure that China will not have incentive to act preemptively to disable U.S. air bases in the region (e.g., hardening facilities, such as Andersen Air Force Base on Guam) (Stillion and Orletsky, 1999). Combined with a robust naval presence off Taiwan’s east coast, these recommended measures would mean that successful invasion would be nearly impossible for the near term, though Chinese force modernization (particularly the acquisition of systems to deny U.S. naval and air assets access to the area around Taiwan) may alter this balance in the next decade.  

The second, more difficult problem is that of coercion. Here, the threat is of compellance by punishment, either through the use of China’s vast arsenal of short- and medium-range missiles or through a submarine-based blockade. There is an ongoing debate about the level of coercive threat posed by submarine blockade. However, most analysts concede that the coercive threat of missile use is fairly potent. Esti-

---

12 An excellent overview of China’s ability to challenge the United States in a Taiwan context is Christensen (2001).

13 See OSD (2008, pp. 22–24). Even this very cautious report noted: “An invasion of Taiwan would strain the capabilities of China’s untested armed forces and would almost certainly invite international intervention. These stresses . . . make an amphibious invasion of Taiwan a significant political and military risk for China’s leaders” (OSD, 2008, p. 44).

14 For both sides of the debate, see Glosny (2004) and Goldstein and Murray (2004).
mates of the number of Chinese ballistic missiles that could threaten Taiwan are presently about 700–800 (a mix of DF-11 and DF-15), with more than 1,000 easily possible by 2010 (“Northeast Asian Missile Forces,” 2006). Even armed with conventional warheads, this force could inflict nontrivial damage on both military and civilian targets in Taiwan.

As with other missile threats, two of the primary options for deterrence by denial are missile defenses and counterforce attacks. Success in either approach will be quite challenging, and they are probably best used in tandem. Theater missile defenses are already being pursued, and that point will not be belabored here. However, there has been only a little open discussion of conventional counterforce strikes against the Chinese missile threat.

The conventional counterforce challenge posed by Chinese missiles is formidable. The launchers are mobile and will be well defended by Chinese air defenses. Recent operations highlight that the target-location problem posed by mobile systems remains high for aircraft even in very benign environments. In nonpermissive environments, it will be even more difficult, and delivering munitions on located targets before they relocate will also pose problems.

One “competitive strategy” for dealing with both Chinese air defenses and mobile missiles is that offered by the U.S. Navy’s Trident-submarine conversion (SSGN) program. Based on converted Ohio-class ballistic-missile submarines, SSGNs replace Trident SLBMs with multiple racks of Tomahawk cruise missiles as well as special operations forces equipment (see Coté, 1999, 2002; and Holian, 2004). Given its low acoustic signature, SSGN would be able to approach and remain near the Chinese coast. As in the nuclear case, SSGN is not vulnerable to preemptive attack, thus mitigating incentives for the Chinese to open hostilities in search of an advantage.

While the current arsenal of 154 Tomahawks per SSGN is impressive, even more could be done to optimize the SSGN for the Chinese missile threat. With only modest adaptation, SSGN could carry tactical ballistic missiles capable of rapidly responding when either Chinese surface-to-air missiles (SAMs) or ballistic missiles are located. Additionally, SSGN could deploy small UAVs to aid in the location of these
mobile targets. SSGN could also carry SAMs itself, giving it a self-defense capability against Chinese ASW aircraft and the potential to target other Chinese aircraft, potentially including Chinese airborne warning and control aircraft. Finally, SSGN could carry antiship missiles, giving it an antisurface capability as well. Countering this challenge would be very expensive for the Chinese navy, which, as noted, will not be likely to totally dominate the Chinese defense budget for geostrategic reasons. Though SSGN is no panacea, it is an example of the potential for competitive strategies with future peer and near-peer competitors.

One other use of counterforce weapons in this context is also worth mentioning. There is presently an effort to modify a small number of Trident SLBMs with conventional warheads. This would enable striking targets almost anywhere on very short notice. Such capability, known as Prompt Global Strike, merits investigation, yet raises many concerns. In the China case, for example, the temptation to use such a weapon to strike deep into China’s interior could make China believe it is in a position of “use or lose” with its strategic nuclear forces, possibly provoking inadvertent escalation. However, this system and other long-range strike options also reduce Chinese ability to disrupt U.S. operations through preemption, possibly reducing first-strike incentives in crisis. RAND alumnus and current director of the Office of Net Assessment in the Office of the Secretary of Defense, Andrew Marshall, has recommended such long-range strike capabilities in part because of their potential for enhancing deterrence (see Sherman, 2006).

Finally, the United States must act carefully to reassure China even as it prepares to potentially deter China. In the case of Taiwan, this will primarily mean deterring Taiwan from taking actions that make it more likely to move toward independence. This could prove challenging, as actions that make it seem less likely that China will act against Taiwan may make Taiwan more confident about independence. This is an almost textbook example of “moral hazard,” or what Timothy Crawford has termed pivotal deterrence (Crawford, 2003; see also Christensen, 2002). In trying to avoid this problem, the United States is best served by continuing to hew to an ambiguous declara-
tory policy about the cross-strait relationship. Though nonoptimal for deterring either Taiwan or China alone, it is probably the best approach for deterring both simultaneously. This would mean avoiding deepening military ties to Taiwan, for example, while maintaining the ability to quickly intervene in cross-strait clashes.

The other issue to bear in mind is that Taiwanese willingness to incur coercive costs is almost surely proportional to the desire to move toward independence. The more the Taiwanese government and population desire independence, the more costs they will presumably be willing to bear. If they have mixed feelings about independence, then the willingness to sustain costs will be much lower. This should make the “balancing act” of deterring both Taiwan and China easier. If Taiwan is very motivated, then the need for the United States to intervene will be limited to deterring and preventing a Chinese invasion. This should be relatively easy at present, though it will become progressively more difficult as Chinese force modernization continues. If Taiwan is not very motivated, then deterring them from taking provocative actions should be quite easy.

**Beyond the “Axis of Evil”: Deterrence and the Regional Power**

In many ways, deterring regional powers should be even easier than deterring peer competitors. After all, regional powers lack anything like the capabilities of peer competitors, and the United States has developed an awesome array of capabilities for force projection, as well as an arsenal of nonmilitary elements of national power. Yet coercive diplomacy of regional powers, including deterrence, is potentially more complex than that of peer competitors.\(^\text{15}\) This section discusses the need both for greater understanding of regional powers and their security environment and for presenting a brief scenario in which conventional counterforce operations might significantly aid deterrence.

\(^\text{15}\) See Watman et al. (1995, Chapters 1–3) for extended commentary on regional deterrence. See also Ochmanek and Schwartz (2008).
Peer competitors, for example, will generally be powerful enough to have limited concern about the threat posed by neighboring states (Russia and China, being neighbors to one another, are a partial exception). Regional powers, on the other hand, must not only be cognizant of threats from great powers, such as the United States, but also from less powerful neighbors. Their calculation of the credibility of threats will therefore be more complicated in many instances, as it will involve balancing between nearby (and thus more credible) threats from neighbors and more distant (yet potentially more powerful) threats from the United States. Some regional powers may face additional threats from within their own borders, further complicating threat assessment and deterrence.

For example, prior to Operation Iraqi Freedom, it seemed highly implausible to most observers that Saddam Hussein would risk invasion by the United States by refusing to fully cooperate with weapon inspectors unless he actually possessed either WMD or the facilities to quickly restart WMD programs. Yet with the benefit of hindsight, it is apparent that this is exactly what happened. U.S. coercive threats seem to have failed in this case for two reasons.

The first is that Saddam was not threatened by the United States alone. He faced many other serious threats, both internal and external. The threat of a military coup against him or of another Shi’a uprising were among the most dangerous internal problems he faced. At the same time, he was confronted with the peril posed by Iran, which was much larger than Iraq and led by an ideologically antithetical regime. He was thus faced with the problem that Steven David has termed omnibalancing, that is, the need to balance against multiple threats, both internal and external (David, 1991b; see also David, 1991a).

To meet the challenge of omnibalancing, Saddam sought to use the threat of WMD for deterrence, relying on a formula that one Iraqi general termed deterrence by doubt.¹⁶ This required maintaining high

---

¹⁶ Al-Hamdani (2004). Al-Hamdani is widely regarded as one of the most nonpolitical, competent Iraqi generals and commanded the II Republican Guard Corps during Operation Iraqi Freedom. Much more detail on Saddam’s strategic calculus is available in Woods et al. (2006).
levels of ambiguity in complying with UN inspectors, both in the United Nations Special Commission and United Nations Monitoring, Verification and Inspection Commission. As the Duelfer report indicated,

The Iran-Iraq war and the ongoing suppression of internal unrest taught Saddam the importance of WMD to the dominance and survival of the Regime. Following the destruction of much of the Iraqi WMD infrastructure during Desert Storm, however, . . . threats to the Regime remained; especially his perception of the overarching danger from Iran. In order to counter these threats, Saddam continued with his public posture of retaining the WMD capability. This led to a difficult balancing act between the need to disarm to achieve sanctions relief while at the same time retaining a strategic deterrent. (Duelfer, 2004, Vol. 1, §1, p. 34)

The need to maintain secrecy regarding the actual extent of the WMD program was so high that many senior Iraq military and political figures were unaware of the extent of disarmament (Duelfer, 2004, Vol. 1, §1, pp. 64–65; Woods et al., 2006, pp. 91–92).

Saddam’s beliefs about the United States further amplified his tendency to downplay the U.S. threat. In particular, he believed that the United States lacked the will to commit ground troops to an invasion and occupation of Iraq. Instead, he believed that the United States would once again use airpower against him, in conjunction with encouraging a Shi’a revolt. At worst, he felt that the United States might launch an armed incursion into southern Iraq. While this was certainly a bad outcome from his perspective, it was much less than the actual peril he faced. Saddam thus sought to maintain ambiguity about his WMD much longer than would have been expected had he realized how serious the United States and the threat was that it posed (see Woods et al., 2006, pp. 25–32).

The case of Iraq points to the critical importance of at least attempting to develop a deep understanding of the way in which regional powers’ governing systems function (see Watman et al., 1995, pp. 84–85). While some states may simply be too opaque (North Korea, for example), most are open enough to at least allow some efforts to be
made. In addition to simply mapping the relationships within the government, different models could be developed for the operational codes of the leadership.

Iran may be the current regional power with which an effort such as this is most critical. The Iranian regime is often portrayed as radical and revolutionary, particularly following the election of populist president Mahmoud Ahmadinejad. His rhetoric in particular, virulently anti–United States and anti-Israel at times, is sometimes taken at face value as indicative not only of his beliefs but also of future Iranian policy. This makes some feel that Iran is simply not deterrable in the way the Soviet Union was.

However, the actual structure of power in Iran as well as the regime’s operational code is murkier. It is far from clear that Ahmadinejad’s rhetoric will actually guide his own actions, much less those of the Iranian regime. He may be president, but Supreme Leader Ali Hoseini-Khamenei is the actual ruler. Khamenei, in turn, seems to look for advice from such men as Ali Larijani, the secretary of Iran’s Supreme National Security Council and now Speaker of the Iranian Majlis (parliament). Larijani, while tough enough to previously serve as acting head of Iran’s Revolutionary Guard, holds a Ph.D. in Western philosophy and has written on Immanuel Kant. He does not fit the stereotype of an undeterrable “mad mullah” that some imagine comprise the Iranian regime.

While it would be impossible to devote to each regional power the level of study that was applied to the Soviet Union by Kremlinologists, the Iranian example points to the need for deeper study of the government and society of key regional powers. Iran would certainly be at the top of this list as it is a potential regional hegemon of the Persian Gulf. However, even less powerful states that are potentially of great concern should be examined. North Korea is a prime example of this category, but others might emerge even as others recede. Libya, for example, only recently shifted away from behavior that was cause for grave concern, in terms of both terrorism and WMD. Even weak states that could harbor terrorists or destabilize neighbors merit more study than they currently appear to garner. Both Sudan and Somalia could enter this category in the near term.
Further, the Iraq case illustrates the importance of multipolar environments for deterrence. Few countries will face only bipolar security challenges that allow them the luxury of attempting to develop an optimal strategy for one threat. Instead, like Iraq, they will face multiple potential threats. This will become particularly relevant for nuclear deterrence if proliferation continues.

For example, India faces two nuclear rivals in China and Pakistan. Crisis mobilization of nuclear forces against one could potentially provoke the other. While this possibility seems relatively remote now, it may not always be. These dynamics can be further exacerbated by the perception of offensive first-strike advantage. As discussed earlier, in the nuclear arena, counterforce options and the ability to strike first are enhanced by an enemy possessing a small arsenal with vulnerable command and control; this is likely to be the case for India and Pakistan, if not China, for some time to come. Additionally, Indian concern about the possibility of being confronted by Pakistan and China simultaneously may in fact make India more likely to alert its vulnerable nuclear forces in crisis.

Similarly, more care will need to be given to alliance politics and extended deterrence in this type of environment. NATO was an incredibly robust alliance that existed in a bipolar security environment. Yet as the debate about tactical and theater nuclear weapons and the neutron bomb illustrated, NATO still had numerous controversies about extended deterrence and its requirements. In contrast, the extension of nuclear deterrence to Japan and South Korea appears to be little discussed, particularly the role of other-than-strategic nuclear forces. Even more embryonic are discussions about extending deterrence to the Persian Gulf states, which may soon be facing a nuclear Iran. Nuclear force posture and response are not issues that can be worked out on the fly in a crisis.\footnote{I thank Jim Dobbins for pointing out the importance of this issue.}
Smothering the “Sea of Fire”: Deterrence and Counterforce on the Korean Peninsula

Perhaps the single most challenging problem facing the United States in terms of deterrence in Asia is that posed by the North Korean artillery threat to Seoul and its environs. While the North Korean nuclear program is apparently not sufficiently advanced at this point to weaponize a design, North Korean long-range artillery could inflict substantial damage on Seoul within a very short period. If the North Koreans used chemical shells, the loss of life could be staggering. This conventional/chemical threat provides the North with a very potent counterdeterrent to the United States and South Korea, which reduces the credibility of deterrent threats intended to check North Korean provocations.

This artillery problem, however, is often overstated to the extent that it seems too daunting to even seriously consider challenging. Much like the case of NATO conventional defense before PGMs, this becomes something of a self-fulfilling prophecy. However, with a combination of first-strike conventional counterforce, counterbattery fires (which can be considered second-strike counterforce), and some level of passive defense of Seoul, this problem need not be insurmountable.

The problem of North Korean artillery is overstated primarily by the conflation of the total number of artillery pieces possessed by North Korea with the number of long-range systems that can actually range downtown Seoul. While the total number is in the thousands, many are quite old, and most lack the range to get across the DMZ and into Seoul, a distance of about 40–50 km. The number of systems that can actually cover this distance is thought to be between 500 and 1,100, about 70 percent of which are self-propelled howitzers, with the remaining 30 percent being MLRSs. This is a far more manageable number of systems from a counterforce perspective.18

Unfortunately, these systems are based in underground shelters with multiple possible exits, making the problem of their destruction much more difficult. It is believed that the North Koreans would move the systems out of the shelter to prepared firing positions, fire quickly,

18 A good assessment of the North Korean artillery threat to Seoul is Bermudez (2003).
and then return to the shelter. This would make counterbattery fire much more challenging, as the window to locate and return fire would be on the order of 1 to 5 minutes (see Matsumura, Steeb, and Gordon, 1998, pp. 22–26). These shelters also make preemptive counterforce more difficult, as the intelligence on locations of shelter openings would have to be very precise to use most conventional weapons, including all but the largest penetrating bombs.

However, a combination of preemptive first-strike counterforce with advanced counterbattery techniques might be capable of rapid attrition of these North Korean artillery pieces. This is in part because even the long-range systems have a relatively limited area in which they can be based in order to range Seoul. This reduces somewhat the target location uncertainty that is generally posed by mobile targets. Combined with the efforts at rapid target location embodied in the Joint Precision Strike Demonstration office along with such efforts as the Counter Multiple–Rocket Launcher Advanced Concept Technology Demonstration, the intelligence problem may be at least partially soluble. While a detailed development of a concept of operations for such an effort is beyond the scope of this study, a brief sketch is presented in Appendix A.

The initial U.S. strike would likely combine a suppression and destruction attack against air defenses by both tactical ballistic missiles (such as the Army Tactical Missile System) and cruise missiles (such as Tomahawk). This would help clear the way for strike aircraft, which would then drop precision penetrating munitions on the suspected shelter locations. In addition, it might also be worthwhile to drop scatterable mines around suspected exits from the shelters. This would help complicate rapid egress and ingress from shelters.

Once this preemptive strike is completed, the emphasis of the operation would shift to counterbattery fire. Rapid target location by both ground-based and aerial assets would allow quick response from artillery. A combination of sense and destroy armor “smart” 155-mm cannon rounds and MLRS rockets would provide an effective means of

---

19 See Bermudez (2003) for a description of these hardened artillery sites.
ensuring prompt destruction of targets, particularly if the use of mines slowed ingress and egress.

It is unlikely that even the combination of preemptive counterforce and counterbattery fire would eliminate all of the long-range systems before they could fire some number of rounds. To minimize damage, it would be worthwhile to undertake some expansion of South Korean civil-defense programs around Seoul. Air-raid drills have long been a part of life in the capital, so this could be done with minimal disruption, assuming that the government in Seoul were politically willing. While this is admittedly a big assumption, North Korea’s continual provocation is rapidly diminishing support for the so-called sunshine policy of engagement with the North. This shift might make expanded civil defense more politically palatable. By taking such steps as providing gas masks and possible chemical counteragents, such as atropine, to citizens of Seoul, the danger of chemical weapons could be mitigated somewhat. Combined with existing drills, this might reduce the casualties caused by a diminished North Korean force to relatively modest levels.

The preceding scenario would need significantly more development before being considered a credible threat and is not without problems. First and foremost is the problem of being subject to North Korean preemption, as assembling the necessary forces would potentially be observable, giving the North Koreans the chance to fire first. However, some discussion of U.S. war plans indicates a willingness on the part of the United States to conduct exercises near North Korea in such a way as to generate “exercise fatigue” as the North Koreans use scarce resources every time they respond by going on alert.20 This, in turn, could be used to disguise preemptive preparations, a technique that helped the Egyptians achieve devastating surprise against the Israelis in 1973. The strategy could also help reveal North Korean plans and even positions of equipment, such as long-range artillery.

20 Auster and Whitelaw (2003). It should be noted that this alleged plan could have been a deception operation or could have been subsequently abandoned. Nonetheless, such an approach may seem appealing in the future if not the present.
This approach, while militarily sensible, has the unfortunate side effect of reducing the United States’ ability to credibly reassure North Korea that it intends only to deter North Korea. Lacking this reassurance, the North Korean leadership, already notably paranoid, may feel that it has no choice but to attack. Conversely, efforts to reassure the North Korean regime, such as a proposed agreement by the United States not to attack the North, might only further embolden the regime, reducing the efficacy of deterrence. This general problem of effective deterrence without excessive provocation is the crux of the security dilemma.

As with the Taiwan scenario, this scenario crucially hinges on the political will of the South Koreans. If South Koreans’ willingness to sustain casualties is moderately high, and their willingness to take passive defensive measures is also reasonably high, then this problem becomes comparatively easy. In this case, deterrence will also be quite easy, as the North Koreans would have to be nearly suicidal to shoot first. On the other hand, if South Korean will is quite low, then virtually no amount of damage limitation would be sufficiently reassuring to them. However, even bearing this in mind, it is important to note that at least part of the point of this effort, like that of PGMs in Europe during the 1970s, is to shore up South Korean political will by making the problem seem less hopeless.

The Men Without Fear? Nonstate Actors and Deterrence

One of the central reasons many argued for the declining importance of deterrence after the Cold War, and particularly after September 11, was that nonstate actors (principally terrorists) were not deterrable. Typically, two main reasons were given for the undeterrability of nonstate actors. First, they were often unknown or at least so elusive that they could not be targeted, and there was little they valued that could be targeted either. Second, many members of these organizations were so highly motivated that they were all but immune to the fear that underpins deterrence. Suicide bombers, for example, were believed to be undeterrable. RAND analysts Paul Davis and Brian Jenkins con-
cluded that, at least in terms of al Qaeda, cold-war deterrence was of little relevance.\footnote{Davis and Jenkins (2002). Davis and Jenkins (2002, Appendix A) did note some aspects of cold-war deterrence they considered useful.}

While there is truth in both of these arguments, particularly in regard to al Qaeda, they should not simply be assumed to apply universally to all nonstate actors, even those that make significant use of suicide attacks. Instead, significant nonstate actors can be divided into two categories of concern: (1) regional armed groups and (2) transnational terror networks and spontaneous terror cells. Both have distinct characteristics that may enable leverage against them in some cases.

Regional armed groups include militias, insurgents, civil-war belligerents, and the like. They most closely resemble states, as they either control territory and population or are attempting to control territory and population. These groups may have regular military units, extensive supply networks, and substantial resources.

Hizballah and the Liberation Tigers of Tamil Elam (LTTE) are perhaps the most successful current examples of this type. Both have made use of suicide terror and are highly competent and well motivated. Yet neither seems undeterrable. Hizballah, while building up its own counterdeterrent, has principally been deterred from undertaking major actions against Israel. The fighting in Lebanon in the summer of 2006 demonstrates that Hizballah’s very strength and control of southern Lebanon means that they have much to lose. Rather than showing the undeterrability of Hizballah, the 2006 fighting shows Hizballah as willing to negotiate when it realized that its actions were not worth the consequences. In fact, the war can be viewed as classic “inadvertent escalation,” in which Hizballah’s kidnapping of an Israeli soldier unintentionally provoked a response they did not anticipate. This makes Lebanon 2006 at worst a failure of deterrence due to unclear Israeli declaratory policy and Hizballah’s beliefs about Israel, rather than an indicator that Hizballah is undeterrable. Another interpretation is that the Israeli government was looking for an opportunity to use military force; if this is true, then the case is not even a deterrence failure.
A similar pattern has held with the LTTE and the government of Sri Lanka. The LTTE has often been willing to engage in talks and maintain cease-fires. Indeed, the most recent outbreak of large-scale violence between the government and the LTTE in July 2006 seems to be another case of either inadvertent escalation or intentional government escalation. A dispute over sluice gates supplying water in the north rapidly escalated into full-scale fighting, despite the LTTE’s apparent willingness to concede control of the sluice gates (see “Sri Lanka Forces Attack Reservoir,” 2006, and “Sri Lankan Aid Workers ‘Shot Dead,’” 2006). The LTTE subsequently agreed to peace talks, which, though unsuccessful to date, indicate a sensitivity to costs that makes deterrence possible (though not easy).

Regional armed groups such as these can thus be considered as akin to regional powers in many important ways. They require the same study of leadership operational codes and the same attention to efficacious and credible deterrent threats. In addition, many of these groups have significant sources of external support, which also merit close attention as both a potential source of leverage for deterrence (the LTTE relies heavily on the Tamil diaspora for funds) and as complicating deterrence calculations (the Iranian relationship with Hizballah may make deterring it more difficult if Israel or the United States were to act against Iran).22

It is worth noting that neither of these groups, despite routine use of terrorism, has acted against an outside power that did not interfere in its conflict. However, both have taken dramatic steps against external forces. The LTTE assassinated Rajiv Gandhi after Indian intervention in Sri Lanka, and Hizballah (or its immediate predecessors) targeted the U.S. Marine barracks in Lebanon during the U.S. intervention. These groups acted to punish intervention and deter future interventions, and the United States should be careful not to lump all regional armed groups together as enemies during the course of the long war.

Transnational terrorist networks are much more troubling from a deterrence perspective. Al Qaeda is the preeminent example of this type, and the anarchists of the late 19th and early 20th centuries were

---

22 On external support, see Byman et al. (2001).
perhaps the first movement or organization fitting this pattern. 23 These groups are characterized by a diffusion of ideology and some operational methods across wide geographic areas (though they may have regional power bases). The core membership is often highly motivated and intensely ruthless. This combination of lack of geographic “address” and high motivation makes these groups relatively insensitive to deterrent threats.

Related to this phenomenon are what can be termed spontaneous terror cells. These are small groups of individuals who have not previously been involved in terrorist movements or are only on the periphery of extremism, who self-organize a small unit to conduct terrorist activities. Sometimes, perhaps even often, these groups receive some type of support or training from the overarching movement. This could include virtual support, such as the use of Web sites to provide bomb-making instruction. Examples of this phenomenon are the groups that conducted the Madrid and London bombings of public transportation. Though perhaps not as highly motivated or trained as the core members of an organization like al Qaeda, these groups can be equally hard to deter.

However, deterrence theory does offer insights into combating these organizations. The first is that, while the threat of deterrence by punishment may be ineffective, deterrence by denial is still an option. 24 This has been, along with offensive action, a major component of the strategy of the United States, seeking to make the conduct of terrorist attacks more difficult and thus deterring would-be terrorists from even attempting an attack. Such efforts might be most effective against spontaneous terror cells, which might simply give up rather than risk failure. This approach does have its limits, as every eventuality cannot be prepared for without incredible cost and societal disruption.

The most vital area for denial is terrorist possession of nuclear weapons. Fortunately, denial in this area, while not easy, is possible. Such steps as attempting to deny nuclear capability to terrorists by

---

23 Several have made this observation. See “For Jihadist, Read Anarchist” (2005).

24 For other interpretations of how deterrence theory can apply to counterterrorism, see Trager and Zagorcheva (2005–2006).
securing fissile material are one clear example of this approach that is not cost prohibitive. Plans to expand nuclear-detection capabilities in both U.S. cities and in cargo ships are other possibilities.

The second insight, however, is one that has only recently been discussed (at least widely and openly). This is the possibility of deterring states from providing nuclear materials to terrorists by expanding nuclear forensics. Nuclear forensics is a technique of determining the isotopic composition of nuclear material (either before or after a nuclear explosion) and then using this composition to determine, for example, the material’s origin and weapon design characteristics.25

If origin can be determined, then the state that provided the material can be held responsible. Given that response to even an attempted nuclear attack on the U.S. homeland is likely to be met with a ferocious response, potentially including nuclear weapons, the ability to identify the source of nuclear material could make deterrence highly credible in this narrow but critically important instance. Moreover, as states and state organizations are virtually the only sources of fissile material (the barriers to entry for production being very high), successful deterrence of transfer of fissile material will all but eliminate the potential for nuclear terror. Some expansion of nuclear-forensic efforts is already taking place, but this expansion should be given both a higher profile and more resources. If both well known and credible, nuclear forensics could contribute greatly to the deterrence of the most catastrophic form of terrorism.

While deterrence in the current and future security environments may look substantially different from deterrence in the past and require different capabilities, the basic concept remains the same. Further, given the desire to preserve the United States as something other than a garrison state, deterrence will continue to be a major component of U.S. grand strategy. While many of the arguments and recommendations presented in this book are open to debate, the central point is that deterrence will be as vital to the long war as it was to the Cold War. Rather than attempting to reinvent the wheel, the huge body of deterrence research produced by RAND provides an excellent starting point for further study and strategic planning.
APPENDIX

Annotated Bibliography


Summary of a briefing of the issues covered later in Builder (1979).


Summary of arguments for first-strike counterforce that is less than “splendid” (completely disarming) but will nonetheless shift the nuclear balance appreciably. Highlights the importance of credibility of first strike for defending Europe.


Study that illustrates the ways in which emerging technology makes conventional counterforce options possible.


A scenario that illustrates the possibilities explored in Builder (1983). Proposes a hypothetical Soviet ICBM capable of delivering conventional submunitions against parked aircraft. The U.S. response is to develop a conventional, silo-destroying ICBM.
http://www.rand.org/pubs/reports/R1088/

Argues that the conventional balance in Europe is not so dire as to be unsalvageable. Proposes a number of changes to NATO conventional spending to address the imbalance.


Early study on the problems and prospects of counterforce.


Early work that points to the potential importance of PGMs for redressing the conventional balance in Europe.


Like Jones (1974), attempts to develop various models of Soviet behavior and assess how various force structures would contribute to deterrence of each model.


Seminal work on modeling the various contingent outcomes of U.S. and Soviet first and second strikes. Also examines various targeting strategies.

http://www.rand.org/pubs/reports/R1065/

Methodologically sophisticated effort to develop profiles that depict various views of Soviet attitudes (such as aggressiveness and cost/benefit perceptions) and then use these models to evaluate the deterrent value of various military forces.

Early study advocating extensive civil-defense measures. Argues that civil defense would boost credibility of U.S. deterrence.


Early work detailing some of the practical problems of deterrence as well as how changing technology in the early nuclear age shaped thinking about deterrence. Also discusses reasons the United States opted for deterrence rather than preventive war against the Soviet Union.


Early work that cogently lays out various ways to try to stabilize potential crisis situations within deterrence relationships.


First work to argue (albeit somewhat obliquely) that counterforce intelligence was not the problem many at RAND believed it to be.


Argues that assumptions about the “bolt-from-the-blue” beginning of any future war may be questionable. Also argues obliquely that counterforce might not face the intelligence problem some believed that it did.


Early work arguing for the importance of dynamic evaluation of military forces. Also notes the importance of nonmaterial factors to military power.

Presents the idea that strategy for long-term competition should be attempt to exploit tendencies in the opponent to overspend resources. One way to do this was to find areas of asymmetric costs, such as air defense, in which the opponent felt it must nonetheless compete.


Argues that the end of U.S. nuclear dominance would make Europeans much less secure. Possible solutions proposed include LNOs and a pan-European nuclear force.


Describes differences in worldview, termed strategic culture, between the United States and Soviet Union. The Soviet worldview is deemed much less amenable to the kind of tit-for-tat exchanges that are the foundation of LNOs.


Early work describing the difficulties in ensuring the stability of deterrence. Emphasizes the importance of a survivable second-strike capability and the problems associated with maintaining this capability.
Active Defense of the United States Against Air Attack, Santa Monica, Calif.: RAND Corporation, February 5, 1948. Not available to the general public.


Augenstein, B. W., Long-Range Surface-to-Surface Ramjet Missiles: Preliminary Investigations and Results, Santa Monica, Calif.: RAND Corporation, RA-15070, March 1, 1948.

———, A Revised Development Program for Ballistic Missiles of Intercontinental Range, Santa Monica, Calif.: RAND Corporation, February 1954.


http://www.rand.org/pubs/reports/R3689/

Davis, Paul K., and John Arquilla, *Deterring or Coercing Opponents in Crisis: Lessons from the War with Saddam Hussein*, Santa Monica, Calif.: RAND Corporation, R-4111-JS, 1991. As of April 25, 2008:
http://www.rand.org/pubs/reports/R4111/

Davis, Paul K., and Brian Michael Jenkins, *Deterrence and Influence in Counterterrorism: A Component in the War on al Qaeda*, Santa Monica, Calif.: RAND Corporation, MR-1619-DARPA, 2002. As of April 27, 2008:
http://www.rand.org/pubs/monograph_reports/MR1619/


http://purl.access.gpo.gov/GPO/LPS54633


George, Peter, Dr. Strangelove, or, How I Learned to Stop Worrying and Love the Bomb, Oxford and New York: Oxford University Press, 1988.


———, *Counterforce IV: Summary Briefing*, Santa Monica, Calif.: RAND Corporation, September 6, 1960. Not available to the general public.


———, *A Study of Bolshevism*, Santa Monica, Calif.: RAND Corporation, 1953.


References


MIT—see Massachusetts Institute of Technology.


Odeen, Philip, “HAK Talking Points: DoD Strategic Targeting Study Briefing,”
July 27, 1972. As of April 23, 2008:

Ofer, Gur, The Relative Efficiency of Military Research and Development in the
Soviet Union: A Systems Approach, Santa Monica, Calif.: RAND Corporation,
R-2522-AF, 1980. As of April 23, 2008:
http://www.rand.org/pubs/reports/R2522/

Office of Management and Budget, “Budget of the United States Government,
FY 2007,” undated Web page. As of April 25, 2008:

———, Budget of the United States Government, Fiscal Year 2007: Historical
Tables, 2006. As of April 23, 2008:

Materials: FY 2007,” undated Web page. As of April 25, 2008:

Office of the Secretary of Defense, Military Power of the People’s Republic of China,
annual report to Congress, 2008. As of April 25, 2008:

Ogarkov, N. V., “Zashchita Sotsializma: Opyt Istorii I Soveremennosti” (“The
Defense of Socialism: The Experience of History and Modernity”), interview in
Krasnaya Zvezda (Red Star), May 9, 1984.

O’Hanlon, Michael, “Why China Cannot Conquer Taiwan,” International

OMB—see Office of Management and Budget.

OSD—see Office of the Secretary of Defense.

OUSD(C)—see Office of the Under Secretary of Defense (Comptroller).

Pedlow, Gregory W., and Donald E. Welzenbach, The CIA and the U-2 Program,
of Intelligence, 1998. As of April 23, 2008:
http://purl.access.gpo.gov/GPO/LPS9941

Hegemony,” International Security, Vol. 28, No. 1, Summer 2003, pp. 5–46. As of
April 23, 2008:
html

Press, Daryl Grayson, Calculating Credibility: How Leaders Assess Military Threats,


Williams, J. D., *Regarding the Preventive War Doctrine*, Santa Monica, Calif.: RAND Corporation, July 24, 1953.


