Part Three
Planning at the Strategic Level
This paper describes a 1991 study applying a strategic planning methodology to sketch the elements of a possible American grand strategy. It then updates the study and draws implications for American defense planning and defense analysis. The analytic methodology used may be appealing to readers even if they differ with the particular assumptions used in the paper. The paper's principal thrust is that uncertainty is a dominating characteristic of the landscape and this should be reflected in the fundamentals of any U.S. grand strategy and defense-planning framework. The first implication is that we should plan flexible and robust military capabilities for diverse contingencies, large and small, not just the contingencies now fashionable for planning. Crucial details of standard scenarios are very unlikely to be correct. Nonetheless, most likely theaters of major conflict can be identified, along with plausible objectives and strategies of protagonists. A second implication is that since the likelihood of a large-scale military threat in the next decade seems to be small, national security strategy should give unusual attention to other concerns, notably economic issues, deterring or coping with second-rate regional aggressors, peacemaking, and peacekeeping. This said, it would be dangerous to ignore the potential early in the next century for larger conflicts or future superpower rivalries, because doing so might create vacuums that would make the emergence of such rivals more likely. Against this background, the paper argues that defense planning should now be centered on a combination of (a) environment-shaping, (b) deterrent, and (c) contingency-fighting needs, and that the nature of defense analysis should change substantially if it is to address these needs rather than become irrelevant. The paper suggests a number of specific implications for defense planning.

INTRODUCTION

Background

Late in 1991, at the request of senior officials of the Bush administration, RAND staff members began writing a series of short papers to inform and stimulate broad discussion of national security strategy (or "grand strategy"), which were subsequently collected and sent informally to policymakers. One
of those papers (Davis, Drezner, and Hillestad, 1992), applied a RAND methodology for thinking about grand strategy in a way encouraging "divergent thinking" and provocative ideas. In this paper I present and update some of that earlier work as a case history illustrating the methodology. I also draw subjective inferences for defense planning, highlighting deviations from tradition. Readers are invited to think about how they could use the same planning approach to come up with conclusions based on their own judgments and values.

The Methodology for Thinking Afresh About Grand Strategy

The work began with a brief exercise attempting to characterize options for U.S. national security strategy in the wake of developments in 1990 and 1991, notably the disappearance of the Soviet Union and the emergence of the Commonwealth of Independent States, which seemed unlikely to be a long-term fixture of the landscape. The methodology used had been developed specifically for planning under uncertainty, and to replace the classic approach of identifying and agreeing upon "threats," deriving "requirements" to cope with those threats, and then deriving related strategies for meeting the requirements. Its components were as follows:

1. Describe the emerging **core environment** and major uncertainties, particularly in the form of **scheduled uncertainties** that can be thought of as future branch points and **unscheduled uncertainties** in the form of potential **shocks** (i.e., potential developments that are quite plausible, but that would never make it onto a list of approved best-estimate threats).

2. Develop a set of alternative grand strategies, referred to by "one-liner" themes for convenience, although all are multifaceted.

3. For each grand strategy, develop a **core strategy**, an **environment-shaping strategy**, and a **hedging strategy**, thereby dealing with uncertainties as a primary matter.

---

1 Other methods are more suitable for carefully structured top-down efforts connecting interests, objectives, strategies, and military missions or tasks. For a description of the "strategies-to-tasks" approach, which is due largely to colleague Glenn Kent, see Kent and Simons (1991) and Thaler (1993).

2 I first developed the methodology with RAND consultant Paul Bracken of Yale University in a study for OSD and the Joint Staff. For a subsequent summary of that work, see Davis (1989). For antecedents in business practice, see Bracken (1989). RAND has used the methods also in studies for USCENTCOM, USPACOM, and the U.S. Army (Levin, 1994).
• Compare the grand strategies in a summary framework based on differentiating attributes of strategy.

• Consider ways to integrate some of the strategies in comfortable ways by exploiting the willingness of the proponents of all strategies to acknowledge the need to adapt strategy over time.

Several features of this approach are particularly important. First, the approach encourages a broad view of the environment rather than one focused on one or a very few threats. Second, it encourages an attitude of environment-shaping rather than the more passive attitude of seeking to "react" to threats as they arise. While simple in concept, this has important implications for national security planning. Third, it specifically requires one to build a hedged strategy rather than one focused exclusively on dealing with best-estimate problems. Last, it moves from a divergent phase of developing strategically different "strategies" to a convergent phase that seeks to combine strategies intelligently, rejecting false dichotomies and false orthogonalties.3 The approach is especially helpful for breaking out of standard mind-sets and encouraging broad and integrative thinking; it is excellent for brainstorming and group discussion. On the other hand, it tends to generate lists that include "apples and oranges" and are neither logically "tight" nor comprehensive. Thus, the approach is a complement to more traditional methods of strategic reasoning and analysis, but not a substitute for them.

Let me now sketch how the approach was used in our 1991 study. The first step was to characterize the environment—both the "core environment" (roughly speaking, a "canonical" view of the environment and trends, one heavily dependent on extrapolation and continuity) and the principal uncertainties of the environment.

THE ENVIRONMENT

The Core Environment

Most features of the emerging core environment were and are familiar (little has changed in the last two years). We referred to "The Great Transition"4 (GT) to describe the structural shift that had taken place (and that may or may

---

3Some of these concepts were used by colleague Zalmay Khalilzad when he headed OSD's policy planning office in 1991–1992. They were adopted by Secretary Cheney, as reflected in Cheney (1992) and other official documents.

4The term "Great Transition" was used by Herman Kahn in The Next Hundred Years, (Kahn, 1976:6). See also Kahn (1979), which elaborates the arguments.
not still be proceeding) at the geostrategic level.\footnote{Opinions differ on the extent of this great transition. The most articulate optimistic appraisal is probably Fukuyama (1992). See also Ajami (1993) and Barley (1993), responding to the more pessimistic discussion of Huntington (1993). For more pessimistic discussions with a structural-realist bent, see Mearsheimer (1991) and Layne (1993). In the wake of events in Bosnia and Somalia, even those of us who believed, cautiously, in the great transition are now less sanguine than we were in 1991. Nonetheless, long-term trends may still be favorable and the United States can affect those trends through leadership.} Table 1 summarizes the environment and trends as we saw them, but includes some additions in italics that I would make today (March 1994). The phrases “Second World” and “Third World,” respectively, represent nations that are not yet “developed” but apparently capable of becoming so, and nations that are in much more trouble and with much less favorable prospects apparent. Examples of the former are Eastern Europe, China, and parts of the former Soviet Union; examples of the latter are other parts of the former Soviet Union, much of the Mideast, and most of Africa. We used the term “developed nations” rather than, for example, “the free world,” “the Western world,” or other previous terminologies.

Another aspect of core environment involves domestic constraints affecting what national security strategies are feasible. Table 2 describes what we saw as some of the most important. Although some can be mitigated by skillful packaging and communication of strategy, constraints will play an even more major role in future planning than in the past, primarily because of the absence of a clear-cut threat, but also because small wars are “different” (Cohen, 1984). Some of the constraints are in tension. For example, while most act to limit U.S. military preparations and activities, the last acts in the opposite direction. In certain types of crisis, there will be strong pressures for the United States to “act like a superpower,” for reasons that include perceived world responsibilities as well as the need for the President to protect our more narrow interests (e.g., our access to Persian Gulf oil or the territorial integrity of formal allies to whom we have given security guarantees). Finally, note a problem of a different sort: There is considerable impatience that takes the form of demands that the President develop a strong vision of the future U.S. role and build a clear-cut and consistent strategy. That is asking too much, and it stems from an implicit nostalgia for the clarity of Cold War planning. We argued, instead, for an adaptive and evolutionary strategy that would avoid overfocusing on particular threats or assumptions.
Table 1
Highlights of the Emerging Core Geostrategic Environment

<table>
<thead>
<tr>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Threat of global war almost nonexistent.</td>
</tr>
<tr>
<td>● Threat of great-power war almost nonexistent.</td>
</tr>
<tr>
<td>● Russian threat to Western Europe almost nonexistent. Eventual reconstitution of threat at most &quot;plausible.&quot;</td>
</tr>
<tr>
<td>● China or Japan as eventual security threat at most &quot;plausible.&quot;</td>
</tr>
<tr>
<td>● Greatest U.S. &quot;threats&quot; now economic and social, and widely seen as such.</td>
</tr>
<tr>
<td>● Many threats to United States are diffuse, ambiguous, and not much affected by military power.</td>
</tr>
</tbody>
</table>

However,

● World filled with danger spots that may require small or moderate military intervention (e.g., Bosnia, Somalia).
● Distinct threat of Russian coercion or reabsorption of Baltic states or Ukraine if conservatives replace Yeltsin.
● Many potentially unfriendly nations and combinations of nations with large armies.
● Worrisome ethnic strife at many points.

World political-economic environment

● Clear trend away from communism.
● Apparent trend in Second World toward liberal democracy and market economies, or at least legitimate governments and relatively unregulated capitalism.
● Trends in Third World less clear-cut; strong strands of fundamenatlist Islamic thinking. In some areas, continued calls for "order" and "a strong hand."
● Many sources of instability throughout Third World and parts of Second World.
● Emergence of large trading blocs.
● Very strong growth in several Asian countries.
● Economic decisions increasingly made by triad of United States, Japan, and European Community through institutions such as G-7, IMF, and even NATO.
● Overall environment still a mix of rich and very poor countries.

U.S. security role in world

● United States no longer dominant in Europe; may become minor player soon, or be invited out.
● United States clearly primary worldwide, with no plausible competitor for the role visible.
● United States constrained to act mostly in coalitions, while remaining able to act quickly and unilaterally in some cases.
● U.S. sense of own role still evolving.

Table 2

Constraints Affecting Strategy Development

- Downward trend of defense expenditures, coupled with diminished sense of threat.
- Aversion to significant U.S. casualties in military actions.
- Instantaneous press coverage, resulting in the need for convincing case to act at all and the need to conclude visible actions rather quickly.
- General concerns about U.S. economic outlook limiting support for external military activities and forward deployments.
- Political obstacles to major changes in reserve-component structure or to depending on the draft for response to strategic warning.
- Obstacles to redirecting the use of the military in peacetime to domestic purposes (e.g., the military cannot be perceived as taking jobs away in teaching, construction, and other areas).
- Moral constraints on use of military (e.g., avoiding civilian casualties).
- Political imperative, at time of crisis, to "behave like a superpower."

Uncertainties as a Central Feature of the Environment

An early conclusion of our 1989 efforts, repeated in the 1991 ministudy, was that the most important feature of the environment from the perspective of planners is the dominance of uncertainty. Most striking is the fact that we do not even know who or what will constitute the most serious future threats (see also Kugler, 1994). This is in stark contrast to the situation over the last four decades. We are in an era in which we are military enemies of almost no one to start with, but during which we will unquestionably find ourselves, from time to time, in confrontation—either diplomatic, economic, or military—with a variety of nations. Which nations will be our partners and which will be our opponents will vary with issue and circumstance. We are in a period of multidimensional competition, which in some respects is a better way to view our problem than focusing on one or two specific military threats.6

A key element of the planning approach, and one that encourages group discussion of a sort that breaks away from traditional conventional wisdom, is recognizing that uncertainties fall into several different classes.

Some uncertainties can be represented as crudely scheduled branches in time. That is, some issues are already on the table, issues that may be resolved in ways we cannot predict now with confidence, but for which we know generally the possible outcomes. Table 3 indicates the ones we identified at the

---

6 This theme has been pursued also by OSD's Director of Net Assessment, Andrew Marshall, by Daniel Goure, David Andre, and others. See, e.g., Andre (1991).
Table 3

Major Uncertainties with Potential Branch Points

- What will be the political, economic, and military futures of the former Soviet Union: Russia, Ukraine, the Baltic states . . . ?
- Regional instabilities: Will the Arabian peninsula settle down politically, or will antimonarchy forces become stronger? If the latter, which type of forces?
- New large threats: Will China become aggressive? Will Japan remilitarize and become a military competitor? Will the Koreas unite peacefully?
- Will the North Korean nuclear program destabilize East Asia and undercuts U.S. counterproliferation efforts generally?
- Will Israel and the Palestinians be successful in their efforts to reach a peace agreement?
- Will NATO expand its membership and charter to deal with Eastern Europe? (Asmus, Kugler, and Larrabee, 1993; Harries, 1993)
- Proliferation: Will potential proliferators see more or fewer incentives? Will the scientists and technicians of the former Soviet Union be a major factor in proliferation?
- Terrorism: Will there be a new wave of terrorism stemming from the Middle East? How will this affect interest in ballistic missile defense?
- Drugs: Will the international drug problem require offensive military actions?
- Economic threats: Will the real or perceived economic threat lead to trade wars and other aspects of political-economic warfare?

...time, plus, in italics, those that I would add today. Some of the principal changes are that even Russia's stability is in question; Israel and the Palestinians are attempting to reach a peace accord; and North Korea's nuclear program looms as a major threat to stability (Bracken, 1993a, and Khalilzad, Davis, and Shulsky, 1993). Perhaps there will be some "resolution" of the Bosnian conflict, although not one that will or should endure.

A rather different type of uncertainty involves possible shocks—events that will come as a surprise when they happen (i.e., we will be completely or partly unprepared for them, even if we have previously recognized them as possibilities) and that will have a significant impact. These are the stuff of crises. Further, they are a crucial factor in prudent defense planning, which must not be captive to the "best-estimate" mentality of the day. Table 4 summarizes some of the many shocks possible in the years ahead, with a few new items added in

7Events in the Soviet Union and Russia since 1989 illustrate well how appropriate the emphasis on uncertainty was in the 1989 study and the 1991 follow-up described here. Political dynamics are to a substantial degree inherently unpredictable during a period of structural change. Such changes are continuing as 1993 draws to a close.
Table 4
Potential Shocks

<table>
<thead>
<tr>
<th>Regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former Soviet Union (<em>and Russia</em>) has civil wars, starvation, dissemination and loss of control of nuclear weapons.</td>
</tr>
<tr>
<td>Cuba collapses: accommodates to United States or has civil war.</td>
</tr>
<tr>
<td>India and Pakistan go to war; nuclear weapons used.</td>
</tr>
<tr>
<td>Islamic or other anti-Western revolution occurs in Saudi Arabia.</td>
</tr>
<tr>
<td>Saddam Hussein regains influence; agents foment coup in Saudi Arabia.</td>
</tr>
<tr>
<td>Sendero Luminoso wins in Peru and exports its revolution (<em>now unlikely</em>).</td>
</tr>
<tr>
<td>North Korea launches last-ditch invasion (<em>or implodes with subsequent chaos and violence</em>).</td>
</tr>
<tr>
<td><em>North Korea disappears from the scene, eliminating one of the two major regional contingencies around which DoD force-planning has been regrettable but nominally structured and reopening the debate, “How much is enough?”</em></td>
</tr>
<tr>
<td>Russia invades or coerces Baltic states.</td>
</tr>
<tr>
<td>Russian invades or coerces Poland or Ukraine.</td>
</tr>
<tr>
<td>United States is asked to leave Europe; NATO dissolves.</td>
</tr>
<tr>
<td>Transition of Hong Kong turns violent.</td>
</tr>
<tr>
<td>Renewed civil war occurs in Cambodia.</td>
</tr>
<tr>
<td>Japan, observing North Korea or for other reasons, decides to develop nuclear weapons.</td>
</tr>
<tr>
<td>United States is asked to leave Korea and Japan.</td>
</tr>
<tr>
<td>United States suffers a deep depression.</td>
</tr>
<tr>
<td>Germany becomes a troublesome and more militant actor.</td>
</tr>
<tr>
<td>Mexico has economic collapse with explosion of emigration.</td>
</tr>
<tr>
<td><em>Peacemaking or peacekeeping operations evolve into wars (e.g., in Bosnia).</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cross-regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrorists actually use or threaten use of nuclear, chemical, or biological weapons.</td>
</tr>
<tr>
<td>Terrorists use or threaten use of ballistic missiles; alternatively they use or threaten use of cruise missiles or satchel nuclear bombs (these would have very different effects on interest in ballistic missile defense).</td>
</tr>
<tr>
<td>Japanese exercise an &quot;economic weapon&quot; (e.g., backing away from U.S. securities) in response to U.S. pressures.</td>
</tr>
<tr>
<td>Islamic fundamentalism accelerates suddenly (e.g., in Iran, the southern republics of the former Soviet Union, Egypt, and/or Turkey).</td>
</tr>
</tbody>
</table>

italics. Humility is appropriate in thinking about such shocks, because the past record of governments in anticipating and preparing for them has been dismal.\(^8\) Doing somewhat better in this regard in the future should be a major

\(^8\)In a spring 1990 application of the strategic planning methodology to Southwest Asia for General Schwartzkopf, then Commander in Chief of USCENTCOM, I listed as mere examples of previous shocks: the Korean War (1950), the Yom Kippur War
priority. This probably means recognizing better the broad range of possible shocks and improving the ability to adapt quickly when shocks occur, rather than improving the quality of strategic warning.

Of the new (italicized) items in the list of potential shocks, the most interesting may be the postulated disappearance of North Korea. If the North Korean threat vanished, then the entire argument about "How much is enough?" would have to be reopened. The DoD would either have to identify a new major regional contingency (MRC) for planning (with no good alternative now visible) or rationalize its force structure without requiring nominal capability for two specific concurrent regional conflicts. This change of paradigm would be difficult to make and sell, but would be desirable in many respects (Davis, 1994b,d).

BASIC THEMES FOR ALTERNATIVE GRAND STRATEGIES

The Concept of Guiding Themes

Having described the environment, the next step was to develop some alternative grand strategies. A bona fide grand strategy must be multifaceted, addressing all the components of strategy in enough detail to establish directions and priorities (see also Davis, 1989). However, it is useful to distinguish among primitive grand strategies by identifying the themes that would be their central feature. If one strategy's theme is to rebuild America, that does not imply an isolationist foreign policy akin to that of the 1920s and 1930s, but it does imply a heavy priority on domestic issues.

The principal factors appearing in the themes we used related to: (a) the degree to which the United States seeks actively to protect and perhaps extend the Great Transition; (b) the degree to which the United States seeks to take the leadership position in world security affairs; and (c) the balance between

---

(1973), the fall of the Shah of Iran (1979), the taking of hostages by the Ayatollah Khomeini (1979), and the Soviet invasion of Afghanistan (1979). My colleague Zalmay Khalilzad and I also discussed how even variants of "standard" planning scenarios can be shocks and noted how difficult real contingencies involving an Iraqi threat to Kuwait could be because of the ambiguity likely about intentions. In that same project we later conducted a contingency planning game in July 1990, which included working through a crisis that turned out to be almost identical to the shock a month later, when Saddam Hussein invaded Kuwait. The point is not that we were prescient, because we did not expect an invasion either (although considering it quite possible), but rather that what later appears as "shocks" can often be anticipated if we move beyond the tyranny of the best estimate.

9A different but related strategic-planning method asks participants to identify fundamental assumptions which, if violated, would seriously invalidate their baseline strategic plans. This approach is described in Dewar, Builder, Hix, and Levin (1993).
emphasis on domestic issues and external security (and, therefore, the allocation of resources between domestic and defense spending).

Alternative Themes

We next developed a baseline strategy and three relatively “pure” grand strategies, which were deemed to be largely orthogonal (i.e., mutually independent). We subsequently compared them more carefully and constructed hybrids.\(^\text{10}\)

A Projection of Then-Current Strategy (Creatively Adaptive “Muddling Along” as a Leader?). It was difficult to characterize the current U.S. grand strategy as of late 1991, but it seemed that it had involved timely and creative adaptation rather than proactive planning. U.S. leadership had been strong and effective, but often reluctant. Except for certain speeches shortly after Saddam Hussein invaded Kuwait, President Bush had seemed rather conservative in his expectations about the world and in his judgments about what could be accomplished. And while the President’s defense program called for impressive across-the-board cuts over the next five years, the changes in the defense program were rather straightforward reactions to an unequivocally reduced threat. There had been little refocusing on the economy and social problems (which some believe cost President Bush the election in 1992) and, many claimed, no strategy to guide details of the military drawdown. As of late 1993, similar criticisms were being made about the Clinton administration, despite Secretary Aspin’s having concluded his “Bottom-Up Review” (Aspin, 1993).

Protecting the Great Transition. This strategy put top priority on the set of foreign-affairs and defense measures necessary to consolidate the Great Transition. It was concerned about assuring the long-term security of Eastern Europe (and even the Baltic states, despite the extreme difficulty of protecting them militarily) and avoiding the reemergence of a major military threat from Russia. Although proponents fully recognized the difficulties in protecting either Eastern Europe (including Ukraine) or the Baltic states militarily, they were even more sensitive to the vacuum created by having a whole belt of newly independent but militarily weak states emerge. These have become buffer states.\(^\text{11}\) Further, proponents believed that for the foreseeable future,

\(^{10}\)More recently, RAND conducted under Army sponsorship a somewhat similar study using the methodology I describe here. In this more extensive and careful work, the alternative grand strategies were characterized in classic terms: realism, multilateral security, democratic internationalism, and strategic independence. See Levin (1994).

\(^{11}\)Proponents saw Saddam Hussein’s invasion of Kuwait as a failure of deterrence caused, in large part, by the ambiguity of U.S. interests and the absence of security
deterrence of any Russian adventures against these states (e.g., under a conservative successor to Yeltsin) should not be hard with a combination of political, economic, and military measures, including prior planning and clarification of Western interests. In the short run, at least, this strategy implied vigorous U.S. leadership, albeit in cooperation with appropriate allies. Proponents disagreed about the appropriate long-run U.S. role, with some believing in a partial handover of security responsibilities to regional states and collective-security organizations such as the UN and Western Economic Union. In general, proponents of this option believed that the Great Transition was more fragile than sometimes recognized: they argued that the debacle in Yugoslavia (evident even in late 1991) should give us pause; future problems between Russia and its neighbors were distinctly plausible.

Protecting and Extending the Great Transition. This strategy went farther than the previous one in that it reflected a belief in the universal values of liberal democracy and the universal effectiveness of market economies. Further, it reflected the belief that U.S. national interests will be served as more nations embrace these values. This strategy argued for the developed nations and international organizations (definitely not the United States acting in isolation) to take a broad range of educational, humanitarian, political, economic, and military measures to encourage extension of the Great Transition to other nations worldwide. Proponents of this strategy would, by and large, favor liberalization even when it came into conflict with other near-term interests. For example, proponents would cautiously encourage liberalization of the Arabian Peninsula nations (at least in the sense of assuring legitimacy of governments), even though that would mean difficulties for the current monarchies. Proponents would seek prompt and effective collective actions against murderous dictators and other enemies of Great Transition values, although they would select their instruments carefully, and did not think of the strategy as propagating a narrow-minded "ideology" or a particular way of achieving the universal values.\(^\text{12}\) It did not involve the missionary zeal or level of hubris the United States was sometimes accused of in the 1950s; it was also more patient.

\(^{12}\)It is undeniably difficult to define where support of universal values slips into zealously propagating an ideology. One example involves requiring that governments derive their legitimacy from the consent of the governed, but without specifying how that legitimacy is demonstrated. Thus, "one man, one vote" is a good deal less "fundamental" than the requirement of legitimacy. Religious freedom is arguably fundamental, but that does not necessarily imply that there cannot be an official state religion. And so on. These are inherently complex matters, but it is striking to note how many values have been adopted by the UN because of their being far more than mere "Western" values.
Nonetheless, it was confident in the ultimate validity of universals, and consistent with that, it pursued “human rights agendas” actively.

**National Renewal.** This strategy put utmost emphasis on revitalizing America: rebuilding the nation’s physical and intellectual infrastructure, addressing a broad range of social problems, and encouraging a high level of international competitiveness. It saw the defense budget as the obvious source of funds for these efforts, and was willing to bet on there being no near-term military threat that would require a large military structure. Proponents thought in terms of reducing the defense budget by 50 percent rather than 25 percent in real terms over the five-year period (see Kaufmann and Steinbruner (1991) for one such budget). Proponents were by no means isolationists, however. They expected the United States to be strongly involved in many aspects of foreign affairs, but they placed top priority on improving economic conditions (establishing a level playing field, avoiding being locked out of the European market, etc.). Proponents saw threats, but the threats they saw were economic. Proponents of this strategy did not believe the United States should be distracted by efforts to establish a new world order, or even to extend arms-control efforts (although they would accept as moderately useful some further arms-control efforts, especially reciprocal unilateral efforts accelerating reductions of defense spending).

**CORE, ENVIRONMENT-SHAPING, AND HEDGING STRATEGIES**

Given grand-strategy themes, the next step was to force proponents to separately define core, environment-shaping, and hedging strategies (Table 5), and in the process of doing so to describe political, economic, military, and perhaps other components. Having done that, we reorganized the results to compare the alternative grand strategies along a number of dimensions (Table 6).

Upon comparing the strategies, it became clear that there were false dichotomies among them. Why, for example, was it assumed that the United States cannot renew itself economically and, at the same time, protect and perhaps even help to extend the Great Transition? The U.S. defense budget was already slated to drop substantially in real terms over the next few years. Further, the defense burden (defense budget as a fraction of GDP) had been dropping for some time and would soon be the lowest that it had been in many decades. And, finally, the cost of accomplishing worldwide security missions would now be lower than in the past, because there were no major threats, the United States had many natural allies, and it had qualitative military dominance in most cases (although it might lack important forward-based infrastructure). This did not appear to be a time when we needed to back away
from a security role because of concerns about overreach.\textsuperscript{13} Instead, we could hope to do as much or more with a good deal less.

Another troubling aspect of the initial strategies was that they accepted the traditional U.S. view that U.S. military forces should be focused almost exclusively on deterrence and warfighting roles (except for occasional humanitarian operations). Why? Do we not have other problems to which military forces are well suited (e.g., by virtue of their discipline, organizational skills, mobility, and diversity of capabilities), and given the need to maintain a significant force structure (greater than needed for "best estimate" military challenges) for environment-shaping and hedging, should we not consider how best to use some of that force structure in peaceful missions?\textsuperscript{14} It is worth noting that the U.S. Army has arguably been the greatest "democratizing agent" of the 20th century as the result of its role in postwar Japan and Germany, and perhaps in the Philippines in the 1930s.

With this and other such items in mind, we constructed a hybrid option called "Renew and Extend Cautiously." By this we meant that the United States would indeed focus on renewing America, but would simultaneously take on the task (with a variety of suitable coalitions and forums) of protecting the Great Transition and, where practical, extending it. To be sure, this strategy would be very cautious about involving the United States in military activities abroad, especially activities that might plausibly be a drain on resources, but it would be proactive in looking for low-risk opportunities to encourage further good trends. Even more important, it would be proactive in developing security arrangements to deter aggression against the newly independent states of Eastern Europe. Since neither these activities nor the most plausible ad hoc military contingencies would require many combat forces (e.g., no more than 3–7 divisions and/or 6–7 wings of air forces), this strategy would establish a new set of missions on which a significant number of regular forces would focus in peacetime, although retaining a secondary combat mission as well. These would include large-scale efforts to improve domestic physical infrastructure (e.g., roads, bridges, and perhaps even urban core areas), to address major environmental problems, and, overseas, to participate in humanitarian activities to eradicate certain diseases and assist development.\textsuperscript{15} We recognized

\textsuperscript{13}This conclusion differs sharply from those of historian Paul Kennedy and various structural realists (e.g., Layne, 1993).

\textsuperscript{14}Implicit here is the fact that my colleagues and I rejected the usual threat-based approach of "deriving" force structure "requirements," because we did not find any of the particular accepted threats to be sufficiently compelling and stressing to justify the force structure we believed was desirable for broader reasons.

\textsuperscript{15}One of the obstacles to such a strategy would be the perception by commercial contractors that they rather than the military should be doing the various jobs (e.g., rebuilding highways). It is plausible, however, that a package deal could be struck in
that this was highly controversial, but we believed it should be looked at seriously: the military could do a uniquely good job, and the problems are of very high priority nationally. We also thought that the symbolic value of such an initiative might be as great as its economic value in that period of gloom (late 1991). Such a strategy would be drastically different from current strategy and had little precedent in recent American practices (especially in its use of the military), but it made sense to us.

Tables 5 and 6 summarize the resulting comparison among the original four options and the hybrid.

CONCLUSIONS OF THE 1991 STUDY

The principal purpose of exercises such as the one we conducted in 1991 was to provide a sense of possible direction and some alternative constructs. We also believed that we could give some conclusions—not about precise directions, but about general guidelines, including new concepts to be looked at seriously. The summary conclusions of this quick-look study were as follows:

Broad Strategy

• Overall national strategy needs to shift substantially toward renewal of America and posturing America for a fierce multidecade global economic competition and improvement in the quality of life for its own sake.

• National strategy should also include a national security strategy maintaining U.S. leadership in global security affairs; it should protect and even extend, opportunistically and selectively, the "Great Transition" in the general direction of liberal democracy and market economies. The United States has a historic opportunity and responsibility, but one that can be discharged only in cooperation with other developed countries and international organizations, not as a matter of imperial will.

• Significantly, in seeking to protect and extend the Great Transition, the United States and its allies can proceed adaptively—taking opportunities, but backing away, without apology, when the price is too high. This is a long-term strategy, not a plan for completing development of a new world order in the next decade.

which the jobs would be shared between the commercial and military sectors. So long as the net effect was to add jobs, such a package might be possible. Given the scope of the challenges available, that seems plausible. Importantly, we believe the investment is desirable.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Encourage liberalization of former Soviet Union and E. Europe.</td>
<td>Encourage liberalization of former Soviet Union and E. Europe.</td>
<td>All items of Option 2.</td>
<td>Reduce security burden although &quot;encouraging&quot; further good trends worldwide.</td>
<td>Protect and extend GT, but choose extension cases very cautiously.</td>
</tr>
<tr>
<td></td>
<td>Pursue Middle East peace and, as minimum, establish U.S. position as honest broker.</td>
<td>Protect newly independent states via collective security arrangements (e.g., CSCE).</td>
<td>Encourage liberalization worldwide through multiple measures, including human-rights agenda in UN.</td>
<td>Refocus resources to domestic agenda of renewing infrastructure and increasing competitiveness.</td>
<td>Emphasize domestic agenda. Establish major domestic renewal role for military. Reduce base force somewhat, especially active ground units.</td>
</tr>
<tr>
<td></td>
<td>Lead developed nations in international security.</td>
<td></td>
<td></td>
<td>Minimize foreign aid. Let Europeans and others fund E. Europe and CIS.</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Go slow on economic aid.</td>
<td></td>
<td></td>
<td></td>
<td>Reduce force structure and reduce presence (e.g., 50%, not 20%).</td>
<td></td>
</tr>
<tr>
<td>Maintain &quot;base force&quot; and forward presence.</td>
<td></td>
<td></td>
<td></td>
<td>Grant leadership roles to regional states or groups (e.g., WEU).</td>
<td></td>
</tr>
<tr>
<td>Invest in economic development of E. Europe and former USSR (possible &quot;Marshall Plan&quot;).</td>
<td>Invest in economic development of E. Europe and former USSR (possible &quot;Marshall Plan&quot;).</td>
<td>Be visibly prepared, with allies, for selective political, economic, and even military intervention.</td>
<td>Invest in infrastructure. Actively seek &quot;level playing field&quot; in foreign trade.</td>
<td>Use military to improve domestic infrastructure and worldwide environment.</td>
<td></td>
</tr>
<tr>
<td>Take unilateral nuclear reductions; eliminate most TNF; seek central control of former USSR's weapons.</td>
<td>Be prepared, with allies, for pol-econ-mil intervention to protect E. Europe. Thus, deter.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td>Base force larger than likely contingencies.</td>
<td>Maintain a base force and forward presence.</td>
<td>In Mideast, be willing to grant international security guarantees.</td>
<td>Identify potential tactics of economic warfare by Japanese and others. Reduce vulnerability.</td>
<td>Maintain benevolent forward presence and leadership.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintain plans and capability for unilateral action in viral areas.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Develop high-readiness reserves.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategy Attribute</td>
<td>Creative Adaptation</td>
<td>Protect GT</td>
<td>Extend GT</td>
<td>Renew America</td>
<td>Renew and Extend Cautiously</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>POLITICAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preeminent objective</td>
<td>Maintain stability in vital regions. Adapt to encourage good trends.</td>
<td>Also, protect GT</td>
<td>Also, extend GT and move toward a new world order</td>
<td>Revitalize America, while reducing security burden.</td>
<td>Revitalize America and protect GT</td>
</tr>
<tr>
<td>U.S. role in international security</td>
<td>Lead cooperatively, reactively, and adaptively.</td>
<td>Lead proactively</td>
<td>Lead proactively</td>
<td>Lead very selectively, on foreign trade issues. &quot;Pass&quot; on optional security issues.</td>
<td>Cooperative and selectively proactive leader</td>
</tr>
<tr>
<td>Extend GT?</td>
<td>Very cautious</td>
<td>Cautious</td>
<td>Strong, proactive, but fairly selective</td>
<td>Encouragement only</td>
<td>Strong, proactive, but selective</td>
</tr>
<tr>
<td>Approach to nonproliferation</td>
<td>Adaptive</td>
<td>Adaptive</td>
<td>Proactive, even interventionist</td>
<td>Adaptive</td>
<td>Proactive, even interventionist, but with emphasis on resolving root security problems also</td>
</tr>
<tr>
<td>Strategy Attribute</td>
<td>Creative Adaptation</td>
<td>Protect GT</td>
<td>Extend GT</td>
<td>Renew America</td>
<td>Renew and Extend Cautiously</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Pol-econ interventionism?</td>
<td>Ad hoc and very cautious</td>
<td>Strong to protect GT</td>
<td>Very willing, in collective setting</td>
<td>Opposed</td>
<td>Very willing in collective setting</td>
</tr>
<tr>
<td>Attitude toward UN and other collective-security organizations.</td>
<td>Arm's-length except to use</td>
<td>Supportive, but not subordinating</td>
<td>Supportive and sometimes with U.S. subordinated</td>
<td>Opposed except for standard humanitarian activities</td>
<td>Expansive and supportive, but with attitude of &quot;using&quot; and &quot;leading.&quot; Variant: hand off in time</td>
</tr>
<tr>
<td>ECONOMIC International economic policy</td>
<td>Ambiguous</td>
<td>?</td>
<td>?</td>
<td>Insist on level playing field; apply sanctions if necessary</td>
<td>Insist on level playing field, but depend more on economic revitalization^</td>
</tr>
<tr>
<td>Federal role in competitiveness?</td>
<td>Modest, via cooperation on regulations, etc.</td>
<td>?</td>
<td>?</td>
<td>Substantial; some industrial policy</td>
<td>Moderate. Negotiations focused</td>
</tr>
<tr>
<td>Foreign aid to former USSR, E. Europe</td>
<td>Modest</td>
<td>&quot;Marshall Plan&quot;</td>
<td>&quot;Marshall Plan&quot; and others (e.g., Mideast)</td>
<td>Minimal</td>
<td>&quot;Marshall Plan&quot; and others. Allies to pay most.</td>
</tr>
<tr>
<td>Foreign aid to basket cases</td>
<td>Modest</td>
<td>Modest</td>
<td>Moderate, via UN</td>
<td>Minimal</td>
<td>Minimal</td>
</tr>
</tbody>
</table>
Table 6—continued

<table>
<thead>
<tr>
<th>Strategy Attribute</th>
<th>Creative Adaptation</th>
<th>Protect GT</th>
<th>Extend GT</th>
<th>Renew America</th>
<th>Renew and Extend Cautiously</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILITARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military interventionism?</td>
<td>Ad hoc and very cautious</td>
<td>Willing but very selective</td>
<td>Willing but selective</td>
<td>Opposed</td>
<td>Willing but selective</td>
</tr>
<tr>
<td>Force structure (army divs. in active/reserve/cadre)</td>
<td>12/4/2</td>
<td>10/6/2</td>
<td>10/6/2</td>
<td>7/4/?</td>
<td>8/4/6 if reserve reform possible</td>
</tr>
<tr>
<td>Approach to planning</td>
<td>Multiple operation plans</td>
<td>Building blocks</td>
<td>Building blocks</td>
<td>?</td>
<td>Building blocks</td>
</tr>
<tr>
<td>Humanitarian use of military</td>
<td>Modest</td>
<td>Modest</td>
<td>Moderate</td>
<td>?</td>
<td>Major</td>
</tr>
<tr>
<td>Use of military to improve domestic infrastructure</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Major</td>
</tr>
<tr>
<td>Use of military for environmental problems</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Optional hybrid national service (mil/domestic)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Structure of NSC</td>
<td>As now</td>
<td>As now</td>
<td>As now</td>
<td>As now</td>
<td>Strengthen economic component</td>
</tr>
<tr>
<td>Domestic NSC?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*The “level playing field” issue has great significance politically, but achieving it fully is less important economically than getting our internal macroeconomic house in order.*
• It is possible to do both of the above (renewal and global security): *The dichotomy between seeking a new world order and attending to our domestic problems is false.* It is false because our economic health is fundamentally intertwined with international affairs; because we do not need to be overstretched economically as the result of our worldwide security activities; and because actions on the one do not contradict actions on the other. It will be necessary, however, to clarify the linkage between national security (broadly construed) and domestic well-being.\textsuperscript{16}

**Political and Economic Strategy**

• Investments in Eastern Europe, countries of the former Soviet Union, and other nations potentially ready for the Great Transition should be seen as national security investments, not as "foreign aid." Their main purpose is to increase the likelihood that current liberalizing movements will be irreversible.

• **The United States, Western Europe, and UN should take special pains to protect the new independence of the Baltic states and Eastern Europe.** Although a Russian threat to either seems remote today, that situation could reverse quickly if conservative elements replaced Yeltsin. Furthermore, clarifying interests and establishing security guarantees now, however unpopular in some political circles, would fill the vacuum that might otherwise lead to trouble some years hence. Further, Russia would today cooperate in establishing such guarantees.

• The United States should be cautiously proactive in protecting and even extending the Great Transition. It should be ready, in cooperation with appropriate allies and international organizations, to intervene in support of movements in states seeking to make the transition, but only if the situation warrants it and the price is acceptable. Most interventions should be political and economic, but should be planned as carefully as military operations—for effect, not cosmetics. West European (and UN) failure to intervene effectively in what was Yugoslavia has been a prime example of how not to do things.

• The United States and its developed-nation allies should be vigorous in deterring or reversing dangerous proliferation. However, this will often require efforts to resolve underlying security problems, which in turn may require developing multinational guarantees for affected countries (e.g., Israel, Pakistan, and even North Korea). The alternative is learning to live in a world with more proliferation.

• The peacemaking effort in the Mideast should continue. In addition to security guarantees for Europe, a satisfactory solution may require significant

\textsuperscript{16}The reader may remember that at the time of the study the alleged overextension of the United States was being widely discussed, due principally to the book on the subject by historian Paul Kennedy.
changes in the nature of Jordan. We should anticipate a substantial price tag to make the solution attractive. Whether the oil-rich Arab states can be persuaded to fund this is unclear. Also, even if peace can be achieved, the economic and social prospects of the region appear dismal without fundamental changes.

- Because of profound cultural differences, which even obscure the universality of some human rights principles, encouraging extension of the Great Transition to Middle Eastern states should be done with the greatest of care and in a language comfortable to the nations involved and through institutions they trust. Finding ways to accomplish this deserves considerable research, since the antipathy toward “Western ways” is deep and broad. We cannot afford to have the Islamic world forever poor and hostile toward the West, and the Islamic world cannot succeed without adopting many of our values and habits, but we must avoid “helping” them in ways they would consider to be threatening.

- The United States should remain involved in diverse formal and informal coalitions, but these will need to adapt to circumstances (e.g., as NATO is attempting to do). Which nations will be our allies and adversaries will vary with issue and instance on political, economic, and military matters.

Military Strategy

- The defense program should be revisited, even though the President’s program in 1990–1991 already anticipated many of the world changes and took many of the scaling-down measures that are needed.

- In contemplating tradeoffs among air, ground, and sea-based forces, the United States should plan to be able to conduct and win one major regional contingency and, more or less simultaneously, to be able to deter a second major regional contingency in circumstances where that can be accomplished with substantial air forces and allied ground forces supporting the regional states in distress (e.g., Baltic or Eastern European states under pressure from Russia, or South Korea under pressure from North Korea). There should be no requirement to be able to commit substantial U.S. ground forces in a second, concurrent, major regional contingency. Should deterrence fail in such a second regional contingency, the full U.S. response might be delayed for months.

- Defense planning should be based on building-block missions, capabilities, plans, and organizations. The mind-sets associated with rigid planning should be eliminated, which will require major reorganizational changes to deemphasize traditional “deliberate planning” procedures and encourage development of teams skilled in adaptive planning. The United States must plan to operate with ad hoc allies, but must be prepared to operate early and unilaterally in some contingencies. Building-block planning will require basic changes in the strategic planning system, including the form of OSD
guidance to the Joint Staff and the form of Joint Staff guidance to the services. (See Davis, 1994b.) Despite the success of Desert Shield and Desert Storm, U.S. military planning is not well suited for contingencies arising quickly.

- Analytical agonizing over the precise size of the new base force for the military is inappropriate. Qualitative judgments are adequate and inevitable. The new base force should be substantial (e.g., 10, 6, and 2 active, reserve-component, and active-cadre army divisions, or something similar with a smaller fraction of active units), not because of major current threats, but because of (1) the size of numerous other national armies; (2) the need to deter other nations from seeing a military vacuum and filling it; (3) the need to avoid rearment of Japan, which might result from a drastic downsizing of American forces; and (4) the need to be able to intervene in force (e.g., with 3–7 divisions) anywhere worldwide, while maintaining a comparable force in reserves (important not only as a hedge against simultaneous crises, but to provide a training base for reservists and to reduce the risks of and therefore increase the credibility of threatening a military intervention). Thus, the size of the new base force should be justified by a combination of environment-shaping (e.g., not creating vacuums) and hedging.

- In fine-tuning the new base force, serious effort should be given to developing high-quality army reserve units capable of effective employment within 120–180 days of call-up that would permit a reduction of active army divisions to approximately 6–7. Similar but smaller reductions should be possible for the other services.

- Even more important than the precise size of the revised base force is maintaining the vigor and quality of specialized high-leverage support activities, such as the C3I infrastructure, work on advanced concepts of organization and doctrine, investment in the R&D necessary to maintain unequivocal technological superiority, and the National Training Center. As a strategy, we should be willing to trade variable-cost components for quality in such support and to have our allies provide most of the manpower.

- To maximize peacetime effectiveness of the force structure sized for environment-shaping and hedging rather than probable threats, the military services need to be tasked with substantial missions for peaceful activities, missions involving perhaps 10 percent of the defense program and military units and requiring many new specialty billets for high-quality officers as part of their normal career development.

- The option should be created for volunteers to join a special national-service program that would provide a combination of military training, education, and participation in peaceful military applications in the United States or abroad. This would provide upward mobility options in some ways comparable to those that have traditionally been provided by the U.S. military.
• Peaceful uses of the military should be based on "thinking big" and should include, in cooperation with private contractors who would otherwise claim they had been improperly deprived of opportunities:
  — Massive reworking of physical infrastructure such as roads, bridges, and ports.
  — Cleaning up environmental hazards.
  — Cleaning up and developing lowest-level infrastructure for major cities in distress.
  — Eradicating selected diseases in target countries.

The principal focus should be on domestic infrastructure, not foreign assistance, especially since foreign-assistance efforts can be counterproductive unless the host governments are reform minded. Importantly, peaceful applications of the military cannot succeed on a large scale without enthusiastic public support that changes institutional perceptions of what is valued.

• Vigorous, broad, and innovative R&D, including prototyping experiments with military units as well as weapon systems, should be seen as a powerful and relatively inexpensive environment-shaping deterrent, as well as the ultimate hedge against reversals in world events. The United States should outclass all competitors with respect to the quality of its weapons, forces, and doctrine. In the event of conflict, achieving success with very few casualties will be a goal. We should pay for this base of R&D and high-technology forces with reduced force structure and readiness of selected units (while maintaining high readiness for key units). In our work with coalitions, we should provide the high-tech leverage and planning skills, while regional states provide, e.g., most of the ground forces.

• In contemplating reconstitution needs, we should be increasingly more concerned with being able to mobilize new forces with best-in-the-world capabilities in the year 2005 than with being able to mobilize 1990-era armored forces in the year 1995.

CRITIQUE AND AN UPDATE

How do the results look today, in March 1994? My answer is, "Not bad, considering the study was done very quickly" (see Levin (1994) for a more recent and extensive effort). Indeed, it looks very much like what has evolved as national policy in many respects. As the result of his Bottom-Up Review, Secretary Aspin has announced (Aspin, 1993) a new force structure and defense plan that is only slightly below the Bush administration's base force and that is very much consistent with the force structure of the 1991 study (but no cause-effect relationships should be assumed). There are some important differences between now and 1991, however:
• The Western nations have done a disastrously poor job in dealing with the former Yugoslavia, and there is still no clear national policy on what should be done now. Indeed, there is no agreement even on what should, in retrospect, have been done at the beginning of the crisis—before “ethnic cleansing” became a household word and a embarrassment for all mankind. While no one argued that the developed world had to succeed with each challenge to the new order, this failure was highly visible and came very soon.

• There is still relatively little strategic appreciation for the importance of developing security arrangements to protect the newly independent states such as Ukraine, Poland, and the Baltic states, and for doing so now rather than later, in five or fifteen years, when crises arise (see also Davis, 1994a). There are, however, some interesting proposals being discussed, particularly proposals that would allow for the expansion of NATO to include those threatened states—and Russia as well. The criterion for entry into “the club” would include acceptance of the club’s standards of behavior.17

• As noted above, Secretary Aspin has announced a new force structure and defense plan that is only slightly below the Bush administration’s base force and that is very much consistent with the force structure of the 1991 study except, importantly, in emphasizing active forces rather than enhancement of reserve forces. I still believe more could be done with the reserves (see also Rostker, Don, and Watman, 1994).

• There has been little progress in tasking the services to take on major peacetime roles of the sort proposed. There has, however, been a string of peacekeeping/peacemaking activities that have sorely stressed some of our force structure, particularly support forces that are primarily based in the reserves.

• The expansion of Asian capitalism has proven even more important than was thought earlier. Chinese economic growth, coupled with the severity and duration of the Japanese recession, imply dynamics in the region that were not adequately anticipated in 1991.

**IMPLICATIONS FOR DEFENSE PLANNING AND DEFENSE ANALYSIS**

**Higher-Level Defense-Planning Objectives**

Grand strategy is inherently rather abstract. What implications does it have for defense planning at a more detailed level? It seems to me that there are many important ones. First, as Table 7 suggests,18 defense planning should

---

17This is controversial; see Khalilzad (1993), Asmus, Kugler, and Larrabee (1993), and Harries (1993).

18This table is adapted slightly from Winnefeld (1992). See also Kugler (1994).
Table 7
Regional Breakdown of Higher-Level Defense-Planning Objectives

<table>
<thead>
<tr>
<th>East Asia</th>
<th>Europe</th>
<th>Middle East</th>
<th>Elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environment-shaping (long term)</strong></td>
<td>Creating conditions where no single power is seen as military hegemon</td>
<td>Creating conditions where no single power is seen as military hegemon</td>
<td>Demonstrating that access to resources is vital U.S. interest</td>
</tr>
<tr>
<td>Making arms races unnecessary</td>
<td>Making arms races unnecessary</td>
<td>Demonstrating that U.S. and Arab security interests are not irreconcilable</td>
<td>Improving security climate</td>
</tr>
<tr>
<td>Encouraging orderly change</td>
<td>Encouraging orderly change</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deterring threats (near and midterm)</strong></td>
<td>Korea, SLOCs, Russia, and rest of CIS</td>
<td>Russia and rest of CIS</td>
<td>Iraq, Iran, Libya, SLOCs</td>
</tr>
<tr>
<td><strong>Responding to contingencies (near term)</strong></td>
<td>Korea</td>
<td>Residual Europe, Libya</td>
<td>Aggression in Gulf, against Israel, U.S. citizens</td>
</tr>
</tbody>
</table>

Focus less on meeting the demands of precisely defined but highly questionable planning scenarios for major regional contingencies and more on "softer" requirements. Note the emphasis on (a) environment-shaping, (b) deterring threats, and (c) responding to contingencies.

**Contrasts Between Classic Planning and New-Look Planning**

Despite uncertainty, it is possible to draw some contrasts between classic defense planning and the planning suitable for the post–Cold War era consistent with the kinds of grand strategy the United States appears to be pursuing and the strategy recommended here. Table 8 (adapted from Davis and Finch, 1993) summarizes differences in emphasis. The number of bullets in each category indicates impressionistically the relative emphasis of the various
Table 8
Contrast Between Old and New Emphases

<table>
<thead>
<tr>
<th>Subject</th>
<th>Cold War</th>
<th>Post-Cold War</th>
</tr>
</thead>
<tbody>
<tr>
<td>Great-power deterrence</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Strategic equivalence (nuclear and conventional)</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Nuclear crisis stability</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Great-power arms-race stability</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Alliance solidarity</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Superpower arms control</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Regional stability</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Nuclear proliferation and counterproliferation</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Ballistic-missile, chemical, and biological-weapon proliferation</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Regional-power deterrence</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Regional military stability and related environment-shaping</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Regional arms control</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Discouraging new military great powers and related environment-shaping</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Coalitional intervention in support of new world order concepts</td>
<td>***(?)</td>
<td></td>
</tr>
<tr>
<td>Crisis management and rapid decisionmaking</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Reactions to or preemption of terrorists and third countries with weapons of mass destruction</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Ability quickly to create and operate militarily with ad hoc coalitions</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Realistic war objectives</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Concern about casualties</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Concern about achieving national consensus before action</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Fiscal restraint</td>
<td>**</td>
<td>***</td>
</tr>
</tbody>
</table>
topics. It suggests a shift away from great-power deterrence, strategic equivalence, and the like toward more emphasis on regional issues and regional environment-shaping. There may be more coalitional intervention in the Third World, but that remains to be seen and may depend on events in Somalia and the former Yugoslavia over the first months of 1994. The table also indicates much more explicit concern about realistic war objectives, minimizing casualties, ground-roots support for military action, and—above all, perhaps—fiscal restraint.

**Special Challenges for Defense Analysis**

It follows, I believe, that analysis for defense planning should give particular emphasis in the years ahead to:

- Measures of how easy or difficult it would be for nations to become military superpowers, and what the United States could do to make those steps less attractive (e.g., by consistently identifying trump cards (see also Bracken, 1993b).

- Theater ballistic-missile defense systems with the potential to be used for continental ballistic missile defense as well.

- Ways to assist newly independent nations such as Ukraine and the Baltic states in establishing a reasonable degree of deterrence-based security despite being bordered by a far more powerful neighbor. This could include transferring or selling highly effective defensive weapons or developing deterrents based on the ability to punish aggressors, probably through the use of precision bombing.

- Weapons and tactics effective in peacemaking and peacekeeping operations.

- Rapid adaptive planning and rapid force deployment for major regional contingencies that unfold in very "nonstandard" ways (see also Davis and Finch, 1993, Davis, 1994b, and Kugler, 1994).

- Creative ways to employ the active force structure in peacetime so as to justify its size.

**ACKNOWLEDGMENT**

I would like to acknowledge the collaboration of Paul Bracken in developing the methods discussed here and of Richard Hillestad and Stephen Drezner as co-authors of the 1991 study. Jonathan Pollack, Paul Bracken, and Stephen Drezner provided useful review comments on the current paper.
BIBLIOGRAPHY


Davis, Paul K., and John Arquilla (1991), Deterring or Coercing Opponents in Crisis: Lessons from the War with Saddam Hussein, RAND, Santa Monica, CA.


Kahn, Herman (1976), The Next Hundred Years, William Morrow and Co., New York.
Khalilzad, Zalmay (1993), Extending The Western Alliance to East Central Europe: A New Strategy for NATO, RAND, Santa Monica, CA.
Khalilzad, Zalmay, Paul K. Davis, and Abram Shulsky (1993), Stopping the North Korean Nuclear Program, RAND, Santa Monica, CA.
NONSTANDARD CONTINGENCIES FOR
DEFENSE PLANNING

Richard L. Kugler

During the Cold War, U.S. defense planning was conducted primarily on the basis of a few canonical scenarios used to define force structures capable of handling other conflicts as well. This practice is now being carried forth into the post–Cold War era with adoption of canonical scenarios for major regional contingencies (MRCs) in the Persian Gulf and Korea. This paper argues that while canonical scenarios should continue to be used, they should be supplemented by serious analysis of nonstandard scenarios involving different versions of the standard conflicts and different conflicts altogether. In some cases, such nonstandard scenarios could not be handled by U.S. forces sized and designed to fight the more canonical scenarios. The paper develops this thesis by first examining the inability of the canonical MRCs to reflect the complex trends underway in international security affairs. It then defines a number of plausible and stressful nonstandard scenarios, ranging from small conflicts, to major conventional war, to regional nuclear crisis. It concludes with a call for major changes in the ways U.S. force planning is conducted, including replacement of "contingency-based" planning with generic "mission-based" planning. These changes aim at producing greater intellectual breadth, an adaptive planning process, and a more flexible force posture.

INTRODUCTION

A new era of international security affairs is dawning, and the United States faces a difficult challenge in planning its future conventional defense strategy and forces. This paper addresses one central aspect of this challenge: the need to be prepared not only for a few canonical cases, but also for a wide range of nonstandard contingencies—i.e., conflicts that are significantly different in political and military terms from the limited set of usual (canonical) scenarios largely used to guide defense planning. The nonstandard conflicts are

---

1This need is a major theme of Davis and Finch (1993) and Winnefeld and Shlapak (1990); Davis (1994a) summarizes major elements of the thesis and proposes fundamental changes in strategic, programmatic, and operations-oriented defense planning.
important because many of them are plausible and they might not be lesser included cases—i.e., the United States might not be able to manage them with the force structure and strategies derived from focusing on canonical scenarios.

The idea that the United States needs to be prepared for nonstandard conflicts is not new. In recent years, the Joint Staff and CINCs have increased the adaptiveness of operations planning, and the Secretary of Defense has exhorted the entire Department of Defense to prepare for diverse contingencies. Nonetheless, the Joint Staff's conversion to adaptive operations planning is not yet complete, and more generally, the staffs of the Department of Defense continue to rely heavily on canonical scenarios, especially for program development, but also for assessment of readiness and many other functions. Nonstandard conflicts are too often regarded as side excursions—i.e., as of marginal importance or as mere lesser included cases.

The attractiveness of canonical scenarios is due to their specificity and the value of having standard cases against which, for example, to measure alternative program options. A canonical scenario helps impart intellectual focus to defense planning by specifying—in place, time, and features—an important future wartime conflict that might actually be encountered. It typically includes a single-point estimate of adversary and allied force levels, mobilization and reinforcement times and rates, strategy and doctrine on both sides, and other key data. Owing to these details, canonical scenarios provide concrete information that can help inform decisions about force requirements and program priorities. The core purpose of these scenarios is to empower the defense effort, not imprison it. But using them in a straightjacket way can purchase in-depth focus on a few events at the expense of grasping a wider set of challenges.

Canonical scenarios first rose to prominence during the Cold War, and although they had a constructive impact, they also left behind a mixed legacy. The canonical scenarios of the Cold War have now been cast aside, but new ones have replaced them. Acting on the strategy goal of being prepared for two roughly concurrent major regional contingencies (MRCs), the Pentagon is now employing MRCs in Southwest Asia and Korea as its primary basis for strategic

---

2Aspin (1993) specifically notes that the canonical scenarios of the Bottom-Up Review (BUR) are illustrative only, and that the wars actually fought are often very different from the ones used for planning. Powell (1991) also stresses the role of uncertainty. The challenge, however, is translating such exhortations into capabilities.

3Another virtue of the canonical scenarios is that they help to establish a credible rationale for U.S. defense policy and programs. Whereas future conflicts in diverse locations are "invisible," today's MRCs are "visible" and therefore "believable." A sense of proportion is needed, however, for the future should not be discounted because of its being unforeseeable (Winnefeld, 1992).
planning. The Persian Gulf scenario involves a repeat Iraqi invasion of Kuwait and Saudi Arabia, launched by a force of 20 divisions and 750 combat aircraft. The Korea scenario envisions a North Korean attack on South Korea, one employing 35 divisions and 800 combat aircraft. These MRCs are being supplemented by several lesser regional conflicts (LRCs), similar to that of Operation Just Cause, when U.S. forces intervened in Panama. Because LRCs impose only modest demands, the two MRCs seem destined to play the dominant role in future defense decisionmaking. To a degree, they will help shape operational doctrine for employing U.S. military power on the basis of decisive force. Equally important, they will help shape plans for building a new force posture (see Aspin, 1993 and Powell, 1991).

This paper accepts and expands upon the thesis that planning for nonstandard conflicts needs to be moved to center stage. The goal should be military plans and planning capabilities that are readily adaptive and a force posture that can perform a broad range of missions in ever-changing situations (Davis and Finch, 1993; Davis, 1994a). These demands cannot be met if defense planning wears intellectual blinders.

To explore the issue of canonical versus nonstandard scenarios, the paper proceeds as follows. First, it surveys the international security environment and demonstrates that it makes no sense for the United States to base planning on the canonical scenarios. Second, it notes that the current MRCs do not even necessarily establish appropriate force-posture goals. Third, it describes with a broad brush a diversity of plausible nonstandard scenarios that should be considered. Finally, it draws some conclusions for defense planning, arguing for an emphasis on generic mission-based planning methods that could support an adaptive planning process and a more flexible force posture.

TRENDS IN THE INTERNATIONAL CONTEXT AND U.S. MILITARY MISSIONS

Regional wars of the type identified by the canonical MRCs are probably the largest plausible military conflicts for the relatively near future. Since the threat of a global hegemon has passed into history and is unlikely to reappear, worldwide war, as encountered in World War II and prepared for during the Cold War, should no longer animate U.S. defense policy. Iraq and North Korea are highly plausible aggressors against vital U.S. interests. The issue is not whether planning should highlight these two threats, but whether the specific canonical scenarios dealing with them are a sufficient basis for gauging military needs for the full spectrum of situations that might lie ahead. The real military contingencies the United States will face may turn out to be very dif-
ifferent scenarios involving Iraq or North Korea, or they may involve altogether different adversaries, theaters, and military and political environments.

The stage can best be set for analyzing this issue by first evaluating trends now underway in international security affairs. If future wars erupt, they will occur as a result of international political dynamics. The trends at work today suggest that the future may offer greater turbulence than was commonly anticipated in the months following the Cold War's abrupt end.

The End of History or Its Rebirth?

As of 1990–1991, popular academic fashion in some quarters held that the end of the Cold War meant not only the disappearance of a rival global hegemon, but also the end of all forms of major military conflict. Several trends were cited: the alleged worldwide triumph of liberal democracy, which was to bring a taming influence to diplomacy; growing economic interdependence; ready access to resources through open trade; the settlement of territorial disputes; the decline of aggressive ideologies; the withering of the nation-state; and the growing power of multilateral institutions. This thesis held that the conflicts of the future would be small-scale affairs that could be handled with only moderate doses of military power. Requirements for these small conflicts presumably could be met through multinational cooperation among the liberal democracies, thereby obviating any need for a powerful U.S. defense establishment.

This optimistic thesis drew inspiration from two books, Francis Fukuyama's *The End of History and the Last Man* and Samuel Huntington's *The Third Wave*. Both were misinterpreted when they appeared. Although they celebrated the triumph of democracy and free markets over Bolshevik communism and command economies, neither predicted permanent global peace anytime soon. Fukuyama's book was more an analysis of political philosophy, including Hegel's long-neglected importance, than a new theory of international politics. Its key point was that the ideological conflict between democracy and totalitarianism has now been settled in ways that allow for the rebirth of 19th century optimism, which had become a casualty of 20th century pessimism.

---

4Davis and Finch (1993) emphasize that it is an error to assume that "multiscenario planning" or "planning for nonstandard scenarios" is adequately accomplished by considering one scenario each for a diversity of theaters and adversaries. Instead, one should consider the vast diversity of plausible scenarios involving the canonical threats (currently, Iraq and North Korea)—e.g., scenarios with nonstandard assumptions about warning, political alliances, and military objectives.

5For elaboration of many of these points, see Kugler (1993), a strategic-planning study conducted for the Joint Staff, one with a relatively long time horizon.
Yet it was far from blind to the stresses ahead in the coming decades. Huntington's book argued that democracy, originally located in Western Europe and North America, is spreading outward, especially in Latin America. But it also noted that democracy now covers only about one-half of the globe, that reversals might be experienced, and that democracy is not yet well installed in many regional hotbeds of future conflict.

The sobering events of 1992 and 1993 underscore that if the presence of liberal democracy increases the prospects for harmony, then its absence weakens these prospects. Moreover, the thesis that democratic states do not wage war against each other merits scrutiny. The 19th century shows many examples of democratizing nations in conflict. For example, Britain and France both became democracies during this century but spent most of this period in rivalry with each other, and often came close to blows. Indeed, democracy's emergence across Europe was feared as a harbinger of war because it brought nationalism, populism, and mass mobilization in its wake. Also, the United States—democracy's shining light—collapsed into a bitter civil war. The Western democracies joined together in alliance when totalitarianism appeared on the scene in the 20th century. But this development, an exercise in realpolitik as well as common values, is no guarantee of global tranquility in regions outside the Western alliance now that totalitarianism is fading.

Because democracy stresses human rights and the peaceful settlement of disputes, it helps buffer against the propensity to go to war. But volumes of textbooks argue that interstate conflict is caused by many factors that can overpower common ideology. This particularly is the case when the tenets of liberal democracy stop at a nation's borders. Democracy helps ensure that the decision for war is made not only by executive institutions, but by parliaments and the mass public as well. Yet parliaments and mass publics have a long history of being as warlike as kings and presidents, often more so. Popular opinion can pressure even unwilling governments to go to war; conversely, governing elites can choose war in the face of domestic dissent if they conclude that the national interest requires that step. Democracy is no impenetrable barrier to warfare among nations that hate each other for reasons that go beyond similarities or differences in governmental structures. After all, the slaughter in the Balkans has not been caused by a dispute over political ideology.

Recent trends suggest that the collapse of European communism is being accompanied in many places not only by democracy and capitalism, but also by a growing emphasis on national self-determination, as opposed to cooperation with other nations. In Europe, not only has the Yalta agreement of 1945 collapsed, but so also has the Versailles peace order created after World War I. The effect has been to wipe away the Soviet empire and the system of multinational states that was established in the 1920s to stabilize ethnic conflict and rampant nationalism in east central Europe, the Balkans, and the Caucasus.
Coming in the wake is a host of new nations that define their identities in exclusionary terms, along with muddled borders, many citizens residing abroad, and the reawakening of long-suppressed animosities. Already, Europe has become one of the globe’s most violent regions. When these developments are fully manifest, the result might be even greater turmoil and violence (Asmus, Kugler, and Larrabee, 1993). What applies to Europe may also apply elsewhere around the globe, where severe strains require no similar collapse of Versailles and Yalta to unleash them.

If the events in Bosnia have not provided adequate refutation to the thesis of global harmony, the emerging academic literature has called its postulates into question and offered a vision whose sobriety approaches outright pessimism. A good example is the landmark article, “The Clash of Civilizations” (Huntington, 1993), which partly repudiates the earlier optimism of The Third Wave. Huntington now forecasts a gloomy world in which the dominant cultures will clash on a global scale, possibly bringing about a plummet into enduring rivalry over the very basics of human life. He foresees a clash in Europe between the Christian West and the Slavic culture led by Russia. In the Middle East, Islam will clash with its old nemesis, infidel Europe. Elsewhere, he asserts, conflicts among Islam, Hinduism, Confucianism, and the Japanese culture will produce a widespread strife that may overpower efforts to fashion order.

Zbigniew Brzezinski (1993) refrains from pessimism this deep, but he foresees a chaotic future marked by sharp conflicts among nations in key regions, abetted by moral weakness in the West itself. The historian John Lukacs (1993) forecasts that communism in Europe will be replaced not by democracy, but by the return of atavistic ethno-nationalism and militaristic fascism. Paul Kennedy (1993) argues that the key fault line will not be in Europe, but between the modern West and the Third World. He points to the Middle East and other regions where an ongoing population explosion will interact with deepening poverty to produce deep hostility to the West. As for the idea that the Western allies will remain cohesive, Jeffrey Garten (1993) worries that the United States, Japan, and Germany will fall into political conflict over dominance of the global economy. In their chilling book, George Friedman and Meredith Leard (1991) go beyond forecasts of U.S.-Japanese economic rivalry to predict a replay of World War II in the Pacific.

To be sure, these predictions should not be accepted uncritically. They may represent the tendency of the academic literature to conform to Hegel’s model of dialectical idealism. The original thesis of optimism is now giving way to an antithesis of pessimism, which might be replaced by a synthesis. Even so, this gloomy literature contains enough penetrating arguments to suggest that even if the future will not be one of pure Hobbesian realism, it will fall far short of Wilsonian idealism.
The Rise of Multipolarity

As many have noted, the stable bipolarity of the Cold War is giving way to a modern version of multipolarity. Despite the ever-present danger of U.S.-Soviet confrontation, the bipolar era proved to be remarkably stable. This outcome was brought about by nuclear deterrence, by a conventional balance of power, and by political equilibrium because both sides already controlled the areas that were important to them. Multipolarity replaces this stability with new and untested global dynamics that, if not carefully managed, can produce chronic tension, conflict, and war.6

Multipolarity is not unstable if the global system is homogeneous (i.e., nations do not harbor incompatible agendas) and a stable balance of power exists. But instability has often been the outcome, owing to deep-seated heterogeneity that produces political-economic tensions. In this situation, the strong have opportunity to prey on the weak undeterred by fear that a collective response will be mounted. Fear of aggression, in turn, can give rise to policies aimed at self-defense that come across as threatening to others. The quest for unilateral security translates into paranoid insecurity for all, further eroding stability. This, at least, was the case during periods when multipolarity was at its height (Aron, 1966).

Owing largely to the status of the United States as a superpower and the Western alliance system, the future system will be far from purely multipolar. Yet interstate anarchy will be a core feature, for the nation-state is not going away. The number of nations is increasing, amidst a host of internal and external constraints that deny governments confidence in their capacity to shape their own destinies. Although the impulse to cooperate is growing in some areas, it is declining in others. The United Nations and other institutions are already showing a weak capacity to impose the norms of international law in today's more multipolar conditions. Much will depend upon whether these collective security bodies acquire greater power and the ability to use it effectively or suffer the fate of the League of Nations.

A bright future for collective cooperation is far from ensured, for the global trends at work today are pulling in the direction of fragmentation. In addition to the chaos brought about by the collapse of the Soviet empire and the Warsaw Pact, the Western alliance system is weakening as many of its members turn inward to deal with long-neglected domestic problems. The world economy is struggling, and even if a collapse into protectionism and mercantilism is not foreordained, the once-bright future of cooperative free trade is uncertain. Equally important, immoderate ideologies are making a comeback. This is the

---

6For discussions of these matters, see articles by Mearsheimer and others in Lynn-Jones (1991); see also Layne (1993).
case in Europe, where resurgent nationalism threatens democratic reforms and peaceful resolution of disputes, but it also applies elsewhere, most notably in the Middle East. There and elsewhere, endemic poverty coupled with population explosion threatens to produce anti-status-quo ideologies and deep anger toward the Western democracies.

If these negative trends spin out of control, the result could be a witches’ brew of 19th century politics, 20th century passions, and 21st century technology. Even if the worst fails to occur, the future may offer widespread turbulence, along with violence and war, at least until democracy and free-market prosperity finally triumph sometime in the distant future. The Persian Gulf and Korea will remain zones of conflict, but the dangers ahead are far from limited to these two regions. Because the future international system will be unstable in important ways, the dangers are global. They will manifest themselves in many different spots, in ways that include, but go beyond, the threat of renewed regional aggression by Iraq and North Korea.

The Proliferation of Military Power

While it is clearly difficult to forecast the future in the currently turbulent international system, one thing can be said: warfare will remain an important instrument of statecraft. A growing number of nations will have the capacity to conduct aggression on their own; Iraq and North Korea are only two among many nations in this category. Moreover, the feasibility of aggression multiplies many times over when alliance formation is considered. Fear of encountering U.S. military opposition will remain a potent deterrent, but as the Persian Gulf war shows, aggressors can misinterpret the signals coming from Washington. Moreover, the U.S. deterrent will remain potent only if American military power remains strong and can be projected to the locations of conflicts.

In response to the political tensions now emerging in the international system, the spread of nuclear weapons and other instruments of mass destruction seems likely to accelerate in future years. Ukraine has been quite ambivalent about yielding the large nuclear arsenal bequeathed by the collapse of the Soviet Union (although a breakthrough was achieved early in 1994), and Iraq and North Korea have been resisting international efforts to dismantle or stop their nuclear weapon programs. Other nations doubtless have similar programs and may escape international sanctions in ways that allow them to succeed. The disturbing case of South Africa shows how a modern nation, even one under international scrutiny, can cross the nuclear threshold. Because delivery systems must be created, more is involved than assembling nuclear devices, but even so, modern industry is giving many nations the assets to become
nuclear powers. If proliferation is to be prevented, this goal will not be accomplished because supply is cut off, but because demand is controlled. In the final analysis, most nations in possession of modern industrial technology can build nuclear weapons. If they remain members of the nonnuclear club, it will be for reasons of their own political choice, not because nuclearization is physically impossible. Demand is best reduced by ensuring security, but the coming multipolar era is making many nations feel insecure. This applies not only to status-quo nations, but also to nations whose dissatisfaction may lead them to act like rogue elephants.

Behind the scenes, modern conventional weapons are also proliferating. This development alone promises to alter the face of warfare and the requirements facing U.S. defense planning. Notwithstanding the collapse of the Warsaw Pact, Europe remains an armed camp, with many nations capable of inflicting immense violence on each other. This is the case in the former USSR, but it also is true in east-central Europe and the Balkans, a region of great security dilemmas and simmering ethnic hatreds. The Middle East and Persian Gulf also remain armed camps, and many nations there, including several that embrace Islamic ideology, have embarked upon efforts to bolster their arsenals. Animating this effort is the realization that, given Iraq’s disastrous defeat, modern defense establishments will be needed if these nations are to stand up to Western-equipped forces.

In Asia, the end of the Cold War has been accompanied by a little-noticed but startling development. Virtually all nations have launched efforts to build stronger military establishments. These efforts owe to growing economic strength, but they also reflect deep-seated worry that the new era will be turbulent. North Korea remains a military threat, but China, benefiting from a decade of sustained economic growth and favorable prospects, has launched a sweeping program of military reform. What lies ahead remains to be seen, but China, long an inward-looking nation, may acquire an impressive capability for power projection.

In response to fear of North Korea and China, South Korea and Japan continue to strengthen their forces. Indeed, Japan today has one of the world’s largest military budgets. Owing to its mutual security treaty with the United States, Japan’s defense posture is focused on protection of the homeland and nearby sea lanes. Regional economic and security dynamics, however, could lead it in the direction of acquiring forces for power projection. The effects are not confined to Northeast Asia. Several Southeast Asian nations are all taking advantage of growing economic strength to bolster their military power. Indonesia, Malaysia, and Singapore are acquiring better military assets, not only to protect their homelands, but also to project forces abroad. Meanwhile, India and Pakistan both possess formidable conventional defense postures that are being slowly modernized. Even absent further nuclear proliferation, South
Asia and Southeast Asia both will house plenty of conventional firepower for major conflicts.

The outcome could be stability if all these nations act in ways that counterbalance each other. But for mathematical reasons alone, balance will be hard to achieve, owing to large differences in national military power and the unstable dynamics of coalition formation. This would be the case even if participating nations are joined together in a global pact in search of military balance. Equilibrium will be doubly difficult to achieve because, apart from the Western alliance system, collective security is in short supply, and many nations will be seeking a margin of advantage over each other.

The looming prospect of major military imbalances amidst a more multipolar system makes the likelihood of tensions all the greater. This prospect further suggests that future conflicts may take quite different forms than anticipated by DoD’s current MRCs for the Persian Gulf and Korea. These two MRCs will need to be guarded against as a central focus of U.S. defense planning. But any literal and mechanical application of them as the sole focus might result in blindness to the much larger spectrum of very different conflicts that could lie ahead.

Peacetime Purposes of U.S. Military Power

Because these political-military trends do not augur well for enduring global tranquility, they confront the United States with the challenge of figuring out what is implied for its foreign policy and defense plans—not only for war, but for peacetime as well. Facing major domestic problems and a troubled but ambiguous external setting, the United States is now struggling with the dilemma of deciding how to strike a satisfactory balance between assertive global leadership and selective restraint. Exacerbating this dilemma is a difficult policy tradeoff. Over the long term, the United States will be unable to exert global leadership if it fails to restore its economic health. Domestic recovery will be hard to achieve if weighty international burdens must be carried during the time this effort is pursued. Yet if the United States casts aside these burdens in some wholesale way, the result could be a free-fall collapse into international turbulence that makes recovery impossible. Prospects for a cooperative world economy—a precondition for U.S. recovery—might be lost. Beyond this, an expensive military rearmament might become necessary, thereby draining scarce resources away from domestic investment.

The task of striking an appropriate policy balance is complicated by the singular American historical experience with foreign policy. Because the United
States arrived on the world scene only in the 20th century, it has no backlog of involvement in dealing with multipolar politics reminiscent of the 19th century. It is seasoned only in pursuing confrontational politics with totalitarian hegemons and in managing tightly knit security alliances to deal with these hegemons. As a result, it is now hard-pressed to develop the policies and grassroots attitudes needed for a multipolar era of fluid, ever-shifting relations with other nations. Also, it suffers from confusion in defining its own interests in a more Palmerstonian era in which its global involvements are growing. There will be fewer permanent friends or enemies, and enduring interests will play an influential role in shaping the policies of all nations.

Throughout the Cold War, the United States was able to reconcile realism and idealism because confrontation with communism allowed it to pursue global power politics on behalf of moralist principles. The coming era of greater multipolarity threatens to deprive the United States of this luxury in ways that will compel adoption of a new synthesis, one that runs against the grain of traditional American thinking. Because the threat of totalitarian hegemonism has been vanquished, pursuit of a global victory for democracy is the only beckoning moral crusade. Yet the future seems destined to be far too complex to allow this crusade to be the only principle to govern U.S. policy. Indeed, democracy is unlikely to flourish unless security can first be guaranteed. International stability amidst multipolar fluidity will have to be pursued. This goal will be achievable only if the core principles of realism are respected: legitimacy, equilibrium, respect for national interests, and a stable balance of military power.

A critical distinction must be made, however, between a “balance of power” policy and a policy that includes recognition of the need for a balance of military power. The former harks back to the days of 19th century imperial conduct, when nations blindly pursued their own interests, thought in “blood and iron” terms, viewed politics as a “zero sum” game, exploited other nations, and intimidated potential opponents by manipulating multipolar rivalries. The latter policy aims for a healthy synthesis of realism and idealism. It recognizes the centrality of national interests, but in acknowledging the common good, it also pursues democracy and cooperative security. It prefers multilateralism, but will behave unilaterally when necessary. It strives for a military balance of power partly in response to realism’s dictates, but also because this balance is needed if idealism’s goals are to be attained.

American military power must remain engaged abroad, for this power will be key to preserving a stable military balance in many vital regions. Its potential can best be realized through alliance relationships and other multilateral institutions, but these organizations will be weak unless they are undergirded by
U.S. commitments. Military power can support enduring commitments to allies, promote cooperation with new friends, and discourage misbehavior by potential adversaries. These functions will have to be performed in regions where aggression might be mounted by medium-sized states, but they will have to be performed in ensuring stability among the great powers as well. Today Germany and Russia are not in conflict, and neither are Japan and China. Long-run stability among these ancient rivals, however, may continue to depend to some degree on the presence of U.S. forces.

If this larger strategic agenda is to animate U.S. foreign policy, the very paradigm of planning for MRCs in a narrow military way may miss the most fundamental point. This will be the case if these MRCs have been forged on the assumption that the international system is stable to the point where U.S. military power can now be viewed as a sword in a sheath, to be hauled out only when wartime threats emerge and regional military crusades must be launched. To the extent this assumption is being employed, it is being rendered invalid by international trends. A broader conception of the purposes of U.S. military power is needed.

Current U.S. defense policy is not bereft of larger purposes, but the difference between today's situation and that of the Cold War is stark. During the Cold War, DoD's canonical scenarios were embedded in a sophisticated framework of global security policy. Notwithstanding the official attention being paid to other dangers—failures in nuclear deterrence, democratic reforms, and economic recovery (Aspin, 1993)—the current MRCs stand outside any similarly elaborate framework for the new era. At a minimum, heavy reliance is still being placed on Cold War precepts. The key task is not really to fine-tune plans for MRCs, but rather to craft a policy and strategy that lays down innovative precepts governing the peacetime use of military power in a very different era of international politics.

The challenge of carrying out this peacetime mission will be demanding because the foundations of U.S. military strategy are changing, as are U.S. capabilities. During the Cold War, the United States carried out this mission through the vehicle of large overseas deployments backed by sizable forces in CONUS. In the era ahead, overseas deployments will be far smaller: present plans call for only about 100,000 troops each in Europe and Asia, and only a modest presence in the Persian Gulf. This situation will necessitate greater reliance on power projection from CONUS, but the drawdowns underway will leave smaller forces for this purpose. Crafting an image of credible peacetime strength will be one of the most important requirements confronting future U.S. defense planning. Being prepared for canonical MRCs is only one part of the solution. Nonetheless, it is an important part. Let us now consider the related requirements.
DO THE CANONICAL MRCs ESTABLISH APPROPRIATE FORCE NEEDS?

Even if international affairs are turbulent and many conflicts are possible, it can be argued that focusing on the canonical MRCs in planning the defense posture is still justified. At issue is whether doing so would provide the requisite flexibility. Involved here are not only the enemies and theaters to be engaged, but also the kinds of U.S. responses to be mounted, and the size and mix of the combat posture needed to mount them.

Force Requirements for Persian Gulf and Korean Conflicts

A strength of the two canonical MRC scenarios for Southwest Asia and Korea is that they call for strong U.S. military forces, thereby barring any wholesale disarmament. They do so by postulating short-warning and roughly concurrent enemy attacks aimed at overrunning Kuwait/Saudi Arabia and South Korea. Sizable mobility forces—airlift and sealift—would be needed to deploy large U.S. forces to these theaters fast enough to save the day. Once these forces arrived, the situations in both theaters would mandate a stylized U.S. military response as envisioned by the Decisive Force doctrine (Powell, 1991). In the Persian Gulf, large combat operations would be conducted in order to stop the enemy attack and then launch a sweeping counterattack, akin to Desert Storm, to eject the enemy, destroy his forces, and attain political objectives. In Korea, U.S. and South Korean forces initially would have to mount a stiff defense to block the North Korean advance. They then would have to launch a counteroffensive to restore the border, destroy enemy formations, and achieve related goals. Following success in both theaters, most U.S. forces would withdraw, leaving behind postures as required for peacetime conditions (see also Frostic and Bowie, 1994).

The amount of U.S. forces required for these MRC scenarios is a matter of debate, but the Defense Department’s Bottom-Up Review (BUR) tabled an MRC “building block” of 4–5 active Army divisions, 4–5 Marine brigades, 10 USAF fighter wings, 100 USAF heavy bombers, 4–5 Navy carrier battle groups, and special operations forces (Aspin, 1993). This posture, DoD asserted, will be adequate to deal with either MRC. As a hedge, the BUR announced a program to increase the readiness of 15 Army Reserve Component (RC) brigades so that they will be deployable after only 90 days of training (although lift capabilities may be inadequate to exploit this readiness). The Pentagon thus implied that, if both MRCs are to be fought concurrently, about 10 Army divisions, 3 Marine divisions, 20 USAF fighter wings, and 10 carriers will be needed, backed up by the high-readiness Army RC brigades.
This total accounts for all BUR forces except 20 Army low-readiness RC
brigades and 1 Navy carrier. Thus, the BUR concluded that while reducing
the Bush administration’s Base Force by about 15 percent was acceptable,
进一步 cuts should be rejected as unsafe (Aspin, 1993:27–31).

Is this theory of requirements anything more than an ex post facto justifica-
tion for decisions taken on other grounds? Critics may levy this accusation,
but the DoD’s MRC “building block” approach serves as a good tool for ori-
entation. One problem, however, is that it does not resolve the debate, because
“requirements” for these MRCs are not reducible to single-point estimates. As
we shall see in the section on nonstandard contingencies, there is a big differ-
ence between the names of the MRCs (Persian Gulf and Korea) and the sce-
nario details, and those details greatly affect “requirements.” The size of the fu-
ture adversary force is now a variable, not a constant, as is the type of military
operation that U.S. forces will be called upon to conduct.

Are Persian Gulf Contingency Requirements Underestimated? Antici-
pating some of the discussion of nonstandard cases, note that the DoD
building block may actually underestimate force needs for the Persian Gulf
MRC. A recent RAND analysis (Bowie, Frostic, et al., 1993), entitled “The
New Calculus,” concluded that a DoD building block could defeat an Iraqi
thrust southward, but was careful to point out that many battlefield dynamics
would have to work in favor of U.S. forces. Rapidly deploying USAF units
would need modern munitions that will be available only in future years. Air
bases in Saudi Arabia already would have to stocked with enough fuel, mu-
nitions, and supplies to permit immediate full-scale air operations. The “new
calculus” assumed an Iraqi ground threat of 20 divisions, ten of which were
lightly equipped motorized infantry units, and readily suppressible Iraqi air de-
fenses. It further assumed that the Iraqi advance would proceed at less than
lightning speed, thereby allowing USAF units time to deploy, and that Iraq
would fail to suppress U.S. air bases. Under these conditions, USAF/USN air
operations were accurately assessed as enabling a successful counterattack by
only 4–6 Army/Marine divisions, provided these divisions could be deployed
fast enough to carry out the task. But do these assumptions reflect the future?

We shall consider alternative versions of the Iraqi scenario later, but even
the BUR noted that additional forces might have to be sent to compensate for
possible failures in the initial defense, to mount a decisive counteroffensive, or
to accomplish more ambitious war objectives. The report did not call for more
air or naval forces, but it did suggest that two additional Army divisions might
be needed, thereby raising the Persian Gulf ground force from 5–6 divisions to
7–9 divisions. This is an important caveat, for it elevates potential U.S. force
needs far closer to what was used in Desert Storm (17 coalition divisions,
roughly 10 of them American), and to what was deployed in Korea and
Vietnam: wars where airpower could not play a dominant role.
Are Korean Contingencies Overstated? If the BUR potentially underestimates Persian Gulf requirements, it may overestimate force requirements for the Korea MRC. North Korea today enjoys a roughly 1.5:1 quantity edge over U.S./South Korean forces in combat units and hardware. However, the U.S./South Korean defense posture benefits from prepared positions, rugged terrain, and knowable axes of advance. The chief risk is not that U.S./South Korean forces will be defeated, but that Seoul will be lost in the early fighting. Thus, in the likely event that South Korean forces acquit themselves well, a U.S. force of only 2–3 divisions, 8–10 fighter wings, and 2–3 carriers might be adequate (but additional forces could be needed if early events do not go well or for a counteroffensive). Thus, the BUR may understate needs for Iraq and inflate needs in Korea in ways that cancel each other out, thereby yielding an overall estimate that is on target. If so, marginal reductions in the Base Force, as discussed in the BUR, will not compromise a two-MRC strategy provided U.S. forces are well prepared, but further reductions could invalidate this strategy.

Force and Posture Needs for Other Conflicts

We have seen, then, that the total force requirements derived from the two canonical MRC scenarios appear roughly right, but with significant uncertainties. Another issue, however, is whether these two MRCs are an appropriate canonical basis for planning military strategy and posturing those forces we maintain.

On the positive side, the two MRCs compel the United States to maintain a sizable force structure, big mobility forces, and well-developed deployment plans. Further, they mandate diversity in that the Persian Gulf MRC calls for heavy armored/mechanized units, whereas the Korea MRC mandates lighter infantry, airmobile, and artillery formations. For both theaters, the required campaign plans demand joint and combined operations, and both defensive and offensive actions. Modern doctrine would have to be employed in both cases, anchored on a coordinated combination of firepower and maneuver carried out by forces that are armed with high-technology weapons, ready, well trained, well led, and fully supported.

Surface appearances suggest, then, that if U.S. forces can deal with these two MRCs, they should be capable of responding to a broad range of challenges, including very different situations. Yet this is not automatically the case, because the proverbial “lesser included case” sometimes turns out to be neither lesser nor included. History shows plenty of cases in which military forces that were well prepared for one type of conflict experienced reversals when war came wearing different clothes. If the U.S. experiences in Korea and Vietnam
do not illustrate this point, then the brutal lesson learned by France in May 1940 should do so. At the time, the French army was regarded as the world’s best, but it had spent twenty years preparing for a canonical scenario of its own: a repeat of World War I. When the German army crafted a nonstandard scenario through attack by blitzkrieg, the French army proved incapable of reacting, and was swept off the battlefield in one month.

One risk is that in preparing a choreographed response for the two canonical MRCs, U.S. forces might be hard-pressed to shift course if events in these two conflicts mandate a different response. Another, and possibly even greater, risk is that war might break out elsewhere: in an entirely different place, against a different enemy, and requiring very different U.S. deployment and campaign plans. If confronted by nonstandard challenges, could U.S. forces deploy fast enough and then carry out the operational campaigns needed for success? Perhaps so, but in the final analysis, the answer can be known only if nonstandard situations are studied.

The need for deep thought and a long view is manifest because uncertainty inhibits our ability to foresee which nations will appear on the scene as enemies. What can be said is that Iraq and North Korea are not destined to be our only military rivals. Politics can and often does change faster than U.S. military forces are altered. Today’s enemies can be tomorrow’s friends, but the converse also is true for nations whose internal politics or external interests can produce a sudden about-face. After all, Iran was once deemed a permanent friend, but almost overnight it became an implacable enemy when Ayatollah Khomeini arrived.

New adversaries can build imposing military forces that permit more ambitious operations than might be feasible today. Force improvement cannot occur overnight; the act of building modern forces that can compete with Western troops and weapons is costly and time-consuming. But especially if outside assistance is provided, buildups can take place, and perhaps faster than is commonly expected. After all, Germany in the 1930s went from being disarmed to becoming the world’s strongest military power in only six years. Much will depend upon the resources, skill, and determination of future adversaries. Nations formerly regarded as military lightweights can achieve at least middleweight status, and perhaps more, in the space of several years. In the interim, their efforts can be observed, but the act of discerning their intentions and ultimate ambitions often is not easily accomplished. This especially is the case for nations whose original agenda is unthreatening but becomes menacing only after military power is built.

Different physical circumstances thus might be encountered that do not offer favorable terrain, a well-developed military infrastructure, prepositioned assets, host nation support, and allied military contributions. Owing to their unique features, Saudi Arabia and South Korea are relatively easy to defend
once U.S. forces have been deployed. Other countries might be harder to protect, and even difficult to reach with sizable U.S. forces—doubly so if the timelines of war do not permit the six-month U.S. buildup that was possible in the Persian Gulf.

Future conflicts might be waged in response to political dynamics that are very different from those postulated by these two MRCs. Whereas these MRCs postulate aggressive enemy attacks aimed at conquering friendly nations, other conflicts might witness aggression aimed at different goals: e.g., seizure of nearby urban areas, destruction of lives and property, or imposition of a new government. One example is a civil war in Saudi Arabia aided by Iraq and Iran. The goal would not be the physical conquest of Saudi Arabia, but rather a takeover of its government. Coalition political dynamics also could be quite different. Whereas these two MRCs postulate the support of many friendly governments, other conflicts might witness neutrality or even opposition. Moreover, U.S. goals might be something other than the rapid destruction of invading enemy forces and restoration of allied borders. These dynamics could call for military operations quite dissimilar from those planned for the two canonical MRCs.

Future conflicts might be waged against adversaries that pose quite different military threats than mountable by Iraq and North Korea. Some adversaries might be less well armed, but others may (eventually) field larger and better-equipped forces. These forces might also be better trained and led, and guided by modern doctrines equivalent to that of Western forces. Whereas Iraq and North Korea today pose primarily ground threats, future adversaries might deploy strong air and naval forces that will have to be engaged. Indeed, some conflicts might be fought in the air and at sea, with little ground combat. Equally troublesome, some adversaries might arrive on the battlefield with nuclear forces and other weapons of mass destruction that could be employed against U.S. forces.

Another issue is whether these two MRCs adequately cover the problem of simultaneity. They presume that at the time of their occurrence, no other conflicts will be in progress. But what if a peacekeeping operation is underway at the same time that tensions rise in both the Persian Gulf and Korea? A peacekeeping operation could consume 1–2 U.S. divisions and therefore stretch thin the BUR’s posture. Or what if an even larger conflict takes place, one that demands more U.S. forces? At issue here is the degree of insurance to be sought from the U.S. military posture. Perhaps the likelihood of three simultaneous conflicts is too low to be taken seriously, or at least not high enough to merit the extra budgetary cost of buying additional forces. Yet the prospect of a peacekeeping mission is not easily dismissed, for during 1993, the United States found itself conducting one peacekeeping operation in Somalia.
and preparing for another one in Bosnia. To the extent that similar requirements arise in the future, the simultaneity problem might not be solved by deploying only enough forces to carry out two MRCs.

To pull things together here, consider that the great advantage of the Cold War for planning was that conflicts were relatively predictable. The focal points of potential aggression could be pinpointed, and entire areas could be discounted. Enemy threats were knowable, as were allied contributions. U.S. goals, strategy, and campaign plans could be decided upon in advance. The need to plan for global conflict ensured that three principal theaters—Europe, the Persian Gulf, and Northeast Asia—were covered. For these reasons, the Department of Defense enjoyed confidence that canonical scenarios left it prepared for other circumstances and sufficiently flexible to adjust to the unexpected. These advantages seem destined to be lost in the new era in which uncertainty abounds.

This important change magnifies the risks of relying on a small set of canonical scenarios. For example, the two MRCs in vogue today do not even consider Europe as a potential site of major regional war. What if this assumption proves invalid? What if major operations must be launched well east of NATO's borders in Europe, or in Central Asia, or anywhere in Asia aside from Korea, or anywhere in the Middle East/Southwest Asia apart from Saudi Arabia? Will the United States be able to respond flexibly if it plans only on the basis of these scenarios? The troublesome answers to these questions illuminate the case for paying careful attention to other scenarios that, while "nonstandard," might be altogether too real. Let us now identify some specific nonstandard scenarios.

NONSTANDARD SCENARIOS FOR THE FUTURE

Any attempt to speculate about specific nonstandard scenarios is a hazardous enterprise, one vulnerable to charges of implausibility and worry-warting. Nevertheless, this section offers a few possibilities. The following analysis is intended to be illustrative. What it offers is an opportunity to break out of current plans by imagining different conflicts that, in one way or another, might be feasible—if not now, then within, say, a decade. The analysis speculates only about future military conflict in the critical theaters of Europe, the Middle East and Persian Gulf, and Asia. It ignores entire regions that could become focal points of conflict: South and Central Asia, sub-Saharan Africa, and Latin America. But coverage of the selected theaters is sufficient to illuminate the central point: the need for broad intellectual horizons, and for flexibility and adaptiveness, in U.S. defense planning.
Conventional Conflicts

Nonstandard Conflicts in Europe. Because Europe does not even play in current MRC planning, this theater is a good place to start. Absent war with Russia, small powers in east central Europe and the Balkans are unlikely to band together to create the large enemy force needed for any single conflict to qualify as an MRC. Yet if regional strains intensify, conflicts far larger than LRCs are possible, for the nine nations there will deploy a total of 60–70 heavy combat divisions and 2500 combat aircraft. Apart perhaps from Serbia, no other nation there qualifies today as an adversary of the United States, but many harbor profound animosity for each other. The possibilities for confrontation are multiple, and if U.S./NATO forces are drawn in, the result could be a new form of warfare: something between an MRC and an LRC.

Although peacekeeping in east central Europe or the Balkans is the mission most likely to be performed by Western forces, a situation might arise in this region in which major peacemaking/enforcement operations, or even large combat interventions, must be mounted. In the case of concurrent conflicts, overall requirements could be for as many as 10–12 divisions and 650–800 combat aircraft. The NATO allies could contribute heavily, but even so, U.S. contributions might be as high as 2–4 divisions, 150–300 combat aircraft, and naval forces. This is less than an MRC requirement, but sizable nonetheless.

The complex politics of the situations to be encountered create many different possibilities for the employment of U.S. forces in a NATO operation. U.S. forces might be used to keep the peace in Bosnia, to pressure Serbia, to protect Romania’s borders, to quell imperial conduct by Hungary, or to defend Poland against Ukraine. U.S. forces might help bring stability to the Baltics, the Caucasus, or south central Asia. All of these situations could require military operations very different from the neat and clean script followed in Desert Storm. Indeed, politics and diplomacy likely would dominate military strategy. Even if forces were initially committed on behalf of clear policy goals, these goals might change in response to new conditions, thereby causing military strategy to shift, perhaps several times over. If so, the prospect would be for a very messy relationship between politics and war.

The coming years might witness NATO’s expansion to include new members in east central Europe and elsewhere. The Partnership for Peace proposal will offer Article 4 guarantees to participating nations, and it holds out the possibility of NATO membership for some. Obvious candidates are Poland, the Czech Republic, and Hungary. This development would extend combined NATO planning eastward. Along with West European forces, U.S. forces would have to acquire the capacity to deploy to these new nations and to defend their borders. The outcome could be a new era of NATO military strat-
egy. If all of this were accomplished well, it would be seen as a step toward rather than against Russia (Asmus, Kugler, and Larrabee, 1993:37). In time, assuming further democratization and development, Russia itself might join NATO.

If democracy failed in Russia and that nation returned to imperial conduct, the prospect of war with it could again have to be factored into U.S. defense planning. This conflict would be an MRC and beyond, for Russia would probably field an army of 50 divisions and equivalent air forces, at least half of which could be committed. The size of the adversary force could increase further if other Commonwealth nations joined the fray. The need to deal with an imposing enemy force would be far from the only troublesome issue confronting U.S. defense planning, for weighty political issues would enter the calculus. For example, where would the war be fought, in Poland, or Ukraine, or the Baltics? What political goals would be pursued? What would be the overall diplomatic context? The answers would have profound implications for force planning, and multiple answers could drive planning in many different directions.

Indeed, war in Poland alone could be fought in many different ways. Western forces might be committed early, or in the middle of an impending crisis, or late, after fighting had already begun. The politics of this intervention could be clear or very muddy, marked by uncertainty about the goals and calculations of many different participants. NATO might respond as a unified alliance through its integrated command, but alternatively, an ad hoc operation, with only a few nations participating, might have to be launched outside the integrated command. If NATO itself had fallen apart, the intervention might have to be mounted on the fly, with only Germany providing an infrastructure and forces. Bases and reception facilities might be available in Poland, but they might instead have been overrun by Russian forces (Kugler, 1992a).

Western forces might be called upon to defend Warsaw and the Bug River, to halt a Russian drive midway through Poland, to launch a counterattack from western Poland aimed at restoring that country’s borders, or even to march into Belarus. The operation might be launched with air forces alone, with large air forces and some ground units, or with large ground formations. The ground campaign might take the form of a linear defense, a defensive maneuver battle, a flanking counterattack, a sweeping counteroffensive, or all of these in sequence. The war might be over quickly, or drag on for weeks and months. The possibilities here are endless, and many are quite different from Desert Storm.

**Nonstandard Contingencies Involving Iraq.** If regional tensions intensified in the Persian Gulf, an equally wide range of possibilities might have to be addressed. For the near future, small-scale operations will remain the order of
the day, e.g., enforcement of the no-fly zone over Iraq coupled with antiproliferation and humanitarian missions. In the midterm and more distant future, a wide variety of MRCs are possible, some like Desert Storm, some others quite different.

An important issue here is whether Iraq would attack with a force as ill-prepared and a political approach as ham-handed as in 1990. As discussed in Bennett, Gardiner, and Fox (1994), there are many possible ways in which Saddam Hussein might hope to improve his prospects for success. For example, a repeat Iraqi invasion might take the form of naked aggression, but it might be mounted amidst complex political conditions such as a domestic upheaval in Saudi Arabia in which a revolutionary movement seized power and called for Iraqi help. An Iraqi invasion might again stop at Saudi Arabia’s borders, but equally likely it could press beyond and aim at overrunning that nation. An Iraqi drive into Saudi Arabia might move slowly, but might unfold at high speed. The United States might respond quickly, but in contrast to Desert Shield, it might react sluggishly.

Sensitivity analyses of nonstandard versions of the Iraqi scenario need to assess how requirements would be affected by a stronger and faster-moving Iraqi attack, or if U.S. air operations were much less effective than postulated by the DoD “building block.” At issue also is the adequacy of a ground posture that, even counting allied forces, would still leave U.S./allied forces outnumbered by 2 or 3 to 1 and fighting on open terrain that invites mobile operations by both sides. Despite the superiority of U.S. weapon systems, some defense planners might assess this deficiency as too great for a confident defense against a capable opponent if U.S. airpower is anything less than truly dominant. To be sure, U.S. airpower could inflict enough damage on an exposed Iraqi army to lower U.S. ground requirements if given 2–3 weeks to operate before the ground battle begins. But what would happen if the next war does not permit a lengthy preparatory air campaign?

The nature of the U.S./allied ground defense campaign also could have a bearing on the adequacy of the planned force. In the unlikely event that a forward linear defense is mounted, a U.S. posture of only 5–6 divisions (along with 3 allied divisions) would be affected by force-to-space relationships and thereby could be hard-pressed to form an adequate line to contain an Iraqi thrust. Much would depend upon the Iraqi army’s ability to move off existing roads and advance across the open terrain. In the more likely event of a mobile defense, force requirements could be elevated by the need to perform pinning maneuvers, frontal assaults, and flanking operations. Mobile defense is far from a cure-all if the enemy is skilled at maneuver (Kugler, 1992b), and the need for adequate forces is all the greater when counteroffensives are conducted. These reasons were influential in shaping requirements for Desert Storm (17 coalition divisions, 10 of them American).
For all these reasons, the range of military possibilities in defending Saudi Arabia alone is quite wide. U.S. forces might again be granted time to build up imposing defenses on Kuwait's border, but they also might have to land in Saudi Arabia amidst an outrush of attacking Iraqi formations. At the outer extreme, U.S. forces might have to launch an invasion of a Saudi Arabia already fully occupied by Iraq. U.S. offensive operations might stop once Saudi and Kuwaiti borders have been restored, but this time they might be followed by an invasion of Iraq aimed at toppling its government. These cases would require substantially different forces and concepts of operations (Davis and Finch, 1993:74-75).

Other Nonstandard Scenarios in the Greater Middle East. Iraq might not be the only adversary nation encountered in the Persian Gulf. Indeed, Iran is now improving its forces under a still-zealous regime, and might transform itself into a well-armed enemy intent on imperial conduct. How would Iran react to another U.S. intervention aimed at inflicting military defeat on Iraq? Would it remain neutral, or might it try to foul the intervention, perhaps by using its air and naval forces to block the Straits of Hormuz and the Persian Gulf? If Iran did insert itself in these ways, U.S. force operations would need to change dramatically away from the Desert Shield/Storm model. Indeed, how would a war with Iran alone unfold? In all likelihood, it would be an air and sea war, with U.S. operations mounted heavily from Saudi Arabia and other Gulf sheikdoms. But it might include a forced U.S. landing on Iranian soil followed by a major campaign into that nation.

Future conflicts in the greater Middle East are also not limited to the Persian Gulf. For example, Israel might be attacked in ways calling for larger U.S. interventions than in the past. Although force commitments probably would not be large for defending Israel, they could be larger if Turkey were to be attacked by a coalition of radical Arab partners: e.g., Syria and Iraq. In this case, large U.S. air and naval forces, with major logistical support, would be needed by Turkey. If the Turkish army proved unable to stop the advance and restore lost territory, sizable U.S. and NATO ground forces might have to be committed. In this event, a military operation akin to Desert Storm could be mounted, but the geographical and logistical conditions would be quite different.

What will happen if Islamic fundamentalism sweeps over the Middle East and North Africa, in ways producing a united coalition of radical Arab governments all angry at the West? The idea that another Muslim invasion of Europe could be launched—akin to that faced by Charles Martel at Tours—is far-fetched. Yet, jihad can be conducted in other ways. In addition to spawning terrorism across Europe, Arab nations might assemble the air, missile, and naval forces needed to contest the Western powers for control of the Mediterranean, the Suez Canal, and the Red Sea. The result could be a long-
running saga of air and naval clashes that would entangle not only West European forces, but also U.S. forces. This conflict would not be an MRC, but to U.S. forces it might seem that way, and current MRC plans would provide few solutions.

Nonstandard Contingencies in Asia. Similar judgments apply to future security affairs and conventional military conflict in Asia. In Korea, a canonical MRC could unfold in many different ways, each having different implications for U.S. forces and defense plans. In the optimistic case, a powerful North Korean attack might run into a stone wall when encountering the South Korean army, its prepared defense positions, and the mountainous terrain straddling the border. In this event, North Korean forces might make little headway while suffering high losses, and their drive might stall after making little, if any, progress. With the South Korean army providing the primary vehicle for containing this unsuccessful attack, only limited U.S. military assistance would be needed: C^3I, tactical air, naval, and logistic support.

A wide spectrum of pessimistic Korea scenarios are equally plausible, however (Bennett, 1993). North Korean forces might threaten to punch through South Korean defenses and seize Seoul. Alternatively, they might actually capture Seoul and surrounding territory. A worse outcome is that the South Korean army might be left too battered to counterattack and recapture Seoul. Even worse, the South Korean army might unravel, thereby leaving all of the Korean peninsula open to North Korean aggression, as happened in 1950.

Each of these pessimistic scenarios imparts a different meaning to U.S. force requirements and defense plans. In the least threatening case, not only USAF forces, but also some U.S. ground units would have to be used to help contain breakthroughts and shore up South Korean defenses near the border. If Seoul were to be lost, larger U.S. ground and air forces would have to be committed to recapture the capital city, restore South Korean borders, and destroy enemy forces—all the more so if South Korean forces themselves were left too depleted to lead the counterattack. In the extreme case of a complete South Korean collapse, quite large U.S. forces might be compelled to launch a large amphibious attack as a forerunner to a sustained drive up the peninsula.

If these possibilities help illuminate the situations that might be encountered in Korea today, future planning might have to undergo a fundamental reorientation if Korea reunifies. Perhaps U.S. forces would withdraw entirely, and the need to plan for any war in Korea would disappear. Alternatively, the U.S.—South Korean military alliance might be redefined to deal with new security challenges after unification. What would these challenges be? The U.S. perspective might focus on the threat to Korea’s northern borders posed by an assault from China or Russia. By contrast, the Korean perspective might call for defense preparations against a potential threat from Japan—a key U.S. ally
in Asia. Alternatively, a reformed U.S.-South Korean alliance might play a regional role in Asia.

Future U.S. defense planning also will need to take into account the many other political-military changes now sweeping over Asia. For the past 20 years, the threat of war with China has not been taken seriously in U.S. defense planning. But China is now undergoing a military buildup, and if that nation were to embark on an imperial course in an atmosphere of mounting political confrontation, an entirely different situation could unfold. Exactly how would a renewed Chinese military threat be manifested? Although traditional examples are threats to Korea, Taiwan, and Southeast Asia, the new era might prove to be untraditional. China might pose a mounting nuclear missile threat to Japan; indeed, if relations with Russia were also to sour, Japan might find itself besieged by new threats from both countries. Another possibility is that China might develop the larger and better-equipped navy that would allow for maritime power projection into the western Pacific. In this event, naval combat might prove to be a core feature of new U.S. defense planning in Asia.

The possibilities would magnify many times over if Japan were to follow the course of expanding its maritime power projection capabilities. Especially in this event, major-power naval rivalry might spread outward from Northeast Asia by expanding into Southeast Asia, the Strait of Malacca, the Spratly Islands, and eventually linking to the turbulent situation in South Asia. Along with distantly deployed naval forces could come networks of new military bases across the region for projecting air and ground power. Inevitably, the Southeast Asian nations would be affected, and new security alliances would form, perhaps in ways destabilizing to the entire region. None of these outcomes are foreordained or even probable, but with Asia changing so rapidly in such profound ways, they are not beyond the realm of the possible.

The new challenges posed to U.S. military forces would necessitate major changes in the entire American defense planning framework for Asia. All the more so if political change were also to bring about a major withering of the security alliances that have provided U.S. forces with invaluable military installations for so many years. Already, U.S. bases in the Philippines have been lost, compelling a difficult search for other alternatives. What would happen if the U.S.-Japanese security alliance were terminated, and if the U.S.-South Korean alliance suffered the same fate? Loss of military bases in these countries could compel a wholesale change in U.S. military strategy for Asia. Provided disengagement was rejected, the United States presumably would be compelled to rebuild its old bases on islands in the Pacific. Even so, it would be left with a military strategy of distant power projection and expeditionary operations.
Regional Nuclear Scenarios

Because these global scenarios deal only with conventional war, the looming prospect of regional nuclear crises adds yet another dimension to the upheavals potentially ahead for U.S. planning. How would the United States act in regional nuclear crises in the Persian Gulf and Korea? Although the answer is not obvious, current plans might have to be radically altered, for any mechanical deployment of large U.S. forces might serve only to supply a target-rich environment to enemy forces. For both MRCs, new plans would have to be crafted, plans capable of dealing with conventional and nuclear threats (Millot, Molander, and Wilson, 1993).

Beyond these two regions, nuclear crises might appear elsewhere, and the possibilities are mind-numbing. The United States might find itself facing a nuclear confrontation between Russia and Ukraine that would threaten all of Europe. Another possibility (less detailed than a "scenario") is a tactical nuclear standoff between nations in east central Europe or the Balkans that had managed to get their hands on these weapons. A third is a nuclear crisis in the Middle East or North Africa. A fourth is a crisis in South Asia pitting India against Pakistan, or one that spreads to Central Asia in ways entangling Russia and China. A fifth encompasses many subcases in Asia that do not originate in Korea. A sixth is even bleaker, for it would fundamentally alter the global security system: Japan and Germany, pressured by mounting insecurity and diminished confidence in U.S. deterrent coverage, might themselves decide to develop nuclear forces.

Even if the threat of destabilizing actions by Japan, Germany, and other responsible nations is discounted, the need to prepare for regional nuclear crises caused by rogue states is growing. To the extent that U.S. combat operations might be undertaken, the prospect of regional nuclear crises will resurrect concerns of the past. During the Cold War, defense planning in Europe viewed conventional and tactical nuclear operations as interconnected. This interconnection would have to be recaptured in new plans, since the old concepts of flexible response, graduated escalation, and massive retaliation no longer apply.

Summary of Contingencies

Table 1 displays the various contingencies discussed here. It is by no means comprehensive, but it illustrates a spectrum of events.
Table 1

Potential Future Contingencies Involving U.S. Military Forces

<table>
<thead>
<tr>
<th>Conventional Contingencies in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peacekeeping or peace enforcement and crisis management in east central Europe, the Balkans, and the former USSR</td>
</tr>
<tr>
<td>Medium-sized warfare in east central Europe and the Balkans</td>
</tr>
<tr>
<td>Defense of new NATO members, and cooperation with other partners</td>
</tr>
<tr>
<td>MRC versus Russia in Poland</td>
</tr>
<tr>
<td>- At Bug, Vistula, or western Poland</td>
</tr>
<tr>
<td>- Linear defense, mobile defense, or counteroffensive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conventional Contingencies in Persian Gulf and Middle East</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRC in Saudi Arabia and Kuwait</td>
</tr>
<tr>
<td>- Forward defense</td>
</tr>
<tr>
<td>- Intervention after Saudi Arabia is partly overrun</td>
</tr>
<tr>
<td>- Reinvansion after Saudi Arabia is conquered</td>
</tr>
<tr>
<td>Civil conflicts exploited by outside powers</td>
</tr>
<tr>
<td>War with Iran</td>
</tr>
<tr>
<td>Defense of Israel</td>
</tr>
<tr>
<td>Defense of Turkey</td>
</tr>
<tr>
<td>Conflict in North Africa and Mediterranean</td>
</tr>
<tr>
<td>Conflict in the Caucasus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conventional Contingencies in Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRC in Korea</td>
</tr>
<tr>
<td>- Defense of border</td>
</tr>
<tr>
<td>- Recapture of Seoul</td>
</tr>
<tr>
<td>- Reinvansion of conquered South Korea</td>
</tr>
<tr>
<td>Defense of unified Korea</td>
</tr>
<tr>
<td>Defense of Japan</td>
</tr>
<tr>
<td>Defense of Taiwan</td>
</tr>
<tr>
<td>Maritime conflict with China</td>
</tr>
<tr>
<td>Maritime conflict in Southeast Asia</td>
</tr>
<tr>
<td>Intervention in South Asia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regional Nuclear Crises</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Europe</td>
</tr>
<tr>
<td>In Asia</td>
</tr>
<tr>
<td>In Middle East/Persian Gulf</td>
</tr>
</tbody>
</table>

**IMPLICATIONS FOR U.S. DEFENSE PLANNING**

**Changing Needs, Changing Requirements, and Flexibility**

The need for national military preparedness stems from reasons far more fundamental than the transient mechanics of contingency analysis. Because the United States will remain a superpower with overseas interests to protect, it
will need strong military forces to underwrite its purposes in peace, crisis, and war. Specific adversaries will come and go, but the requirement for military power will remain—a product of enduring geostrategic realities. What is required now is a true strategic concept, one portraying the rich set of premises and postulates that are the appropriate foundation for the U.S. defense posture and budget. During the Cold War, the United States was blessed with a coherent strategic concept; today it needs one for the post–Cold War era. Once this strategic concept is fashioned, canonical scenarios can be retired to the more limited roles for which they are suited.

If one thing is obvious, it is that defense planning will need to be quite dynamic. The days are gone when the Department of Defense could erect an elaborate analytical framework that would endure untouched for many years. Because static thinking will no longer be feasible, basic military strategy may have to be uprooted every few years, and specific plans might have to be altered more often yet. Therefore, the emphasis should not be on reestablishing a timeless edifice of plans for the new era, but rather on creating an energetic planning process that is capable of handling regular upheaval.

To the extent that scenario analysis continues to be employed, the emerging military environment is rendering invalid many of the nostrums that animated defense planning during the Cold War. The challenge will be to construct a new planning framework that casts out old invalid thinking but avoids misinterpreting the new era. Desert Storm overturned many of the analytical approaches that derived from the old NATO–Warsaw Pact confrontation in Central Europe (Davis, 1994c). These include the practices of valuing quantity over quality, of discounting differences in readiness and modernization when there are large disparities in mass, of assessing ground requirements in terms of a fixed ratio in comparison to enemy forces, and of striving for a predetermined mix of ground, air, and naval forces. Yet if these old practices are no longer appropriate, Desert Storm itself was a unique war in many ways, owing to the many advantages enjoyed by U.S. forces (Record, 1993). The challenge will be to assess future conflicts on their own merits, and to anticipate a broad range of possibilities (Davis and Finch, 1993; Davis, 1994a).

The need to prepare for the twists and turns of events has special implications for DoD’s operations plans (OPLANS), which determine how forces are deployed, and for CINC campaign plans, which determine how forces are to be employed on the battlefield. In the past, these plans often offered only one fixed blueprint for operations. In the future, as the Joint Staff and CINCs are aware, these plans need to be modular, and based on building blocks capable of being rapidly adjusted to handle different situations. Modularity can be achieved only if scenario analysis examines a broad spectrum of events and determines the alternative responses that might have to be mounted. Analysis of individual contingencies should employ, e.g., decision tree analysis—composed
of multiple branch points—to determine the alternative paths that military operations might have to take in any single conflict.

Planning for Generic Missions

Future uncertainty is so great that U.S. defense policy might be well advised to shift away from threat-based planning to mission-based planning. For all its quantitative appeal, threat-based planning is an unreliable instrument for dealing with a change-filled era. If the Israeli-Arab dispute is settled and Korea reunifies, today's canonical MRCs could disappear overnight. But this development would not remove the need to be prepared for major regional wars, which might occur in different places for different reasons. What is needed is a planning mechanism that enables the United States to be prepared for these and other conflicts, wherever they might occur.

Under the new approach contemplated here, defense plans would not focus on specific threats posed by specific nations as the final arbiter of decisions. Rather, it would identify the generic military missions that will need to be performed in order to deal with many different threats. The U.S. force posture and defense program would then be designed to perform these missions. In this approach, canonical scenarios would still be employed, but not in an exclusive sense. Rather, planning would consider several categories of scenarios for determining the multiple military missions that might have to be performed in many different situations (Kugler, 1992a).

Table 2 illustrates eleven different categories of contingencies that should be taken into account in force planning. Because all these categories may be encountered in three or more theaters, and in several different forms, the number of permutations is quite large. Mathematics alone suggests that a narrow planning framework of only two canonical MRCs might fail to address all of these permutations; even inclusion of a few LRCs might not solve the problem. Common sense says likewise.

Each of these classes of contingency imposes unique requirements for missions to be performed by U.S. forces. "Peacetime stability/presence" refers to the role that U.S. forces, especially those deployed overseas, will play in reassuring friends, dissuading adversaries, and otherwise guiding the international system toward stability. "Humanitarian assistance" refers to the delivery of food and other supplies needed in regions suffering devastation from war or natural causes. "Peacekeeping" refers to the mission of deploying forces in noncombat missions to help maintain an existing agreement. The missions required for "counterterrorism and hostage rescue" are obvious from the titles, and the missions imposed by "lesser regional contingencies" have been discussed above. "Crisis management and resolution" refers to the use of forces to exert politi-
Table 2

Types of Future Contingencies Requiring Commitment of U.S. Forces

1. Peacetime stability/presence
2. Humanitarian assistance and peacekeeping
3. Counterrorism and hostage rescue
4. Lesser regional contingencies (LRCs)
5. Crisis management and resolution
6. Peacemaking and peace enforcement
7. Medium-sized regional contingencies
8. Canonical MRCs
9. Nonstandard MRCs
10. Greater-than-expected MRCs
11. Regional nuclear conflict

cal-military pressure on adversaries for the purpose of achieving U.S. goals in a confrontational situation short of full-scale war. “Peacemaking and peace enforcement” refer to the use of force in combat missions to either transform an existing conflict into peace, or to ensure that an existing accord continues to be carried out.

Whereas these first six categories deal with the use of military forces in the gray area between war and peace, the final five categories deal with military missions in wartime settings. “Medium-sized regional contingencies” refers to conflicts that fall between LRCs and MRCs. Whereas the category of “canonical MRCs” refers to the Persian Gulf and Korean scenarios now employed in DoD planning, “nonstandard MRCs” refers to either major variants of those scenarios or entirely different conflicts of similar magnitude. “Greater-than-expected MRCs” refers to potential conflicts in which stronger enemy forces are employed than projected in typical MRC planning. At the outer extreme are potential “regional nuclear conflicts” that escalate beyond conventional fighting.

In the coming years, U.S. forces will be required to perform many of these missions. The canonical MRCs offer the promise of enabling the U.S. posture to wage a repeat of the Persian Gulf war while deterring aggression in Korea. But will this posture be able to perform the peacetime mission of undergirding international stability? Will its forces be able to conduct the small-scale but important missions of humanitarian assistance, peacekeeping, counterrorism, and hostage rescue? Will it be capable of performing in lesser regional contingencies? How about the thorny operations that might have to be conducted under the rubric of crisis management and resolution, peacemaking, and peace enforcement? What about a medium-sized wartime contingency in
east central Europe or the Balkans, far away from existing bases and infrastructure? What about a truly nonstandard MRC: a major naval tussle with China in Asia? What about a greater-than-expected MRC that involves a major war with the Russian army fought in Poland, the Baltics, or Ukraine? And what about a regional nuclear conflict in any of three different theaters?

If sufficient forces are available to deal with the two canonical MRCs, in all likelihood the U.S. posture will provide enough combat forces—in overall size—to address any single nonstandard conflict in the other categories. The real problem is that for many of these nonstandard contingencies, specific capabilities might be lacking because they are not mandated by the canonical MRCs. For example, peacekeeping can require specialized training and equipment that is quite different from that needed for major warfighting. Crisis management/resolution, peacemaking, and peace enforcement normally require combat forces similar to those needed for warfighting, but they normally mandate employment doctrines very different from Decisive Force. A medium-sized regional contingency in east central Europe or the Balkans might require logistic support forces larger than, and quite different from, the support forces needed for the Persian Gulf and Korea. A regional nuclear contingency might require specialized forms of C²I and air-strike operations quite different from that needed for a canonical MRC. These are but a few examples of how unique capabilities might be needed for nonstandard contingencies, but they illustrate the basic point.

Perhaps force sizing will continue being based on the two canonical MRCs, but the act of developing military plans and future-looking programs should be guided by a much broader framework. In essence, planning should begin by taking a careful look at all eleven categories. For each category, potential contingencies in all three theaters should be considered. In each case, the robustness of solutions should be tested by examining significant deviations from expected norms. The resulting analyses should be used to gauge the adequacy of force plans and programs. The goal should be a robust and flexible military posture that can perform many different kinds of operations, aided by deployment/employment plans and military doctrines that are sufficiently adaptive to respond to the situations at hand.

Analyses of this sort admittedly would complicate the defense planning process. The goal of planning, however, is comprehensiveness, not simplicity. The drawback of a few canonical scenarios is that they buy detailed appraisals of a few major events at the expense of roughly accurate assessments of a far larger class of conflicts that might actually occur. They thus help give the Department of Defense penetrating vision for expected conflicts, but may leave it largely blind to the surprises created by the unexpected. The outcome of considering nonstandard scenarios would be greater confidence that the
Defense Department will not be flying blind in a coming international era seemingly destined to create a fog of confusion of its own.

BIBLIOGRAPHY


Bennett, Bruce (1993), *Global 92 Analysis of Prospective Korean Conflicts in 2002*, RAND, Santa Monica, CA.


Record, Jeffrey (1993), *Hollow Victory: A Contrary View of the Gulf War*, Brassey’s (U.S.), Washington, D.C.


Chapter 8

IMPROVING DETERRENCE IN THE POST—COLD WAR ERA: SOME THEORY AND IMPLICATIONS FOR DEFENSE PLANNING

Paul K. Davis

This paper describes an interdisciplinary theory for conventional deterrence of regional aggressors in post—Cold War crises. A basic assumption is that the opponent is attempting to make rational decisions by comparing options on the basis of their prospective payoffs (as viewed in the opponent’s personalized and context-sensitive scheme of values) in what he judges to be the most likely, most favorable, and worst cases. Despite attempting to be “rational,” the opponent is quite likely to make errors of judgment and assessment because of uncertainties, misperceptions, and a variety of standard psychological phenomena such as ignoring altogether risks below some threshold of perceived likelihood. At the same time, the nation attempting to deter aggression is also likely to make serious mistakes, which include underestimating how unacceptable the aggression would be once it happened and being unwilling to take effective deterrent actions early because of the political costs of doing so and rationalizations about why the aggression won’t occur. The theory has been applied to analyze why deterrence of Saddam Hussein failed and whether it would have been possible to deter him with different actions. It has also been applied to some much earlier historical cases. After discussing those applications, the paper examines the generic challenge of avoiding future aggression against weak states such as Kuwait, Saudi Arabia, Poland, Lithuania, and Ukraine. It ends by noting significant implications for defense analysis, intelligence assessments, and the process of crisis deliberation.

INTRODUCTION

Conventional deterrence is difficult. Extended conventional deterrence, where one nation attempts to influence the actions of a second nation concerning a third, is even more difficult. Further, it is poorly understood, perhaps because during the Cold War so much theoretical attention in the West was focused on the very special case of deterring invasion of NATO by the Warsaw

This paper draws on a chapter in Huber and Avenhaus (1993). It also benefited from reviews by Kenneth Watnman and Paul Bracken and comments by John Arquilla.
Pact. That problem was special because: (a) the requirement to defend was widely accepted; (b) the threat appeared to be real, massive, and monolithic; (c) large military forces were stationed close to each other and maintained at high levels of readiness; (d) nuclear weapons played a central role; and (e) the opposing sides knew a great deal about each other's strategic thinking and military culture. The situation is quite different when we contemplate deterring the kinds of aggression that seem most likely in the next decade or so. Iraq's invasion of Kuwait was a massive failure of deterrence. It can be argued that the United States did not try seriously to deter the invasion, and that Saddam's invasion was therefore less a failure of deterrence than a U.S. failure to recognize and appreciate national interests in advance. However, the evidence is that the United States did try, ineffectually, to deter invasion, as evidenced by the President's message to Saddam, the ineffective warning provided by Ambassador Glassie, and an ill-conceived naval exercise to indicate U.S. concerns. The ongoing war in the former Yugoslavia and its despicable "ethnic cleansing" activities are more recent examples of failure and a setback for those hoping for something approaching a "new world order." It is not difficult to identify many other potential trouble spots for the years ahead. It follows, then, that we should be interested in a theory of deterrence so that the future can be better than the recent past.

ASSUMPTIONS

There are many ways to approach the issue of deterrence. These include: statistical analysis of political, military, and economic factors present in historical crises (Huth and Russett, 1990); attempting to generalize from anecdotal accounts in the biographies and autobiographies of participants in crisis (Khrushchev, 1990; Garthoff, 1989); applications of game theory (Schelling, 1980; Dixit and Nalebuff, 1991); military stability analysis (Huber, 1990; Huber and Avenhaus, 1993); applications of classic balance-of-power theory (Walt, 1987); observations from history (Payne, 1992); and other theories (e.g., George and Smoke, 1974; George, 1991; Watman and Wilkening, forthcoming). All of these approaches have much to offer, but the one I have been pursuing focuses on an aspect of the problem that has been largely neglected: understanding and attempting to influence the reasoning of the potential aggressor (who may not think of himself as an aggressor). Thus, I am concerned with human perceptions, arguments, and logic—all of them affected by psychological considerations. However, I am concerned less with making

1See Arquilla (1992) for discussion of diverse approaches to understanding aggression and deterrence. It includes a good bibliography of both classic and current writings.
mere observations about psychological factors in crisis than with describing such reasoning, *analytically*—i.e., with building models of reasoning that can be used not only to improve insights retrospectively, but also to guide strategy *prospectively.*

In attempting to describe reasoning analytically, one could structure the problem in any of several ways. My approach assumes *limited rationality* and *universal classes of reasoning patterns.*

By “limited rationality” I mean that the relevant leaders: (a) attempt to relate means to ends (i.e., their decisions and actions have purpose); (b) consider a range of options; and (c) evaluate those options in terms of likely outcome, most favorable outcome, and worst-case outcome. Thus, the leaders *attempt* to be rational and even take uncertainty into account. However, their decisions may be flawed for a wide range of reasons that include incomplete or incorrect information, the mental frames through which information is viewed, anxieties, extreme dissatisfaction with the status quo, oversimplified and sometimes highly erroneous mental models of the other protagonists, and other factors. Perceptions may even shift wildly during a fast-moving crisis. Further, leaders have very different attitudes about risk taking.

It is controversial to assume rationality, but while it is surely possible for a national leader to be a lunatic making random decisions, there are few examples of that in history, despite references to “crazy states” in some of the literature. Most of the national leaders who have sometimes been described as irrational (e.g., Adolph Hitler, Joseph Stalin, the Ayatollah Khomenei, and Saddam Hussein) were quite rational in the sense defined above. Some of them suffered from severe psychological problems and exhibited bizarre and abhorrent behavior, but their most strategically significant decisions can be understood in terms of their objectives and perceptions. It is also important to recognize that all of us are subject to making a wide variety of perceptual and reasoning errors, but we do not consider ourselves irrational. “Limited rationality” allows for a wide variety of such cognitive “errors.”

---

2For a more detailed discussion, see work with my colleague John Arquilla in Davis and Arquilla (1991a,b) and Arquilla and Davis (1992). These studies were sponsored by the Joint Staff. They contain citations to relevant literature in psychology, political science, international relations, and economics, including citations to work on the origins of flawed reasoning in crisis by Robert Jervis, Robert Axelrod, Alexander George, Daniel Kahnemann, Amos Tversky, and others. Shafer and Pearl (1990) is a good compilation of classic articles on uncertain reasoning.

3See, e.g., Jablonsky (1991). Despite framing his work in terms of “crazy states,” Jablonsky also recognizes (e.g., Chapter 2) that there can be much that is rational about crazy-state behavior, if merely we understand its perspective.

4There are many examples associated with terms such as framing, anchoring, attributional inference, and group think (Davis and Arquilla, 1991b). Humans sometimes
The second assumption is that it is useful to structure the theory around universally observable types of reasoning rather than culture-specific concepts such as the so-called "Arab, Oriental, Latin, or Western minds." To be sure, cultural factors can have profound effects that must be reflected in any application of theory, but the approach I describe has such factors entering along the way in context-dependent ways rather than as part of basic structure. My principal reason for the assumption is that the relevant behaviors of historical leaders can be found in all cultures, albeit with different frequencies. For example, the Arab world has produced Anwar Sadat and Saddam Hussein, and the Western world has produced George Bush and Adolph Hitler.

MODELING OPPONENTS AND THEIR ASSESSMENT OF OPTIONS

Assessment of Options

Let me next describe a way that many aggressors may, in effect, have reasoned about their options in the past and how many others may do so in the future. To repeat, they are attempting to make rational decisions. They are considering options and are also examining likely and possible consequences of those options, as suggested in Table 1. The format here is that for each option the reasoner estimates the likely outcome, most favorable outcome, and worst-case outcome. He then makes an overall assessment of the option based on these estimates. Each outcome is characterized by a value in the set {very bad, bad, marginal, good, very good}.5

---

5To be sure, humans seldom reason in so linear and reductionist a manner in the course of their decisionmaking process. Ideas arise and are considered; competing ideas arise; some ideas are forgotten, others are championed; organizational and group-dynamic considerations dictate how options are presented and when; procedural methods may or may not exist to enhance "rationality"; and so on. However, the assumption
Table 1

Generic Decision-Table Format for Assessing Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Likely Outcome</th>
<th>Most Favorable Outcome</th>
<th>Worst-Case Outcome</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Option 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This basic structure is generic, but estimates of the various outcomes depend sensitively on perceptions and values. To understand how a potential opponent might reach individual judgments about, e.g., the worst-case outcome (would it be very bad, bad, marginal, good, or very good?), we need:

- Alternative mental images of the opponent,
- An understanding of what factors are most likely to affect the opponent’s reasoning, and
- A way to go systematically from the image and factors to estimates of the opponents’ various judgments.6

Alternative Images of the Opponent

Developing alternative images is crucial as an antidote to some of the problems associated with the normal focus on best-estimate thinking (Davis and Arquilla, 1991b). To develop alternative “images” of the opponent’s reasoning, my colleagues and I have used a combination of essay writing, attribute lists, influence diagrams, and cognitive maps. In one image, for example, the opponent may be pragmatic and incrementalist; in another, he may be exceedingly ambitious and frustrated; in yet a third, he may feel cornered, surrounded by enemies, and desperate.

Figure 1 shows contrasting “cognitive maps” (closely related to what others call “influence diagrams”) used in a study of Saddam Hussein (Davis and Arquilla, 1991a). They represent schematically two very different images of

---

6This approach stems from earlier research modeling decisions in nuclear crisis. See, e.g., Davis (1987).
Saddam's perceptions about the economic situation in mid-1990. The convention in such diagrams is that when an arrow connects two items, an increase or improvement in the first leads to an increase or improvement in the second, unless there is a negative sign, in which case an increase or improvement in the first leads to a decrease or worsening in the second. Negative signs are usually used to indicate a troublesome influence. For example, at the bottom of Figure 1a, we see that an increase in U.S. trade sanctions would worsen Iraq's economic status.

Figure 1a represents the cause-effect relationships emphasized in the intelligence community's "best-estimate" understanding of Saddam prior to the invasion. Figure 1b represents an alternative image that could readily have been formulated and disseminated at the time, except for the pressures to focus on a single best estimate. It includes some very important additional factors, factors such as Saddam's perception that his problems were the direct result of Iraq being squeezed deliberately by his enemies (the United States, Kuwait, and Saudi Arabia among them). It also highlights the connection between his eco-

Figure 1—Saddam's Image of the 1990 Economic Situation: Two Models
nomic travails and his grandiose ambitions. Note that while nearly all the experts would have agreed on all the factors in either diagram being "significant," the dominant mental image (Figure 1a) was one in which some of the factors weren't given much emotional weight. The purpose of the diagrams is merely to highlight differences of perspective, in this case differences in perspective about how Saddam might be viewing the world. We used a number of such diagram pairs in depicting our two images or models of Saddam Hussein. Although we started our work after the invasion and therefore had no trouble constructing a model to explain it, our models proved both insightful and predictive for Saddam's subsequent behavior through February 1991 (i.e., his failure to pull out of Kuwait in the kind of compromise American strategists feared).

Table 2 illustrates a different method for clarifying distinctions between images of the opponent, one based on an attribute list. Again using the example from our study of Saddam Hussein, model 1 is painted as being essentially pragmatic and relatively risk averse. Model 2 is more ambition driven and risk accepting.

**Identifying the Factors Affecting Judgments and Decisions**

Suppose we have used methods such as the cognitive maps, attribute lists, and other devices to develop strong alternative mental images of the opponent—i.e., suppose we have reached the point of being able to say, in a discussion, "No, no, that's not the way model 1 would look at the situation. Where he's coming from is . . . and what he would be concerned about is . . . . Model 2 might reason the way you're talking, but not model 1!" What next? The next step is identifying the factors (i.e., variables) most likely to contribute to the opponent's judgments, notably judgments about the likely, best-case, and worst-case outcomes of various options. It is not very useful to attempt this in abstract terms, because so much of what seems to matter is exquisitely context dependent. My approach is to brainstorm the problem with an interdisciplinary mix of regional experts and strategists, to identify key factors in concrete "natural" language (e.g., Saddam Hussein's assessment of President Bush's resolve), and to develop *hierarchies* of such factors (or variables). This reflects the observation that people make their most reasoned judgments on the basis of only a few "high-level" variables, but these variables, in turn, sometimes reflect many subordinate judgments about "lower-level variables" (Davis, 1987).

To illustrate this, consider how Saddam Hussein may, in mid-1990, have assessed his worst-case outcome (i.e., his "risks") for an option in which he invades Kuwait. Is the worst-case outcome (risks) very bad, bad, marginal, good,
Table 2
Comparing Attributes of Models 1 and 2 of Saddam Hussein

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruthless, power-focused; emphasizes realpolitik</td>
<td>●●</td>
<td>●●</td>
</tr>
<tr>
<td>Ambitious</td>
<td>●●</td>
<td>●●</td>
</tr>
<tr>
<td>&quot;Responsive&quot;; seeks easy opportunistic gains</td>
<td>●●</td>
<td>●</td>
</tr>
<tr>
<td>Impatiently goal seeking; likely to seek initiative</td>
<td>●</td>
<td>●●</td>
</tr>
<tr>
<td>Strategically aggressive with nonincremental attitudes</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Contemptuous of other Arab leaders</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Contemptuous of U.S. will and staying power</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Financially strapped and frustrated</td>
<td>●●</td>
<td>●●</td>
</tr>
<tr>
<td>Capable of reversing himself strategically; flexible (not suicidal)</td>
<td>●●</td>
<td>●●</td>
</tr>
<tr>
<td>Clever and calculating (not hip-shooter)</td>
<td>●●</td>
<td>●</td>
</tr>
<tr>
<td>Pragmatic and once-burned, now cautious</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Still risk taking in some situations</td>
<td>●</td>
<td>●●</td>
</tr>
<tr>
<td>Grandiose, ambitious</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Paranoid tendencies with some basis</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Concerned about reputation and legitimacy in Arab and Islamic worlds</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Concerned only about being respected for his power</td>
<td>●●</td>
<td></td>
</tr>
<tr>
<td>Sensitive to potential U.S. power not immediately present</td>
<td>●●</td>
<td>●</td>
</tr>
</tbody>
</table>

or very good? Well, in the summer of 1990 as Saddam Hussein contemplated this matter, it is likely that he considered the risks to be due to two principal possibilities: the possibility that the United States would defend Kuwait directly and immediately (i.e., that war would begin when he crossed the border) and the possibility that even though the United States didn’t defend Kuwait itself, it would deploy forces into Saudi Arabia and change the balance of power in the region (both militarily and, through sanctions, economically). We don’t know that Saddam thought about the problem this way, but it is likely that however he expressed the issues, these possibilities were on his mind explicitly or implicitly. To assess risks, then, he would be concerned about the likelihood of each of these possibilities and the consequences. The consequences of
an immediate war with the United States would obviously be very bad, but the likelihood of that (i.e., the likelihood of the United States defending Kuwait) probably did not appear large. The United States was more likely to deploy into Saudi Arabia, although the Saudis probably wouldn’t permit it, but even if such a deployment and related sanctions occurred, it is likely that Saddam judged the likely consequences to be tolerable: the Saudis would tire of the U.S. presence, other regional states would deplore it, and economic sanctions would probably not last longer than six months or so.

Figure 2 illustrates how we characterized Saddam’s likely risk assessment hierarchically when he contemplated the particular option of conquering Kuwait. For example, the figure suggests that Saddam would have seen larger risks if there had been strong and credible political warning of U.S. intervention, warning evidenced by strong and credible diplomatic messages along with other indications of resolve by President Bush and Congress. Saddam would also have seen higher risks if there were reason to believe that the United States considered Kuwait to be a vital national interest. Indicators of that might have been a defense agreement, the presence in Kuwait of U.S. forces, or “objective” considerations such as the expectation that Iraq would cut off Kuwaiti oil to the West. Diagrams such as Figure 2 can be worked out in group discussions, and then embellished with subsequent analysis.
Estimating the Opponent’s Judgments and Decisions

Given alternative images of the opponent and an understanding of likely options and major variables or factors, it is possible to estimate how the opponent might reason in a wide variety of circumstances—not merely today’s circumstances, but those that might exist tomorrow or next year. For each image of the opponent, we can develop what I call judgment tables and decision tables. Judgment tables represent how the opponent might look at each of several factors and reach an overall judgment about, say, the most-likely or worst-case outcome of a given option. A decision table is similar, but relates specifically to evaluating the options in a common format.

Table 3 illustrates a judgment table for model 2 of Saddam Hussein evaluating risks of a conquer-Kuwait option in mid-1990 consistent with the factors identified in Figure 2. Note that Table 3 covers a wide variety of possible world situations. The line indicated in bold letters shows the situation that Saddam probably believed best characterized reality in mid-1990, with the result that he probably considered risks to be marginal rather than bad or very bad.

Table 4 now shows a decision table for model 2 of Saddam Hussein evaluating strategic options in late July 1990. Its net assessment for the conquer-Kuwait option is very good. (By contrast, model 1’s assessment was very bad.)

Where do the judgments and decisions (i.e., the values in the last columns of Tables 3 and 4) come from? The answer is that they are subjective estimates made by analysts who have studied the alternative images and tried to “get inside their minds.” However, there is a great deal of logic connecting the elements of the image (e.g., the cognitive maps and attribute lists) with the individual judgments. Indeed, some of this can be treated mathematically to improve “rigor.” It is also possible to build artificial intelligence models to formalize the logic.

To illustrate one way to express the “combining logic” mathematically, consider the following formula calculating the net assessment of an option (Davis and Arquilla, 1991a, app. A):

\[ N = \frac{R\left\{aL + bM + cW\right\}}{\left\{a + b + c\right\}}. \]  

Collectively, these and some additional lines of code can constitute a complete formal model of the opponent’s reasoning. Indeed, one can construct computerized artificial intelligence models. The price of the formality, rigor, and completeness is very high, however. Models of this sort developed to better understand possible nuclear crisis decisionmaking (Davis, 1987) involved roughly 15,000 lines of code for each of two models of the Soviet Union and United States. Even these were not complete.
Table 3

Model 2's Late-July Risk Assessment for the Conquer-Kuwait Option

<table>
<thead>
<tr>
<th>Likelihood of U.S. Deploying into Saudi Arabia</th>
<th>Attitudes in Arab World About Invasion (Arbitrarily Toward Iraq)</th>
<th>Risks(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low</strong></td>
<td>Very bad</td>
<td>Very high</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad</td>
<td>High</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>High</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>High</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>High</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad or marginal</td>
<td>Marginal</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad</td>
<td>Marginal</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Bad</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>Marginal</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Marginal or good</td>
<td>Low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Good or very good</td>
<td>Very low</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>Good or very good</td>
<td>Very low</td>
</tr>
</tbody>
</table>

\(^a\)Values of variables: Likelihood . . . : [high, marginal, low]; Consequences . . . : [very bad, bad, marginal, good, very good]; Attitudes . . . : [bad, mixed, good]; Risks . . . : [very high, high, marginal, low, very low].

Here \(N\) is the net assessment of an option; \(L, M,\) and \(W\) represent the likely outcome, most-favorable outcome, and worst-case outcome; \(R\) is a rounding operator; and \(a, b,\) and \(c\) are weighting factors. If reasoning itself is qualitative, then the formula can be used by first mapping the qualitative values into numbers (e.g., very bad \(\rightarrow -2,\) bad \(\rightarrow -1,\) . . . ), computing the net assessment numerically, and then remapping the result back into qualitative values. This approach creates a preference order for the options.

One model's logic might correspond to \(a = 1, b = 2,\) and \(c = 1,\) and \(R = R^a,\) where \(R^a\) corresponds to rounding up (e.g., 1.5 becomes 2, \(-1.5\) becomes \(-1,\) etc.). The rounding operator is necessary because if reasoning includes discrete concepts such as very good and good, then outcomes may sometimes be "in be-
Table 4
Model 2's Assessment of Saddam's Options, Late July 1990

<table>
<thead>
<tr>
<th>Option</th>
<th>Current Status</th>
<th>Likely Prospects</th>
<th>Risks (Worst-Case Prospects)</th>
<th>Opportunity (Best-Case Prospects)</th>
<th>Net Assessment of Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coerce Kuwait</td>
<td>Very bad</td>
<td>Bad</td>
<td>Very high</td>
<td>Marginal</td>
<td>Bad</td>
</tr>
<tr>
<td>2. Occupy part of Kuwait</td>
<td>Very bad</td>
<td>Marginal</td>
<td>Very high</td>
<td>Good</td>
<td>Marginal</td>
</tr>
<tr>
<td>3. Conquer all Kuwait</td>
<td>Very bad</td>
<td>Very good</td>
<td>Marginal</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>4. Invade Kuwait and Saudi Arabia</td>
<td>Very bad</td>
<td>Very bad</td>
<td>Very high</td>
<td>Very good</td>
<td>Bad</td>
</tr>
</tbody>
</table>

tween. This type of reasoning would give most weight to the likely outcome and lesser weights to the most-favorable and worst-case outcomes. On the margin, the reasoning would lean toward optimism.

It is useful to postulate several types of reasoning that differ primarily in attitudes toward risk, and that they assume a higher willingness to take risks when the current and projected situations are deemed to be very bad, and a reduced willingness to take risks when the current situation and prospects are deemed to be reasonably good. This reflects the well-established (and intuitively familiar) psychological phenomenon described in "prospect theory," developed largely through the work of Daniel Kahneman and Amos Tversky. Psychologically, the reasoning styles might better be characterized as having a predisposition to "go for it" or "take no chance," depending on perceptions about the goodness of the current situation and current trends. Another point I have emphasized in my own work is the role of thresholds: below some level of perceived probability, risk is treated as zero, despite the consequences of the

---

8Kahneman and Tversky (see their work and others’ in Shafer and Pearl, 1990) have based most of their work on empirical testing of individuals. In this work I am claiming that a similar prospect-theory phenomenon occurs at the level of national decisions, which are not usually the result of a single individual making choices unilaterally. As noted by colleague Kenneth Watman, there is interesting theoretical and empirical work to be done in testing the claim. In any case, I mention the Kahneman-Tversky work not as "proof" of the approach, but rather as providing some corroborating evidence.
risk. That is, not only do we often underestimate risks, we often go farther and ignore those we have judged "low." The reverse also happens: we sometimes rule out options because we see them as involving a level of risk beyond some threshold of acceptability.\footnote{Such thresholds, coupled with a grasping-at-straws phenomenon, were crucial in earlier studies using Soviet and U.S. models to understand possible escalation and de-escalation in nuclear crises. See, e.g., Davis (1992).}

The point here is that we can not only construct formal models to reflect best-estimate notions about how the opponent is and may in the future be reasoning, we can also construct alternative models to reflect fundamental uncertainties about the nature of that reasoning. The principal question, of course, is whether we have to consider an infinite number of such alternative models. The answer appears to be no. Indeed, having two or at most three models appears to go a very long way, especially since one can also do sensitivity analysis within a given model. This is crucial, because it means that the technique, which is surely good for getting groups to confront uncertainty and be more humble about any "best estimate," should also be workable in practice. Formal intelligence estimates and high-level meetings should be able to cope with two, or conceivably three, very different perspectives on how the opponent may be thinking.

FACTOR TENDING TO INCREASE RISK TAKING

Since risk-taking propensity is such an important issue in determining aggression, it is worthwhile to review major factors tending to increase willingness to assume risk (see Figure 3). Starting at the top and moving clockwise, we see first the previously mentioned role of the current situation. The next factor is the degree to which the decisionmaker can make decisions unilaterally, without broad discussion that might mitigate perceptions and introduce new considerations. The next factor is ambition. This is often underestimated in thinking about adversaries in crisis and conflict. Status-quo powers fairly comfortable with their own circumstances are especially likely to underestimate others' ambitions. So it is that Saddam Hussein was erroneously assumed to be "pragmatic" and to be merely looking for a way to improve Iraq's economic situation "somewhat," when in fact he had grandiose goals. Similarly, the United States applied incrementalist compulsive logic to Ho Chi Minh, when he was an idealist revolutionary. Other factors include opportunities for reaching important goals, the abstractness of risk factors (the more abstract the risk factor, the more it may be underestimated by someone who is yearning for
Figure 3—Factors Contributing to Risk-Taking Behavior

action, pain tolerance, and the degree to which the protagonist believes he is in control of events and therefore able to "make his own luck." All of these factors should be familiar to us from everyday life supplemented by a knowledge of history.

It should perhaps be obvious that in applying the theory described above, one considers the presence or absence of the factors in Figure 3 when estimating how a given type of decisionmaker might judge the worst-case outcome of a given option. One also uses these factors in judging which reasoning models to employ (e.g., in choosing parameter values \(a, b,\) and \(c,\) and rounding rules of equation 1, to correspond to more or less risk aversion).

**A GENERIC SITUATION ENCOURAGING AGGRESSION**

Let me now sketch how the theory applies to real-world problems of defense planning and foreign policy. To do so, let me first describe, using a decision table, a remarkably generic situation to be avoided, one in which aggression is possible and deterrence is difficult. Table 5 is the decision table that we do not want potential aggressors to have in mind. That is, we want to avoid situations in which, when the potential aggressors consider options, they end up reasoning as shown in Table 5.

---

10See Arquilla (1992) for examples of how aggressors have often underestimated the power and will of maritime powers and naval forces.
Table 5
Dangerous Assessments Encouraging Aggression

<table>
<thead>
<tr>
<th>Option</th>
<th>Likely Outcome</th>
<th>Best-Case Outcome</th>
<th>Worst-Case Outcome</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Continue peaceful policies</td>
<td>Very bad</td>
<td>Bad</td>
<td>Very bad</td>
<td>Very bad</td>
</tr>
<tr>
<td>2. Coerce target</td>
<td>Bad or marginal</td>
<td>Marginal</td>
<td>Very bad</td>
<td>Bad</td>
</tr>
<tr>
<td>3. Take limited military action for limited gains (e.g., conquer a portion of target’s country)</td>
<td>Marginal or good</td>
<td>Good or very good</td>
<td>Bad or marginal</td>
<td>Marginal or good</td>
</tr>
<tr>
<td>4. Invade; conquer target country</td>
<td>Very good</td>
<td>Very good</td>
<td>Bad or marginal</td>
<td>Marginal or good</td>
</tr>
</tbody>
</table>

The salient features of this somewhat generic dangerous situation are:11

- The perception that the current situation is very bad (implicit in the conclusion that a continuation of peaceful policies would have a very bad likely outcome).
- The perception that continuing current or other peaceful policies will not improve the situation.
- The perception that mere coercion may have a payoff, but not much, and might make things worse (e.g., by strengthening the coalition of hostile interests and by causing the potential target of aggression to increase its defenses).
- The perception that military action is likely to pay off, may pay off handsomely, and involves risks that are not outrageous and perhaps only marginal.

Importantly, national leaders have their own standards in evaluating “current situation” and the outcomes of various options. These often differ substantially from the standards that leaders of other nations might expect. As suggested above, it is easy to underestimate ambitions (and emotions) of adver-

---

11This discussion overlaps with classic expositions on the causes of war (e.g., Blainey, 1983 and Howard, 1984).
saries by assuming that they will behave “pragmatically” or “reasonably,” by which is meant being satisfied with only marginal improvements in their situation. See Arquilla and Davis (1992), Payne (1992), and Arquilla (1992) for historical cases in which deterrence “should” have worked as judged by relative power, but did not.

It is also a profound mistake to believe that adversaries necessarily reason in a way that decision theorists would describe as attempting to maximize expected utility. In my view, exceedingly ambitious goal-driven people often behave so as to maximize the likelihood of success, which is quite different psychologically from maximizing expected utility. That is, utility theory is a poor way to represent such reasoning even though one can look at behavior and infer effective utility functions.

DANGEROUS POLITICAL-MILITARY CONTEXTS

Having discussed briefly how the reasoning of potential aggressors may be characterized analytically, and having described in that framework the kind of situation that breeds aggression, let us now consider what kinds of political-military context would cause a potential aggressor to reach the conclusions suggested in Table 5, even when there is the material potential for deterrence.

Figure 4 indicates schematically one such context. Here one country is considering aggression against a target country and another country or group of countries are potential protectors. The potential aggressor, the potential protectors, and the target all have allies or potential allies. They also have images, or perceptions, of how all the other actors reason and behave. Even this complex set of interactions is incomplete, because each country has an image of itself that may or may not be correct, and each ally has an image of all the actors and itself. Further, each actor may have an image of how each other actor views each other actor, and so on recursively to infinity. It is often adequate, however, to simplify to the level of detail indicated.

If the potential protectors regard the target country as “vital national interests,” then they will ordinarily communicate that view to the potential aggressor, who will then take that into account in making decisions. If deterrence fails, it may then be because of an unstable military balance (see relevant discussion in Huber, 1990 and Huber and Avenhaus, 1993). The case I am most interested in, however, is the one in which the potential protectors are materially capable but ambivalent. They are not clear in their own minds about how important the security of the target country is to them. It may be, for example, that one of the target’s regional neighbors is a vital interest and that the target itself is a buffer between the potential aggressor and those interests, but is not
itself important (and perhaps may not even be friendly). Or it may be that what happens to the “unimportant” target country could somehow affect events in their own countries (e.g., through the spread of revolution, the disparaging of new-world-order concepts, or the enhanced reputation and stature of the aggressor, who is attempting to foment revolution in their own countries), but the relationship seems abstract.

Will deterrence work? Not easily. Consider that not only is the potential protector ambivalent, it must also have cooperation from the target and its own allies in order to do anything. Indeed, it is worse than that, because the protector is often a democracy with internal political processes. Even if its executive branch believes in protecting the target, it may well be that the legislative branch and public do not. Even the military may be unwilling to get itself embroiled. And, because the executive branch is aware of these reluctances, it may be self-deterred.

So let us now suppose that the potential aggressor looks at the situation. It may observe that the protector is doing little, that there are internal political divisions within the protector’s country, that some of the protector’s allies are unwilling to participate, and so on. It sees the risks associated with its aggression as being low or at least tolerable, thereby creating one of the principal features of the situation in Table 2.

Or consider the target country. It may see the lack of cohesion, will, and support in the protector country and its allies. It may also see lack of support
and cohesion in its own regional neighbors. One thing it does know, however, is that making the potential aggressor mad could bring serious trouble. Under these circumstances, how hard should the target country even work to request assistance from the protector? If the request becomes known (which is likely) and fails, which is very possible, the result may be to decrease rather than increase its security.

It is presumably evident that this illustrative case corresponds to specific instances in history and prospective history. In the case of Iraq threatening Kuwait, the situation obtained almost precisely (Davis and Arquilla, 1991a). Before August 2, 1990, the United States had no policy or internal political consensus about whether Kuwait was a vital interest worth protecting; Kuwait and Saudi Arabia did not believe Iraq would invade, but when they thought about it they questioned the reliability of the United States as a protector. Saudi Arabia felt that it did not want to further enrage Iraq. Saddam Hussein looked for indicators of Western resolve and saw nothing to cause him much worry: no military buildup, only equivocal warnings, no indications that Saudi Arabia would permit the United States to come to Kuwait's aid even if it chose to do so, and no evidence that Kuwait was even considered vital.

But the case of Kuwait was not unique. Consider now the situation in Europe where both Western Europe and Russia have at one time or another viewed Eastern Europe as a buffer. One might naively assume that both would consider the security of Eastern Europe to be very important, but it is hardly surprising that, in peacetime, little is being done to create that security. Would it be so surprising if at some point in the future a crisis arose between parts of the former Soviet Union and parts of Eastern Europe—or between Russia and other parts of the former Soviet Union such as Ukraine? Would it be surprising if military action occurred and that after that action occurred, observer nations decided that their own stakes were higher than they had previously realized? Consider here the case of Yugoslavia. Originally, it was not seen as a "vital national interest" of any of the NATO nations, or of Russia. But after seeing what has happened, with a level of willful and brutal aggression against civilians that has not been witnessed in Europe for decades, many in those nations wish they had acted firmly at the outset of crisis. It is now too late for deterrence, and compellence is proving as difficult as always.

In reality, there are numerous examples of potential future crises that fit the picture sketched here (Table 6). The generic features of the crises considered here are:

- The stakes are likely not appreciated beforehand by the potential protector and its critical allies;
Table 6
Possible Challenge Cases in the 1990s

<table>
<thead>
<tr>
<th>Target</th>
<th>Aggressor</th>
<th>Potential Protector</th>
<th>Potentially Critical Regional States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kuwait</td>
<td>Iraq</td>
<td>U.S.</td>
<td>Saudi Arabia, Iran</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>Iraq</td>
<td>U.S.</td>
<td>Egypt, Oman, UAE, Iran,...</td>
</tr>
<tr>
<td>Poland</td>
<td>Russia</td>
<td>Germany, U.S.</td>
<td>Rest of NATO</td>
</tr>
<tr>
<td>Baltic states</td>
<td>Russia</td>
<td>NATO</td>
<td>Other states of the former Soviet Union</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>China</td>
<td>U.S.</td>
<td>Japan, South Korea, Russia</td>
</tr>
</tbody>
</table>

- There is reluctance to believe that the aggressor will take military action of a sort that will prove unacceptable;
- There is widespread failure to communicate candidly about the situation and options beforehand;
- There are major perceived domestic-political and alliance-related costs to attempting deterrence; and
- The potential aggressor perceives weakness and lack of either cohesion or will.

In addition, the potential aggressor may feel compelled to action by strategically adverse trends and may be overconfident because of recent payoffs for decisive action or recent failures of its adversaries to take actions when they should have done so.\(^{12}\)

**THE SPECIAL PROBLEMS CAUSED BY PROLIFERATION**

Although I shall not dwell on these issues here, the proliferation of weapons of mass destruction (and delivery systems for those weapons) makes deterrence and compellence even more difficult, as becomes evident when one "games

---

\(^{12}\)The tendency toward overconfidence in human reasoning is well documented in the empirical literature and should be familiar from military history. Conquerors and other successful risk takers often come to believe that they are lucky or even invincible. There is a fine line between that poor judgment and the more correct one that vigorous and decisive people often "make their own luck" or the one that in warfare it is important to achieve and maintain the initiative.
out” various crises using the methods I have described only briefly here.\textsuperscript{13} For example, if Saddam Hussein had had survivable operational nuclear or biological weapons, and if there had been reason to believe he would use them, it is less than obvious that Kuwait would exist today. This subject deserves far more attention than is possible here, but two observations are especially salient based on work to date:\textsuperscript{14}

- The United States is more likely to intervene to deter an invasion by an aggressor armed with weapons of mass destruction than it is to intervene to roll back an invasion that has already occurred. Thus, early crisis response is more critical than ever before (Davis and Finch, 1993).

- It is highly desirable that the United States possess credible options to respond with chemical and nuclear weapons to attacks on U.S. forces or allies with weapons of mass destruction. In some cases at least, it does not seem likely that conventional attacks with precision-guided munitions would have comparable deterrent or compellent value.

IMPROVING APPROACHES TO DETERRENCE AND COMPELLENCE

Having described situations and reasoning patterns that undermine stability and efforts to deter aggression, it is now natural to suggest improvements of approach that would help avoid these situations and patterns.

Elevating the Recognized Importance of Indirect Threats and Buffer States. A challenge for statesmen is to recognize, and to persuade their governments to recognize, that many indirect threats to security are threats to “vital national interests.” This is difficult, because it is not even intuitively obvious. The challenge will increase as those who grew up during the 1920s and 1930s retire from the political scene and new figures take their place, figures who will not have the same memories about appeasement.

I have seen related issues in my own graduate-level classes in the United States, and in discussions elsewhere involving serious mid-career military officers, when the people in question think about what is and ought to be the

---

\textsuperscript{13}See Millot, Molander, and Wilson (1993) for discussion of an extensive set of exploratory games conducted on such matters. These, the “Day After” games, were designed primarily to get analysts and policymakers to think seriously about the consequences of proliferation for future crises and the consequences for U.S. defense planning.

\textsuperscript{14}A study (Watman and Wilkening, forthcoming) discusses these and other issues in more detail. It makes important distinctions between the effects of weapons of mass destruction on the U.S. willingness to intervene and its likely war aims given that it does intervene.
President’s national security strategy. Such discussion often begins by reviewing what are considered to be fundamental values and enduring interests. It is common for arguments to break out almost immediately, because there are disagreements about whether the security of allies and other nations is a fundamental interest of the United States. It seems evident that the security of other nations is important, but is not a “vital” interest.\textsuperscript{15} If so, however, the tendency is to conclude that any strategy to maintain that security should be “political” and “tentative,” without creating obligations that might lead to war. That undercuts deterrence. Statesmen have understood this problem and overcome it in many instances, most notably by developing treaties such as those that bind the United States to its NATO allies and Japan, even though the treaties create obligations that could lead to war, and even to nuclear war. Statesmen have been much less successful, however, in dealing with threats less monolithic and less menacing than that of the Soviet Union during the Cold War.

**Recognizing Uncertainty in Assessing Intentions of Potential Aggressors.** As argued elsewhere (Davis and Arquilla, 1991a,b), there is a chronic tendency for governments to focus unduly on “best estimate” assessments of what potential aggressors are up to, even though there is a long history of the best estimates having substantially overestimated or underestimated threats. The tendency to focus on the best estimate, and to filter information through the lens of that estimate long after the contrary evidence is strong, is typical of human reasoning in general. It can, however, be mitigated by changing the processes by which governments study potential crises in advance and the processes by which they write national estimates and conduct crisis deliberations. In particular, I recommend requiring staff agencies to develop, maintain, and treat seriously alternative models of other-country reasoning, and to have national-security staffs develop strategies that hedge against the less popular of the models proving to be correct. Despite cynical claims to the contrary about the inability of top-level political leaders to deal with complexity, discussing such hedged strategies should be quite feasible at the highest levels (e.g., in the U.S. National Security Council). Is there really any reason to believe that presidents, prime ministers, cabinet members, congressional and parliamentary leaders, and four-star generals cannot deal with uncertainty and hedging requirements?

**Reducing the Costs of Hedging.** I am confident that political and military leaders can do better in thinking about uncertainty and the need to hedge. For

\textsuperscript{15}A major problem here is semantic. People read the phrase “vital national interest” and interpret it literally, when a better interpretation might be “an interest over which the nation would go to war, whether or not doing so would be required for the nation’s survival.” That is, “vital” interests are often not, in practice, “vital.”
this thinking to be fruitful, however, it is also necessary that it be easier for them to take actions consistent with their analyses. Currently, it is extremely difficult for statesmen to take actions to underwrite deterrence or to make competence effective in crisis. Until events reach some point of clarification, many of the very steps that are most needed are considered—by other policymakers, elements of the news media, and the man in the street—as potentially provocative and very dangerous. Consider a President contemplating vigorous military, economic, and political actions to avert aggression (including internal aggressions such as that occurring in the former Yugoslavia). He must worry about tomorrow’s headlines, which may claim he is recklessly leading the United States into war, or recklessly committing the nation to be “policeman of the world,” even in portions of the world where other countries should be providing police functions. These costs of action are real, tangible, and certain, whereas the actions themselves may be in the name of “hedging” against “possible” intentions by other countries, intentions that may in fact be much more benign.\textsuperscript{16}

What can be done to reduce the costs? I see several generic approaches:

- Create political obligations and doctrines that compel actions (e.g., security agreements committing the United States to the protection of other nations).
- Have top aides (including relevant generals), and subsequently policymakers, “play” through the political-military crises when there is strategic warning of crisis, so as to sensitize them to the need for early actions to reinforce deterrence if the crisis worsens. Use modern technology such as videoconferencing to assure participation by all appropriate people.
- Have them do so in cooperation with allies and other organs of government so that necessary political alliances will form quickly in the event of crisis.
- Prepare in advance for how to explain necessary actions to the public.

All of this would require governments to do a far better job than their predecessors in reacting to strategic warning, which almost always precedes serious crises.

Preparation for Early Intervention. The best time to deter an opponent is before events go too far. Doing so may require only a moderate application of power (political, military, and economic), because the tactical objective is to change perceptions about stakes—i.e., to demonstrate unequivocally one’s commitment to protect interests. This, unfortunately, runs directly counter to military conservatism, which agonizes about the risks posed to military forces.

\textsuperscript{16}The critical role of hedging costs was clarified for me in discussions with colleague Arnold Horlick.
that may be sent as political signals, and which objects deeply to intervention that precedes clear expressions of national objectives and the development of a political consensus assuring support for what will be necessary. The concept of a tripe wire force is anathema to many military officers, and the notion of piecemeal commitment of military forces before developing political and moral support from the nation as a whole is even more so (Summers, 1984). Currently, the U.S. military emphasizes the importance of “overwhelming force” and de-emphasizes early actions beyond mere signaling of a sort that does not greatly endanger U.S. forces. My own recommendations here (for the United States and NATO in particular) are to develop rapidly deployable forces that are far more lethal and mobile (and hence survivable) than traditional initial projection forces such as the 82nd Airborne Division. Such forces might include a mix of tactical air forces with modern munitions and highly mobile ground forces such as the 101st Air Mobile/Air Assault Division (augmented with additional attack helicopter units), or armored cavalry regiments. Such forces could be put in harm’s way early, so as to demonstrate commitment and protect politically important boundaries or locations. They could also provide a formidable defensive challenge and hope to do so while surviving. This is especially plausible because of the likelihood that the United States would have or quickly gain air supremacy, after which time an attacking armored force would be extremely vulnerable to airpower.

Let me end this item on a sober note by observing that it is the height of folly to establish tripwires without having the will and capability to respond massively to attacks on such a force. Tripwires can be, and often have been, tripped. The nation has moral obligations to any of its military personnel that it places in jeopardy.

Paying Attention to Theory When Developing Plans of Action. Perhaps the most remarkable feature of the deterrence concepts discussed here is how readily some of them are accepted in casual conversation and consistently overlooked or rejected in practice. Let me mention a few examples here (Davis and Finch, 1993):

- The doctrine of overwhelming force is widely accepted, but is antithetical to timely deterrence.
- The flexible deterrent options (FDOs) developed by the Joint Staff in recent years are excellent in name, but may be counterproductive in practice if the options are seen as evidence of timidity by the aggressor or our allies. Examples of such timid actions in crisis include: (a) President Carter’s 1979 deployment to Saudi Arabia of unarmed tactical aircraft, intended to show support for Saudi Arabia during a period of turbulence after the fall of the
Shah; and (b) the late-July 1990 exercise in the Persian Gulf involving naval forces and KC-135s, intended to be a signal to Saddam of U.S. support for the United Arab Emirates and Kuwait.

- Nations consistently attempt deterrence and compellence with threats of gradually escalated political and economic sanctions, only seldom backed up with military options. Gradualism is seldom effective and threats that are not backed up with military power are often not taken seriously by the most dangerous types of aggressor.

A New Emphasis on “Punishment Options.” The cold reality is that the United States and its allies are not likely to be able or willing to defend many of the weak states that may, in the decade or two ahead, be the subject of coercion or invasion. Kuwait was a close call; the Balkans were a failure; Ukraine and the Baltic states are far away and isolated; Hong Kong is far away and isolated; and so on. My own view is that the United States and its principal allies should move toward deterrent strategies that would focus on severe and immediate political, economic, and military punishment of aggressor states. The image here should be one of both attacking the aggressor from the air and sea to charge him a high price, and following up with political and economic isolation. Would it be possible for the major nations to arrive at such an approach? Is it conceivable that it might achieve some status of legitimacy in the United Nations over time? Could such a deterrent strategy be used against a nuclear-armed country such as Russia? As I have attempted to demonstrate, deterrence is exceedingly difficult to accomplish. We need deterrent strategies that are better than those nations have used historically. This, it seems to me, is a profound issue, because the alternative to a new strategy may be a continuation of history.

---

17See Woodward (1991:240) for a description of how counterproductive Prince Bandar felt that deployment had been, causing the Saudis to worry about American “guts.”

18Not only was the exercise militarily meaningless and very ambiguous as a signal, it was approached timidly because even this action was considered by some (including the UAE and other Arab states) as provocative. As General Powell later noted, the exercise got Saddam’s attention, but didn’t scare him; it did scare our allies.

19Richard Haas sharpened this point in my mind by emphasizing strenuously that the key to a punishment strategy is deliberately not establishing compellent objectives. It is within our power to punish, and even to do so severely. It is often not within our power to successfully compel the behavior we seek. The credible threat of punishment, however, might well deter the action in the first place; the exercise of punishment might deter those who would copy the actions of a “successful” aggressor.
BIBLIOGRAPHY


Arquilla, John, and Paul K. Davis (1992), *Extended Deterrence, Compellence, and the "Old World Order,"* RAND, Santa Monica, CA.


Davis, Paul K., and John Arquilla (1991a), *Deterring or Coercing Opponents in Crisis: Lessons from the War with Saddam Hussein*, RAND, Santa Monica, CA.


Jablonsky, David (1991), Strategic Rationality Is Not Enough: Hitler and the Concept of Crazy States, Strategic Studies Institute, U.S. Army War College, Carlisle Barracks, PA.


Shafir, Glenn, and Judea Pearl (eds.) (1990), Uncertain Reasoning, Morgan Kaufman Publishers, San Mateo, CA.


Warman, Kenneth, and Dean Wilkening, with John Arquilla and Brian Nichiporuk (forthcoming), U.S. Regional Deterrence Strategy RAND, Santa Monica, CA.

Unprecedented population growth, urbanization, and population displacement will not only lead to conflict in the developing world, but will change the nature of conflict. This has strategic, operational, and tactical implications for U.S. military forces that may become involved in such conflicts. Case studies of U.S. operations ranging from direct intervention to peacekeeping to disaster relief yield many practical lessons for U.S. forces operating in the more urban, heavily populated, and politically constrained conflict environments likely to predominate in the developing world. Examinations of U.S. efforts in Panama, Lebanon, Somalia, and Bangladesh also reveal that the operations most likely to succeed are those conducted in a stable operational environment (whether peace or war), where the United States has clear-cut political and military objectives. This was true in both Panama and Bangladesh, despite the fact that one was a combat operation and the other a noncombat effort. In Lebanon and Somalia, on the other hand, the political and military objectives were less clear and the operational environment was fluid. In both instances, U.S. forces were deployed to accomplish a humanitarian mission—for which they were carefully selected, prepared, and equipped—only to find themselves in a more threatening situation and taking a more aggressive stance. U.S. military planners will have to better prepare U.S. forces for future such efforts, where shifting operational environments and changing political goals will require concomitant shifts in military objectives, missions, and capabilities.

INTRODUCTION

For more than 40 years the U.S. military concentrated its attention, energy, and resources primarily on defending Western Europe from Soviet aggression and, in turn, on preparing for total victory through the application of decisive force against the Warsaw Pact. Even when U.S. strategists and planners were forced to shift their focus to Korea in the 1950s and Indochina in the 1960s, the American military's fundamental operational imperatives and basic approach to warfighting remained unchanged.

Although the U.S. military's primary mission remains to fight and win the nation's wars, the increasing salience of military operations other than war (OOTW) in the years since the Cold War's end have imposed additional, and
in some cases new, missions and responsibilities. Operations other than war include peacekeeping (PKO), peace enforcement (PE), counterinsurgency (COIN)/insurgency, counternarcotics (CN), counterterrorism (CT), noncombatant evacuation operations (NEO), arms control, support to domestic civilian authorities, humanitarian assistance (HA) and disaster relief (DR), security assistance (including training), nation assistance (including civic action), shows of force, and attacks and raids (U.S. Army, 1993). In retrospect, planning and mission preparation during the Cold War appears almost straightforward compared to the complex variations of the OOTW missions of the new era. Such missions will often be conducted within a highly fluid operational continuum: ebbing and flowing between peace and conflict, combat and noncombat operations, and conventional and nonconventional activities.

OOTW, moreover, are likely to occur in parts of the world where the United States has traditionally had few critical interests, areas that therefore have received little attention or interest from the American military (recent OOTW conducted in the Kurdish enclave in northern Iraq, in Somalia, and in Bangladesh bear out this point). The global nature of new post–Cold War political pressures will possibly increase demand for OOTW missions such as disaster relief, humanitarian assistance and/or support, peacekeeping, etc.—with the attendant likelihood that the U.S. military will be called upon to deploy to unfamiliar and, in many cases, physically demanding corners of the globe.

Within the context of the intensive population growth, migration, and urbanization now unfolding throughout the developing world, this paper focuses on how these emerging OOTW missions will affect strategic, operational, and tactical planning for future conflicts (limited conventional combat or nonconventional combat operations such as insurgency), noncombat missions (humanitarian assistance and disaster relief), and peacekeeping and peace enforcement operations.

THE CHANGING NATURE OF CONFLICT

Although the overall worldwide population growth rate is slowing, the populations of 37 of the world’s most populous countries—which are all located in less-developed parts of the globe (i.e., the nonindustrialized regions of Asia, Africa, Latin America, and the Middle East as well as certain parts of Central Europe and the southern regions of the former Soviet Union)—are continuing to grow by more than 3 percent annually (The Economist, 1990).1

1The projected world population of eight billion in 2020 will double the population of 1976 and quadruple that of 1930. By comparison, it took from the dawn of time until 1830 for the world to acquire just one billion people. See Green (1989:2).
Between 1990 and 2025, the fastest population growth will occur in the world's least-developed countries, which will experience a 143 percent increase in population. In the rest of the developing world, the population will grow by 75 percent over that same period, while the developed regions will have only a 12 percent increase. Thus, the countries with the greatest population growth are already among the world's poorest, least developed, and most economically deficient.²

Admittedly, problems of population growth, poverty, and hunger are not new to the less-developed world and have been throughout history the fulcrum for war, revolution, and subversion. But what sets today's developments apart are the monumental urbanization and mass migration processes that are together transforming the less-developed societies from predominantly agrarian to urban.³ As rapidly as these countries' populations are growing, their urban populations are increasing at more than twice that rate (Rogers, 1982:1 and Yeung and Belisle, 1988:99).

For the first time in history, more people today live in cities in the developing world than in the industrialized world.⁴ By the turn of the century, 264 of the world's 414 cities with a million or more people will be located in the less-developed parts of the globe.⁵ Africa, the world's poorest continent, alone will have more than 50 of these cities (compared to only 19 in 1980 and just 6 in 1950),⁶ and by 2025 it will have an urban population three times that of North America (Camp, 1990:1). The number of million-plus cities in Asia will double by the turn of the century from 81 to 160 (McAuslan, 1985:127),

²Within the next decade, at least 65 countries (including 30 of Africa's 51 countries) will be completely dependent on food imports. Africa, for example, currently holds 28 of the world's 42 poorest countries and will experience a threefold population increase within the next 35 years. See The Economist (1990) and Crosette (1990).

³By the year 2000, for example, half the world's population will be urban—compared with 30 percent today and just 17 percent in 1950. See Rogers and Williamson (1982:1) and Rogers (1982:466, 486); see also McAuslan (1985:127). Although slightly different figures are cited, the same general pattern is identified in Yeung and Belisle (1988:99).

⁴Although demographic change and population growth have long been staples of fertility, health, economic, and sociological studies, their effects on international politics or their potential as catalysts for intranational and transnational conflict have not been scrutinized to the same degree. Exceptions are Chourchi (1983, 1974) and Sarkesian (1989).

⁵Rogers and Williamson (1982:1). Some 58 of the world's largest metropolitan areas, for example, are now located in developing countries. See Camp (1990:1).

⁶McAuslan (1985:127). Indeed, by 2025, Africa will have 36 cities of four million or more residents, with an average of nine million, more than greater London today. By contrast, 40 years ago, no African city between Johannesburg and Cairo had even one million inhabitants (The Economist, 1990).
and although less dramatic urban growth rates are projected for Latin America, by the year 2000 more than three-quarters of its population will live in urban areas (compared to 50 percent today), representing the highest proportion of urban dwellers on any continent (McAuslan, 1985:129; Regional Conflict Working Group, 1988:12; and Yeung and Belisle, 1988:100). Moreover, as fast as cities in the developing world are growing, the slums and shantytowns are growing twice as fast, thereby contributing to the uncontrolled geographical—as well as numerical—expansion of these urban centers.

Finally, the implications of population displacement go beyond the problems of urbanization. Refugees (who migrate across international borders) and internally displaced persons (who migrate within their own countries) frequently relocate without even a minimal amount of clothing, food, and shelter to places where, without assistance, they could not survive. They may carry diseases, expand existing or create new slums and shantytowns, and exacerbate racial, religious, and ethnic prejudices. They drain the local or national host government’s limited resources for social services, infrastructural development, and policing, often creating resentment that can lead to violence. Unsurprisingly, perhaps, the world’s poorest countries typically experience the heaviest flows of displaced persons. For instance, the average number of displaced people (including refugees and internally displaced people) for countries with a low per-capita gross domestic product (GDP) is 465,000, in contrast to the 222,000 mean for high per-capita GDP countries.

The Demographics of Future Conflict

The massive expansion of both the population and geographical dimensions of the developing world’s urban centers will have a profound effect on both conventional and unconventional conflict. On the one hand, this radiating urbanization process—consuming increasingly large tracts of a country’s territory and inevitably encompassing key lines of communication, transportation nodes, and road and rail links—will make it difficult in conventional warfare for ground forces to bypass or maneuver around cities as they have traditionally attempted and as warfighting doctrine mostly stresses. On the other hand, the proliferation of weapons—including small arms, machine guns, rocket-propelled grenades (RPGs), and plastic explosives, as well as weapons of mass destruction—throughout the developing world will facilitate the growth of insurgency and heighten the potential for other forms of political violence.

Conventional MOUT (Military Operations on Urban Terrain). Most traditional warfighting doctrine has stressed the desirability of avoiding combat in cities. As urban warfare is often intensive, protracted, and conducted among a surrounding civilian population, emphasis has long been given to avoiding
urban areas whenever possible. However, combatants have a natural attraction to centers of political and economic activity as a means of controlling populations at large. In addition, the massive urbanization described above not only increases the probability of urban warfare by sheer chance, but has related consequences that can themselves lead to conflict, including dissatisfaction that grows as city infrastructures prove increasingly incapable of supporting rapid influxes of people.

Because of the historical emphasis on avoiding urban combat, the U.S. military has not stressed MOUT in its doctrine, training, or equipping, and its capabilities for conventional operations on urban terrain are thus arguably inadequate to operate in this new threat environment. Furthermore, military MOUT capabilities that do exist emphasize the U.S. advantage in massive firepower. Little doctrine or training is therefore relevant to combat in the developing world's heavily populated urban areas, where strict rules of engagement are likely to limit the utility of a firepower advantage. In this environment, the need for combined arms, artillery, snipers, and armor is paramount. Tanks and armored personnel carriers, for example, can be invaluable in MOUT—provided they are defended by dismounted infantrymen. Similarly, artillery can provide direct-fire support or be used more for psychological effect (see McLaurin et al., 1987).

Unconventional MOUT (Insurgency). Urbanization in the developing world will also fundamentally alter the nature of insurgency. As rural populations, uprooted by poverty, hunger, and conflict, continue to migrate to the cities, the guerrilla forces dependent upon them for food, information, concealment, and support will have no choice but to follow, adjusting their strategies and tactics along the way. Thus, the future killing grounds of the developing world will not be the impenetrable forests or remote mountain areas where guerrilla wars have traditionally been fought, but increasingly the crowded slums in and around the less-developed world's burgeoning urban centers, whose residents will become inextricably enmeshed in insurgent-government conflict.

These conflicts will differ from the ultimately unsuccessful urban insurgencies conducted in Argentina, Bolivia, and Brazil in the 1960s, insofar as a variety of hospitable conditions enables today's guerrillas to better adapt their rural insurgency strategy to an urban environment. Urban guerrillas have the same benefits and advantages that they enjoyed in rural areas: control over territory, the allegiance (whether voluntary or coerced) of a considerable part of a country's population, inaccessibility to security forces, and a reasonably secure base for operations around the heart of the government and its administrative and

---

7 See, for example, O'Connell (1992); Dewar (1992); McLaurin, Jureidini, McDonald, and Sellers (1987); and Mahan (1983).
commercial infrastructure. They also have more opportunities for media coverage and international attention that would be unobtainable in isolated jungles or mountains.

Because of their warren-like alleys and unpaved roads, the slums have become as impregnable to security forces as a rural insurgent's jungle or forest base. The police are unable to enter these areas, much less control them. Urban insurgents in Lima today and in San Salvador during the late 1980s, for example, have been able to seize control of defined geographic areas in the band of slums and shantytowns ringing those cities, sometimes establishing shadow governments that exercise a crude form of sovereignty and provide public services or scarce commodities while repulsing government efforts to reassert control over these so-called "liberated zones." The insurgents thus pursue a deliberate strategy to sever the government's authority over its urban centers and thereby weaken both its resolve to govern and its support from the people, the aim being to eventually take power, first in the cities and then in the rest of the country.

This phenomenon is no longer restricted to Latin America. As other regions of the less-developed world begin to sustain the high rates of urbanization that occurred in Latin America two or more decades ago, they too are experiencing the same—arguably inevitable—rise of urban insurgency. The renewed insurgency in Angola, for example, has been fought in and for that country's urban centers. The most vicious combat has taken place in the resource-rich cities of Huambo, Uige, Menongue, Kuito, and Luanda. Using a strategy of capturing some urban areas while bottling up government forces within others, the rebel organization UNITA has now effectively deprived the government of control in three-quarters of the country. Similarly, insurgents in Liberia have focused their efforts in the capital city of Monrovia, while guerrillas in Sierra Leone have battled the government repeatedly for the diamond-mining hub, Koidu. Shiite rebels in Afghanistan have brought their conflict with government troops into the heart of Kabul.

Even if insurgents choose not to base their operations in urban areas, they can nonetheless take advantage of urbanization. Rural-based insurgencies are finding cities increasingly attractive and lucrative targets. Whereas cities were once the culmination of a revolution, with the new proliferation of urban areas—and the inability of governments to defend them all—cities have become relatively simple targets that yield high dividends for low cost. Insurgent groups can disrupt energy and telecommunications facilities, draw international attention, demonstrate the government's inability to protect its people, and recruit from among the disaffected population. Thus, even the insurgencies that remain based in rural areas can take advantage of urbanization by increasing their reliance on urban terrorism.
Likely Impact on Military Operations Other Than War

Most MOUT operations in the post–Cold War era, however, are likely to be conducted in the "gray world" of OOTW, where the operational environment can be highly fluid and able to shift quickly from peace to conflict. The environment may also require the full range of OOTW missions, including peacekeeping, peace enforcement, disaster relief, and humanitarian assistance as well as counterinsurgency, counterterrorism, noncombatant evacuation, and nonproliferation operations. The problem for many U.S. military forces is that they are well versed in the missions at either end of the OOTW operational continuum—with skills in urban combat honed through conventional MOUT courses on the one hand and familiarity with disaster relief and humanitarian assistance on the other—but they are hampered by inexperience with the OOTW tasks falling somewhere in between, such as crowd control, riot duty, and other activities more akin to police work.

For example, traditional MOUT training has only recently—if at all—included any of the preceding tasks in the context of those activities doctrinally associated with OOTW, such as nation assistance, combating terrorism, peacekeeping operations, peace enforcement, support to domestic civil authorities, or counterinsurgency. Nor has that training typically been linked to noncombat missions or oriented to close cooperation with the variety of civilian government agencies or nongovernment humanitarian and relief organizations likely to be present in the OOTW environment.

Also, given the relatively small size of most typical OOTW deployments, units tasked with these missions (a light infantry unit, for example) may find themselves performing tasks and executing responsibilities that would otherwise fall to various special operations forces (such as Special Forces, Civil Affairs, and Psychological Operations personnel) or support units (such as Military Police) if sufficient personnel from such units are not available. Accordingly, given the dynamic operational environment inherent in most OOTW deployments, U.S. military forces will need to be trained to perform tasks across the OOTW operational continuum, including those usually performed by special or other support units.

Greater attention in both doctrine and training therefore needs to be paid to MOUT operations in the OOTW operational environment. Account must be taken of operations in densely populated, built-up areas; restrictive and specially imposed ROE (rules of engagement); responsibilities more akin to police than to traditional military operations; and work relationships with civilian government agencies and nongovernment humanitarian assistance and relief organizations. Moreover, while considerable attention is already given to training American light infantry in some specialized tasks such as MOUT, disaster relief, and noncombatant evacuation operations, the scope should be still
broader. Light infantrymen in OOTW environments will be working with or taking on the responsibilities of Special Forces, Civil Affairs, Psychological Operations, and Military Police units, and should receive the requisite cross training or joint training. Finally, units designated for, or likely to be tasked with, OOTW need to be ready to make the best use of limited assets and specialized equipment in dynamic conflictual environments and also to function effectively without such assets and equipment if need be.

Although urban insurgent activities are usually restricted to sabotage, sniping, barricades, roadblocks, and terrorist or criminal acts, such activities cannot be easily checked, since counterinsurgent options in cities are severely circumscribed. In most cases it would be counterproductive to use massive firepower, indirect fire, orairpower in urban environments, given the difficulty of avoiding civilian casualties. Indeed, insurgents can deliberately conceal themselves within the population, trading the concealment of the jungle for the anonymity of crowds, emerging only to attack.

Given that urbanization and population growth in most developing countries are creating conditions that can lead to internal conflicts, U.S. forces likely to become involved (i.e., light infantry, SOF, MPs, etc.) should be schooled in the political and military requirements of successful OOTW. They should be prepared to respond with appropriate measures in both the cities and the countryside. This will require a diversity of capabilities, as well as an appreciation for the causes of internal conflict and the political requirements for resolving it.

Urban OOTW requires a unique combination of the doctrine, training, and equipment appropriate for military operations on urban terrain, counterterrorism, and traditional counterinsurgency operations. On the one hand, troops must be prepared for the kind of urban offensive that took place in San Salvador in 1989, which required conventional combat skills tempered by stringent rules of engagement. On the other hand, effective use of intelligence, civic action, psychological operations (PSYOP), and population protection could perhaps prevent urban insurgency from developing to the point it did in San Salvador. Such activities must be supported by appropriate doctrine, training, and command-and-control arrangements, and must be undertaken in support of a broader political and legislative effort.

As they learn the specialized skills required by each environment (urban and rural), forces must also be trained to treat the civilian population with respect, so as to prevent popular alienation and to improve the conditions for gathering human intelligence. They may also have to coordinate police and military responsibilities with the civilian government’s political countermeasures, if any, as well as efforts by other agencies.

---

These general observations concerning future requirements for American armed forces, given the massive population flows and urbanization unfolding throughout the world, are brought into sharp focus by the four case studies that follow. These case studies of past U.S. military OOTW illuminate through specific examples the critical operational requirements already created by these demographic changes. They also shed light on the range of capabilities and further adjustments in doctrine, training and equipment that the U.S. military will need to perform well in a highly fluid future operational environment.

THE CHANGING NATURE OF MILITARY MISSIONS: INVITATION VERSUS INTERVENTION

As discussed above, conflicts such as that in Somalia—characterized by humanitarian components, urban and rural combat, refugee movements, fractionated social structures, dilapidated or destroyed infrastructures, and fragile or nonexistent governments—are increasingly likely in the developing world, where exponential population growth, unprecedented rates of urbanization, and massive migration flows combine to create a volatile situation where governments cannot meet the needs of their people. Even when conflict does not arise, the marginal survival of a country’s populace in such circumstances can easily be threatened by natural disaster, unexpected immigration flows, or even shifts in the international economy.

Where and when the United States will provide assistance or intervene is difficult to predict. U.S. interests in pursuing the drug war, for example, have brought U.S. military forces into Colombia, where they are currently involved in humanitarian assistance efforts. Concerns about migration flows led to the foreshortened deployment of U.S. armed forces to Haiti in 1993. Relations with Europe precipitated sending U.S. troops to Macedonia as peacekeepers. And U.S. soldiers remain part of the multinational force (MNF) in the Sinai. Although the United States will probably proceed cautiously before involving itself in further conflicts or humanitarian crises in the developing world, the recently released Draft Presidential Directive does not preclude deploying U.S. forces for future unilateral or coalitional operations other than war.

The U.S. military must therefore be prepared for any such future operations other than war. Even though U.S. military doctrine, training, and equipment for these operations remain nascent, their development slowed by greater attention to (and funding for) more traditional conventional military requirements, lessons from similar operations in the past can help guide future efforts.

In the following pages, we will examine the lessons learned from the failed U.S. peacekeeping operation in Lebanon, the humanitarian assistance efforts in
Somalia, the direct intervention in Panama in 1989, and the disaster relief operation in Bangladesh. Lessons gleaned from these various OOTW can serve as guideposts in developing doctrine, training, and equipment for future efforts.

Peacekeeping and Mission Shifts: Lebanon and Somalia

Although some would argue that U.S. peacekeeping in Lebanon failed and U.S. humanitarian assistance in Somalia succeeded, the two efforts yield similar lessons for future operations. They were both multinational efforts; both evolved, respectively, from simple, well-meaning peacekeeping and humanitarian assistance to something more aggressive, costly, and far less neutral; and both took place in complex political environments riven with territorial claims and power struggles between longstanding factional rivals. Moreover, in both Beirut and Mogadishu, urbanization, population growth, and population migration compounded, and to some extent caused, the conflicts that arose and clearly affected the nature of the conflicts as they unfolded.

In Lebanon, for example, the conflict was based in and around Beirut, where Palestine Liberation Organization (PLO) fighters had set up operations specifically to take advantage of a more defensible—and hence offensively frustrating—urban environment. The migration of Palestinians into Lebanon and the establishment of refugee camps were causes of concern for both the Lebanese government and the Israelis, who worried that the PLO could operate freely out of the camps. The peacekeeping forces in Lebanon functioned in an almost entirely urban environment, dealt daily with issues related to the refugee camps and the Palestinians' migration, and thus found themselves in the middle of a volatile, densely populated operational environment.

Similarly in Somalia, cities became the hub of the confrontation between the country's rival warlords. Relief supplies were blocked and militiamen from the various contending factions controlled the streets from their "technical" (trucks and jeeps mounted with machine guns). Thus, when the U.S. Marines first arrived in Somalia in December 1992, they had to stabilize the urban areas first before moving into the countryside. Throughout the American and UN operations, tensions and violence remained most prevalent in the cities, especially Mogadishu. Additionally, the peacekeepers conducted operations in heavily populated areas and helped control the flow of Somali refugees out of country, and they were involved in repatriating Somalis back into the country from neighboring states.
Lebanon. U.S. military forces were deployed to Beirut alongside French, Italian, and, eventually, British troops between 1982 and 1984 as part of the First and Second Multinational Forces (MNF 1 and MNF 2). MNF 1 was a peacekeeping force, intended on one hand to prevent the Israeli Defense Force (IDF) and Christian Lebanese militia from further attacks against Palestinians in Lebanon, while on the other to supervise the withdrawal of the PLO from Beirut to Syria by sea and land. These efforts were meant to satisfy the competing demands of the Israelis, the Palestinians, the predominantly Christian Maronite government in Lebanon, Moslem militias, and the Syrians. Although MNF 1 apparently succeeded in its efforts—nearly two weeks ahead of its own four-week deadline—tensions quickly resurfaced upon the coalition’s withdrawal. Syria allegedly sanctioned the September 14, 1982, assassination of Lebanese President Bashir Gemayel; the IDF entered West Beirut where Palestinian and other Moslem civilians lived; and the Christian Phalange militia, with alleged Israeli complicity and on the premise of ousting hidden PLO guerrillas, entered the Palestinian refugee camps in Sabra and Shatila, massacring hundreds of noncombatants.

The MNF coalition members hastily returned to Lebanon as MNF 2, with more and heavier military forces, although they were far outnumbered by the Sunni, Shiite, and Druze militiamen in the capital, and the Syrian and Israeli armies stationed around Beirut. MNF 2’s mission remained peacekeeping. It established a presence and interposed itself between the warring parties. Yet the peacekeeping efforts were never completely neutral: the Lebanese Armed Forces (LAF) were involved with every aspect of the coalition effort, while contact with the Moslem militias and other contending factions involvement was far less institutionalized.

As the security situation in Lebanon continued to deteriorate, the MNF 2 coalition itself splintered. The Americans and the French committed themselves to support LAF attempts to restore and maintain order, thus unwittingly becoming partisan players in the conflict. Meanwhile, their relations with the Israelis and the Moslems continued to worsen. By contrast, the British and the Italians strove to remain more evenhanded in their dealings with the government and the Moslem militias, remaining effectively neutral and avoiding the complications and violence that were to follow.

Not surprisingly, the U.S. and French troops became inextricably identified with the Lebanese government, and were targeted in a series of deadly attacks by Moslem factions, beginning with the April 18, 1983, bombing of the American Embassy—which claimed 17 American and more than 40 Lebanese

---

9This section is in large part derived from research conducted by John Schmeidel, a RAND consultant.
lives—and culminating with the October 23, 1983, bombings of the French MNF encampment and American Marine barracks, which, in total, left three hundred dead. Four months later, the British, Italian, and American members of the MNF withdrew their forces from Lebanon. A month later, the French followed suit.

A number of lessons can be gleaned from MNF 2’s experiences in Lebanon. The three most important issues relate specifically to multinational peacekeeping: neutrality among peacekeepers, the clarity and practicability of the mission, and the nature of the coalition arrangement.

MNF 2’s rapid establishment precluded careful prior arrangements with the various parties to avoid the semblance or substance of bias: when the coalition included the LAF in the peacekeeping efforts, it relinquished neutrality from the outset. Subsequent American and French decisions to support the Maronite government and military only exacerbated an existing problem. Without neutrality, the peacekeepers lost all legitimacy, and simply became participants in the conflict.

MNF 2’s mission, too, was unclear. Even the force structure demonstrated the general confusion about MNF 2’s objectives: although the heavy forces sent into the country were too aggressively oriented for peacekeeping, they were numerically insufficient to make a viable show of force against the welter of militia factions and massed troops either within or surrounding the city. Once the forces were in the country, MNF 2’s mission changed quickly, but without official acknowledgment, from peacekeeping to a combination of counterinsurgency, nation assistance, and security assistance, with the coalition forces acting in support of the Christian government and armed forces. Indeed, U.S. Marine and Army units actually trained elite LAF units into a quick-reaction force. This shift in mission left U.S. forces in a precarious position: prepared and equipped for peacekeeping, they nonetheless became participants in the conflict, with all the attendant risks and none of the requisite planning or preparations.

On a more positive note, the multinational arrangement—in which each national contingent operated in parallel, within discrete and contiguous zones of responsibility, and with informal liaison between the forces—worked remarkably well. Though MNF 2’s contingents did not always provide a united front, their efforts were not plagued by the language difficulties, political considerations, and logistical confusion that have affected similar UN operations.

Other, more general, tactical and operational lessons emerged from MNF 2, applicable to future OOTW activities. U.S. command and control (C2), for example, was deeply stratified, leading to slow decisionmaking. In contrast, the British, French, and Italians established direct, real-time radio links with their cabinets and defense ministries back home, substantially speeding the decisionmaking process.
Another lesson involved the rules of engagement (ROE) followed by U.S. forces in MNF 2, which failed to evolve, much less keep pace, with the growing threat in Lebanon, leaving the Marines in Beirut insufficiently armed or ready for the hostilities that eventually engulfed them.

Also, the Marine Amphibious Units that provided troops for the U.S. contingent in MNF 2 did not have enough intelligence personnel, including translators. Moreover, the information gathered was not efficiently or effectively shared between the various U.S. agencies, nor passed in a timely manner to the tactical commanders on the ground in Beirut.

Finally, U.S. forces in Lebanon could have benefited from adopting some proven British tactics for peacetime operations on urban terrain: cooperating directly with civilian police and civil administrators in urban combat areas; deliberately using patrolling as a means of gathering intelligence by observation and contacts; and educating soldiers in ground-level negotiation skills and sensitizing them to the civilian population they would be interacting with.

Each of these operational and tactical lessons has applicability beyond multinational operations or peacekeeping. Future U.S. involvement in OOTW activities would therefore benefit from a close examination of the C3I, force structure, training, ROE, and intelligence efforts in MNF 2.

Somalia. U.S. involvement in Somalia began in 1992 as a strictly humanitarian effort. Operation Provide Relief (OPR) was a small, predominantly Air Force operation for lifting food and other assistance into the war-torn country from neighboring Kenya. While the United States conducted these aid lifts, the United Nations began its own peacekeeping efforts in Somalia (UNOSOM). Few countries signed on with the UN, although Pakistan sent 500 troops to serve as peacekeepers.

It soon became clear that the American and UN efforts were having little effect in ameliorating either the hunger or the conflict endemic to the country. Assistance packages were continually stolen, aid workers were harassed, and starvation in certain areas persisted. In December 1992, therefore, the United States launched a coalition effort under the operational control (OPCON) of the U.S.-led United Task Force (UNITAF). The U.S. component of the coalition effort was known as Operation Restore Hope (ORH). The mission of ORH, and UNITAF more generally, remained humanitarian, although with a larger security component than OPR had: secure ports and food distribution points, provide open and free passage of relief supplies, ensure security for relief convoys and operations, and otherwise assist the United Nations and nongovernment organizations (NGOs) in providing humanitarian relief under UN auspices (Freeman, Lampert, and Mims, 1993:64).

From the outset, a date was set for transferring control of the coalition forces from the United States to the UN, which would then consolidate, expand, and maintain security in Somalia to allow the process of humanitarian
assistance, economic assistance, and political reconciliation to proceed. On May 4, 1993, the transfer of control took place, and UNOSOM II superseded both UNOSOM I and UNITAF.

Although U.S. forces were supposed to withdraw following the transfer of control to UNOSOM II, a number of U.S. support units nonetheless remained in Somalia under UN operational control. Elements of the U.S. Army 10th Mountain Division (Light Infantry) also remained in the country, reorganized as a Quick Reaction Force (QRF) under the operational control of the U.S. Commander-in-Chief, Central Command (CENTCOM) (with a few narrowly defined exceptions, where they fell under the operational control of the UN Force Commander). These U.S. efforts during UNOSOM II are known as Operation Continue Hope.

Under the United Nations, the mission in Somalia continued to shift from humanitarian assistance to security. Tensions between the followers of Somali warlord Mohammed Farah Aideed and UN forces quickly increased, culminating in the June 5, 1993, ambush of a Pakistani convoy in Mogadishu, during which Aideed militiamen killed 24 Pakistani soldiers. Tensions continued unabated, and a month later the UN issued a call for the arrest of General Aideed. U.S. Rangers and other elite special forces arrived in Somalia in August and, on September 22, captured Osman Ato, Aideed’s chief financier. Less than two weeks later, on October 3, 18 Americans were killed and 84 wounded when an attempt to capture more of Aideed’s top officials went tragically awry.

Reevaluation of American policy in Somalia began immediately. Four days after the incident, President Clinton canceled the hunt for Aideed and his followers, ordered the deployment of 1700 additional Army troops to Somalia and the offshore emplacement of two amphibious groups with 3600 Marines, and set a March 31, 1994, deadline for the complete withdrawal of U.S. forces from Somalia.

The patchwork of military efforts in Somalia yields valuable lessons for future OOTW. In particular, the lessons relating specifically to multinational peacekeeping closely parallel and reinforce those from Lebanon. For example, problems arose in Somalia regarding neutrality in peacekeeping and the clarity and practicability of the mission. Furthermore, whereas coalition arrangement in Lebanon was relatively effective and practicable, the arrangements in Somalia were much more problematic, providing a useful contrast for analysis.

UNOSOM II lost all semblance of neutrality when the UN criminalized Aideed and called for his capture. At that point, the UN forces simply became another faction in the conflict, just as U.S. and French forces had in Lebanon ten years before. Even though they continued to maintain the pretense of

---

10 UNOSOM II briefing, 1993.
11 For more discussion of the failed raid, see Atkinson (1994a,b).
neutrality, the UN forces were now unquestionably a faction in the conflict and thus became perceived as legitimate targets for Aideed’s militia. Yet those forces had been deployed for only a limited security mission, and were neither prepared nor equipped for a more aggressive stance. As in Lebanon, the forces in Somalia thus found themselves in a dangerous and frustrating situation: their ROE remained limited, they were armed and equipped for peacekeeping, but the Somali militias clearly considered them opposition forces.

This dilemma was exacerbated by the changing mission over the course of operations in Somalia. However, although the mission was said to “creep”—and did to some extent during ORH at the operational level, when bored U.S. forces began to conduct limited nation assistance and civic action in nearby villages—the mission shift was in fact far more deliberate than that at the strategic level. The political intent to adjust the mission over time from strict humanitarian support to more security tasks and greater nation assistance activities existed from the moment the decision was made (prior to ORH) to conduct—under UNOSOM II—a broader security and nation assistance operation in the interests of political reconciliation and eventual Somali elections. The political intent, however, remained garbed in the language of peace operations, and was never translated into explicit military requirements or plans. Thus, as mentioned above, UNOSOM II forces were still ostensibly in Somalia to conduct peacekeeping operations, and were equipped, staffed, and prepared for that mission rather than the more aggressive one that emerged.

In addition, the coalition arrangements in Somalia actually exacerbated and were accentuated by the problems with neutrality and mission shift. Under both UNITAF and UNOSOM II, coalition forces yielded their operational control to a central command, led by the United States and United Nations, respectively. Although the forces cooperated effectively under UNITAF, their mission was far clearer and less controversial than it would be under UNOSOM II. During UNOSOM II, as operations became more dangerous and the mission more controversial, coalition forces increasingly began to turn to their national commands for guidance rather than to the UN, and they frequently refused or demanded revisions to UN orders. Lacking adequate contingency plans for such situations, planned UN operations thus often had to be canceled or postponed. Also, as the preferences and concerns of different national contingents surfaced in response to the mission changes, any semblance of a united UN front vanished.

Other operational and tactical problems surfaced during ORH and UNOSOM II. Communications problems between the contingents, owing to

---

incompatible equipment and language differences, slowed planning and made combined operations more difficult and dangerous. Widely divergent doctrine, training, and general practices frequently meant that combined operations had to be geared to the lowest capability rather than the highest. U.S. forces could not use tear gas in riot control, for example, until the United States donated gas masks to the UN for distribution to other coalition members in Somalia; also, although the United States uses air support in its ground operations, many of its coalition partners in Somalia were unfamiliar with such operations, limiting its utility and use. Coalition members also developed different relations with the Somalis, who learned to distinguish between contingents and respond accordingly. The June 5 attack, for instance, deliberately targeted the Pakistanis, one of the most aggressive national contingents.

U.S.-specific lessons emerged from Somalia as well: human intelligence (HUMINT) was insufficient prior to ORH and was difficult to cultivate given the dearth of U.S. military or intelligence personnel proficient in Somali languages or knowledgeable about the Somali culture. The rapid development of links with Somali scholars in the United States and Somalia would have been helpful. Maps, too, were outdated and depicted Somalia’s vast expanses at useless scales for urban combat. Too few special operations forces, especially Civil Affairs and Psychological Operations personnel, were sent to Somalia. Uniformed troops therefore had to take up the slack and perform civil affairs duties. There were also far too few liaison officers to interact with the various nongovernment organizations, local civilians, and other national contingents in Somalia, so uniformed personnel were tasked with that responsibility as well. Had hostilities broken out under UNITAF, there clearly would not have been sufficient surplus personnel to perform these tasks, as the fire support and other support personnel temporarily assigned to such duties would immediately have had to resume their primary responsibilities. In the future, accordingly, similar operations will need to be staffed with greater numbers of liaison and SOF personnel. Finally, the U.S. military would have benefited from better and more standardized rules for command and control of U.S. forces in coalition activities. Had the lines of command and control been clearer in Somalia, for example, the October 3 debacle with the Rangers might have been avoided. As it was, because separate chains of command and control had been established for the Rangers and the QRF, the QRF was not aware it might be needed to support the Rangers’ operation and was therefore unavailable when the contingency arose.\textsuperscript{13}

Again, as in the case of Lebanon, the practical lessons learned in Somalia have relevance beyond multinational peacekeeping operations to a range of fu-

\textsuperscript{13}From interviews with 10th Mountain Division (Light Infantry) personnel deployed home from Somalia, Fort Drum, August 1993.
tural OOTW activity. Where the lessons are redundant, their validity is reinforced; valuable lessons can also be drawn, however, when different approaches to the same objective yielded different results, as in the case of C^2 or coalitional arrangements.

**Intervention: Operation Just Cause in Panama**

The U.S. operation in Panama provides a useful comparison to the operations in Lebanon and Somalia, insofar as the military objectives were clearly established from the outset and (despite rhetoric to the contrary) closely matched the Bush administration’s political goals. The mission was clear, and success would be easily recognizable. The emphasis in the operation was on a traditional U.S. military strength: firepower. And although some operations other than war were planned for the postcombat phase of the operation, they were not emphasized in either planning or preparation. Nonetheless, the conflict in Panama shared characteristics with the conflicts in Lebanon and Somalia insofar as the operations conducted there were almost entirely urban and undertaken in heavily populated areas. Indeed, because the U.S. objectives included decapitating the Panamanian Defense Force (PDF), capturing Panamanian leader Manuel Noriega, and preserving the country’s infrastructure, operations were necessarily conducted on urban terrain, where the PDF was based and where most of the country’s key infrastructure is located. Issues of refugee flow and presence were also important in Panama City, where the U.S. forces had to control population movement and limit civilian casualties. Thus, in Panama, as in Lebanon and Somalia, population growth, migration, and urbanization were each critical factors in shaping the nature of the conflict.

The U.S. intervention in Panama, Operation Just Cause (OJC), followed more than a year of increasing animosity between the United States and Noriega. When U.S. forces finally invaded the small country, their mission was to protect American lives, assist the democratically elected government that Noriega had denied power to, seize and arrest Noriega (who had been indicted in the United States for drug trafficking), and defend the integrity of U.S. rights under the Panama Canal treaties. The United States achieved all these objectives: the PDF was dismantled after the attack and neutralized; during the operation, U.S. forces swore in the Endara government, whose victory in the May 1989 election had been annulled by Noriega; Noriega himself was eventually caught, brought to the United States, tried, and imprisoned; and

---

U.S. concerns about control over the Panama Canal reverting in the year 2000 to an unsavory Panamanian government were arguably allayed.

Operation Just Cause was strictly a unilateral intervention to pursue American political and strategic objectives, and it had only a limited pretense of humanitarian objectives. Yet the operation could almost be considered fortuitous, in that it afforded the United States the opportunity to conduct OOTW under extremely advantageous circumstances. The errors and miscues that did take place were to be expected: OJC represented a significant departure from both the conventional battlefield warfare that U.S. forces have trained for since the end of World War II and the unconventional jungle operations of the Vietnam War. Indeed, it was most similar to the U.S. stability and peace operations in the Dominican Republic (1965–1966) and Grenada (1983).

OJC was nonetheless unique: planning and operations were fully integrated across all four services; much of the operation was conducted on urban terrain; a large number (the largest number since Vietnam, but surpassed during Desert Shield) of U.S.-based forces were rapidly deployed; special operations forces played a highly visible and critical role in the operation; rules of engagement were uncommonly restrictive; soldiers were expected to apply minimum use of force; indirect fire and aerial bombing were limited; surgical strikes were necessary; and the preservation and defense of infrastructure and public utilities were key objectives.

These characteristics are also common to recent OOTW in Somalia, Bosnia, Iraq (among the Kurds), and Bangladesh. At the same time, however, OJC was also simpler than subsequent military operations other than war in Kuwait, Iraq, Bangladesh, Bosnia, and Somalia. Because it was a unilateral effort, no coalition issues or problems complicated or slowed U.S. operations. The communications, logistics, planning, and command-and-control issues that arise in multinational operations never surfaced during OJC. Nor did the United States have to coordinate its efforts with nongovernment organizations (NGOs) or humanitarian relief organizations (HROs). As recent events attest, future U.S. OOTW operations are unlikely to be unilateral, and as much as they will benefit from both coalition and NGO/HRO support, Desert Storm and Somalia demonstrate that U.S. forces will also have to adjust training and doctrine to accommodate such cooperation.

Moreover, in Operation Just Cause, U.S. armed forces outnumbered the PDF by more than two to one. The Panamanian military was also a known quantity to its American counterparts, who had trained it and were consequently well versed in its doctrine, skills, and capabilities. Nor did the U.S. military face angry crowds, violent uprisings, or even passive popular resistance: the people of Panama welcomed the Americans and provided little, if any, support to the PDF. Communication with both the public and the PDF was not
a problem, because many members of the U.S. military speak Spanish as a first or second language. Finally, tension between the United States and Panama had escalated for more than a year, allowing sufficient time for planning and practicing an operation such as Just Cause. Each of these factors contributed to the ease and speed with which the PDF was defeated and U.S. military objectives met.

Despite all these unique advantages, OJC offers a number of wider, practical lessons for application in current and future OOTW. One very good example involves postcombat stability operations and the role of civilian agencies in OOTW. Planning was conducted separately for conventional offensive operations, special operations, and stability and civil-military operations. This led to problems during OJC as it transitioned from the combat phase to posthostilities stability operations, when insufficient numbers of Civil Affairs personnel were available to assume responsibility in the operation’s wake. There was also very little civilian input into planning the stability phase of OJC. Yet because military planners were not always aware of what civilian agencies could realistically be expected to step in and do (especially with little advance notice), both military personnel and civilians were overwhelmed with actual postcombat requirements. Coordination and cooperation between the military and civilian agencies requires improvement for OOTW in which civilians may be expected to assume responsibility for such diverse tasks as assisting with the demobilization of militaries, reestablishing or building infrastructure, setting up or reconfiguring judiciaries or electoral systems, and assisting in the procurement and distribution of humanitarian aid.

OJC also demonstrated that U.S. military training in military operations on urban terrain was inadequate and that more units needed to include MOUT in their mission essential task lists (METLs). Intelligence also was a problem: electronic intelligence, though essential in conventional battlefield warfare, is often irrelevant in OOTW, and must be supplemented by human intelligence. The military and CIA must therefore develop arrangements allowing greater military access to the civilian intelligence community’s HUMINT. Efforts to streamline intelligence in joint operations also must not overlook unit-specific needs and therefore should take particular care to maximize use of special operations forces by employing them in the specialized tasks for which they are trained. Equipment was another issue, as OJC clearly demonstrated the special requirements of MOUT. Advancements in technology applicable to OOTW have been made since 1989, but the military’s priority in research and development remains conventional weaponry and materiel.

Perhaps the most interesting thing about OJC in the context of OOTW, however, is that it demonstrates the value of a clear and sustainable mission (especially in contrast to operations in Lebanon and Somalia), while at the same time illustrating precisely the shortfalls in the U.S. military’s ability to
plan or prepare for operations other than war. On the one hand, Operation Just Cause was an unqualified success: the planning, C², force structure, equipment, and logistics were ideal for the operation, and all of the mission objectives were met. On the other hand, planning for the postcombat phase was deemphasized and underfunded, with no civilian input, creating problems in the transition from combat to stability operations and only partial success in reconstruction and stability efforts. Clearly, this has implications for operations in fluid environments, where the need to adapt U.S. efforts swiftly to changes from combat to noncombat (or vice versa) might be required.

Disaster Relief and Humanitarian Assistance:
Operation Sea Angel in Bangladesh

To complete the circle of comparisons, a case study of a pure and simple noncombat effort—in this case, disaster relief—is extremely useful. At one end of the scale is OJC: a straightforward conventional military intervention, with some characteristics of OOTW. In the middle lies MNF 2 and the operations in Somalia: fluid operations with shifting missions, each of which began as a purely humanitarian effort but—in the face of conflict and changing political goals—became much more aggressive and complicated operations. And finally, Operation Sea Angel (OSA) in Bangladesh lies at the other end: a straightforward humanitarian OOTW in a benign and stable environment. Again, however, the demographic trends that shaped the crises in Lebanon, Somalia, and Panama also affected OSA. Indeed, the government’s inability to respond adequately in the aftermath of a natural disaster was due in large part to the country’s limited resources and relatively huge population.

When Cyclone Marian struck Bangladesh on April 29–30, 1991, much of the country’s infrastructure was destroyed or damaged. Transportation became nearly impossible, ships sank and blocked the port, water supplies were disrupted, millions were left homeless, over a million cattle died, and crops on tens of thousands of acres were completely wiped out. The fledgling civilian government was young, inexperienced, poor, and yet had to avoid any appearance of weakness or incompetence.

The United States responded within 10 days with practical emergency assistance and short-term recovery operations. A U.S. Contingency Joint Task Force (CJTF) was established to run OSA. The bulk of the U.S. force in Bangladesh was from Amphibious Group 3 and the 5th Marine Expeditionary Brigade (5th MEB), supplemented by Army and Air Force elements (in

---

15For a more detailed discussion of OSA, see McCarthy (1994:4–6), from which this section is heavily drawn.
particular, a special operations forces Damage Assistance Relief Team—DART—from Okinawa).

The operation was not unilateral: British, French, Japanese, and Pakistani units acceded their own operational control to the CJTF, and Indian and Chinese forces, while not officially OPCON to the CJTF, nonetheless worked closely with the task force to deliver aid. The CJTF also cooperated effectively with the nongovernment organizations in Bangladesh, providing them with transportation and communication assets and acting as a liaison between them and the government of Bangladesh (with which they had longstanding poor relations). The NGOs, in return, helped identify the regions in greatest need, procured supplies, ran reconstruction programs, and generally operated efficiently and competently in providing a full range of assistance.

CJTF’s mission in Bangladesh was to provide command, control, and coordination in support of humanitarian assistance efforts in the country. U.S. forces helped distribute food, water, and medicine, conducted area assessments, established secure communications between outlying areas, and offered limited medical assistance.

Operation Sea Angel is unique among the cases discussed in this paper, insofar as it began and in fact concluded as a purely humanitarian mission. Yet in other aspects, it is very similar to other OOTW: it was a joint U.S. operation with heavy SOF involvement; it was a multinational effort; and nongovernment organizations played a large role in the disaster relief effort.

Nonetheless, the differences, not the similarities, between OSA and other OOTW determined the success of the operation. Three factors in particular warrant discussion: the disaster relief efforts took place in a benign environment; the government of Bangladesh was intentionally left in charge, and transition of the operation to the government began three weeks after the start of OSA, exactly as planned; and neither mission creep nor mission shift took place.

Perhaps the most important factor in the operation’s quick success was the absence of conflict. U.S. forces were able to work with the Bangladesh government, rather than trying to function in either a vacuum or an anarchical situation like Somalia’s. Accordingly, they faced no threat, and did not have to worry about force protection, combat, or rules of engagement. The populace was supportive of the U.S. efforts and extremely cooperative. The nongovernment organizations, for the most part, were also supportive and cooperative—in direct contrast to the often problematical military-NGO relations in Somalia, where NGOs were critical of the military’s role and actions in the country. With no controversy or concern about threats or mission objectives, national contingents were readily prepared to accede their operational control to the U.S. task force, thus simplifying command and control and coordination. Finally, the U.S. forces were in Bangladesh simply to clean up after a
natural disaster; they were not involved at all in the politics of the country, and their limited disaster-relief activities were unlikely to spark political controversy.

A second and perhaps equally important factor in the success of OSA was that U.S. forces in Bangladesh very deliberately did not exceed their mandate, difficult as it sometimes was for troops to limit their assistance in the face of the country's overwhelming poverty and need. Offshore basing, for example, allowed a minimum U.S. "presence" in the country, thus helping the government to retain relatively greater visibility in the relief operations. Restoration of the country's infrastructure was not part of the mission, nor was nation assistance or civic action; these were left to the Bangladesh government and the NGOs. Where construction or reconstruction took place, things were built back to previous standards, not to U.S. standards. And throughout the operation, Lieutenant General Stackpole, the CJTF commander, made a point of not overstating what U.S. forces would do, so as to maintain realistic expectations among the Bengalis, who would have to assume full responsibility for operations once the Americans withdrew. OSA is thus a positive example of how careful definition of, and adherence to, a mission can facilitate an operation. U.S. forces could easily have become embroiled in trying to provide nation assistance to Bangladesh, but they carefully limited their involvement to disaster relief from the outset.

Other lessons for application in future OOTW can be drawn from Operation Sea Angel. For example, the concept of a two-tiered command-and-control system was implemented during the operation for the first time, and worked extremely well. Under the two-tiered system, built around existing commanders and their staff, the chain of command (especially at the higher levels) was clearly defined, and Lieutenant General Stackpole was able to add or delete units fairly rapidly. Coordination and cooperation between the military and local government officials and NGOs was also a critical factor in the success of the operation and depended, in part, on the military's self-imposed limitations. Each agency worked to its comparative strength, and the military, far from interfering, coordinated its own activities to complement those of the other agencies.

Of course, OSA had its problems. Resupply lines were established late, and many parts requested simply never arrived in the country. Some of the units involved in the operation did not have standard operating procedures (SOPs) for disaster relief/humanitarian assistance and therefore failed to anticipate some logistical and communications problems. Participating aviation units, for example, did not develop load plans and lists of equipment and personnel until

16 Sea basing also helped U.S. forces avoid problems related to health risks, lack of infrastructure, and force protection.
the day the unit received its alert notification. Intelligence and maps, as in other OOTW, were insufficient and inadequate in OSA. Finally, communications were a significant problem throughout the operation, with links required between CINCPAC (Hawaii), ships anchored offshore, two major regional airfields, the Main and Forward task force headquarters, helicopter landing zone teams, NGOs, and the DARTs. Satellite and SOF communications networks were helpful, but they quickly became overloaded. The situation improved only after the 4th Combat Communications Group arrived on May 16. Even then, air-to-ground communications remained problematic, because Marine aircraft did not have a VHF (AM) radio, which many foreign countries, including Bangladesh, use for aircraft control. Nor was all USMC and Navy communications equipment compatible, further exacerbating the problem.

Summary

Each of the brief case studies above—representing peace operations, direct intervention, and disaster relief—demonstrates the U.S. military's lack of familiarity with OOTW in general and OOTW requirements more specifically. The cases also provide practical tactical and operational lessons applicable to the range of OOTW missions. These lessons can help the U.S. military in its nascent efforts to adapt doctrine, training, equipping, and planning to take these requirements into account.

However, it is perhaps more instructive to note how these cases demonstrate the distinction between straightforward operations conducted in easily identifiable operational environments and more flexible operations conducted in fluid situations with the potential to shift from peace to conflict with little warning. Although OJC and OSA were respectively straightforward combat and noncombat operations, they both benefited from well-articulated political and military goals and objectives—which translated into clear and steadfast missions—that were carried out in well-defined and easily gauged operational environments: Panama was clearly conflictual and Bangladesh was clearly benign. In contrast, the operations in Lebanon and Somalia both began as humanitarian efforts, but the missions could not be as clearly defined in the absence of straightforward political objectives or stable operational environments. Moreover, political rhetoric was often inconsonant with the actual political

---

17It is true that the operational environment in Panama was not typically conventional. Yet, strict ROEs and concerns about collateral damage notwithstanding, planners knew that U.S. forces needed to be able to defend themselves and their mission, and could plan accordingly. In other words, planners knew the threats posed by the environment for which they had to equip and prepare the forces.
goals—but rhetoric nonetheless still often drives military planning. Planners,
accordingly, could not adequately select, prepare, or equip forces for opera-
tions. In both cases, forces found themselves in much more threatening situa-
tions than they anticipated.

This distinction has implications for every aspect of planning for future
U.S. military involvement in operations other than war, from C² to intelligence
efforts to force mix. Military planners already know how to prepare for either
end of the OOTW continuum—OJC and OSA—but they must nonetheless
reevaluate planning and preparation for less clear-cut efforts in the changeable
OOTW operational environment that is often neither strictly peace nor strictly
war.

CONCLUSION AND RECOMMENDATIONS

The United States is seeking to adapt to the "new world order" and to move
from a national strategy of containment to one of collective engagement (i.e.,
involving the world community in such transnational issues as the environ-
ment, drug enforcement, economic development, democratization, humanitar-
ian relief, and conflict resolution). The changing landscape of population dis-
tribution throughout the less-developed world and the likely urban cynosure of
future conflict in these regions may require changes in U.S. military training,
doctrine, and equipment.

The history of U.S. and multinational military efforts in peacekeeping,
peace enforcement, urban operations, disaster relief, humanitarian assistance,
humanitarian intervention, and urban warfare seems to show that although the
United States is capable of succeeding in some operations other than war, most
such operations suit neither its culture nor its military.

The United States seems best able to involve itself in operations other than
war when those operations have clearly defined goals and endpoints, can be
relatively quickly achieved, are likely to have limited loss of American soldiers' 
lives, and are not intended to resolve any longstanding social, political, or eco-
nomic issues. For example, the U.S. military has proven adept at disaster relief,
as Operation Sea Angel in Bangladesh has shown. Disaster relief requires no
combat or warfighting equipment, and it is carried out by light forces focused
on extensive communication and cooperation with local civilian officials and
humanitarian organizations. The United States has also proven itself very ca-
pable at direct intervention, the form of conventional warfighting that it is
most likely to undertake in the less-developed world in the post-Cold War pe-
riod. Direct intervention requires not just combat but massive firepower and
maximum force, extensive planning and coordination, a mix of heavy and light
forces, warfighting materiel, rapid deployment, signal and human intelligence, force protection, and a myriad of other requirements.

However, U.S. forces are not ideally prepared to operate in the gray area between these two extremes, i.e., more urban, less conventional conflict environments. If ever they are called upon to do so, their mission must be based on a solid understanding of the conflict environment and it must be clearly defined and delimited from the outset. Otherwise, U.S. forces may be sent to resolve a simple problem and become engaged in the treacherous causes underlying it. This was true in Beirut in 1983, and it was equally true in Somalia in 1992. Moreover, strictly defined and limited missions are of great importance given the U.S. public’s reluctance to become militarily involved in operations other than war unless they directly affect U.S. interests, pose little or no threat to U.S. soldiers, are of short duration, and are relatively easily achieved. Peace enforcement, peacekeeping, humanitarian intervention, and urban counterinsurgency are thus questionable missions for U.S. forces.

Faced with similar dilemmas in the past, the United Nations learned how to sidestep overwhelming, longstanding problems while helping resolve conflicts. Peacekeeping evolved precisely because it allowed this kind of flexibility. UN forces did not have to engage combatants to quell them—indeed, they avoided such direct involvement. Yet this solution requires that peacekeepers accept long-term, perhaps indefinite, operations (such as on Cyprus since 1964), which is an unpalatable option for U.S. forces. The one time UN peacekeepers did become actively involved in the conflict around them (the 1960 operation in the Congo), they were immediately considered a faction and their mission ended disastrously—much like the U.S. mission in Beirut 20 years later.

Although both the U.S. public and the U.S. military are reluctant to participate in OOTW, such involvement may be unavoidable. The demographic pressures in the developing world will inevitably lead to conflicts, some of which can threaten U.S. interests. Although the United States may avoid future humanitarian interventions, other concerns may well motivate U.S. involvement. Massive population flows into the United States from beleaguered countries, conflict within or between countries with nuclear weapons or other weapons of mass destruction, or increased control of a country by drug traffickers could be sufficient cause for U.S. participation in urban, unconventional conflict.
BIBLIOGRAPHY


Camp, Sharon (1990), Cities: Life in the World's 100 Largest Metropolitan Areas, Population Crisis Committee, Washington, D.C.


Dewar, Colonel Michael (1992), War in the Streets: The Story of Urban Combat from Calais to Kheiji, David & Charles, Newton Abbot, Devon. Distributed by Defense Technical Information Center, Cameron Station, Alexandria, VA.


Hoffman, Bruce, and Jennifer Morrison Taw (1991), Defense Policy and Low-Intensity Conflict: The Development of Britain's "Small Wars" Doctrine During the 1950s, RAND, Santa Monica, CA.

——— (1992), A Strategic Framework for Countering Terrorism and Insurgency, RAND, Santa Monica, CA.

Hoffman, Bruce, Jennifer M. Taw, and David Arnold (1991), Lessons for Contemporary Counterinsurgencies: The Rhodesian Experience, RAND, Santa Monica, CA.

Latimer, John C. (1985), Considerations for Operations on Urban Terrain by Light Forces, Command and General Staff College, Fort Leavenworth, KS. Distributed by Defense Technical Information Center, Cameron Station, Alexandria, VA.


This paper discusses how the United States may wish to approach programs for ballistic missile defense during the next ten years. It starts by summarizing the history of strategic defense, particularly ballistic missile defense (BMD). It then turns to the changing international situation, with special emphasis on the proliferation of weapons of mass destruction and ballistic missile systems. These threats are cause for serious concern—for U.S. troops that might be employed overseas and for friendly nations neighbored by threatening countries. The paper next discusses options for theater ballistic missile defense (TMD). It emphasizes the importance of pressing forward with development of such defenses as well as their limits. The paper then turns to defense of the United States and what would be required to protect the country against a variety of threats such as unauthorized attacks from Russia or other major nuclear states, or from small numbers of ballistic missiles launched from Third World countries. In contemplating national (and global) missile defenses as something that might eventually be deployed cooperatively, the paper considers implications for nuclear stability in the context of START treaties. Finally, the paper expands the discussion and addresses the central underlying issue: What role should ballistic missile defenses have in our overall national security strategy? Because the answer appears to be that “it depends on how the world develops,” the paper ends by suggesting a series of programmatic and other actions that would establish TMD-related priorities in the near term, but provide important hedges for the mid and long term that would permit future national and global missile defenses.

INTRODUCTION

Early in the Clinton administration, as part of reductions in DoD’s budget, it was decided to reduce the funding available for ballistic missile defense (BMD) from the planned expenditure of $39 billion over a five-year period to $18 billion. This very sharp reduction has important implications for the future of BMD, raising questions about the appropriate balance across the various goals that Congress set for BMD in the Missile Defense Act of 1991. The

The MDA mandated a program that would (1) assure early deployment of a limited national military defense that would be highly effective but consistent with the
question of priorities was addressed in Secretary Aspin's Bottom-Up Review (BUR). While the BUR essentially codified the priorities already in place, placing first priority on theater missile defense (TMD), with national missile defense (NMD) a distant second, the funding tied to these priorities was far from balanced. The result was a TMD budget still two-thirds of that planned by the Bush administration, but a budget for NMD only one-fifth of its previously planned amount. Other activities, e.g., various technology efforts, were similarly affected.

Ironically, this shift in program balance and funding levels comes after a decade of prodigious BMD-related research and development focused primarily on NMD, i.e., on protection of the United States. This decade of research appears to have placed the United States in the position of having the technical capability to field an effective defense against sizable, albeit limited, ballistic missile attacks. While the shift toward focusing the research on TMD began several years before the Clinton administration took office, and was well in place in January 1993, the effort on NMD was at that time still substantial, with planned initial deployment dates within about a decade. At BUR support levels, it is no longer clear when an initial limited NMD might be deployed.

The focus on TMD is certainly justified by current international events. And it is difficult to argue that budget priorities for DoD don't demand sharply reduced budgets for BMD in general. Is the issue therefore resolved? A central thesis of this paper is that interest in (and even demand for) NMD may well increase in the future as the result of events that seem very hypothetical today. It follows then that DoD's BMD programs should strike a balance between focusing budgetary support on obvious near-term priorities (TMD rather than NMD) while maintaining and creating executable options for national-level defenses when their need becomes obvious. This paper seeks to explain what analysis can say about how to proceed.

Let me begin by defining and distinguishing more carefully among three types of ballistic missile defense:

ABM treaty (a difficult feat); (2) maintain strategic stability; (3) develop a highly effective theater missile defense (TMD) system as quickly as possible to protect U.S. forward-deployed forces and U.S. friends and allies; and (4) give space-based defenses (notably including the Brilliant Pebbles concept) low priority (DoD, 1992).

2Not all of the skeptics, the author included, believe that BMD technology has reached the point where defense can handle sophisticated, albeit small, attacks.

3The material discussed in this paper comes from a variety of sources. I greatly appreciate the cooperation and support of Mike Miller, Susan Everingham, Herb Hoover, and Richard Mesic. And I owe a special debt of gratitude to Ambassador Henry Cooper for his review. While I am solely responsible for the contents of this paper, many of the good ideas and original analysis came from them, and I am in their debt.
• *Theater missile defense* (TMD) seeks to defend against short- and intermediate-range missile threats such as those that may be faced by U.S. projection forces and by friendly nations. The options for TMD consist presently of THAAD (an advanced high-altitude terminal-defense system) and an enhanced version of Patriot, which will act as an augmentation to and underlay of THAAD. One of the current debates is the extent to which activities focused almost exclusively on target-area defense should be augmented by boost-phase defenses (seeking to kill theater ballistic missiles (TBMs) while they are in powered flight) or counterforce attacks (seeking to kill TBMs or their transporter/erector/launchers (TELS) on the ground before launch).

• *National missile defense* (NMD) seeks to defend the United States against missiles of all ranges, including intercontinental. The current system concept for NMD involves one or more ground-based radar (GBR) sites, one or more ground-based interceptor (GBI) sites, and other long-range surveillance sensors (including the possibility of new acquisition radars at current BMEWS sites). The GBI are nonnuclear hit-to-kill missiles that engage targets before reentry into the atmosphere.

• *Global missile defense* (GMD) seeks to provide a comprehensive ballistic missile defense, not only for the United States, but also for its allies and selected other countries as appropriate. The GMD initial-system concept currently consists of hit-to-kill space-based interceptors (Brilliant Pebbles or BP), aided by space-based sensors. Later variants might include space-based beam weapons (e.g., lasers) and/or designators to aid midcourse tracking and discrimination as well as to affect kills. Global missile defense would be an overlay to theater- and national-level missile defense systems.

The paper discusses all of these, as well as how the United States may wish to build programs to develop and, in some cases, to deploy them. The paper proceeds as follows. It provides historical background and then discusses the many implications for BMD of the changed security environment, especially the need to defend U.S. projection forces and allied nations in various regions. The paper then discusses TMD, NMD, and space-based missile defense (SBMD) issues in successive sections, touching on issues of feasibility, need, and other matters. After this survey, the paper reviews some of the most trou-

---

4 TMD should concern itself with cruise missile threats as well as ballistic missile threats, but attention has centered so far on the latter. I will discuss the former only briefly in this paper, despite its importance.

5 Another system, CORPSAM, would augment TMD with a low-altitude defense layer, but the CORPSAM program has been deferred (Aspin, 1993).

6 GMD usually is a surrogate for space-based BMD. We will use it in this context.

7 Space-based sensors can support all defense deployments and should not be categorized as part of GMD.
blesome issues and dilemmas in attempting to define a sensible long-term approach to BMD. The paper ends with some personal observations and recommendations.

BACKGROUND

The advent of nuclear weapons changed U.S. security in a fundamental way. In his 1948 *Foreign Affairs* article, Bernard Brodie captured the fears of future U.S. presidents when he wrote the following:

It is now three years since an explosion over Hiroshima revealed to the world that man had been given the means of destroying himself. Eight atomic bombs have now been detonated . . . and each was in itself a sufficient warning that the promise of eventual benefits resulting from the peacetime use of atomic energy must count as nothing compared to the awful menace of the bomb itself. The good things of earth cannot be enjoyed by dead men, nor can societies which have lost the entire material fabric of their civilization survive as integrated organisms (Brodie, 1948).

As Brodie warned, nuclear weapons have put the survival of civilizations at stake. Faced with the awful consequences of nuclear weapon employment, all presidents from Truman until today have sought the means to provide at least some protection for the United States. Political and budgetary support for strategic defenses have waxed and waned over the years, depending on attitudes, DoD budgets, etc. However, the underlying rationale for strategic defense never wavered: *Provide protection for the United States, its people, and its institutions in the face of foreign nuclear capabilities over which it has no direct control.* Debates about strategic defenses, and more particularly ballistic missile defenses, and their relevance to nuclear deterrence and national survival should be viewed in this context.

**Strategic Defenses Before Ballistic Missiles**

In 1950 the Soviet Union detonated its first atomic bomb, and the race between strategic offensive and defensive forces commenced. In response to first the threat and then the fact of USSR possession of nuclear weapons, Presidents Truman and Eisenhower undertook extensive active and passive strategic defensive measures. Because the only delivery vehicles available were aircraft, the United States (and similarly the Soviet Union) heavily invested in strategic air

---

8 York (1970) estimated that by 1970 the United States had spent about $30 billion on air defenses and the Soviets about $75 billion. In today's dollars, the figures would be roughly $90 billion and $225 billion.
defenses. While by no means leakproof, the U.S. air defenses nonetheless provided some confidence that a Soviet bomber attack would not totally devastate the United States, nor disarm it to a degree that would threaten its ability to carry out a devastating retaliatory blow. Further, the air defense system was thought to be capable of handling a partially damaged and poorly coordinated Soviet attack of the sort that might follow a preemptive strike by the massive U.S. bomber force, perhaps in the context of Soviet preparations for nuclear attack on the United States following an attack on Western Europe.5 This preemptive capability, in turn, underwrote the U.S. policy of extended deterrence (i.e., U.S. deterrence of a Warsaw Pact invasion of Western Europe by threatening a retaliatory nuclear attack on Warsaw Pact forces and possibly the Soviet homeland as well). The 1950s was an era of U.S. strategic nuclear superiority in militarily meaningful terms, and U.S. strategic defenses played a central role.

More broadly, the U.S. strategy for protecting the country evolved during this period into a combination of three approaches: (1) first and foremost, punishment-oriented deterrence, through threats of massive retaliation; (2) prevention-related deterrence (or counterforce capability), through the credible threat of a substantial damage-limiting preemptive first strike; and (3) direct defense, through a combination of "passive" measures such as civil defenses and aircraft shelters and "active" defenses in the form of multilayered air defenses.

None of these were believed to be infallible. Nor were they viewed as satisfactory to protect the United States against future threats from other countries that might acquire nuclear weapons. Thus, in addition to being a decade that fostered major expenditures on strategic air defenses, the 1950s saw continuing efforts toward controlling the spread of nuclear weapons and limiting global access to technologies that could hasten both that spread and the development of ever-more-difficult-to-counter nuclear delivery systems. Policymakers continued to hope, however wistfully, that the security of the United States could be kept under U.S. control. While nuclear war might be awesomely destructive, it was not unreasonable—if controversial—to argue that with appropriate

5The commander of the Strategic Air Command (SAC), General Curtis LeMay, once shocked Robert Sprague, deputy head of the Gaither Commission, by explaining his (LeMay's personal) preemption planning: "If I see that the Russians are amassing their planes for an attack, I’m going to knock the [expletive] out of them before they take off the ground" (Kaplan, 1983:134). At that time, in the mid-1950s, SAC had an extensive fleet of long-range and overseas aircraft constantly observing the Soviet Union to provide strategic warning. There were routine penetrations of Soviet airspace. Ironically, had the Soviet Union been able to accomplish a risky surprise attack, SAC’s forces would also have been exceedingly vulnerable. That was the essence of the famous RAND “basing study” (Wohlstetter, Rowen, Hoffman, and Lutz, 1954). For a good history of this era, see Kaplan (1983).
strategy and preparations, the United States might survive a nuclear war (see, e.g., Kahn, 1960). That policymakers thought similarly was evident from the major efforts of that era in civil defense programs.

The Introduction of ICBMs and the Need for BMD

The introduction of large numbers of ICBMs into the Soviet weapon inventory in the early 1960s began to change all this dramatically, as had been anticipated by the Gaither Commission in 1957. Air defenses would no longer suffice, and early U.S. efforts to transform existing surface-to-air antiaircraft missile systems (SAMs) into crude BMD systems were unsuccessful. The strategy choices facing the Kennedy administration included (a) developing more effective BMD systems, (b) developing an effective first-strike counterforce capability against both missile and bomber threats, or (c) forgoing serious hope of defending the United States well against large-scale attack. All of these possibilities had strong proponents. It is important to note, however, that the original primary objective of strategic defense remained protection of the United States—not deterrence or some notion about military balances. That was the "natural" objective, in contrast to what came later.

This natural objective still underlies interest in BMD today. To be sure, other uses for BMD have been suggested, emphasized by strategic analysts, and even pursued for some years—e.g., protecting Minuteman missiles against a Soviet counterforce attack, protecting U.S. national leadership, and protecting bases and troops in the field from attacks by short-range tactical ballistic missiles. None of those rationales, however, has had the staying power or continued persuasiveness regarding core issues. In the 1960s, 1970s, and again in the 1980s, BMD systems were proposed and even planned for other purposes, but the rationale was not convincing once people concluded that BMD could not accomplish its potentially unique mission of directly defending core values. Minuteman missiles could be protected in other ways and at less cost (or done away with in favor of more survivable systems). So also could the leadership be defended in other ways. But the devastating effects of nuclear weaponry on population centers could not be denied without actively preventing the weapon from detonating over the target.

The lack of an effective BMD also raised serious questions about the vulnerability of the air defenses themselves. Without protection from ICBM attack, the air defense command-and-control network was highly vulnerable to even small attack sizes. This fact alone led Robert McNamara to cease large expenditures on existing air defense systems and initiate new programs (e.g., AWACS) that might better deal with this situation. Consequently, funds for air defense dropped sharply in the 1960s.
Damage-Limitation Studies

With the introduction of Soviet ICBMs and SLBMs in the early 1960s, the United States had to rethink its nuclear strategy. Because the population was no longer protected even if the United States initiated a preemptive attack, damage limitation as originally conceived (and initially supported by the Kennedy administration)\(^\text{11}\) was no longer workable. To recapture some degree of population protection, Secretary of Defense Robert McNamara put forward in his famous Ann Arbor speech of June 1962 the novel strategy of city avoidance, i.e., in case of a nuclear war, each superpower would pledge not to target the cities of the other. This strategy died quickly when the Soviet leadership rejected the notion out of hand.\(^\text{12}\) Faced with mounting costs to maintain some semblence of damage limitation, McNamara directed the Office of the Director, Defense Research and Engineering (DDR&E) in early 1964 to undertake a comprehensive reexamination of the feasibility of preserving damage limitation capabilities over the long term. Colonel Glenn Kent (then in DDR&E) prepared a seminal report on damage limitation that examined counterforce, BMD, and passive defense options for various survival criteria (Kent, 1964). At the direction of McNamara, this study examined the following questions:

- For any proposed level of expenditures on “damage limiting” forces, what is the “optimum” allocation of the total among the various means that contribute to this function: (1) civil defense; (2) terminal ballistic missile defense and terminal bomber defense; (3) area bomber defense; (4) strategic offensive forces; and (5) defense against Soviet missile-carrying submarines?
- What are the possibilities available with regard to limiting damage to the United States and its allies? For example, what is the “percent surviving” in the United States as a function of the total expenditures on damage limiting for various contingencies?

---

\(^\text{11}\) Early in his administration, President Kennedy was faced with a serious crisis over the future of Berlin. The prospects of nuclear war were all too real, as was the vulnerability of the United States to Soviet nuclear attack. Kennedy took damage limitation seriously, urging Americans to become prepared through civil defense measures, and he sharply augmented DoD’s budget (with added money for civil defense). See Reeves (1993).

\(^\text{12}\) The Soviet leadership was also interested in damage limitation, but their nuclear systems did not have the requisite capabilities (in either accuracy or numbers) to seriously threaten the U.S. nuclear arsenal, and they too were having difficulty in developing an effective ABM defense. Thus, they saw McNamara’s initiative as a U.S. attempt to regain strategic superiority.
Figure 1, drawn from that report, shows the extent to which the U.S. population could be protected as a function of the ratio of U.S./Soviet costs to reduce/impose this level of damage. Two curves are shown: one in which the United States executes a successful first strike, the other in which the Soviets strike first. The calculations behind these curves assume optimum choices for the United States, i.e., the best mixture of counterforce, active defenses, and passive defenses. The Soviet responses were confined to simple ICBM proliferation; more stressful responsive threats (e.g., MIRVing of the ICBMs) would have made the curves significantly worse for the United States. Even under the optimistic assumptions of a U.S. first strike and a mundane Soviet responsive threat, the figure shows that the Soviet Union could insure holding at risk at least 20 million Americans by spending only one dollar for every three spent on U.S. damage limiting systems. If the United States sought even fewer casualties, the ratio would have been even higher. According to Kaplan (1983), these charts confirmed McNamara's fear that achieving meaningful levels of damage

![Graph showing cost/exchange ratios as a function of population survival.]

SOURCE: Kent (1964).

Figure 1—Cost/Exchange Ratios as a Function of Population Survival
limiting was impossible, and trying to do so would lock the United States into a very unfavorable and unwinnable arms race (see also DoD, 1965). Reluctantly but inexorably, U.S. defense policy shifted away from notions of damage limitation and toward an explicit and nearly complete emphasis on deterrence. That emphasis held throughout the remainder of the 1960s and 1970s, although the strategy for accomplishing deterrence matured—going far beyond assured destruction and toward the assured-retaliation capability of denying the Soviet Union (or any other attacker) the ability to achieve its war aims (mainly believed to be expanding its control over Europe and Asia). Thus, U.S. nuclear targeting strategy increasingly embraced a comprehensive set of military targets, both "strategic" and "general purpose" (e.g., armies, navies, air forces, and their vast infrastructure), as well as the "political control structure" consisting of Communist Party leadership installations.

Early U.S. Ballistic Missile Defenses

Active BMD was the big loser in Kent’s study. Nonetheless, proponents of active BMD did not abandon their fight. Many arguments were put forward for BMD’s utility. These included responding tit-for-tat to the Soviet deployment of a BMD system around Moscow, defending against light attacks (e.g., from what was described as an emerging Chinese threat), defending U.S. strategic offensive forces, and hedging against a possibly much more extensive Soviet deployment of BMD systems. Also, some believed that NATO’s policy of graduated escalation, and especially its reliance on nuclear escalation, lacked credibility because of the growth in Warsaw Pact capabilities; BMD, as a means to protect the United States against limited Soviet attacks, could strengthen that credibility. As for capability, technology had advanced from attempts to adapt air defense systems to the much more advanced systems (e.g., the Spartan and Sprint missiles and a family of phased-array radars) that might indeed enable defenses to successfully defend the United States against limited and technically unsophisticated attack. These arguments did not convince Robert McNamara, but President Johnson, apparently for domestic political

13Although not part of Kent’s study, these curves are sensitive to the absolute size of the threat, smaller being better in terms of limiting damage at specific cost/exchange ratios. And they are very sensitive to defense performance.

14For a good review of this approach to deterrence, under the rubric of the countervailing strategy, see Slocombe (1981). In practice, U.S. targeting policy under the Reagan administration was substantially consistent with that developed under the Carter administration, which in turn drew heavily on conclusions from Ford administration studies. For discussion of continuity, see Nolan (1989).
reasons, decided in 1968 to deploy the Sentinel system at a small number of sites around the periphery of the United States.

Soon upon taking office, President Nixon decided to cancel the Sentinel system and to deploy instead a less expensive Soviet-oriented “new system” called Safeguard that would focus on defending U.S. Minuteman silos. In fact, Safeguard used the same hardware as Sentinel, but the change in mission and cost was important because Nixon saw ABM strictly as a bargaining chip to be used in attempting to gain control over the development of Soviet strategic offensive systems. In his evaluation of SALT’s Summit I, he had this to say:

The ABM treaty stopped what inevitably would have become a defensive arms race, with untold billions of dollars being spent on each side for more ABM coverage. The other major effect of the ABM treaty was to make permanent the concept of deterrence through “mutual terror”... Each side therefore had an ultimate interest in preventing a war that could only be mutually destructive (Nixon, 1978:618).

Most analysts judged the planned Safeguard deployment for defense of ICBMs to be technically plausible and at least a modest addition to U.S. deterrent potential. It was also viewed as stabilizing (and thus acceptable to many of the earlier ABM opponents). However, the ABM treaty, signed in 1972, confined Safeguard deployments to a single site selected to be at Grand Forks and limited both countries to defensive interceptor numbers not to exceed 100. Safeguard became operational in 1973 and was decommissioned almost instantly on grounds of operational cost and effectiveness. The sole vestige of this deployment, a long-range acquisition radar, still exists at Grand Forks.

Note that arguments for BMD slipped from full protection of the country, to protection of the population against only small attacks, to protection of only our retaliatory forces to augment deterrence, and finally to bargaining leverage for banning ABM altogether. By the late 1970s, a balance of nuclear capabilities was commonly accepted as a stable and desirable state for nuclear forces. While heated debates continued between those who supported a nuclear deterrent based on mutual assured destruction (MAD) and those who argued for an array of city-sparing nuclear targeting options, countrywide BMD was, at

---

15The proponents of BMD also viewed its deployment as necessary to maintain the scientific and engineering personnel who had been attracted to ABM research in the 1960s, as well as a potential precursor to a countrywide deployment. In this there are many parallels between then and now.

16In fact, Safeguard augmented U.S. retaliation capabilities only marginally and thus would have made little difference if disabled.

17In reality, the SIOP (Single Integrated Operations Plan) offered a wide array of options from which the President and his advisers could choose. These debates had more to do with basic national strategy than with nuclear employment planning.
best, placed on the back burner. Indeed, at one point in the 1970s, the Congress actually enjoined DoD and the Army from pursuing BMD technology that might contribute to broad-area BMD, i.e., population protection.

The Issue Is Reopened: The Star Wars Speech

On March 23, 1983, President Reagan reopened the BMD debate. In his famous Star Wars speech, Reagan called on the technical community to examine prospects for making (nuclear-armed) ballistic missiles "impotent and obsolete," in conjunction with a call for deep reductions in nuclear forces. In strategic terms, he brought into question the efficacy of any long-term strategy of reliance on mutual assured destruction. The result was substantial ferment, with the debate tending to fall along ideological, and often partisan, lines. This was unfortunate, because the questions Reagan was raising deserved to be asked and answered. Interestingly, some of the original negotiators of the ABM treaty agreed and were by no means hostile to reopening the question of BMD, even if it meant changing the treaty. They were, however, skeptical about whether BMD was more attractive than it had been in 1972.

After a comprehensive study of the technical factors that might make a comprehensive defense possible (the so-called Fletcher Report), the Strategic Defense Initiative Organization (SDIO) was formed to pursue the more promising technological opportunities. SDIO reported directly to the Secretary of Defense and lasted until 1993, when it was disbanded by the Clinton administration. During its ten-year existence, it spent something in excess of $30 billion (then-year dollars), perhaps half of which was over and above what would have otherwise been spent in a less centralized and focused set of service programs. Interestingly, during this ten years, history was repeated. As in the 1960s, initial enthusiasm revolved around population defense, but scientific and technical experts concluded once again that this was not yet in the cards. Slowly, the focus of BMD under the Star Wars program was shifted, almost inexorably, toward defense of the deterrent force and protecting population against a variety of vaguely defined light attacks. Then, in

---

18 Many opponents of MAD as policy based some of their arguments on the immorality of attacking innocent civilians. The Catholic bishops heatedly debated the morality of assured destruction, almost (but not quite) confronting the Reagan administration in the early 1980s (van Voorst, 1983). Questions were raised about the legality of nuclear attacks against civilians (Bulter and Graubard, 1982), and even supporters of the existing policy toward nuclear deterrence felt the need to address its ethics (Nye, 1986). These ethical and logical issues were used by the proponents of BMD in trying to make the case to reopen debates on nuclear strategy.

19 For discussion of the history of Star Wars, see Nolan (1989) and Shultz (1993).
1991, it shifted toward defense against theater ballistic missiles (TBM), a shift that was codified by the Bottom-Up Review. 20

**Current State of Missile Defense Technology**

Given all the development effort on BMD, where do we stand technologically? As of late 1993, most defense analysts would agree to the following:

- Technology has advanced sufficiently so that *limited defenses against nuclear-armed ballistic missiles are both feasible and affordable*. While studies suggest that the offense can still claim some cost-exchange edge over the defense, that advantage has been reduced. 21

- *Comprehensive ballistic missile defenses*, i.e., those that would protect the entire United States to a degree that would satisfy President Reagan’s aim of making ballistic missiles “impotent and obsolete,” are still unachievable within realistic budgets against large and sophisticated attacks (i.e., attacks that might come from Russia or a future superpower). Even so, the estimated acquisition costs for such a defense have fallen from estimates as high as $1 trillion to $100 billion or less.

- *Defenses against a large attack of air-breathing nuclear delivery systems (bombers and cruise missiles)* might cost as much again as comparably capable BMD deployments. Without balanced defenses against all delivery means, comprehensive countrywide protection would not exist, thereby casting into serious question the value of the expenditures. 22

These conclusions may seem to argue against continued pursuit of ballistic missile defenses, but there is more to the story, as I shall now discuss.

---


21 By the late 1980s the exchange ratio was three or less, at which point it was approaching the ratio of the U.S./USSR GNP’s, making it possible for BMD proponents to argue that the United States could win an offense-defense race even if the ratio was adverse—especially since the Soviet defense burden was already strangling the economy. With technology now on the horizon (e.g., Brilliant Eyes and Brilliant Pebbles), even lower ratios are at least plausible, except that the attacker could “open a new front” by increasing the emphasis on cruise missiles.

22 Some advocates of BMD have argued that, as “slow flyers,” air-breathing delivery vehicles do not pose a risk to nuclear stability. However, the grass-roots interest in defenses that arose after President Reagan’s speech was concerned with protection, not marginally improving some theoretical notion of stability. Reagan was calling for making nuclear weapons “impotent and obsolete,” not merely “slow.” Further, if improving stability were the issue, there were many other less-expensive ways to proceed (i.e., increasing the survivability of offensive forces).
THE NEW INTERNATIONAL CONTEXT FOR BMD

With this brief background of BMD history, we must now consider the drastically new political-military context before thinking about appropriate direction for BMD programs. The most important elements of the new context for BMD appear to be (1) proliferation of weapons of mass destruction and ballistic missiles, (2) the START process and the opportunity it provides for cooperative U.S.-Russian efforts to draw down offensive forces and perhaps cooperate on defenses, and (3) the many fundamental uncertainties about the more distant future of international relationships. Let me discuss each of these in turn.

Proliferation-Related Challenges

There is no question that the feared proliferation of weapons of mass destruction (WMD), coupled with the ongoing proliferation of ballistic missiles and the technology associated with them, forms the primary motivation for current TMD efforts (TMD because, in the near term at least, the proliferation will probably not involve missiles that could strike the United States). Table 1 (based mostly on data from Systems Planning Corporation (1992) collates information on which countries are pursuing WMD and ballistic missiles. The data are sobering.

Desert Storm made it crystal clear that even conventionally armed TBMs could significantly affect U.S. national security interests as well as military capabilities. Without discounting the tragic loss of U.S. service personnel at the barracks near Dhahran, it must be said that the direct military consequences of the Scud attacks against targets in Saudi Arabia and Israel were insignificant. But the political impacts were large, leading to a massive and sustained effort to locate and suppress TBMs and their TELs.

Armed with nuclear weapons, future TBMs could well alter the outcome of future confrontations. A RAND study dealing with the consequences of confronting a nuclear-armed adversary (Molander and Wilson, 1993; Millot, Molander, and Wilson, 1993) reported the following results after a series of political-military contingency games:

- Many participants questioned whether U.S. national interests were truly at stake, suggesting American nonparticipation (or, to put it differently, evidence that the United States might well be deterred from intervening).23

23When asked to state the most important lesson from Desert Storm and the Gulf War, India’s defense minister reportedly said that it was not to go to war with the
Table 1

Countries Pursuing Ballistic Missiles or Weapons of Mass Destruction

<table>
<thead>
<tr>
<th></th>
<th>Missile Producer</th>
<th>Missile Possessor</th>
<th>Space Launcher</th>
<th>Nuclear Program</th>
<th>Nuclear R&amp;D</th>
<th>Chemical Weapons</th>
<th>Bio-Weapons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Burma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Libya</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>North Korea</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>South Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Syria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Yemen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

- Many participants seriously doubted the support of key allies in the contingencies.
- Rather generally, there was a call for restructuring power projection forces, to include greatly enhanced counterforce and active defense capability against TBMs, in order to prevent significant U.S. casualties.

This suggests that U.S. willingness to use its military capabilities to protect its interests and those of its allies would be significantly affected by nuclear proliferation. TMD can be only a partial answer because it can provide neither absolute protection nor complete coverage against all threats. Although this paper deals almost exclusively with defending against use, we should realize that a wide range of counters to TBMs are probably needed, including (1) dissuading possession, (2) deterring use, (3) devaluing threat of use, and (4) defending against use (BMD and passive measures).

United States unless armed with nuclear weapons. This lesson is almost certainly appreciated by most Third World countries.
START II and Implications for BMD

If proliferation offers challenges, U.S.-Russian nuclear arms agreements offer opportunities. Starting about a decade ago with the treaty to eliminate intermediate-range ballistic missiles, strategic arms agreements have progressed to START II (see Wilkening, 1994), which will limit the total number of strategic weapons per side to between 3000 and 3500, about one-third of those that exist today. Figure 2 depicts one plausible set of force levels in the year 2002 when the treaty comes into full force (Gershwin, 1993).

The following points are especially relevant to BMD issues:

• The Russian ballistic missile threat under START II will be sharply reduced, to roughly 2200 reentry vehicles (RVs) total (substantially fewer on alert).

• As a consequence, BMD systems would face much smaller nominal threats, be they launched by accident or deliberately, e.g.:
  – 2000 RVs from an all-out attack of fully alerted forces.
  – 100 RVs from an unauthorized attack by a “mad submarine commander,” and numbers much smaller from ICBM fields.
  – 10 RVs or less from accidental attacks.24

These latter two are small in size (smaller than, e.g., the Chinese arsenal and possible new threats from Third World countries).

It is also worth noting that many of the residual missiles under START II will either be dMIRVed variants of current missiles, or will otherwise have fewer RVs than they are capable of carrying. This means that either side could relatively quickly expand its offensive forces if it so desired. This could improve confidence during a coordinated deployment of defenses, because if one side sought to achieve unilateral advantage, the other could respond by deploying more weapons. On the other hand, it means that a defense sized against nominal levels could quickly find itself too small in the case of “breakout.” In today’s world, breakout scenarios seem much less plausible than only a half-dozen years ago, but the world could change again.

24 Some policymakers are considering trying speeding up START II’s full implementation date or negotiating with the Russians (and other nuclear states of the former USSR) for de-alerting or other measures that might take off active status those parts of the nuclear force that are scheduled for elimination because of START II. If such agreements can be reached, threat reductions could occur earlier, to the apparent benefit of both the United States and the former Soviet states. See Wilkening (1994).
More generally, it is important to note that with the United States and Russia working cooperatively, and with advances in technology, it is plausible that coordinated actions on reducing nuclear inventories and building countrywide defenses could create a situation in which President Reagan's goal of "impotent and obsolete" might at least be approximated. Additional technical "miracles" might still be needed, but at least a number of the technical challenges have been met. What is plausible (i.e., not impossible) now is that the United States and Russia could cooperatively shape their offensive and defensive forces so that nuclear weapons would play an increasingly negligible role in U.S.-Russian affairs, even if the two countries became adversaries again. The proverbial nuclear genie would not be "back in the bottle," but (mixing metaphors) the sword of Damocles would not be hanging so conspicuously over our heads as during the Cold War. Further, both nations would have defenses against accidents and from modes: Third World threats.

Uncertainties: Political Stability of Russia and Other Issues

All of this is fine, but it depends on the current positive relationship between the United States and the states of the former Soviet Union. That positive relationship, however, is not immutable. The leaderships of those states face difficult problems, including friction between Russia and Ukraine and the growing popularity of leaders arguing for a recovery of empire. While it is difficult to imagine that even a new leadership in Russia would revert to a full-scale military competition with the United States, it is quite imaginable that a
new regime might see the nuclear arms agreements as biased toward the United States, causing it to back away from them. It is also possible that the international situation could come apart if a future Russian government began aggressively seeking to recover the territories of its “empire” (the Baltic states, Belarus, Ukraine, etc.). The truth is that more generally we do not know what the world will look like twenty years hence (some would say even five years is a stretch). Where will the conflicts occur? Will nuclear or other WMD have been used? And so on. Reminding us of what could happen, Japan recently announced (and subsequently recanted) that it would seriously reconsider its commitment to the nuclear nonproliferation treaty if it can be proved that North Korea possesses nuclear weapons.

Against the backdrop of this changed international environment and uncertainties about the future, let me now discuss in turn issues and options for theater missile defense, national missile defense, and space-based missile defense (TMD, NMD, and SBMD).

THEATER BALLISTIC MISSILE DEFENSE (TMD):
PROTECTING U.S. PROJECTION FORCES AND ALLIES

Any approach to TMD should depend on the targets to be defended, the size and character of the threat, the technical feasibility of the defense options, and a host of other factors.

Likely Targets for Enemy TBM

What must be defended? On the one hand, U.S. opponents might target a variety of military sites (e.g., airfields and seaports) to discourage or prevent U.S. force projection into a theater of operations, as well as to disrupt force employment once the forces were deployed. To the extent U.S. power projection capabilities are at risk, these sites must be protected. In addition, Congress has already mandated that an important objective of TMD must be to protect forward-deployed U.S. forces.

However, protecting U.S. forces is only part of the problem. Protecting our friends and allies adds a complex and sometimes subtle political dimension to the TMD problem, as we saw in the Persian Gulf war. It is also easy to imagine a situation where a neighboring country of the aggressor—i.e., a country “under the gun” of its TBM—would believe that even a U.S.-favorable war outcome would not deter the aggressor from “getting even” after the war was over and U.S. forces withdrew. Under such circumstances would it cooperate with the United States and the country being invaded? Would the protection of U.S. TMD change the equation? Aside from fundamental humanitarian
reasons, these questions strongly imply that protection for our friends and allies is also needed. Congress has also mandated that this be a central objective of our TMD developments.

The potential need to defend allied territories creates substantial coverage challenges for TMD. Indeed, much of the concern about the proliferation of ballistic missile technology has to do with likely increases in TBM range, which would dramatically increase the number of politically interesting targets within reach of an adversary. Missiles expected to have ranges of 1000 km or longer (e.g., North Korea’s No Dong 1 and Iraq’s Al-Abbas) are particularly worrisome. Figure 3 illustrates what extended-range missiles might threaten if deployed in Iraq and Libya. Given the large size of the areas to be protected, defense will be neither easy nor inexpensive.

Types and Numbers of Threat Weapons

The challenge posed to defense systems also depends significantly on the type of warhead employed and the size of the TBM threat. Table 2 provides some crude estimates, for each of several weapon types, of the area of effect against a standard military target such as an airfield with personnel in the open (see Gold and Welch, 1993). Nuclear and biological weapons have very large areas of effect and therefore stand above the others. By contrast, the large-area disruptive effects of chemical weapons can be substantially mitigated, assuming adequate protective suits and facilities at the base under attack. However, the persistence of these weapons is cause for concern, and the importance of chemical attacks against airports and seaports of entry should not be minimized. Unless decontamination equipment were available and successfully used, chemical attacks with persistent chemicals could close critical ports and airfields for a considerable time. Table 2 also indicates rough numbers for the inventory of threat weapons. There could be hundreds or thousands of conventional TBMs akin to Frog and Scud missiles, but the number armed with WMDs will be much less.

Feasible Goals for TMD Systems

Given the potential lethality of nuclear and biological weapons, it is natural to seek a TMD goal of near-zero leakage for small TBM attacks. Whether such a goal is practical is debatable (especially for attacks in which small num-
Figure 3—Target Coverage for Extended-Range TBM's
Table 2
Comparative Effects of Different WMD

<table>
<thead>
<tr>
<th>Warhead Type</th>
<th>Difficulty in Obtaining</th>
<th>Likely Numbers</th>
<th>Submunitions</th>
<th>Affected Warhead Area (square miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional (500 kg)</td>
<td>None</td>
<td>Many 100s</td>
<td>Maybe</td>
<td>(10^3) (10^2)</td>
</tr>
<tr>
<td>Chemical (500 kg)</td>
<td>Some</td>
<td>10 to 100s</td>
<td>Yes</td>
<td>(10^6) (10^2)–(10^5)</td>
</tr>
<tr>
<td>Biological (240 kg)</td>
<td>Some</td>
<td>1 to 10s</td>
<td>Yes</td>
<td>(10^9) (10^6)–(10^8)</td>
</tr>
<tr>
<td>Nuclear (20 kT)</td>
<td>Substantial</td>
<td>A few</td>
<td>No</td>
<td>(10^7) (10^5)–(10^7)</td>
</tr>
</tbody>
</table>

NOTES:
1. Single Scud ballistic missile attack versus airfield.
2. Affected area calculation assumes disabling of more than 50 percent of personnel.
3. Protection against chemical attack assumes MOP 5 suits and masks.
4. Protection against biological weapons assumes no inoculations.
5. Protection against nuclear weapons assumes facilities normally found on foreign airfields.

bers of WMD-armed TBMs are mixed in with and indistinguishable from a larger number of conventionally armed TBMs) but it is not mathematically unrealistic. It might also be possible to achieve small leakage against somewhat larger numbers of ballistic missiles with chemical warheads—at least at selected sites, if the attacker did not know which sites were being preferentially defended or how successful his previous attacks had been.

The situation is quite different for a conventionally armed TBM threat (i.e., ballistic missiles armed with conventional explosives rather than WMD). In many cases, passive defense measures are adequate by themselves because TBMs have relatively small payloads and poor accuracies, and many targets (e.g., airfields) are tough to destroy. Even heavy and persistent allied bombing attacks against Iraqi airfields did not close those bases, despite the use in some cases of special runway-busting munitions. This outcome was consistent with numerous past studies (by RAND and others) predicting that air bases would

\[25\] Straightforward math offers the following insight. Assuming a shoot-look-shoot terminal defense capability (or, equivalently, a two-layer defense), zero leakage against a 10-RV attack can be achieved with greater than 90 percent confidence if the interceptor's single-shot kill probability is 0.9. While 0.9 is high, it falls within the design goals for many defensive systems and cannot be ruled out technically.
be hard to close and would be readily repairable except when attacked by the best of runway-busting munitions under circumstances maximizing munition effectiveness.\textsuperscript{26}

While there is no similar wealth of vulnerability analysis of ports, army bases and the like, a simple consideration of the effects radius versus the area of such targets tells us that with the exception of, e.g., large cranes and gantries, most of these targets are not critically vulnerable to conventionally armed TBMs.

**Terminal/Midcourse Defenses: The Canonical Approach and Its Limitations**

In theory there are many approaches to TMD, many of them complementary. The current DoD program, however, focuses on systems to destroy the incoming weapons in the terminal or late-midcourse portions of their flight. Plans call for upgrading the Patriot and for developing and deploying the THAAD system (Aspin, 1993). Both Patriot and THAAD would be ground-based Army systems. These defenses are achievable in the relatively near term and would form an essential element of any longer-term U.S. TMD strategy. However, a good deal of RAND analysis, much of it led by colleague David Vaughan, indicates that such terminal/midcourse defenses, if forced to operate on their own, are likely to be inadequate against difficult threats. Among the most important limitations are the following:

- **Unspectacular single-shot kill probabilities (SSPKs).** It has long been customary for studies of defense systems to postulate SSPKs on the order of 0.7, 0.8, or even 0.9. But the experience with Patriot in Desert Storm and subsequent analysis raises significant questions about what can be achieved and how confidently one should accept assertions about system effectiveness. Recent analyses suggest that Patriot's success in Desert Storm was no greater than 50 percent and perhaps much smaller.\textsuperscript{27}

- **Availability.** Although THAAD is being designed to be easily deployed into the theater of operations, that deployment may not happen until after the

\textsuperscript{26}Because these problems were well known to the Americans and British, the attacks had other goals. In particular, precision strikes against shelters probably destroyed many Iraqi aircraft and helped to motivate the flight of Iraqi aircraft to Iran.

\textsuperscript{27}Although the Army reported immediately after the war that Patriot had achieved 45 hits in 47 engagements within its defended areas over Israel and Saudi Arabia, subsequent analysis (Postol, 1991) suggested that the number of successful engagements was very small. Israeli reports, during the war and in late 1993, supported that view (see especially Atkinson, 1993:277–278).
conflict has begun. If no other defensive capability existed, the enemy could use the unchallenged TBM threat to (1) coerce regional countries not to allow U.S. entry, (2) compel the threatened country to sue for peace (especially if WMD were involved), and (3) deny suitable ports and airfields to U.S. power projection forces.

- **Coverage.** As discussed above, the United States may need to defend allied countries in the general region. Systems like Patriot and the baseline, ground-based THAAD may not have the requisite coverage areas to cope with extended-range TBMs.

- **Collateral damage.** Again, the data from Desert Storm are illuminating. Spent boosters, maldirected interceptors, and even damaged warheads can still cause significant damage (Postol, 1991). In the case of WMD, the mere release of the weapon products is cause for serious concern, even if the weapons do not hit their intended targets.

- **Sensitivity to details of threat, including fractionation.** Even if a BMD system has a very high SSPK against simple targets (e.g., a single warhead to be engaged well within the atmosphere), it may do poorly against actual attacks in which the targets are rather different technically than anticipated. Further, the early release of bomblets or chemical canisters, as suggested in Postol (1991), would cause serious problems for any terminal-phase system (see also Mesic, 1994a, Larsen and Kent, 1994, and Canavan, 1993).

Acting on their own, terminal and late-midcourse defenses also suffer from a number of problems: battlespace constraints, which limit their shoot-look-shoot opportunities and their ability to achieve very low leakage rates even with relatively high SSPKs; limited coverage per site, suggesting either a very large deployment for full regional area coverage or leaving some targets undefended; uncertain lethality against some targets, such as chemical warheads detonated high in the atmosphere; and compatibility with the existing ABM treaty.

---

28 There are a variety of reasons why this is likely to be true. Perhaps the most important is the enemy's interest in starting the war prior to that deployment, i.e., the political and military value of its TBMs would be greatest in this case and the initiative for the start of the war rests with the enemy.

29 The Bottom-Up Review recognized this problem and that of early availability; it called for continued development of a ship-based midcourse intercept capability, called the sea-based upper tier, which could be deployed on station prior to conflict outbreak (Aspin, 1993).

30 For example, Iraq's Al-Hussein missile broke into pieces during reentry. This unintended countermeasure confused the Patriot's tracking algorithms and caused it in some cases to lock on to false targets. A second problem was that the Patriot fuze's timing mechanism was set for slower incoming objects (e.g., shorter-range missiles), thereby causing late detonation. Deliberate countermeasures could be even more troublesome. Whether such failures can be avoided in THAAD and other terminal defense systems remains uncertain.
(although the Russians may well agree to changes). And, not least, THAAD and Patriot will require very significant funding before their full developments are finished after the turn of the century.

None of the above suggests that THAAD and Patriot aren’t essential constituents of a robust TMD system. However, it does suggest that the United States needs to consider additional options for TMD, augmenting terminal/midcourse defenses based on Patriot and THAAD.

Counterforce, Counterbattery, and Boost-Phase Options

In examining complementary TMD options, RAND work has focused on concepts of operations that can be implemented and has suggested two imperatives: (a) the need for *timely* active defense with the potential for substantial geographic *coverage*, and (b) the need to address *complex targets* (e.g., targets amid debris or protected by countermeasures) *that must be destroyed, not merely diverted*.

*Timely Active Defenses With Significant Coverage.* Active defenses need to be in place early in crisis to protect regional allies, deploying U.S. forces, and critical infrastructure. The approaches to accomplishing this that have been proposed most often are as follows, with the first two being the most relevant in the near term, and the ones being most seriously considered by the DoD.

- **Ship-based terminal or midcourse defense systems** using either a variant of THAAD or the Standard missile with a LEAP upper stage.\(^{31}\) Deployed on Aegis cruisers, the system would use an upgrade to the SPY-1 radar for target acquisition and tracking. To be effective, the cruiser would need to be deployed near the country being attacked, limiting the defended area or potentially placing the ship in harm’s way.\(^{32}\) Nonetheless, ship-based defenses have great strategic and operational advantages, particularly sustainability.

- **Airborne midcourse interceptors** using SRAM as a booster and LEAP as the upper stage. The missile could be carried by several aircraft, including most bombers and the F-15E.\(^{33}\) Its sensor support still needs to be determined.

---

\(^{31}\) THAAD operates early in the terminal phase or late in midcourse. LEAP provides more midcourse capability; it is a missile and homing kill vehicle that has been under development by SDIO for several years. LEAP is also being considered as an upper stage for other ground-based defense interceptors and for SRAM in an airborne mode.

\(^{32}\) Among the weapon systems that have been widely proliferated into Third World inventories are antiship cruise missiles and various forms of antiship mines. These could constrain Navy operations and the area defended.

\(^{33}\) If required, bombers can operate from bases (including those in CONUS) well outside the theater in question. Buchan (1994b) describes RAND’s recent work on the potential employment of long-range bombers in major regional conflicts. This mission
with options including the use of Cobra Ball aircraft, forward-deployed ground-based radars, and space-based sensors. To be effective, the weapon platform would have to be somewhere between the target and the launch point. Depending on the extent of early air defenses, this might or might not place the aircraft in danger. Maintaining such operations during crisis would be very expensive.

- **Space-based kinetic-kill vehicles (KKVs)** (Brilliant Pebbles). Brilliant Pebbles would be predominantly an exoatmospheric interceptor, with the same limitations as LEAP.

- **Proliferated TMDs in threatened nations.** Still another option would be to deploy TMDs in all the countries likely to be threatened. Although there are competitors such as the Israelis, Russians, and Europeans, the United States is technically well placed to build and sell such systems to its friends and allies.

**Addressing Complex Targets and the Need to Destroy Them Early.** The above options provide capabilities that partially hedge against potential countermeasures, but they do not engage the threat until well past burnout; thus, they are still sensitive to many of the problems that afflict THAAD. The following capabilities would help a great deal (Larsen and Kent, 1994, and Mesic, 1994a):

- **A boost-phase intercept capability.** Various combinations of platforms and interceptors are under consideration, all of which would exploit the immediate detectability of a TBM’s rocket plume. Most require overflight of enemy territory to be effective, raising questions about their availability early in conflict. Lasers and hypervelocity interceptors might not require overflight, but they are not feasible in the near term.

- **A capability to locate, identify, and kill the TBM or its TEL on the ground (called counterforce).** A variety of concepts of operations that might enable this capability are under consideration, but high-confidence approaches remain elusive.\(^\text{34}\)

Obviously, killing TBMs while still on the ground would be a highly attractive capability, were it operationally feasible and affordable. However, as one might expect, the most attractive options are usually the most difficult to acquire. Figure 4 is a simple matrix showing relative difficulty in obtaining the capability versus relative attractiveness in having the capability. At the top of the attractiveness and difficulty axes is counterforce. Next in attractiveness and difficulty is boost-phase kill, and then terminal or midcourse kills. The last

---

\(^{34}\)There are no confirmed cases of successful Scud hunts in Desert Storm.
item is killing the TEL after the missile has been launched. Without its mis-
sile, the TEL may not appear to be a very attractive target, but if there are
multiple reload missiles per TEL, destroying one would limit the number of
TBMs launched subsequently.

Unfortunately, none of the approaches provides by itself a robust capability.
Table 3 summarizes attractiveness for the various systems along a number of
dimensions. A “+” indicates major strength for the option, a “−” indicates a
negative impact, and a “0” lies somewhere in between. This table reflects my
judgments; the reader should feel free to substitute his or her own.

Whether or not one agrees with the pluses and minuses of Table 3, it re-
 mains true that no single approach is likely to simultaneously provide high kill
probabilities and robustness to countermeasures or other technical or scenario-
related uncertainties. A combination of approaches is therefore needed in the
long run, one that would, among other things, provide substantial layering of
the defenses.

Having surveyed TMD issues, let us now turn again to NMD and then to
global missile defense.

NATIONAL MISSILE DEFENSE

The Bottom-Up Review (Aspin, 1993) has decisively put NMD develop-
ments onto the back burner of DoD priorities. Nevertheless, NMD remains
Table 3
Strengths and Weaknesses of TMD Options

<table>
<thead>
<tr>
<th>Available</th>
<th>Counter-force</th>
<th>Boost-Phase Intercept</th>
<th>Midcourse or Terminal Area Intercept</th>
<th>Postlaunch Counterbattery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relocatable precrisis</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Rapidly deployable</td>
<td>+</td>
<td>+</td>
<td>–/0</td>
<td>+</td>
</tr>
<tr>
<td>Exportable</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Effective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detectable</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Lethality</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>Insensitive to intel. data</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>Insensitive to responsive</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>threats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limits collateral damage</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>Provides good BDA</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>Synergistic</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large coverage</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Subject to attrition</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Multiple mission potential</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Available BM/C³I</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In DoD plan</td>
<td>0</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Operationally sound</td>
<td>–</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Overall comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technically risky;</td>
<td></td>
<td>Relies on control of air; opportunity cost may be large</td>
<td>Best near-term option; important as underlay; not robust re responsive threat</td>
<td>Uncertain value; couples with boost-phase and counter-force</td>
</tr>
<tr>
<td>Potential unclear c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Depends on availability of friendly bases in neighborhood of threatening country.

*Assumes shipborne or airborne midcourse defenses.

*Depends inherently on "finder's" success in hider versus finder competition. Fiscal and operational costs of a high-confidence solution would be extremely high unless finder's ability to successfully search for and locate TBM's in large areas is very high.

an issue within DoD and Congress. After all, what is the long-term logic of defending abroad, but not at home? The near-term goal of NMD would, presumably, be to protect the entire U.S. population against an accidental or unauthorized attack from one of the states of the former Soviet Union (e.g.,
Russia), and from other adversarial states or terrorist groups that might develop ballistic missile threats to the United States. This goal will not easily be met. And in meeting it, the United States will have to face additional problems such as renegotiating the ABM treaty.

The ABM treaty limits ABM deployments to no more than 100 interceptors and related tracking radars at a single site. This section first examines the capability of a treaty-limited NMD, assuming that Grand Forks continues to be the U.S. site. Then it considers noncompliant deployments, where ground-based radars (GBRs) and ground-based interceptors (GBIs) could be located elsewhere.

**Treaty-Limited Deployments at Grand Forks**

The capability of an SBM interceptor launched from a particular site to intercept a warhead aimed at a specific target in the United States rests on a number of important variables. Among the most important are:

- The energy of the GBI (i.e., its “delta-V”), which translates into range and altitude versus time.
- The acquisition and target-tracking sensor coverage (obviously, the earlier the detection of the RV, the earlier the interceptor can be launched and the greater the coverage).
- The command-and-control decision time.
- The defense system track-handling capability, including numbers and accuracy of the track.
- The midcourse target-discrimination capability.

The capabilities of single-site deployments are also sensitive to the launch location of the ICBM (or SLBM) in relationship to the intended targets. Figure 5 shows this by indicating the coverage of a typical GBI site if located at Grand Forks.\(^{35}\) The coverage represents the area that is protected against an attack from any point in Russia. Thus, a single site at Grand Forks can in principle *guarantee* that 55 percent of CONUS (the contiguous states, not including Alaska or Hawaii) is protected against an attack from Russia, regardless of the location of the ICBM launch.\(^{36}\) For any specific launch location of the ICBM, however, the actual fraction of CONUS protected would be

---

\(^{35}\)Most of the figures in this section are derived from earlier RAND work by Michael Miller, Herbert Hoover, and Susan Everingham.

\(^{36}\)These calculations are representative of, but not identical to, the estimated performance of current Ballistic Missile Defense Organization GBI concepts.
substantially higher. For example, most of the east coast of the United States would be protected against attacks from central or eastern Russia, while much of the west coast would be protected from attacks from launches near Moscow. Thus, the coverage area shown is the "worst case" coverage in terms of launch location.

Not surprisingly, as the geographical extent of the combined ICBM threats increases, the coverage worsens. In the limit, threats that can come from any direction—as would be approximately true for Russian SSBNs if they reestablished patrols off the coasts of the United States—pose the most serious coverage problem. It is intuitively obvious that attacks from the south or from close to the United States cannot be engaged by a single GBI site in North Dakota. Furthermore, and also fairly obviously, a GBI site at Grand Forks cannot provide protection for either Hawaii or Alaska. If the entire United States is to be protected, or if threats take on a global character, restricting U.S. BMD to the structures of the ABM treaty cannot be permitted.

CONUS territorial coverage (or its equivalent in fraction of CONUS population protected\(^{37}\)) is but one measure of defense performance. A second major consideration is leakage. Even assuming small attacks, achieving low leakage rates is likely only if adequate battlespace exists for shoot-look-shoot opportunities.\(^{38}\) Imposing a requirement for shoot-look-shoot opportunities sharply reduces the coverage of any particular BMD site. As a result, a single site at Grand Forks provides coverage and shoot-look-shoot opportunities only for that part of CONUS lying between the Appalachian and Rocky Mountains. The majority of the population that lives near the two coasts gets neither.\(^{39}\)

---

\(^{37}\)Because the population of CONUS is more heavily concentrated near the coasts, the fraction of area coverage does not map directly into population coverage. This is an important factor for single-site deployments, but gets less important if additional GBI sites are deployed.

\(^{38}\)Shoot-look-shoot (SLS) assumes that one or more GBIs are launched against an RV and engage at some specified distance from the launch site (the first "shoot"). After that engagement, a determination is made as to the success of the first engagement (the "look"). If success is doubtful, then a second launch of one or more GBIs would occur. This second launch is possible only if there is sufficient "battlespace" for the first shoot and look opportunity to occur before the RV passes through the engagement envelope of the GBI site. Single look-shoot capabilities are usually judged inadequate for high kill probabilities because of the belief that high single-shot kill probabilities are technically difficult to achieve, and multiple salvo shots may not be adequate (because of correlated error sources) to raise the single-shot kill probability to levels needed for very low leakage rates against multiple RV attacks.

\(^{39}\)Typical calculations from prior RAND research estimate that the 55 percent coverage of CONUS from a Russian attack yields population coverage of less than 40 percent. Whether SLS will be sufficient to achieve the desired high kill probabilities rests on other factors (e.g., multiple shots may still have correlated error sources).
Extending CONUS Coverage

One relatively cost-effective way to improve both the area coverage and the shoot-look-shoot opportunities is to augment the sensor support to the GBR. Figure 6 shows how both percentage of CONUS coverage and the shoot-look-shoot capabilities increase for various BMD system sensor enhancements. Three enhancement options are shown: (1) placing additional GBR sites around the periphery of CONUS, (2) adding to this option radar upgrades to existing BMEWs sites in Greenland and England, and (3) deploying a space-based target acquisition and tracking system (one such option would be Brilliant Eyes). All three options are arguable violations of the ABM treaty, but because Russia has its own motivation for wanting extended sensor-coverage capabilities, it might not challenge these enhancements seriously and might entertain treaty changes more generally.

Particularly attractive on grounds such as cost-effectiveness would be a space-based targeting system. The Bottom-Up Review recommended continued development of a long-wavelength infrared (LWIR) sensing satellite system (e.g., Brilliant Eyes (BE) or a less expensive theater version) that could do the needed job nicely.\(^{40}\) BE would track an RV early in its midcourse flight, permitting GBI commitment before the RV comes over the horizon. This early-launch opportunity substantially increases coverage opportunities. In the extreme, it leads to calculations that show intercepts occurring over Miami by GBIs launched from Grand Forks, a feat that probably needs to be demonstrated before reasonable people would accept these calculations as realistic. The Brilliant Eyes program will almost certainly slip to have a timeline consistent with that of NMD.

Effective NMD Deployments and Implications for Stability

Figure 7 shows how CONUS defense protection improves as a function of the number (and location) of additional GBI sites deployed in CONUS (Alaska and Hawaii each require an individual GBI/GBR site for their protection). Both territorial and population coverage are displayed. Additional sites are deployed around the periphery of the country, mainly augmenting protection against attacks that might come from directions other than the north.

\(^{40}\)BE would also help TMD, particularly against longer-ranged TBMs, and as such could be a valuable complement to defenses protecting U.S. friends and allies (see, e.g., Best and Bracken, 1993). The version needed would not need to be so "brilliant" or expensive as that envisioned for GMD (Canavan, 1993:7).
Figure 6—Providing Grand Forks GBI Deployment with Enhanced Sensor Support
Figure 7—Multi-GBI Coverage of CONUS Territory and Population
Two cases are shown on each chart; one that assumes Grand Forks as the first site and then adds new sites in an incremented way (the solid curve), and one that allows sites to be selected optimally (the dashed curve). In all cases the fractional coverage of CONUS approaches one after a deployment of about four sites.

The effectiveness of the defense is also a function of the size of the attack, and especially so because of the ABM treaty limit on GBI deployments of 100. Figure 8 displays the total number of reliable GBIs that would need to be (optimally) deployed to successfully engage all but 3 percent of the RVs. Two attack sizes are shown, one for 50 RVs, the other for 200. The larger size exceeds the 100-RV threat that represents a nominal maximum for an unauthorized attack from a Russia compliant with START II. The indicated stockpile sizes, ranging from 800 (with only GBR sensor support on U.S. territory) to about 400 (with space-based sensors), should suffice to handle all future attacks.

![Figure 8—GBIs Required to Achieve No More Than 3 Percent Leakage](image-url)
except those that might be deliberately launched from a newly hostile Russia.\textsuperscript{41}

Of course, even 400 GBIs exceed the treaty limits. One hundred GBIs could provide low leakage protection against attack sizes of less than about 40 RVs—e.g., small accidental attacks from Russia as well as foreseeable Third World attacks from all countries except perhaps China. The treaty would still have to be altered, however, because of the need for geographical coverage.\textsuperscript{42}

How likely is it that the Russians would accept a renegotiation of the ABM treaty, allowing GBI deployments of 400 or more? Obviously, 400 interceptors would not stop a fully alerted Russian ICBM or SLBM attack from devastating the United States, even after full implementation of the START II treaty. However, if such GBI deployments were combined with a surprise U.S. first strike against Russia's (nonalerted) nuclear forces, then it is \textit{analytically} plausible that the follow-up Russian ballistic missile attack might be "manageable."\textsuperscript{43} Thus, there would be the potential for first-strike instability (see also Wilkening, 1994), by which I mean that from a mathematical perspective there might be a significant incentive for conducting a first strike because if the second side attacked first, the first side would have less than a minimum retaliatory capability.\textsuperscript{44} In some situations (i.e., with some combinations of defense capabilities), only one side would have such a first-strike incentive; in others, both would; and in still others, neither would.

\textsuperscript{41}These force sizes are sensitive to the permitted maximum leakage levels. The value of 3 percent assumes low, but not zero, leakage. In a real deployment, this value would be weighed against other nuclear delivery options available to the opponent. Of course, for any given size deployment, lower leakage rates can theoretically be achieved with a given stockpile by "doubling up" on interceptor launches, salvoing two or more GBIs for each "shoot" opportunity. Thus, for most anticipated attacks, the leakage rate could be closer to zero than shown here.

\textsuperscript{42}Because complete coverage is required, some GBIs cannot participate in engagements against threats launched from unfavorable locations. This "absentee" factor varies with interceptor capability, the sophistication of the sensor coverage, etc. For many cases, the factor is about 30 percent, i.e., about 30 percent of the GBIs are in the wrong location to engage a specific attack.

\textsuperscript{43}"Manageable" is in the eye of the beholder. It is difficult to imagine that a U.S. president would find even a few nuclear detonations on U.S. cities acceptable under any circumstance where another option is available. To a Russian leader, however, the inability to threaten wholesale damage to the United States as a counterbalance to a similar U.S. capability may be cause for serious concern.

\textsuperscript{44}For discussion of first-strike stability focused more on decisionmaker psychology, notably facts of desperation and fear, see Davis (1989), which disparages the purely mathematical approach to stability assessments but notes how the mathematical characterization enters the problem.
Figure 9 illustrates this using a “defense-domain plot.” The discussion here is qualitative, but for more sophisticated and quantitative results see Shaver (1986), Warman and Wilkening (1986), and Kent, DeValk, and Thaler (1988). In Figure 9, which is drawn for a particular assumed set of offensive capabilities more or less of the sort likely after START II takes effect, the abscissa and ordinate axes represent U.S. and Russian defense capabilities, respectively. They extend from none (no capability) to perfect capability to stop all attacking weapons. Note the existence of different zones. The

---

45I first used this particular graphic description for examining first-strike stability in a 1983 RAND study with James Thomson.

46These charts can be drawn with alternative variables for the axes. For example, Kent, DeValk, and Thaler (1988) proposed use of a variable called “defense potential.” This variable is calculated by multiplying the number of defense interceptors times their expected kill potential against the threat. This variable has the attractive feature of allowing comparison of dissimilar offensive force postures, showing that “equal” defense deployments do not lead to balanced capabilities. See also the later publications Kent and Thaler (1989, 1990).
shaded zones correspond to conditions where the combination of Russian and U.S. defense deployment provides to one or both sides some advantageous first-strike capabilities. In a severe crisis these advantages may be cause for instability. Instabilities are greatest wherever the zones overlap. The situation is stable in the unshaded zones, i.e., where defenses are absent or where both sides possess near-perfect defenses.

Figure 9 is merely illustrative. The actual plots depend sensitively on many details of both offensive and defensive posture, the assumed exchange scenario, and other factors, but a few general observations can be made:

- The instability regions tend to grow larger as the total size of the nuclear arsenals shrinks. Thus, START II force postures are more likely to be destabilizing with defensive deployments than those planned under START I. Further force reductions could exacerbate this situation.

- Ameliorating this is the fact that the regions shrink with reduced vulnerability of the offensive forces. Stable transition regimes are readily obtainable if both countries pay attention to insuring the invulnerability of their nuclear forces.

- The regions examined so far pertain mainly to ballistic missile forces. Bombers and air defenses interact in slightly more complicated ways, tending to further reduce the instability domains. Most mid-to-late 1980s calculations using realistic postures and planning factors for nuclear exchanges showed only small or no instability regimes, primarily because of the air breathers. It is noteworthy, however, that the air-breathing forces may have very low alert rates in the future.

So, reasonable actions to care for the survivability of the offensive forces sharply diminish instability concerns, obviously lessening any incentives for first strike while leaving ample room for both countries to deploy defenses that can provide the desired protection against Third World threats. In real-world terms, this means that U.S. defense deployments are only as destabilizing as Russian leaders are willing to allow them to be. After all, Russia controls the vulnerability of its forces to a U.S. first strike. Obviously, the United States can affect this situation by reducing its own first-strike capabilities. But a Russian commitment to higher alert rates and secure operating bastions for its SSBN fleet should mitigate if not outright eliminate any concern about U.S. GBI deployments numbering in the many hundreds.47

These defense domain curves also say something about the prospect of cooperative defense deployments leading toward the defense-dominant world of Present Reagan. Just being cooperative on defenses wouldn’t prevent instabil-

47See Best and Bracken (1993) for an elegant discussion of n-sided stability calculations (i.e., calculations that consider many "sides" rather than merely the United States and Russia).
ity regions, but if instability regions can be avoided through careful management of both offensive and defensive deployments, then it is possible to reach the offensive-impotent region wherein neither side has lost its assured retaliation capability but each has denied to the other all options save attacking urban areas. This region would be highly stable. It is a possible target point for cooperative defensive deployments that would avoid the difficulty of finding a way for both sides to abandon their nuclear deterrence capabilities altogether.

But to what point? Defenses against Third World threats can be achieved with relatively small defense deployments, as can protection against small attacks from Russia. Stability is not an issue; the current situation of little or no defense is itself quite stable. And there is no assurance that large deployments are a positive step toward total protection from nuclear attack; at the least, covert weapon delivery capabilities would probably still exist. Moreover, defenses to reach the upper right portion of the defense domain would be extraordinarily expensive.

All of this suggests that middling defense deployments, short of those that provide meaningful damage limitations, are difficult to justify. Limited BMD deployments can be quite valuable for limiting damage from small attacks, but deployments of a much larger magnitude may only be more costly and more dangerous.

GLOBAL BALLISTIC MISSILE DEFENSES

So far, we have discussed theater and national ballistic missile defenses. Let me now also discuss the kind of global missile defenses (GMD) sometimes envisioned by President Reagan a decade ago. These would not only protect the United States, but provide protection for selected nations worldwide. As a practical matter, GMD virtually implies a layer of space-based defenses, because a number of studies over the last decade have indicated that only this could provide the desired comprehensiveness and robustness to countermeasures. The attractiveness of getting ballistic missiles in boost flight is obvious, but especially so when the missiles are outfitted with multiple RVs or penetration decoys. Without space-based defense overlays such as envisioned in the Brilliant Pebbles (BP) system, it would also be nearly impossible to achieve

---

48 Among the options that merit consideration is the possibility of adapting excess Minuteman missiles as exatmospheric interceptors. Coupling Minuteman with the LEAP vehicle and an early version of the GBR could provide the United States with a credible limited NMD system early (Cooper, 1993).

49 While these studies strongly suggest that space-based defense would be a necessary element of a comprehensive and robust strategic defense, there remain serious questions about whether such defenses would be sufficient.
an acceptable cost-exchange ratio of U.S. defenses over Russian sophisticated offenses.

Although the Clinton administration has reduced Brilliant Pebbles and other space-based weapon concepts to at best technology status, killing any near-term prospects for GMD, it is quite possible that interest in GMD will come back in time, perhaps in the context of cooperative U.S.-Russian efforts.

To have an effective boost-phase kill potential, space-based kinetic kill (KKV) satellites (e.g., the Brilliant Pebbles system) must be in low earth orbits. This results in a significant satellite absentee factor—that is, a major fraction of the satellites are located over parts of the earth from which no missiles are being launched, and thus they cannot participate in the defense.\textsuperscript{50} The absentee factor is sensitive to the total time that the booster is in powered flight. These times vary widely, being shorter for shorter-range missiles (e.g., TBMs) and for solid-propellant missiles. The large Russian missiles (e.g., the SS-18) have long burn times and are the most attractive targets for BPs. Figure 10 displays the potential effectiveness of deployments of BP in terms of the number of salvo-launched boosters that a constellation of orbiting BPs could intercept. Because of the absentee factor, the number of kills is small compared to the number of BPs deployed. Note also that TBMs can be intercepted, although the number of shorter-range missiles that could be handled is quite small.

Some immediate conclusions are possible. First, Brilliant Pebbles by itself does not constitute a very robust defense, at least in deployment numbers up to 2000.\textsuperscript{51} Based on substantial RAND work by Michael Miller, Susan Everingham, and Herbert Hoover, it seems that ground-based defenses are almost certainly a more cost-effective approach to defend against limited attacks than a Brilliant Pebbles approach. The most likely future roles for space-based defenses are (1) as augmentation for ground-based defenses if defense against large and sophisticated attacks is required, and (2) as gap-fillers for both TMD

\textsuperscript{50}The absentee factor is especially severe for kinetic kill vehicles, exceeding 95 percent under many circumstances. The factor is greatly reduced if these vehicles are capable of killing RVs after booster burnout. When most ICBMs were MIRVed, that MIRVing inflicted its own multiplier penalty on midcourse killers, but with START II, MIRVed ICBMs (not SLBMs) should be eliminated.

\textsuperscript{51}Earlier studies by SDIO and others looked at BP deployments of 20,000 and larger. Even with an absentee factor of 95 percent, 20,000 BPs could theoretically engage 1000 ICBMs, even if launched nearly simultaneously. Given the small likelihood that this many boosters would be launched within a small time period, along with the fact that the remaining 19,000 BPs would still be in orbit and available for employment (unless the attacker timed its salvos to correspond with “holes” that the initial engagement created in the space-based defense), deployments of this magnitude could in theory provide a very robust strategic defense capability.
Figure 10—Impact of Constellation Size on BP Interception Potential

Salvo of 20 missiles; boost/boost/postboost/midcourse intercepts permitted

Averaged over threat launch time and direction

Expected missile payloads negated

Worst-case threat launch time and direction

Expected missile payloads negated
and NMD deployments where coverage gaps exist. As already discussed, one can imagine scenarios where other TMD systems either cannot or will not be in position to defend critical theater targets when needed; BP could play a (perhaps limited) role in filling this need. However, where other options exist it is doubtful that space-based systems are the most cost-effective approach to provide this coverage. Space-based sensors such as BE are more clearly valuable (see, e.g., Figure 8).

ISSUES, TENSIONS, AND DILEMMAS IN CONTEMPLATING THE FUTURE OF BALLISTIC MISSILE DEFENSES

Given the Bottom-Up Review, it is highly doubtful that Congress will authorize sufficient funds to support full development of both NMD and TMD systems. It is easy to anticipate that for the next few years, concern about Third World TBM threats will carry the day for TMD developments. Whether NMD will be resurrected depends on whether there are credible threats and credible ways of dealing with them.

Is There a Credible Threat Justifying National Missile Defense?

The argument against NMD goes something like this and depends on balancing U.S. national interests: The TBM threat is already real and growing. Especially if outfitted with weapons of mass destruction, TBMs in the hands of our adversaries clearly threaten U.S. national security interests. Furthermore, it is far from clear what will be required to deter some of our potential adversaries from executing this threat. In contrast, ICBM threats to the U.S. population and its infrastructure, while real, are diminishing. Moreover, the countries that currently possess ICBMs (and SLBMs) that could threaten the United States are judged to be deterrable. Thus, although lacking BMD, the United States is protected by its offensive retaliation forces, and strategic defenses hopefully are not needed, just as they weren’t required for U.S. survival in the past.

The response to this argument, which is an argument for at least laying the base for a future NMD, starts with the assertion that even a single nuclear detonation on a U.S. city would be a disaster of unprecedented proportions for the United States. Regardless of cost-effectiveness considerations, American voters would demand an explanation as to why the United States was spending many billions on theater missile defenses to defend foreign countries but not equal or greater amounts for protection for itself. This is a question that is likely to arise even before such a detonation occurs, and reasonable answers will be required.
Ballistic missile threats against the United States can come from three sources: (1) the residual nuclear stockpile of Russia (or, conceivably, other states of the former Soviet Union), either deliberately or accidentally launched against the United States; (2) the growing but still small inventory of nuclear delivery vehicles in other nuclear states (e.g., China); or (3) a now existing, but plausible, presence of ICBM threats from as-yet unidentified Third World countries or cross-national terrorist organizations. Because the last source of threat is likely to be the most difficult to deter, a few additional words about it are warranted.

It may be ironic, but successful U.S. counter-TBM deployments could actually motivate Third World development of ICBMs. Frustrated by being unable to prevent U.S. intervention, countries like Iraq may see ICBMs as an effective means to deter the United States. The technologies needed for (inaccurate) ICBMs are widely proliferated around the world, and it is to be expected that space-launch vehicles will at some time be converted for this purpose. The Indian official’s observation about the main lesson from Desert Storm being “don’t fight the United States unless you’re armed with nuclear weapons” is a view probably shared by many Third World leaders, including Khadafy. If the United States could defend regionally against TBM’s possessed by antagonistic Third World nations, what would deter these nations from seeking capability against the United States itself?

How would the United States react to a Third World country that built an ICBM with an apparent nuclear weapon on top? Even if we could find the missiles (i.e., even if they were not protected by mobility, basing in caves, or whatever), would we preemptively strike the launch or storage site in hopes of destroying the threat before it could be launched? Would we do this before the onset of a crisis involving our vital national interests, or would we wait until a crisis arose? And if the President had doubts about the certainty of our capabilities to destroy the threatening missiles before their launch, would he nevertheless press ahead with plans to deploy our forces into the region of crisis? Importantly for this discussion, how would answers to these questions change if the United States possessed a limited BMD system?

With or without defenses to underwrite U.S. commitments, it is hard to state how future U.S. presidents would answer these questions. But a not unreasonable speculation is that only with high-confidence defenses would they

---

52See Garrity (1993) for a wide variety of foreign views about the lessons of the Gulf War, including the role of nuclear weapons.

53When I posed some of these questions to my colleagues, the answers varied widely, ranging from “we would be absolutely deterred” to “we would turn their country into glass.” Whatever the answers, one observation was common: We need to develop strategies for dealing with countries whose leaders hold different values than our own.
feel free to press forward with U.S. power projection assets, protecting our vital interests and those of our friends and allies around the globe.

Countering Other Nuclear Delivery Means

Of course, ICBMs are not the only means to threaten the United States with nuclear attack. The above list of questions applies as well to nuclear attacks delivered by aircraft or ships. If we deploy a NMD to stop a hypothetical Third World ICBM threat, what is to prevent that threat from constructing an aircraft threat, or a cruise missile threat? And if it did, would the United States be prepared to defend against it?

The answer to the last question rests in large part on the adequacy of intelligence. The United States has ample aircraft and ships to protect itself against small attacks by either airplanes or ships, so long as it knows that it is under attack and which vehicle (or collection of vehicles) constitutes the attack. Such defenses already are part of the existing military structure and come at no cost.

Not as clear-cut is whether adequate intelligence exists or will exist to the extent needed to provide U.S. decisionmakers with the needed warning to implement appropriate defensive measures. The warning must be both timely and highly credible. How this intelligence requirement can or will be met is unclear.

But Must We Live Forever Under the Threat of Nuclear Attack?

Against this background of history, new and emerging threats, and technological options, how should we think about managing the nuclear threat and the role BMD could play? Those who set long-term priorities for BMD developments should certainly address how best to manage the fact of the existence of nuclear weapons (and, increasingly, the threat of biological weapons) around the globe. Although START II promises to sharply reduce global nuclear inventories, international politics and human nature argue that abolishing nuclear warheads altogether may not be possible. Does this mean, therefore, that our children, our children's children, and their children will have to live under the constant threat of nuclear annihilation? And how do we deal with the reasonable fear that, sooner or later, some zealot or crazyman will gain access to such weapons and use them despite efforts at deterrence? Such an occurrence may be very unlikely, but events with very small probabilities happen nevertheless.

It may be useful to review the approaches available to the United States to manage this threat.
• **Active defense.** This is, of course, the principal subject of this paper. It is a subset of a more general category, denial. Denial includes direct counter-force attack as well as active defense, essentially denying to the threatening country the ability to successfully attack the United States.

• **Deterrence.** This promises the potential attacker that nothing can be gained and much will be lost by attacking the United States. It tries to influence the attacker's view of both his cost from the attack as well as his benefit from not implementing the attack.

• **Disarmament.** In its broadest sense, this seeks to deny nuclear arms to those who don't currently possess them (through the Non-Proliferation Treaty or other means) and the abandonment of nuclear weapons by countries that do have them (by international agreement, perhaps along the lines of the Baruch plan of 1946).

Active defense has the attractive feature that, to some extent, the possessor is in control of his own fate. However, as studies over the last 40 years have clearly demonstrated, it is extremely difficult and quite possibly impossible to stop all nuclear attacks by a dedicated opponent. Defenses might stop the first attack, the second, and the third, but sooner or later something is likely to get through. Thus, so long as nuclear weapons exist, active defenses need to be underwritten by deterrence, i.e., by retaining capable nuclear offensive forces. This creates the burden not just to build and maintain near-leakproof defenses, but also to maintain offensive weapons that are at least as capable as any enemy's of getting to their targets. This is a far cry from the pure notion of defense dominance, where the moral dilemma of relying on nuclear retaliation would not exist.

Deterrence is hardly a sturdy reed for the indefinite future, resting as it does on the rational behavior of our opponents. But a true psychopath rarely reaches a leadership position, even in the Third World. Deterrence has worked for the past 40 years and may work for 1000 more. Who knows? But for those who would abandon it, what will replace it? Not active defense by itself, as explained above.

Deterrence also rests on the credible threat of a response that will raise the cost of the attack above its benefits as seen by the side that one is trying to deter (see Davis (1994a) for discussion of deterrence in the post-Cold War era). But it is far from clear just what threatened response the United States would use to deter the use of WMD against its forces in the field or against its friends.

---

54 This may seem contradictory, but it is obvious once we recognize that the country under attack must force a price on the attacker for his first attack, even if that attack is unsuccessful. Without such a price, the attacker has no incentive to stop such attacks until success is realized.
and allies. Except for direct attacks against the homeland of the United States, the credibility of U.S. nuclear use is subject to doubt.\footnote{Even in this case, some would question whether a U.S. president would order a nuclear attack against the offending country, especially if extensive collateral damage to that country's population was likely. Such a decision would be particularly difficult if it were felt that the country's leaders did not reflect the sentiment of its people, and that the people should not be held responsible for the actions of their leaders.}

That leaves us with nuclear disarmament. The attempts to control nuclear technology and weaponry immediately after World War II were doomed to failure from the start. No country, including the United States, was ready to empower an international organization with such authority (Bundy, 1993). However, without the involvement of an international body to monitor and enforce any agreement, nuclear disarmament is most likely impossible. The process will almost certainly fall victim to power politics.

Thus, proponents of total nuclear disarmament face a serious dilemma: push for a process that the United States and other major nuclear states will almost certainly reject, or accept the likelihood of failure. Under current international circumstances it is hard to foresee conditions where total nuclear disarmament can be achieved.

Finally, there are a number of compelling reasons to believe that the spread of nuclear weapons will continue. Realistic approaches to stopping this spread are lacking. Draconian approaches, wherein some state (e.g., the United States) takes a unilateralist approach and acts to confront, deter, and if necessary disarm those states that brandish such weapons, have many flaws. Defenses will help delay the holocaust, but direct action may be needed to deny it.

"In for a Penny, in for a Pound"

All the above suggests that some combination of policies—active defense, deterrence, and (perhaps in the long term) setting the conditions for the formation of an international organization to help reduce nuclear weapon stockpiles—needs to be applied to manage the nuclear threat, hopefully leaving our children more safe and secure than we are today. Active defense can and should play an essential role, providing needed protection where deterrence fails or where the credibility of it is weak.

However, it is worth raising a question asked earlier: "How much is enough?" There is no logical stopping point for deployments of either TMD or NMD, short of near-perfect defense. Nor does there appear to be a high likelihood of turning such deployments around once started. Much like nuclear deterrence, active protection will not be abandoned easily, even if the
costs of maintaining it start to grow sharply. Once we are firmly down this path, there may be no turning back. Figures 11 and 12 portray two alternative outcomes associated with starting down the BMD path. One is generally favorable. The other is the opposite. Both are plausible. In either case, the world will not be the same.

FINAL OBSERVATIONS AND SUGGESTIONS

The above discussions on BMD lead to the following observations and recommendations, which are unabashedly subjective in some cases.

- The TBM and WMD threat already exists and is likely to become worse over time. It must be challenged, or U.S. global interests will be severely threatened. TMD is an important ingredient in countering that threat and deserves the priority it received in the Bottom-Up Review.

- However, the Bottom-Up Review’s focus on terminal or late-midcourse defense needs to be broadened. The shortcomings of such defenses can be

<table>
<thead>
<tr>
<th>Threats</th>
<th>Responses</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near-term threats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBMs vs. U.S. forces</td>
<td>Defend with TMD</td>
<td>Protects U.S. forces</td>
</tr>
<tr>
<td>TBMs vs. U.S. friends and allies</td>
<td>Nonproliferation of WMD</td>
<td>U.S. power projection remains practical</td>
</tr>
<tr>
<td>ICBMs vs. U.S. (unauthorized launch)</td>
<td>Reduce threat through de-alarming, permission action links (PALS), etc.</td>
<td>Further proliferation of WMD and TBMs deterred</td>
</tr>
<tr>
<td>Far-term threats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICBMs vs. U.S. (from Third World countries)</td>
<td>Nonproliferation of ballistic missile technology</td>
<td>Protects U.S. population vs. limited attacks</td>
</tr>
<tr>
<td>ICBMs vs. U.S., friends and allies (from new major powers)</td>
<td>Defend with NMD</td>
<td>Eliminates antagonistic Russian response</td>
</tr>
<tr>
<td></td>
<td>Global cooperation on BMD deployments</td>
<td>Protects friends and allies</td>
</tr>
</tbody>
</table>

Figure 11—One Outcome of BMD Developments: U.S. Strategic Interests Preserved and Strengthened
Figure 12—Another Outcome of BMD Developments: U.S. Strategic Interests Harmed with Little Protection for U.S. Population or Deployed Forces

exploited by future aggressors, greatly diminishing TMD effectiveness. Counterforce, counterbattery, boost-phase, and early-midcourse intercept options should all be pursued. If necessary, THAAD development should be slowed to provide additional funds for other high-leverage counter-TBM options.

- There are reasons to be concerned about potential unauthorized attacks emanating from a turbulent Russia. However, the likelihood of such attacks is very small, given Russian security measures. Deliberate attacks from Russia and other countries (e.g., China) are also very unlikely, based on deterrence. On this basis, the risks associated with abandoning plans for near-term deployments of NMD are acceptable.

- However, the appearance of a Third World ICBM threat, while far from a certainty, could also severely threaten U.S. global interests. In this case, deterrence would work against the United States. A NMD may be a prerequisite to managing this threat. Thus, NMD developments should be kept on a timeline consistent with having a limited NMD deployment option available at the time that such a threat might appear.

- Because it provides unique capabilities, a space-based midcourse tracking system (e.g., Brilliant Eyes) should be developed. The timing of such a development should be paced to match threat and budget concerns.
• The importance of space-based BMD weapons rests on their ability to add important capabilities to already deployed TMD and NMD systems. It is unclear when and if these capabilities will be needed. Their current status as a technology demonstration effort seems appropriate.

These first six observations largely support current DoD policy, but they suggest a reexamination of the strategy being used to obtain robust TMD and NMD capabilities on the time scale needed. The next two suggestions cover broader issues associated with managing the long-term nuclear problem.

• Work with Russia to (a) augment the ABM treaty and (b) implement meaningful confidence and security building measures that would lower the likelihood of either accidental or unauthorized attacks from Russian soil or submarines.

• Develop a policy toward nuclear weapons that combines concerns about nuclear proliferation and both TMD and NMD deployments. That policy should address the most fundamental questions about the long-term role of nuclear weapons in U.S. national security objectives.

In conclusion, technology has offered the United States the opportunity to start down a new path—improving national security through deployment of active ballistic missile defenses—at a time when the need for BMD, at least TMD, was never clearer. However, not all of the consequences of BMD deployments are well understood. Over the next several years it will be important for U.S. policy analysts to explore these consequences, looking toward a future in which the global interests of the United States are protected and the fear of nuclear warfare is lessened.

BIBLIOGRAPHY

In addition to the references shown here, I made extensive use of current trade journals, in particular Aviation Week and Space Technology and Defense Week. Listing all the pertinent articles here would have would needlessly lengthened the bibliography.


Delaney, William P. (1992), “How Do We Respond to Theater Ballistic Missile Attacks?” MIT Lincoln Laboratory, February.


——— (1993), *BMDO Boost Phase Intercept Study*.


——— (1990), *First-Strike Stability and Strategic Defenses: Part II of a Methodology for Evaluating Strategic Forces*, RAND, Santa Monica, CA.


Millot, Marc Dean, Roger C. Molander, and Peter A. Wilson (1993), *The "Day After..." Study: Nuclear Proliferation in the Post-Cold War World* (summary volume), RAND, Santa Monica, CA.


Watman, Kenneth, and Dean A. Wilkening (1986), *Strategic Defense and First-Strike Stability*, RAND, Santa Monica, CA.


Wohlleber, Albert, Harry Rowen, Fred Hoffman, and Robert Lutz (1954), *Selection and Use of Strategic Air Bases*, RAND, Santa Monica, CA.

FUTURE U.S. AND RUSSIAN NUCLEAR FORCES: APPLYING TRADITIONAL ANALYSIS METHODS IN AN ERA OF COOPERATION

Dean Wilkening

With the end of the Cold War, the perceived need for a large U.S. nuclear arsenal has virtually disappeared. Thus, U.S. policymakers must confront the question of how low the nuclear arsenal should be cut, as well as whether strategic nuclear forces should be taken off alert as a symbol of the improved political atmosphere with the former Soviet Union. At the same time, interest in a limited nationwide ballistic missile defense system has increased because of concern with possible accidental, unauthorized, or deliberate third-country ballistic missile attacks. This paper applies a classical analysis of deterrence to the evolving U.S.-Russian strategic nuclear relationship. The aim is to identify potential problems that might be created during cooperative times as force levels are reduced, alert rates drop, and ballistic missile defenses deployed, so they can be avoided if tensions resurface between these two former Cold War adversaries. In general, deterrence appears to be robust with START II force levels so long as U.S. and Russian alert rates are not reduced dramatically from their current levels. However, relatively modest nationwide ballistic missile defense deployments (on the order of 600 ground-based interceptors or more) can seriously erode Russia's deterrent when U.S. and Russian forces are on day-to-day alert. This might cause Russia to react in ways making future crises difficult to control, thereby increasing fears (or the actual risk) of inadvertent nuclear war. These reactions could include rapid force generation early in a crisis or placing Russian ballistic missiles in a launch-on-warning posture, neither of which is attractive to the United States. The Russian deterrent posture is less robust than the U.S. posture on day-to-day alert because a smaller fraction of Russia's strategic nuclear forces would survive a hypothetical counterforce attack.

INTRODUCTION

With the Cold War over, U.S. relations with the states that once comprised the republics of the Soviet Union are undergoing rapid change. As part of this

For a more expansive treatment of the issues addressed in this paper, see Wilkening (1994).
change, the United States and Russia are reassessing the role that nuclear weapons play in their respective national security strategies. Under the rubric of “nuclear disengagement” the United States and Russia have called for deep cuts in their respective nuclear arsenals, as well as reductions in the peacetime readiness posture for the remaining forces.

The magnitude and speed of these reductions is obviously stunning by Cold War standards.\(^1\) In the emerging environment, it is natural to ask what impact these and possible future cuts will have on U.S. security. Or, put another way, how should U.S. and Russian leaders go about future nuclear force reductions so as to minimize the likelihood that old problems associated with the U.S.-Soviet Cold War nuclear standoff will reappear? One problem to be avoided is that deep cuts, if taken inappropriately or too deeply, could weaken the stable deterrent relationship that has existed between the United States and the former Soviet Union for decades. This paper explores ways in which the U.S.-Russian strategic relationship can be modified in cooperative times so as to avoid such problems if the U.S.-Russian relationship once again becomes adversarial. Specifically, it analyzes the impact of deep cuts (as embodied in the START II Treaty), reduced peacetime alert postures, and limited nationwide ballistic missile defenses (deployed to protect against accidental, unauthorized, and deliberate third-country ballistic missile attacks) on the deterrent relationship between the United States and Russia.\(^2\)

First, it is necessary to say a word about the possibility of future conflicts (not necessarily hot wars) between the United States and Russia, because if one truly believes that conflict is impossible, then the analysis that follows is irrele-

---

\(^1\) On July 31, 1991, the START I Treaty was signed, calling for a ceiling of 6000 accountable weapons in the U.S. and former Soviet strategic nuclear arsenals (a reduction of approximately 25–30 percent from the arsenals in existence in July 1991). Unilateral initiatives were declared by Presidents Bush and Gorbachev in the fall of 1991, reducing the tactical nuclear arsenals on each side to approximately 1600 weapons (a reduction of about 80 percent in the U.S. case) and removing all heavy bombers from heightened states of alert. On May 23, 1992, the Lisbon Protocol to the START I Treaty was signed, obliging Belarus, Kazakhstan, and Ukraine to eliminate all nuclear weapons formerly deployed or stored on their territory. And most recently, on January 17, 1993, the START II Treaty was signed, though it has yet to be ratified, further cutting U.S. and Russian strategic nuclear arsenals to between 3000 and 3500 total weapons. The START II Treaty represents the latest and deepest cut in both sides’ strategic nuclear forces, reducing the U.S. strategic nuclear arsenal to its lowest point since the early 1960s and the former Soviet arsenal to its lowest point since the early 1970s.

\(^2\) The analysis presented here could be extended in a straightforward manner to include other interesting nuclear interactions, for example, the United States and China, Russia and Great Britain or Russia and France, as well as Russia and Ukraine (assuming Ukraine remains a nuclear power despite its declaration to give up all former Soviet nuclear weapons stationed on its territory).
vant. Despite current cooperative trends between these two great powers, there are reasons to be cautious about concluding that conflicts could not occur. Indeed, U.S.-Russian relations have improved dramatically over the past several years. The Cold War is over. Global conflict, especially global nuclear war, is no longer a major threat. This is all for the good.

However, Russia is in the midst of a profound political and economic transformation—the outcome of which is far from predetermined. Major powers have rarely, if ever, undergone changes comparable to those currently underway in the former Soviet Union without a war or revolution. That these changes are occurring peacefully is truly remarkable, as well as highly desirable. Nevertheless, economic conditions continue to deteriorate, nationalism is on the rise, ethnic conflict is brewing, and crime is rampant. Ultimately, conditions could become so desperate that social cohesion within Russia may come apart. The strong turnout for Vladimir Zhirinovsky (a neofascist leader favoring restoration and expansion of the Soviet empire) in the 1993 parliamentary elections is sobering.

On the military front, the recent announcement of a new post-Cold War Russian military doctrine is some cause for concern, though most of it appears to be a straightforward extrapolation of former Soviet doctrine. The new doctrine notes that the possibility of global nuclear war between Russia and the West is no longer a major threat. Instead, the emphasis is on future regional conflicts on the Russian periphery and on possible internal conflicts. The use of force (including offensive and defensive operations) is justified, according to the doctrine, to protect Russia’s interests beyond her borders, e.g., to protect Russian-speaking peoples, as well as to quell domestic disturbances internally. Finally, the doctrine formally renounces the “no nuclear first use” pledge taken by previous Soviet governments, except against non-nuclear states that are not allied with a nuclear power. Consequently, in the future it is quite possible that Russia will increase its reliance on nuclear weapons to compensate for the

---

3 Given that virtually no one predicted the end of the Cold War, it is curious that so many people predict with confidence that the current cooperative trends will not come to an end. Few people have ever predicted with great accuracy when, where, and under what circumstances great powers will come into conflict. For the foreseeable future, Russia will remain a great power with interests that may diverge from those of the United States, and it will also have interests in common.

4 Also, some economists predict the imminent collapse of the Russian economy. See Lelyveld (1993).

5 The misleadingly named New Democratic Party captured 24 percent of the popular vote. The next-closest contender was “Russia’s Choice,” the party supporting Yeltsin’s reforms, with 14 percent of the vote. See Erlanger (1993) and Schmemann, (1993c).

6 See, for example, Hiatt (1993) and Schmemann (1993a,b).
weak state of its conventional military forces. In some respects, the Russian nuclear debate may come to resemble the U.S. nuclear debate during the Eisenhower administration, when such slogans as “defense on the cheap” and “more bang for the buck” reflected a greater emphasis on nuclear weapons in U.S. national security strategy to compensate for the high costs of maintaining robust conventional forces.

Under these circumstances it is premature to assert that U.S.-Russian relations could not take a significant turn for the worse, quite apart from the desire that this not occur. Such a reversal does not necessarily imply a return to Cold War tensions, if only because certain conflict scenarios seem virtually impossible today (for example, a Russian conventional invasion of Western Europe—a scenario that animated so much of the Cold War U.S. nuclear debate). However, the United States and Russia might find themselves embroiled in a future crisis (for example, over the territorial integrity of the Baltic states or Ukraine) where the underlying military balance, the strategic nuclear balance in particular, would be at least one factor influencing the dynamics of the crisis. Recognizing that nuclear weapons are less important today does not mean they are irrelevant, or that U.S. and Russian leaders should become blasé about the future of the U.S.-Russian deterrent relationship.

Therefore, though U.S. attention is focused currently on the problems associated with maintaining security and control over the vast former Soviet nuclear arsenal and nuclear weapons industrial complex, one must not forget that an enduring U.S. national security objective is to maintain a stable deterrent relationship with Russia while pursuing a foreign policy of increased cooperation on a range of issues vital to the security of both countries. In particular, U.S. foreign policy should promote political and economic reform within the former Soviet Union because, were it not for this transformation, the United States would still be embroiled in the Cold War, with its associated political/military tensions and high defense spending. Therefore, the United States should retain sufficient strategic nuclear force to deter any resurgent Russian threat, either against the United States or its allies, while avoiding threats that would undermine Russia’s strategic nuclear deterrent, since this could upset U.S.-Russian political relations and encourage a military backlash within Russia.

Nuclear deterrence rests on the ability to deliver a crushing nuclear blow against an adversary, even after absorbing a massive nuclear attack. Hence, as Albert Wohlstetter pointed out over 30 years ago, deterrence rests on the ability to deliver nuclear weapons to the opponent’s homeland and not on the mere existence of nuclear weapons in one’s arsenal (Wohlstetter, 1959). Therefore, deterrence requires forces that can survive an opponent’s counter-force first strike, operate reliably in a postattack environment, and penetrate any defenses the adversary has in place in sufficient quantity to threaten that
which the adversary values most. In short, deterrence depends on the existence of a secure second-strike capability.

As a general proposition, deterrence becomes weaker as force levels drop. The question is: How low is too low? There is no simple answer to this question because it depends on whom one is trying to deter, from doing what, under what circumstances, and what U.S. threats appear credible under these circumstances. Deterring a risk-averse, status quo power that values the survival of its cities (i.e., its population, industrial base, etc.) from attacking the U.S. homeland with nuclear weapons does not require a sophisticated strategy. Deterring a revanchist regime from reconquering territory it believes is part of its homeland with a large conventional invasion would be much more difficult—especially if the U.S. commitment to the country being attacked is not highly credible.

As a corollary, the only rational incentive for a premeditated nuclear attack against a nuclear-armed adversary occurs when the attacker's counterforce capabilities and defenses are so robust that the attacker can significantly improve his chances for survival by striking first; i.e., he can destroy enough of the opponent's nuclear arsenal so that the ragged retaliatory strike cannot do significant damage to the attacker (particularly if it has to penetrate defenses). Hence, this paper analyzes the effectiveness of future U.S. or Russian damage-limiting first-strike options as both sides reduce their strategic nuclear arsenals, reduce the alert rates associated with these forces, and consider deploying limited nationwide ballistic missile defenses.

The perspective advanced in this analysis is not that a weak deterrence posture necessarily leads to deliberate nuclear attacks, but rather that it forces the strategically disadvantaged side to react in ways to redress its perceived vulnerability. This reaction may appear provocative to the other side in the midst of a crisis, thus setting off a spiral of mutual suspicion and mistrust that increases the chance for misperception and miscalculation of the opponent's intentions. Though the likelihood that this spiral dynamic leads to intentional nuclear attacks may be small, unlike conventional scenarios, it will make crises more difficult to control, thereby increasing the chance of inadvertent nuclear war. Specifically, a vulnerable day-to-day alert posture does not imply that the

---

7Counterforce attacks are defined here to be attacks against strategic nuclear forces. In other words, they include attacks against ballistic missile submarine bases, long-range nuclear bomber bases, ICBM silos, mobile-ICBM garrisons, and possibly barrage attacks against mobile nuclear-delivery systems (e.g., submarines at sea, mobile ICBMs in the field, or bombers in flight). Attacks against strategic nuclear command-and-control systems and nuclear weapon storage and support facilities are not analyzed here, nor are attacks directed against theater or tactical nuclear forces.

8For a good discussion of spiral dynamics, largely drawn from historical examples of conventional conflict, see Jervis (1976).
opponent will launch a "bolt-out-of-the-blue" attack, because as a crisis develops, the vulnerable side can increase its alert rate, thereby increasing the fraction of its force that would survive a surprise attack. On the other hand, rapid force generation tends to be politically provocative because it often appears to the other side as a prelude to attack.\footnote{In this regard, the interesting strategic interaction is not the performance of U.S. and Russian arsenals when they are fully generated (the typical scenario for determining the overall size of the required force structure) but rather the mobilization dynamics as the United States or Russia move their forces from low to high states of alert; specifically, the extent to which either country might feel pressure to generate its forces rapidly in the early stages of a crisis, or perhaps to threaten to launch its vulnerable forces out from under an attack, because the peacetime alert posture does not provide an adequate deterrent.\footnote{To the extent such pressures exist, future crises may be more difficult to control. Suffice it to say that prudent planners should avoid creating relatively vulnerable U.S. or Russian strategic nuclear postures as both sides' forces are reduced. The likelihood of conflict with Russia may be small in the future, but with events this fateful, leaders should err on the side of caution.}} In this regard, the interesting strategic interaction is not the performance of U.S. and Russian arsenals when they are fully generated (the typical scenario for determining the overall size of the required force structure) but rather the mobilization dynamics as the United States or Russia move their forces from low to high states of alert; specifically, the extent to which either country might feel pressure to generate its forces rapidly in the early stages of a crisis, or perhaps to threaten to launch its vulnerable forces out from under an attack, because the peacetime alert posture does not provide an adequate deterrent.\footnote{For discussion of how psychology and perceptions could complicate a nuclear crisis, including assessment of force generation and efforts to de-escalate, see Davis and Wolf (1991).} To the extent such pressures exist, future crises may be more difficult to control. Suffice it to say that prudent planners should avoid creating relatively vulnerable U.S. or Russian strategic nuclear postures as both sides' forces are reduced. The likelihood of conflict with Russia may be small in the future, but with events this fateful, leaders should err on the side of caution.

**FORCES AND ALERT RATES**

Table 1 shows illustrative U.S. and Russian START II force structures. Both arsenals have been constructed to equal 3500 total weapons, despite the fact that President Yeltsin indicated in June 1992 that Russia might deploy only around 3000 strategic nuclear weapons. The larger Russian arsenal has been assumed because treaty ratification may require parity with the United States. The U.S. force structure is assumed to consist of 500 single-warhead Minuteman III ICBMs deployed in silos, 18 Trident submarines deployed with a mix of C-4 and D-5 SLBMs downloaded from 8 to 4 warheads each, and a bomber force of 114 nuclear-capable heavy bombers. The B-52H is assumed to be an air-launched cruise missile (ALCM) carrier, carrying both the
Table 1
Illustrative START II Forces (Weapons)

<table>
<thead>
<tr>
<th></th>
<th>United States</th>
<th>Russia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICBMs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minuteman II</td>
<td>0</td>
<td>SS-18</td>
</tr>
<tr>
<td>Minuteman III/3 RV</td>
<td>0</td>
<td>SS-19/1 RV</td>
</tr>
<tr>
<td>Minuteman III/1 RV</td>
<td>500</td>
<td>SS-24</td>
</tr>
<tr>
<td>Peacekeeper</td>
<td>0</td>
<td>SS-25 silo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SS-25 mobile</td>
</tr>
<tr>
<td><strong>Subtotals</strong></td>
<td>500</td>
<td>795(^a)</td>
</tr>
<tr>
<td><strong>SLBMs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-3 (Poseidon)</td>
<td>0</td>
<td>SS-N-18 (Delta III)</td>
</tr>
<tr>
<td>C-4 (Poseidon)</td>
<td>0</td>
<td>SS-N-20 (Typhoon)</td>
</tr>
<tr>
<td>C-4 (Trident)</td>
<td>8×24×4</td>
<td>SS-N-23 (Delta IV)</td>
</tr>
<tr>
<td>D-5 (Trident)</td>
<td>10×24×4</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotals</strong></td>
<td>1728</td>
<td>1744(^b)</td>
</tr>
<tr>
<td><strong>Bombers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-52H/ALCM-B</td>
<td>44×8</td>
<td>Bear-H6</td>
</tr>
<tr>
<td>B-52H/ACM</td>
<td>50×12</td>
<td>Bear-H16</td>
</tr>
<tr>
<td>B-1</td>
<td>0</td>
<td>Blackjack</td>
</tr>
<tr>
<td>B-2</td>
<td>20×16</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotals</strong></td>
<td>1272</td>
<td>960(^c)</td>
</tr>
<tr>
<td><strong>Total weapons</strong></td>
<td>3500</td>
<td>3499</td>
</tr>
</tbody>
</table>

\(^a\)The eventual number of mobile SS-25s that Russia might deploy is subject to debate, though 600 seems like a reasonable estimate. Russia may also deploy more silo-based ICBMs because they are less expensive and easier to maintain, and there are fewer problems providing adequate security and control over nuclear warheads in silos.

\(^b\)The Russian SLBM force was constructed on the assumption that most of the Delta IIIIs will be retired by the year 2003 because of their advanced age. No new SSBNs are believed to be under construction. The Typhoon is shown deployed with its full load of 10 warheads per SS-N-20 to keep the total number of warheads in the SLBM force relatively high. If the Russians download the SS-N-20, for example to six warheads each, then more Delta IIIIs (i.e., SS-N-18s) could be retained. In either case, the SLBM force contains approximately the same number of warheads, though in the latter case these weapons are spread over more submarines. If all Delta IIIIs are retired and the Typhoon is downloaded, then the SLBM force would contain closer to 1200 warheads and the total Russian force would be around 3000 weapons.

\(^c\)This assumes that the Blackjacks currently deployed in Ukraine will be returned to Russia and that some Bear-H bombers will be retired, leaving a total force of 80 heavy bombers carrying around 960 nuclear weapons.
ALCM-B and advanced cruise missiles (ACMs). The B-2 is counted as a nuclear-capable heavy bomber with a nominal load of 16 nuclear weapons.

Greater uncertainty surrounds possible Russian force structures. For the purposes of this analysis, the Russians are assumed to deploy around 800 ICBMs, 195 of which are single-warhead ICBMs (90 SS-19s downloaded from 6 warheads to 1 and 105 SS-25s deployed in former SS-18 silos) and 600 mobile SS-25s. The Russian SLBM force is assumed to consist of 15 missile-carrying submarines (SSBNs, or subsurface ballistic nuclear vessels) deployed with 264 SLBMs carrying a total of 1744 warheads. Finally, the Russian bomber force is assumed to contain around 1000 weapons. Despite uncertainties about the exact Russian force structure under START II, it will closely resemble the U.S. force mix—with the exception of mobile ICBMs. Note that the Russian land-based ICBM force is quite small by former Soviet standards.

Quantitative analyses of deterrence must calculate the number of retaliatory weapons that can effectively be delivered to an opponent’s homeland after absorbing a nuclear first strike. In general, far fewer retaliatory weapons arrive on target than are contained in a country’s strategic nuclear arsenal. Of the weapons in the total inventory, only those not in overhaul are available for retaliation (approximately 10 percent of the delivery systems at any one time). Of the available forces, only those that survive the opponent’s counterforce first strike, that operate reliably, and that penetrate defenses can threaten targets in the opponent’s homeland and, hence, contribute to deterrence. The number of arriving weapons (as opposed to inventory weapons) is then compared to the number of targets one believes should be held at risk to judge the adequacy of deterrence. Table 2 provides illustrative “planning factors” that capture the aggregate character of each side’s nuclear force posture on different states of alert.11

Of the available forces (those not in overhaul), the fraction that survive an opponent’s counterforce first strike is a function of the scenario one chooses to analyze. For this analysis, we examine three different alert postures: day-to-

---

11 It is important to note that these numbers represent averages over a given delivery system. As such, they gloss over numerous operational details that may affect the actual performance of specific systems, particularly for the bomber force. For example, the penetration probabilities for bombers and cruise missiles vary widely depending on the flight path, whether the targets attacked are located in heavily defended regions of the opponent’s country or are terminally defended with sophisticated surface-to-air missiles (e.g., the SA-10), the degree of defense suppression, and, finally, assumptions one makes regarding the future modernization of the opponent’s air defenses (an unlikely event for either country in the near term). The penetration probabilities shown in Table 2 reflect the fact that neither side is assumed to have highly effective air defenses in the next several decades.
<table>
<thead>
<tr>
<th>Weapon System</th>
<th>Availability</th>
<th>Day-to-Day Alert</th>
<th>Partial Alert</th>
<th>Generated Alert</th>
<th>Reliability</th>
<th>Penetration Probability^a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICBMs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM III/1 RV</td>
<td>0.95</td>
<td>0.20^b</td>
<td>0.20^b</td>
<td>0.20^b</td>
<td>0.85</td>
<td>1.0</td>
</tr>
<tr>
<td>SLBMs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-4</td>
<td>0.90</td>
<td>0.67</td>
<td>0.67</td>
<td>1.0</td>
<td>0.80</td>
<td>1.0</td>
</tr>
<tr>
<td>D-5</td>
<td>0.90</td>
<td>0.67</td>
<td>0.67</td>
<td>1.0</td>
<td>0.80</td>
<td>1.0</td>
</tr>
<tr>
<td>Bombers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-52H/ALCM-B</td>
<td>0.90</td>
<td>0.00</td>
<td>0.33</td>
<td>1.0</td>
<td>0.85</td>
<td>0.70</td>
</tr>
<tr>
<td>B-52H/ACM</td>
<td>0.90</td>
<td>0.00</td>
<td>0.33</td>
<td>1.0</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>B-2</td>
<td>0.80^c</td>
<td>0.00</td>
<td>0.33</td>
<td>1.0</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Russia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICBMs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS-19/1 RV</td>
<td>0.95</td>
<td>0.20^d</td>
<td>0.20^d</td>
<td>0.20^d</td>
<td>0.85</td>
<td>1.0</td>
</tr>
<tr>
<td>SS-25 silo</td>
<td>0.90</td>
<td>0.20^d</td>
<td>0.20^d</td>
<td>0.20^d</td>
<td>0.85</td>
<td>1.0</td>
</tr>
<tr>
<td>SS-25 road</td>
<td>0.95</td>
<td>0.25</td>
<td>0.35</td>
<td>1.0</td>
<td>0.85</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 2—continued

<table>
<thead>
<tr>
<th>Weapon System</th>
<th>Availability</th>
<th>Day-to-Day Alert</th>
<th>Partial Alert</th>
<th>Generated Alert</th>
<th>Reliability</th>
<th>Penetration Probabilitya</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLBM Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS-N-18</td>
<td>0.85</td>
<td>0.25c</td>
<td>0.40d</td>
<td>1.0</td>
<td>0.75</td>
<td>1.0</td>
</tr>
<tr>
<td>SS-N-20</td>
<td>0.85</td>
<td>0.25c</td>
<td>0.40d</td>
<td>1.0</td>
<td>0.80</td>
<td>1.0</td>
</tr>
<tr>
<td>SS-N-23</td>
<td>0.85</td>
<td>0.25c</td>
<td>0.40d</td>
<td>1.0</td>
<td>0.80</td>
<td>1.0</td>
</tr>
<tr>
<td>Bombers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bear-H</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
<td>1.0</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>Blackjack</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
<td>1.0</td>
<td>0.85</td>
<td>0.90</td>
</tr>
</tbody>
</table>


aThese penetration probabilities are illustrative. They do not include the presence of ballistic missile defenses. The air defenses are notional.

bBased on the assumption that one Russian warhead equivalent to an SS-18 attacks each U.S. ICBM silo. If U.S. ICBMs are launched out from under the attack, the survival probability is assumed to be 0.95.

cThis availability is based on an inventory of 20 B-2 bombers.

dBased on the assumption that one U.S. warhead equivalent to the Peacekeeper ICBM attacks each Russian ICBM silo. If Russian ICBMs launch out from under the attack, the survival probability is assumed to be 0.95.

eAssuming approximately three SSBNs are at sea on day-to-day alert.

fAssuming approximately five SSBNs are at sea on partial alert.
day alert, partial alert, and generated alert. We also assume that launch-under-attack scenarios are possible on partial and generated alert, but are less likely on day-to-day alert unless a concerted effort is made to implement this option.

The survival probability for submarines and bombers is essentially determined by the number of submarines at sea and the number of bombers on strip alert. The United States is assumed to keep approximately 11 out of 18 SSBNs at sea on day-to-day and partial alert. Russia is assumed to keep approximately 3 SSBNs at sea on day-to-day alert and 5 on partial alert. Moreover, for the purposes of this discussion, possible SSBN attrition due to antisubmarine warfare has not been included, bombers are assumed to receive adequate tactical warning to escape from their bases, and each side’s strategic command-and-control system is assumed to survive well enough to make retaliation likely.

---

Day-to-day alert is the normal readiness posture of strategic nuclear forces in peacetime. Generated alert is the readiness posture of U.S. and Russian nuclear forces on the brink of war. An intermediate readiness level (partial alert) has been included here to represent situations where the alert rate has been increased in response to international tensions, but a full wartime footing has not been ordered. There are numerous intermediate alert postures one could assume. For the purposes of this analysis, the partial alert rates are taken to be roughly equivalent to the alert rates the United States and Russia maintained in peacetime during the latter stages of the Cold War—this being an alert rate that could be sustained for a prolonged period, e.g., during a crisis. One could as well define intermediate alert rates higher or lower than these values.

One must distinguish between SLBM alert rate and SLBM prelaunch survival. Submarine survival is determined by the fraction at sea. This is slightly different from the “alert rate” because submarines are not “on alert” unless they are patrolling within range of their targets and are standing by to receive instructions from the National Command Authority. Submarines transiting to or from their alert stations may survive but are not considered to be on alert. Similarly, submarines may be at sea on training missions but not standing by to receive instructions—in which case they too are not considered to be on alert, though they would survive attack. The difference between the number of SSBNs at sea and the number on alert may be small, particularly with long-range SLBMs, because submarines can be on alert shortly after leaving port. Submarines in port may also be on alert if the SLBMs have sufficient range to strike their targets from port. However, these SLBMs would not survive unless they are launched out from under the attack.

These are reasonable assumptions for analyzing the basic U.S.-Russia deterrent relationship. However, one should note that scenarios can be invented where these assumptions may not hold, e.g., an aggressive U.S. conventional antisubmarine warfare campaign that precedes a nuclear crisis, inadequate tactical warning for Russian bombers owing to gaps in their early-warning network created by the Soviet Union's dissolution, or attacks against the strategic command-and-control network that effectively prevent retaliation for some period. These scenarios are important if one is interested in worst-case analyses.
The survival of silo-based ICBMs depends on whether they are launched out from under the attack (LUA) or whether they ride out the attack.\textsuperscript{15} Table 2 gives the survival probability assuming silo-based ICBMs ride out an attack with one hard-target-kill warhead allocated to each silo. The survival probability is fairly low because, by assumption, the United States and Russia upgrade the accuracy and yield of their single-warhead ICBMs over the next decade to give them a high single-shot kill probability (assumed to be 0.8) against the opponent’s silos. If silo-based ICBMs launch out from under the attack, their survival probability is assumed to be 0.95. Mobile ICBMs are less vulnerable than silo-based ICBMs, provided they obtain sufficient warning to disperse from their garrisons. The Russian mobile-ICBM alert rate in Table 2 assumes the majority of these systems are kept in garrison during peacetime, with a slight increase in the number out of garrison on partial alert. In addition, it has been assumed that the United States cannot localize these mobile ICBMs once they are deployed in the field.

Of the forces that survive an initial counterforce attack, the fraction that arrive on target is determined by the percentage that operate reliably (i.e., the delivery platform functions properly and the warhead detonates reliably) and by the fraction that penetrate the opponent’s defenses. The impact of ballistic missile defenses is accounted for separately in this analysis. Hence, the penetration probabilities for ballistic missiles is 1.0 in Table 2. The impact of strategic air defenses is explicitly taken into account by the attrition factor associated with different airborne platforms.\textsuperscript{16}

Multiplying these factors together for each delivery system and summing over the entire force structure gives the expected number of arriving or "effective" weapons for a given alert posture. The number of targets that can be destroyed for a given alert posture depends on how the attack is allocated.\textsuperscript{17}

\textsuperscript{15}If they are launched under attack, the survival probability is equal to the launch reliability. If not, the survival probability is determined by the accuracy, yield, and number of attacking warheads.

\textsuperscript{16}The U.S. bomber penetration probabilities in Table 2 reflect the fact that the United States will deploy a modestly advanced bomber force by the year 2003 and that Russian strategic air defenses may be relatively weak for decades due to the loss of one-time Soviet early-warning and other air defense assets located outside Russia. The Russian bomber penetration probability has been set at 0.90 to reflect the fact that the United States has a thin strategic air defense system that could present a threat to some Bear-H and Blackjack bombers.

\textsuperscript{17}Several operational factors not used in this analysis complicate the calculation of the number of targets each side can hold at risk. For example, cross-targeting with warheads from different delivery platforms improves the confidence with which high-priority targets can be destroyed. Cross-targeting was left out because single-warhead ICBM silos (a target often assumed to require two warheads) are assumed here to be targeted with only one warhead. There are relatively few high-priority targets other
It also depends on the probability that an arriving weapon destroys the target of interest. With the exception of ICBM silos, most targets in this analysis are soft. Hence, the probability that an arriving weapon actually destroys its intended target is assumed to be close to 1.0.\textsuperscript{18}

Finally, some weapons may be withheld even from an all-out retaliatory response so the United States (or Russia) is not completely vulnerable to nuclear coercion by other nuclear powers in the wake of a U.S.-Russian nuclear war. The size of this secure reserve force—not to be confused with a secure second-strike capability—is debatable; however, numbers in the range of 100–500 survivable weapons seem reasonable under START II, depending among other things on the alert posture. The weapons withheld for a secure reserve force have not been subtracted from the number of “effective” weapons shown in the subsequent analysis, though it is not difficult for the reader to make the necessary adjustments.\textsuperscript{19}

**DETERRENCE WITHOUT BALLISTIC MISSILE DEFENSES**

The extent to which the United States and Russia can maintain a secure second-strike capability for deterrence can be viewed pictorially by means of a “drawdown” curve. Drawdown curves are simply plots of the number of

---

\textsuperscript{18}Under START II, relatively few hard targets remain in the United States and Russia because most ICBM silos will have been removed. To the extent that some hard targets exist, gravity bombs on heavy bombers and high-yield ICBM and SLBM warheads can achieve high damage expectancies. Therefore, the probability that a given inventory weapon can destroy a target is essentially the same as its arrival probability. The number of arriving weapons (i.e., effective weapons) is therefore equivalent to the total number of targets that can be destroyed, assuming one weapon is allocated against each target.

\textsuperscript{19}The number of “effective” weapons withheld for a secure reserve force is determined by multiplying the number of weapons withheld of each type by their reliability and penetration probabilities given in Table 2. Subtracting the sum from the total number of effective weapons that survive in a given scenario yields the maximum number of effective weapons that are available for an immediate retaliatory strike. This difference is the size of the secure second-strike capability. Since SLBMs and mobile ICBMs will likely be the weapon of choice for the secure reserve force, reserves between 100 and 500 surviving weapons imply that approximately 80–400 effective weapons would be subtracted from the effective weapon totals, depending on the alert posture, to provide a more accurate estimate of the size of the secure second-strike capability.
weapons remaining on both sides after one side initiates a counterforce attack against the opponent's strategic nuclear forces. As such, they provide a convenient pictorial representation of the extent to which the defender's retaliatory strike is secure. Whether a given retaliatory capability is adequate for deterrence can be answered only by reference to a particular deterrence strategy for a given adversary (and its accompanying targeting doctrine). The virtue of drawdown curves is that they capture the quantitative aspects of each side's arsenal, leaving the reader to judge whether deterrence remains robust or is becoming "delicate" in a given scenario, based on his own criteria for sufficiency and other exogenous factors he believes are relevant.

The U.S. Strategic Nuclear Deterrent

Figure 1 illustrates the drawdown curve for a hypothetical Russian counterforce first strike against U.S. START II forces on day-to-day alert. The curve's initial point, in the upper right-hand corner, represents the total U.S. and Russian arsenals before the hypothetical attack begins. The slope of the curve at any point determines the marginal counterforce exchange ratio, i.e., the ratio of U.S. warheads destroyed to Russian warheads expended at this point in the attack. Hence, the slope indicates the relative attractiveness of this particular portion of the counterforce attack. The curve is plotted so that the most lucrative counterforce options occur first, the least attractive options last.

Initially, the drawdown curve is nearly vertical because a Russian attack against U.S. submarine bases (four are assumed in this analysis) and bomber bases (five are assumed on day-to-day alert) could destroy a large number of U.S. weapons off alert (i.e., U.S. SSBNs in port and nonalert bombers on the ground) with the expenditure of very few Russian weapons (three or four weapons are allocated to each base). Hence, by attacking a total of nine targets with approximately 30–40 weapons, the Russians could destroy a large fraction of the U.S. strategic nuclear arsenal, at least on day-to-day alert.

The next most lucrative counterforce option is an attack against U.S. single-warhead ICBM silos. This portion of the drawdown curve is broken into two segments. The first part represents the drawdown of the Minuteman force if one Russian weapon is targeted against each silo. The second segment represents the drawdown of the surviving Minuteman ICBMs if a second warhead is allocated to each silo.

After the attack on U.S. Minuteman silos, no attractive Russian counterforce options remain. On day-to-day alert, the entire U.S. bomber force is off alert, otherwise one might consider barrage attacks against bomber flyout corridors. Obviously, barrage attacks against mobile ICBMs are not considered because the U.S. force structure contains no mobile ICBMs. The only theoret-
Figure 1—Russian Counterforce First Strike on Day-to-Day Alert

Figure 1 illustrates the potential for Russian counterforce attacks on U.S. strategic forces. The vertical axis represents the number of U.S. weapons remaining, while the horizontal axis represents the number of Russian weapons remaining. The diagram shows various attack scenarios and their impact on the number of weapons remaining on both sides.

The critical counterforce option left is a barrage attack against U.S. submarines at sea. This attack is ineffective because it has been assumed that the Russians cannot localize U.S. submarines at sea. Hence, a large number of Russian weapons would be expended with the destruction of very few U.S. weapons, i.e., the drawdown curve in Figure 1 is essentially horizontal.

The Russian counterforce attack in Figure 1 comes to an end with approximately 1200 Russian weapons remaining because these weapons—namely, bomber weapons and a few weapons associated with ballistic missiles in overhaul—cannot be used in the counterforce first strike.20

So far, the drawdown curve illustrates theoretical counterforce attack options. Nothing has been said about the number of weapons the Russians would actually allocate to a counterforce attack, i.e., how far down the theoretical drawdown curve they might actually proceed. Nor is it clear how many

20Bomber weapons cannot be used for counterforce attacks because it is assumed that their slow delivery provides sufficient warning for the opponent’s forces to escape.
ballistic missile weapons the Russians actually have on alert (assuming their forces are also on day-to-day alert) that could participate in the counterforce attack. To determine an optimal stopping point for the counterforce attack, one must assume some objective function the Russians attempt to maximize; for example, the postattack ratio of remaining forces.\footnote{Several authors have devised different objective functions for counterforce first strikes in an effort to determine “optimal” counterforce allocations. See, for example, Kent and Thayer (1989), Canavan (1991), Bracken (1990), and Grotte (1980).} Using this criterion, the optimal stopping point can be found by drawing a straight line from the origin tangent to the drawdown curve (i.e., the dashed line in Figure 1). The tangent point is the optimal stopping point for this objective. This occurs after the Russians have allocated one warhead to each Minuteman silo, but before they have allocated two warheads. Therefore, single-warhead ICBM silos, though they are less attractive targets than MIRVed ICBM silos, will be attacked in a counterforce first strike. Throughout this analysis, one warhead is assumed to be allocated to each single-warhead ICBM silo. From Figure 1, one might expect intuitively that the Russians would stop after allocating one weapon against each Minuteman silo but before two weapons are allocated, since this appears to include the most attractive counterforce attack options without wasting too many weapons.

The fraction of the opponent’s arsenal that must be destroyed before an attacker has an appreciable incentive to strike first is, of course, debatable. Again, the drawdown curve simply shows the number of weapons that survive, leaving it up to the reader to decide whether one side, under some circumstances, might be tempted to attack.

The classic incentive for nuclear attacks occurs when one side believes it can substantially limit damage by striking first. Whether or not one side has a significant damage-limiting counterforce capability can be determined easily using drawdown curves. The distance between the drawdown curve and the x-axis in Figure 1 indicates the size of the U.S. second-strike capability at any point along the drawdown curve. As one observes, approximately 1100 U.S. weapons survive the maximal Russian counterforce attack on day-to-day alert. Later we will see exactly how much damage this surviving force can inflict. Nevertheless, it seems clear from the figure that the U.S. deterrent is fairly robust even on day-to-day alert. If the drawdown curve came much closer to the x-axis, then the U.S. deterrent could be judged to be weak.

With this introduction, we can now use drawdown curves to display the size of the U.S. or Russian secure second strike under a wide range of possible scenarios. Figure 2 illustrates Russian counterforce attack options on day-to-day alert, partial alert, and generated alert. In the partial and generated alert scenarios, alternate drawdown curves are shown for the cases where the United
States launches its silo-based ICBMs out from under the attack. The solid line in each drawdown curve illustrates the number of Russian weapons that actually are available (i.e., on alert) for a counterforce first strike. In the partial and generated alert scenarios, the Russian barrage attack is directed against U.S. bomber fly-out corridors because they represent more lucrative targets than U.S. SSBNs at sea—though neither barrage target is particularly attractive from the Russian perspective.

The Russian counterforce attack against the most lucrative targets, including a one-on-one attack against Minuteman silos, involves only about 530 warheads. This is a factor of six less than the number of weapons that would

---

22In addition to the alert forces, it has been assumed that mobile ICBMs in garrison can be generated covertly and, hence, can be used in the counterforce first strike regardless of the fraction that are actually in garrison. Note that all Russian ICBMs (mobile or silo-based) are assumed to have sufficient hard-target-kill capability to threaten U.S. ICBM silos. This may be a generous assumption. No Russian SLBMs are credited with hard-target-kill capability.
have been required for comparable attacks throughout most of the Cold War. The collateral damage associated with these attacks would, therefore, be less severe. Whether this makes counterforce attacks more “thinkable” under START II, because they appear to be more limited or surgical in character, is debatable. Nevertheless, one should recognize that as the number of counterforce targets shrinks, it may be easier for leaders to convince themselves that escalation can be controlled after “limited” counterforce attacks. If ICBM silos are left out of the attack, then the Russian counterforce first strike would involve only a few tens of weapons (perhaps with another ten or so allocated against U.S. strategic C3 targets). Hence, in the post–Cold War era it may be easier for a desperate leader to convince himself that nuclear war could remain limited, making nuclear attacks more likely.

Figure 2 shows that even under the worst-case scenario in this analysis (day-to-day alert), the United States still has over 1000 nuclear weapons that survive an all-out Russian counterforce attack. This surviving arsenal could threaten devastating retaliation—even if this force might be insufficient to meet comprehensive U.S. targeting objectives. The U.S. second strike thus appears to be secure under all scenarios. On generated alert, approximately 2700 weapons would survive. This provides enough retaliatory capability to hold at risk a wide range of Russian military, economic, and leadership targets. Hence, on generated alert, deterrence is quite robust. In no scenario is the United States limited to retaliation against Russian cities alone. Eleven hundred surviving weapons provides the United States with substantial capability to threaten a large number of the most highly valued Russian military, economic, and leadership targets, though comprehensive damage cannot be accomplished simultaneously against all target categories, even if several hundred weapons are withheld for a secure reserve force.

Having weapons that survive a counterforce first strike is not the only requirement for maintaining a secure second strike. In addition, one must have weapon systems that are reliable and that can penetrate the opponent’s defenses. Taking weapon system reliability and penetration probabilities into account, one can replott the drawdown curves in an “effective” weapon domain as shown in Figure 3. “Effective” weapons are simply weapons that arrive on target in the opponent’s homeland. The number of effective weapons equates to the number of nuclear detonations that one side can inflict on the adversary. If one weapon is allocated to each target, then the number of effective weapons essentially equals the number of targets that can be destroyed, assuming the probability with which an arriving nuclear weapon destroys a target is essen-
Figure 3—U.S. Secure Second Strike (effective weapons)

...tially unity, which is true for most targets. Hence, the effective weapon domain shows the residual retaliatory capability, at any point along the drawdown curve, measured in terms of the number of targets each side can hold at risk in the opponent’s homeland (minus the number of effective weapons that would be withheld for a secure reserve force).

In Figure 3, the inventory point is at 2800 effective U.S. weapons and a comparable number of effective Russian weapons. Therefore, of the 3500 weapons in each side’s inventory, only about 2800 would actually arrive on the opponent’s homeland if the entire strategic nuclear inventory was used in the attack. After a Russian counterforce first strike on day-to-day alert, the residual U.S. retaliatory capability is around 900 effective weapons. U.S. forces on generated alert can hold at risk approximately 2200 targets, or around 2500.

\(^{23}\)Most targets are relatively soft. To the extent hard targets exist, each side will have sufficient hard-target-kill capability in its bomber force and prompt hard-target-kill capability in its ICBM force (and, for the United States, in its SLBM force).
targets if the United States launches its single-warhead Minuteman III ICBMs out from under the attack on generated alert.

Many people believe the essence of deterrence rests on the ability to hold at risk the opponent’s population. In other words, instead of threatening to destroy a certain number of military, leadership, and/or economic targets, as implied by Figure 3, deterrence rests on the ability to target each other’s cities. Put another way, when people speak of damage limitation as an incentive for striking first, it is damage to one’s urban population that is to be limited. Figure 4 replots the same drawdown curves, but this time the residual retaliatory capability is measured in terms of the opponent’s urban population at risk. In other words, Figure 4 translates effective weapons into urban fatalities (resulting from prompt nuclear effects alone).

Therefore, even though only 1100 U.S. weapons survive a Russian counterforce first strike on day-to-day alert (Figure 2), approximately 900 of these weapons arrive on target within Russia (Figure 3). If these weapons are targeted at major Russian urban areas, approximately 60 million people would be killed (Figure 4). Clearly, deterrence appears to be much more robust if one assumes that political leaders believe deterrence rests on the ability to

---

24 Some people argue that as arsenals shrink, countries will tend to emphasize targeting of the opponent’s cities because this creates the greatest deterrent effect with a limited nuclear arsenal. This is particularly true if a country has only a few nuclear weapons, as the United States had in 1945, or as small nuclear powers such as Pakistan, Israel, and India might have today.

25 One should note that this calculation only includes prompt nuclear effects. No account has been made for collateral damage resulting from attacks against the opponent’s counterforce targets. During the Cold War, the collateral damage associated with U.S. and Russian counterforce attacks was thought to be substantial, primarily because of the fallout created when thousands of weapons detonate over each country's ICBM silos. Under START II, counterforce attacks involve relatively few weapons by Cold War standards. As a result, the collateral damage due to fallout would be less by a factor of 5 to 10 compared to comparable attacks at the height of the Cold War. Using the calculations of Daugherty, Levy, and von Hippel, the fatalities resulting solely from the fallout created by a Russian counterforce attack against the United States would be on the order of 0.7 to 6 million deaths (the total casualties would be higher). See Daugherty, Levy, and von Hippel (1986) and Levy, von Hippel, and Daugherty (1987/88). A similar calculation for a U.S. counterforce attack against Russia produces approximately 0.8 to 2 million fatalities due to fallout alone. This is small relative to the tremendous fatalities that could be inflicted by direct attacks against each other’s population. In addition, the collateral damage associated with counterforce attacks against submarine ports and bomber bases, which may be located near major urban areas, is also relatively small due to the small number of warheads used in these attacks (around 30–40) compared to Cold War attack scenarios. Hence, one can ignore these collateral effects as well, at least for the purpose of this discussion.
inflict casualties on the opponent's population. Even a small surviving arsenal can do tremendous damage, as illustrated in Figure 4.

The Russian Strategic Nuclear Deterrent

So far, the discussion has focused on possible Russian incentives to strike first. We now turn to hypothetical U.S. first-strike options, particularly as they might appear to Russian leaders. Figure 5 illustrates U.S. counterforce first strikes under a range of possible scenarios. Once again, the inventory point in the upper right-hand corner represents the total U.S. and Russian arsenals prior to the attack. The most lucrative U.S. attacks are directed against Russian SSBNs in port and Russian bombers at their bases. The next most attractive option would be an attack against mobile ICBM garrisons, followed by a one-on-one attack against Russian ICBM silos. After this, no attractive counterforce options remain. In principle, the United States could allocate a second warhead against each Russian single-warhead ICBM silo, or it could conduct a barrage attack against submarines at sea or mobile ICBMs out of
garrison. The latter is illustrated in Figure 5. As one can see, this barrage attack (against mobile SS-25s out of garrison) would consume a large number of U.S. weapons and destroy very few Russian weapons (according to the assumptions used in this analysis). Hence, the likely stopping point for a U.S. counterforce first strike would be after one weapon has been allocated against each Russian single-warhead ICBM silo. This also happens to be the point at which the ratio of remaining U.S. weapons to surviving Russian weapons is at a maximum. As in Figure 2, the solid line indicates the number of weapons the United States actually has available for a counterforce attack in each scenario. Clearly, the United States has ample capability to carry out the most lucrative counterforce options.

Figures 6 and 7 illustrate the size of the Russian secure second strike measured in terms of effective Russian weapons remaining after a U.S. counterforce attack and in terms of the U.S. urban population these effective weapons could hold at risk in a retaliatory strike if Russian leaders target their retaliatory strike to maximize U.S. urban fatalities. Again, weapons withheld for a Russian se-
cure reserve force should be subtracted from the numbers shown in Figure 6 to provide a more accurate representation of the maximum number of effective weapons the United States would absorb in an all-out Russian retaliatory strike.

Several interesting points emerge from this series of figures. First, a U.S. counterforce first strike against the most attractive Russian counterforce targets requires only about 350 weapons—far fewer than comparable U.S. counterforce attacks during the Cold War. Consequently, the collateral damage associated with these attacks would be relatively small when compared to comparable Cold War scenarios.\textsuperscript{26} Hence, from the Russian perspective, U.S. counterforce attacks may become increasingly "thinkable."

\textsuperscript{26}At the height of the Cold War, U.S. counterforce attacks against roughly 1500 Soviet ICBM silos would have involved around 3000 warheads fused to detonate close to the ground (thus creating significant fallout). Under START II, the comparable scenario involves only 195 U.S. weapons ground burst against 195 Russian ICBM silos. Hence, the fallout from this counterforce attack would be approximately 15 times less severe, assuming weapons of comparable yield.
Second, on generated alert, approximately 3000 weapons would survive (Figure 5), which equates to approximately 2400 effective weapons landing on the U.S. homeland (Figure 6). In terms of countervalue damage, the lives of over 100 million Americans could be threatened (Figure 7). Hence, deterrence under these circumstances would be quite robust.

Third, at low alert rates, few Russian weapons survive by Cold War standards. In particular, on day-to-day alert, approximately 550 Russian weapons survive: about 375 weapons aboard submarines at sea, around 135 mobile ICBMs out of garrison, and about 40 silo-based ICBMs (assuming the planning factors in Table 2). All Russian bombers are assumed to be destroyed on the ground on day-to-day alert. Of these surviving weapons, approximately 450 weapons would actually operate reliably and could arrive on target within the United States (Figure 6).

Whether 450 effective weapons constitute an adequate deterrent in the minds of Russian political and military leaders cannot be answered with certainty. On paper, it seems adequate. However, these effective weapons exist on relatively few submarines at sea and a small number of mobile ICBMs out of garrison. If the Russian general staff believed the United States had a modest antisubmarine warfare (ASW) capability, or a rudimentary capability to
threaten mobile ICBMs in the field, then Russian leaders might become concerned about the adequacy of this relatively small surviving arsenal. One should also recall that, though attacks against strategic nuclear command-and-control nodes have been ignored in this analysis, such fears might add to the Russian sense of vulnerability. Finally, Russia would likely withhold some weapons for a secure reserve force to deter other possible threats from China, France, Great Britain, and perhaps Ukraine (assuming Russia abides by START II even if Ukraine retains some nuclear capability). Therefore, on day-to-day alert, Russian leaders may conclude that they have only several hundred effective weapons with which to retaliate against the United States. Again, it is debatable whether Russian leaders would believe this is adequate for deterrence.

Figure 7 shows that even a relatively small surviving arsenal can do tremendous damage to U.S. cities (approximately 50 million people could be killed with 450 effective weapons). In terms of a counterforce deterrent, then, the Russian secure second strike seems adequate. Of course, for Russian leaders to derive any comfort from this, they must actually be willing to threaten U.S. cities and target their forces accordingly.

This analysis also sheds light on arms control treaties with lower ceilings or further reductions in the peacetime alert rates. Treaties that reduce the aggregate weapon total simply shift the inventory point of the drawdown curves toward the origin. The exact shape of the curves depends on the force structure and the alert rates. Without illustrating the myriad of possibilities, suffice it to say that further cuts may be reasonable if one assumes that generated alert is the only plausible planning scenario. On the other hand, further reductions in weapon inventories or alert rates may not appear attractive if one focuses on day-to-day alert scenarios. It is true that one can always strengthen deterrence by generating the force, but early force mobilization then becomes necessary for adequate deterrence. Creating a situation where one or both sides must mobilize their forces to ensure adequate deterrence may not be desirable because rapid force generation is often perceived by the other side to be provocative, especially early in a crisis.

To summarize, under START II, counterforce attacks are smaller in size and have less collateral damage compared to similar Cold War scenarios. This tends to increase the chance that leaders might believe deliberate counterforce attacks are rational instruments of policy. Nevertheless, with offensive forces constrained by START II, neither the United States nor Russia can significantly improve its chances for survival by striking first. Hence, the mutual hostage relationship that characterizes mutual deterrence will remain unless radical force reductions are implemented. This is true even when U.S. or Russian forces are on day-to-day alert, barring further reductions in the U.S., and especially Russian, peacetime alert posture.
As a secondary matter, the U.S. deterrent appears to be more robust than Russia’s because the United States maintains a larger fraction of its forces on alert in peacetime. If several Russian SSBNs could be sunk by U.S. antisubmarine warfare, or if a modest number of Russian mobile ICBMs could be attacked out of garrison, then Russia’s deterrent could become quite weak at low alert rates. If the United States reduced the alert rate for its SLBM force, for example by halving the number of SSBNs at sea, then it would be in much the same position as Russia. But aside from these worst-case scenarios, deterrence again appears to be quite robust under START II. This is much less true if both sides deploy limited nationwide ballistic missile defenses, as the next section will demonstrate.

THE IMPACT OF LIMITED NATIONWIDE BALLISTIC MISSILE DEFENSES

Until now, the only defenses that have been included are U.S. and Russian strategic air defenses (captured by the attrition rates associated with different aircraft in Table 2). We now add limited nationwide ballistic missile defenses (BMD). For the purposes of this analysis we assume that both sides deploy ground-based BMD interceptors symmetrically. Space-based defenses are not modeled. The effect of asymmetric defense deployments can be inferred from the symmetric cases. This analysis focuses on a case where 600 ground-based interceptors are deployed on both sides, since this level of nationwide defense represents a turning point for deterrence. One should also note that theater BMD interceptors may contribute to a country’s strategic BMD capability. For example, if the United States builds a large number of theater BMD interceptors and stores them in the United States, Russian leaders might fear that these theater BMD interceptors could be deployed around the continent

---

27 In this analysis, the defenses are assumed to be capable of “shoot-look-shoot” tactics with a maximum of four interceptors allocated to an incoming warhead. Shoot-look-shoot tactics are much more effective because the second shot(s) are directed against only those warheads that leak through the first layer of the defense. In addition, it has been assumed that all interceptors can engage the incoming attack; i.e., interceptor range limitations have been ignored. Moreover, the defenses are assumed to survive defense suppression attacks and the interceptors are assumed to be highly effective; i.e., the interceptor single-shot kill probability is assumed to be 0.8—a level of effectiveness which, in principle, could be achieved for ground-based interceptors, though a lot depends on reactions the opponent might take. These may seem like heroic assumptions for the defense; however, the point of this analysis is to examine the strategic impact of defenses, not to prejudge the question of whether such defenses could be built or whether they would be cost-effective in light of an opponent’s countermeasures. If the opponent deploys effective penetration aids or develops effective tactics to suppress the defense, this would change much of the subsequent analysis.
short notice, thereby contributing substantially to the U.S. strategic BMD capability.28

Ballistic missile defenses affect the analysis presented earlier in two ways. First, they interfere with counterforce first strikes. This is illustrated in Figure 8 by the kink at the beginning of the drawdown curves. For example, the Russians have to expend several hundred warheads in a counterforce first strike before sufficient warheads leak through the U.S. defense to destroy nonalert submarines and bombers at their bases. After the defense is saturated, the drawdown curves appear more or less as they did without ballistic missile de-

---

fenses. Second, the first-striker’s defense can block a certain number of the opponent’s surviving weapons, thereby reducing the number of effective weapons available for retaliation. If the surviving arsenal is relatively small, limited defenses may be able to significantly limit damage from the retaliation. This second effect is illustrated notionally by the shaded regions close to the axes. The distance between the edge of these shaded regions and the drawdown curves illustrates the number of retaliatory weapons that can penetrate the first-striker’s defense; i.e., it closely approximates (apart from reliability factors) the number of effective weapons available for retaliation. If the drawdown curve moves into one of the shaded regions, then the country attacking first can significantly improve its chances for surviving a nuclear war, provided it strikes first.

U.S. Secure Second-Strike Capability

Figure 9 shows Russian counterforce first-strike options on three levels of alert assuming 600 ballistic missile interceptors are deployed on both sides. Note that it takes around 400 weapons to saturate the U.S. defense before enough leak through to destroy U.S. nonalert submarines and bombers. The solid line in each drawdown curve represents the number of weapons on alert and, hence, available to conduct the counterforce first strike in each scenario.

The first point to notice in Figure 9 is that with 600 interceptors, the Russians barely have enough weapons on alert to conduct a complete counterforce attack against U.S. ICBM silos on day-to-day alert (note that the drawdown curve is still sloping downward, albeit slowly, when the Russians run out of alert weapons). Comparing this drawdown curve to the comparable curve without ballistic missile defenses (i.e., Figure 2) one observes that without defenses, sufficient weapons are available to attack each U.S. ICBM silo with one warhead on day-to-day alert. Hence, the first impact of a U.S. defense is to interfere with the Russian counterforce first strike (at low levels of alert).

---

29Due to the defenses, the attacker cannot determine with certainty which warheads will arrive on target. As a result, the drawdown curve cannot be easily segmented into attacks against separate types of targets (i.e., submarine bases, bomber bases, ICBM sites, etc.). Nevertheless, comparing the general shape of the drawdown curve in Figure 8 to the cases without ballistic missile defenses, one can surmise which targets are predominantly being attacked at different points along the curve.

30As an aside, one should note that these defenses effectively eliminate small ballistic missile attack options. In other words, “Limited Nuclear Options” using ballistic missiles, options that supposedly were included in U.S. strategy during the Cold War, would no longer be effective.
Therefore, defenses strengthen deterrence by making it more difficult for the Russians to accomplish a comprehensive counterforce attack at low alert levels.

It is difficult to tell exactly how many weapons the Russians might allocate to a counterforce first strike. Suffice it to say that if they wanted to reduce the U.S. arsenal to the maximum extent possible on day-to-day or partial alert, they would have to launch their entire alert force in the attack. Obviously, the Russians would not do this, because then a U.S. counterforce second strike could, in principle, disarm them, since their remaining forces would be highly vulnerable. Russian counterforce attacks thus would likely stop several hundred warheads short of the end of the solid line in Figure 9; for example, several hundred SLBMs at sea and/or mobile ICBMs out of garrison would be held in reserve.

As mentioned above, deterrence requires that weapons not only survive but operate reliably and penetrate any residual defenses. In Figure 9, the U.S. ballistic missile defense is exhausted early in the Russian counterforce first strike.
However, the Russian defense is still intact to defend against the U.S. retaliatory strike. Therefore, one must examine the drawdown curves in the effective weapons domain to see the true residual retaliatory capabilities after the Russian first strike. Figure 10 plots these same drawdown curves in terms of the effective weapons remaining on both sides, i.e., the number of weapons that would arrive on each country's homeland after any residual defenses are taken into account. These curves are virtually identical to the ones shown in Figure 9; however, they are shifted down because the Russian ballistic missile defense absorbs around 500 surviving U.S. warheads. Therefore, at any point along the drawdown curves in Figure 10 one can read off the residual retaliatory capability measured in arriving weapons. For example, after a Russian counterforce first strike on day-to-day alert, the United States retains the capability to deliver approximately 600 weapons against the Russian homeland (assuming the Russians withhold approximately 200 warheads). U.S. retaliatory capability on generated alert is more substantial (approximately 2000 ef-
fective weapons). As mentioned in the previous section, these estimates do not account for weapons that are withheld as part of a U.S. secure reserve force.

If these effective weapons were targeted entirely against Russian cities, approximately 45 million fatalities would result in a day-to-day alert attack scenario, as illustrated in Figure 11. This represents an awesome countervalue retaliatory capability (assuming the United States would actually target its forces this way, contrary to past pronouncements). Hence, the U.S. secure second-strike capability appears to be fairly robust even when 600 ground-based interceptors are deployed on both sides. As the alert rate increases, this countervalue retaliatory capability becomes even more deadly.

Russian Secure Second-Strike Capability

Figure 12 illustrates U.S. counterforce first-strike options in several scenarios, assuming both sides deploy 600 ground-based interceptors. From this fig-
Figure 12—U.S. Counterforce First-Strike Options: 600 BMD Interceptors

One can see that the United States has ample weapons available to saturate the Russian ballistic missile defense and strike at the most lucrative Russian counterforce targets. The actual number of weapons the United States would allocate to a counterforce attack is debatable. Nevertheless, if 1000 to 1400 weapons are allocated to this attack on day-to-day alert (approximately 1500 are available), Russia would be left with approximately 500 surviving weapons (for the same reasons that they had about 500 surviving weapons in Figure 5).

Again, the adequacy of deterrence can only be observed in the effective weapon domain, as illustrated in Figure 13. If virtually the entire U.S. alert force is allocated to a counterforce attack on day-to-day alert, with 600 ballistic missile defense interceptors deployed on both sides, then hypothetically only about 20 Russian weapons could penetrate the U.S. defense in retaliation. That is, of the 500 or so surviving weapons illustrated in Figure 12, only around 20 could get through the U.S. defense. This represents a better damage-limiting first-strike capability than either side ever achieved throughout the Cold War (except arguably on the U.S. side around 1960–1961). Therefore,
600 ground-based interceptors (deployed on both sides) would virtually eliminate the Russian secure second strike on day-to-day alert if the United States chose to preempt in a crisis with a massive counterforce attack. By contrast, a Russian counterforce first strike on day-to-day alert cannot limit damage to a comparable degree because approximately 600 effective U.S. weapons remain after the attack (recall Figure 10).

As noted before, a relatively small number of effective weapons can cause enormous civilian casualties if the retaliatory attack is directed against urban areas. Figure 14 illustrates the residual Russian capability to hold at risk the U.S. urban population when 600 BMD interceptors are deployed on both sides.\(^3\)

Even though a U.S. counterforce attack on day-to-day alert can

\(^3\)Collateral damage has again been ignored. While this was reasonable in a purely offense-dominant nuclear balance (i.e., Figures 4 and 7), it clearly is a poor assumption here because the number of nuclear detonations associated with U.S. and Russian counterforce attacks has grown. Therefore, the number of fatalities shown in Figure 14
hypothetically reduce the number of effective Russian weapons to around 20, these weapons can still hold at risk on the order of 7 to 12 million American lives. While a U.S. President may believe that such large civilian casualties are sufficient to deter the United States (under most circumstances), it is doubtful that Russian leaders would feel equally confident. Hence, they would inevitably feel pressure to improve their deterrent posture, perhaps quickly.

On partial alert, the Russian secure second-strike capability equals about 250–300 effective weapons. Whether this retaliatory capability is sufficient, in the Russian leadership’s mind, to deter the United States is debatable. What is not debatable is that the introduction of relatively limited defenses (i.e., 600 ground-based interceptors) begins to interfere significantly with the Russian secure second strike at low levels of alert.

should be treated as a lower bound. Collateral fatalities could add from 1 to 8 million additional fatalities.
The problems created by limited nationwide ballistic missile defenses occur first with the Russians because they have a smaller number of weapons deployed in survivable basing modes on day-to-day alert. As the defense level increases, the next scenario to encounter problems would be a U.S. counterforce first strike against Russian forces on partial alert. As defenses increase further, the U.S. secure second strike eventually becomes vulnerable to a Russian counterforce attack on day-to-day alert.32

Finally, one should note that on generated alert the Russian secure second strike is quite robust. If Russian forces are fully generated, they could retaliate with approximately 1900 effective weapons. The U.S. deterrent was similarly robust on generated alert (see Figure 10). Therefore, if one believes that U.S. and Russian forces will always be fully generated in a crisis, then deterrence is quite robust even in the presence of limited nationwide ballistic missile defenses. However, it is the transition from peacetime to generated alert that concerns us. Since the Russian deterrent posture at low alert rates is compromised in the presence of 600 BMD interceptors, Russian leaders have a strong incentive to react to redress this vulnerability before a crisis with the United States reaches the acute phase. This is the principal strategic problem that occurs when limited ballistic missile defenses are deployed.

Thus, at a level of around 600 ground-based interceptors deployed on both sides, one can observe two effects from limited nationwide ballistic missile defenses. First, the U.S. defense obstructs the Russian counterforce first strike, thereby strengthening the U.S. deterrent. However, the second and more pronounced effect is that a U.S. defense, in conjunction with a U.S. counterforce first strike, virtually eliminates the Russian secure second strike on day-to-day alert, thereby eroding Russia’s deterrent.

Obviously, U.S. leaders are less concerned with the integrity of Russia’s deterrent capability than they are with that of the United States. However, this perspective can be short-sighted because it ignores the fact that the United States and Russia are involved in a strategic interaction. Undermining Russia’s deterrent will adversely affect political relations between these two countries, upsetting the current warming trend, and may stimulate another arms race as Russia attempts to shore up its deterrent. If Russia cannot afford the expense,

32As a rule, the size of the defense required to substantially interfere with an opponent’s secure second strike can be determined by equating the defense potential (the number of interceptors times their single-shot kill probability or SSPK) to the number of surviving ballistic missile weapons on the opponent’s side. Thus, if 500 Russian ballistic missile weapons survive a U.S. first strike, and the U.S. interceptors have an SSPK of 0.8, then the Russian secure second strike virtually disappears when the United States deploys approximately 625 interceptors, as we have seen. Using this same line of reasoning, the U.S. secure second strike would virtually disappear on day-to-day alert if Russia deployed approximately 1300 ground-based interceptors (with an SSPK of 0.8).
then it may opt for less expensive quick fixes that may be inimical to U.S. interests. Hence, as a general proposition, U.S. leaders should be concerned about Russia's deterrence capability if they wish to maintain stable relations with this former adversary.

The Russian strategic vulnerability at defense levels of 600 ground-based interceptors can easily be avoided if Russia keeps a higher fraction of its forces on alert in peacetime. However, this obvious solution may not be an option if Russia does not have the financial resources to implement it. Deploying more submarines to sea and mobile ICBMs out of garrison increases peacetime operating costs as well as maintenance costs because submarine reactors fail more frequently and mobile ICBMs experience increased problems with guidance system failures, etc. It may also affect the operational reliability of the force. In short, it is reasonable to believe that the strategic problems suggested by this analysis will come to pass. Therefore, if a crisis ever occurs between the United States and Russia in which the United States has a defense potential on the order of 500 warheads worth of intercept capability, Russia will likely feel pressure to rapidly redress its strategic vulnerability.

There are essentially three approaches the Russians might take to avoid this problem: (1) threatening to launch their vulnerable ballistic missiles out from under a U.S. counterforce attack, (2) rapidly generating their forces early in a crisis to improve their deterrent posture, or (3) limiting the size of nationwide ballistic missile defenses so the strategic problem outlined above never comes to pass.

Russian threats to launch a large fraction of their vulnerable forces (principally land-based ICBMs and perhaps some SLBMs in port) out from under a U.S. counterforce attack is perhaps the easiest way to redress this vulnerability. On day-to-day alert, this would provide approximately 600 additional surviving weapons. If Russia threatened to launch its vulnerable ICBMs under attack on day-to-day alert and target these weapons on U.S. cities, they could increase U.S. fatalities from around 10 million to approximately 40–45 million. Hence, launch under attack becomes an effective means to quickly

33Another Russian concern with mobile nuclear forces has been maintaining secure control over the nuclear warheads when they are deployed off their main base. The fear of sabotage, confiscation, or accidental/unauthorized use made Soviet leaders reluctant during the Cold War to disperse too many nuclear weapons in peacetime. Instead they preferred to keep them locked up at secure storage sites under the assumption that they would receive adequate strategic warning to generate their forces. Whether this same concern will cause Russian leaders to keep most mobile nuclear forces on their bases in peacetime in the post-Cold War era remains to be seen.

34For a day-to-day alert–LUA scenario, the total number of retaliatory weapons is the same as in the partial alert–LUA case except that two fewer SSBNs are at sea. This reduces the number of retaliatory weapons by approximately 230, and the number of
convince the United States that it cannot improve its chances for survival by striking first.

On the other hand, implementing a launch-under-attack policy increases the chance of an accidental or unauthorized attack in the midst of a crisis, thereby increasing the likelihood of inadvertent nuclear war. This should be of particular concern to the United States given the relatively poor condition of Russia’s ballistic missile early-warning network in the wake of the collapse of the former Soviet Union—though obviously this system may be improved by the time START II is finally implemented (i.e., around 2003). In addition, Russia must decide which targets to aim these LUA warheads against. If they are targeted against other U.S. military targets or urban/industrial areas, Russia could obtain substantial deterrent benefits but at the risk of losing escalation control if they actually launched these weapons against these targets.35 Nevertheless, threatening LUA would certainly give U.S. leaders pause, particularly if they are contemplating preemptive counterforce attacks.

It is perhaps surprising that Russia may continue to rely on LUA under START II, despite the fact that the treaty eliminates MIRVed ICBMs—those weapons thought to be particularly dependent on LUA to survive. The reason is simply that limited defenses can undermine Russia’s secure second strike to such a degree that it may be compelled to rely on launch under attack to deter a U.S. first strike. In other words, reliance on launch under attack should not be associated with a particular weapon type (e.g., MIRVed ICBMs); rather, it occurs when one’s entire arsenal is not sufficiently survivable to provide a secure second strike. Hence, in a world with reduced offensive forces and limited ballistic missile defenses, one would expect to see an increased emphasis on launch under attack whether or not MIRVed ICBMs have been eliminated.

A second approach would be for the Russians to increase their alert rate relatively early in a crisis. Figure 13 shows that increasing the Russian alert rate from day-to-day to partial alert increases the Russian secure second strike by approximately 250–300 weapons (assuming the planning factors in Table 2). Increasing the readiness level further would provide a larger margin of security. Strategically, force generation is a stabilizing act. Politically, however, it can be

effective weapons by around 180, compared to the partial alert—LUA case. Using this information, one can interpolate between the partial alert and partial alert—LUA cases to find the approximate fatalities associated with a day-to-day alert—LUA scenario.

35 If Russian leaders implement a LUA option, they must decide which targets to send these missiles against before the U.S. counterforce attack arrives. Sending them against U.S. counterforce targets in a tit-for-tat exchange may keep the war limited; however, most of these weapons will be wasted because few U.S. counterforce targets remain that are worth attacking after a U.S. first strike. On the other hand, if Russia sends these LUA weapons against other military targets or urban/industrial areas, it runs the risk of escalating the war.
quite provocative because it is difficult to separate offensive from defensive intent.

Increasing the alert rate may be especially provocative early in a crisis, because if one side successfully generates its forces while the other side remains at a low alert level, then the side on higher alert may have an incentive to strike first before the opponent brings its forces up to heightened states of alert. Whether first-strike incentives actually exist in this circumstance is arguable. Nevertheless, increasing the alert rate early in a crisis may send the wrong signal. When one side observes the opponent’s force mobilization activities, it will feel pressure to bring its own forces rapidly up to high states of alert. Thus, the instability most likely to develop under START II with limited nationwide ballistic missile defenses is not so much first-strike instability but “mobilization instability,” where both sides feel pressured to generate their strategic forces quickly as a crisis develops to strengthen their deterrent posture. This mobilization spiral does not necessarily lead to war, as has been argued was the case prior to World War I. But it will exacerbate mutual mistrust, increase misperceptions of the opponent’s intent, and thereby make crises more difficult to control. Ultimately, the risk of inadvertent nuclear war increases.

Throughout the Cold War, the Russians practiced force generation less frequently than the United States. In the future, if Russian leaders believe that their day-to-day alert posture is insufficient for deterrence, they will either have to maintain a higher fraction of their force on alert in peacetime or will have to be prepared to generate their forces relatively quickly at the first signs of a crisis. Moreover, the United States should expect the Russians to take these measures. U.S. leaders will have to become accustomed to viewing such actions as defensive and not as signs of offensive intent—provided the United States can match Russian mobilization rates. Once fully generated, neither side has an incentive to strike first.\footnote{Throughout the Cold War the Soviets did not increase their alert rate very often, in part because their larger arsenal provided a secure second strike even on day-to-day alert. The United States, on the other hand, used its alert rate as a form of signaling. The point to bear in mind here is that under START II, with around 600 ground-based interceptors deployed nationwide, the Russians may be forced to increase their alert rate early in a crisis in an effort to shore up their deterrent.}

\footnote{As an aside, one should note that strategic bombers are the first force element that should be placed back on alert in a crisis. There are two reasons for this. First, bomber weapons are ineffective counterforce weapons because of their long flight time. They thus pose less threat of a surprise counterforce attack. Second, placing bomber weapons on alert increases the number of effective weapons to a greater extent than an equivalent number of ballistic missile weapons placed on alert, because bomber weapons face a less menacing defense. If either side’s strategic air defense system is thought to be highly effective, this argument could be reversed—though with the U.S. and Russian strategic air defenses most likely to be in existence over the next several decades, this is not likely}
A third option, which turns out to be the most attractive one under these circumstances, is to limit the size of nationwide ballistic missile defenses so they do not undermine the effectiveness of the U.S. or Russian deterrent, even on day-to-day alert. However, limiting the size of the defense may conflict with the U.S. objective of protecting against accidental, unauthorized, or deliberate third-country ballistic missile attacks. This tradeoff is discussed below.

How Much Ballistic Missile Defense Is Too Much?

Figures 15 and 16 capture the tradeoff between the level of defense required for an effective accidental launch protection system and the level that interferes with the U.S.-Russian strategic nuclear balance. Figure 15 illustrates the results of a simple model that calculates the size of the defense required to completely block an accidental, unauthorized, or third-country launch as a function of the attack size, for different levels of ground-based interceptor effectiveness (i.e., single-shot kill probability). The calculations shown here assume a 90 percent confidence level that no warheads (or objects, since some could be decoys) leak through the defense. Obviously, if one demands a higher level of confidence, more interceptors are required for a given attack size. As one can see from the figure, 100 ground-based interceptors can block (with 90 percent confidence) an attack containing up to around 60 objects, and 200 interceptors to be the case. Hence, both countries should maintain and practice the capability to generate their bomber forces in a crisis.

The defense is modeled here with a shoot-look-shoot firing doctrine using two shot opportunities, where one interceptor is assumed to engage each object in the first shot opportunity (i.e., the first "layer") and the remainder are fired in the second layer. This is close to the optimal firing doctrine for such a defense. The calculation of defense leakage is performed by using the expected number of leakers through the first layer as the input to the second layer of the defense. The binomial distribution is then calculated for the number of leakers through the second layer, assuming the shots are independent. The probability that zero objects leak through the second layer is the confidence level that the attack can be completely blocked. The errors introduced by these approximations are small compared to the factors that have been ignored in this analysis, e.g., interceptor range limitation, actual ballistic missile trajectories, etc. Nevertheless, the number of interceptors calculated using this model is probably accurate to within 10–20 percent (based on a comparison with more detailed simulations conducted at RAND by Mike Miller that include the effects of geography, interceptor range limitations, actual ballistic missile trajectories, etc.). This is certainly accurate enough for the conclusions drawn here. Finally, we have ignored factors such as the probability of establishing an initial track on a reentry vehicle, i.e., factors that are not independent between shots. If the detection probability is too small, e.g., less than 0.995, then it is not possible to achieve high confidence that no warheads will leak through the second layer of the defense. If these factors dominate the leakage rate, a different criterion than zero leakers must be selected for defense effectiveness.
could block an attack containing on the order of 115 objects, assuming the interceptors have a 0.8 single-shot kill probability. As the BMD interceptor effectiveness degrades, more interceptors are required to block a given attack size. For example, if the interceptors have a 0.6 single-shot kill probability, then only about 30 objects could be blocked by 100 interceptors at the 90 percent confidence level.

There is no unequivocal way to determine the size of possible accidental, unauthorized, or deliberate third-country ballistic missile attacks. However, most threats probably fall below 50–100 objects, unless one assumes a large number of decoys that cannot be discriminated by the defense. Therefore,

---

Most accidental Russian ICBM attacks would probably involve fewer than ten warheads, particularly after 2003, when all MIRVed ICBMs have been eliminated from the Russian arsenal—though the number of objects could be larger if decoys are deployed. Unauthorized Russian ICBM or SLBM attacks could be larger, depending on the size of the conspiracy one assumes. An unauthorized attack from a mobile SS-25 battalion could involve up to nine warheads, a single SS-18 launch control facility could launch up to 100 warheads, and a single Typhoon boat could launch 200 warheads (120 warheads if the SS-N-20 is de-MIRVed to six warheads apiece under START II).
limited nationwide ballistic missile defenses on the order of 100–200 ground-based interceptors should provide sufficient protection for a wide range of accidental, unauthorized, or deliberate third-country threats, provided the interceptors are fairly effective (i.e., have a single-shot kill probability greater than 0.7).

One should note that this calculation is approximate because it ignores interceptor range limitations (i.e., it assumes each interceptor has sufficient range

Fortunately, highly MIRVed ICBMs will be eliminated after 2003. The “mad submarine commander” is the only scenario left that, in principle, generates large unauthorized attacks. However, this scenario should be discounted, since the Russians have recently revealed that they have Permissive Action Links on their submarines to prevent unauthorized SLBM launches. Consequently, most remaining accidental or unauthorized attacks after 2003 will likely contain fewer than 10 warheads, with perhaps as many as 50–100 objects if advanced decoys are deployed. Third countries are not likely to develop large ICBM arsenals, much less SLBMs or MIRVed ICBMs, for some time. Hence, attacks from these countries will probably involve fewer than 10 warheads. The debate about possible attack sizes will ultimately revolve around the number and types of decoys an adversary might deploy, as well as on the ability of the defense to discriminate decoys from actual warheads.
to engage any of the incoming objects), and it assumes effective warning and track information early in the ballistic missile's flight trajectory to cue the interceptors. Such an effective battle management capability requires upgrading the existing U.S. ballistic missile early warning network, possibly with space-based detection and tracking sensors such as "Brilliant Eyes." If more than a single site is required to provide coverage of the entire continental United States (due to interceptor range limitations or the lack of accurate track information early in the ballistic missile's flight), then the number of interceptors required for an effective limited nationwide defense would be larger than the numbers shown in Figure 15.\textsuperscript{40} The exact amount depends on the number of sites, the degree of site overlap, and the extent to which one believes the attack may be focused only on one site as opposed to being spread across the entire United States (in which case, having multiple sites would not significantly increase the interceptor requirements from those given in Figure 15).

We now return to the question of how much defense is too much. From the preceding discussion, it is clear that the first strategic problem to occur under START II is the weakening of the Russian deterrent on day-to-day alert. Figure 16 plots the size of the Russian retaliation on day-to-day alert (measured in effective weapons) as a function of the number of ground-based interceptors deployed on both sides. The number of effective weapons is shown for three levels of defense effectiveness. (Until now, the size of each side's retaliation has been shown only for an interceptor SSPK of 0.8.)

This calculation assumes the United States tries to maximize its counterforce effectiveness by allocating 1400 out of the 1520 available nuclear weapons to the counterforce first strike on day-to-day alert. If 600 interceptors are deployed, the size of the Russian secure second strike drops to around 20 effective weapons (for an SSPK of 0.8), assuming none are withheld for a secure reserve force. This is the result discussed in connection with Figure 13. If more than 600 interceptors are deployed, the size of the effective Russian retaliation could be reduced, on average, to as low as 2–3 weapons. Clearly, defense deployments over 600 interceptors provide the United States with a splendid damage-limiting first-strike option on day-to-day alert—assuming the defenses can survive, that decoys can be discriminated, and that the BMD interceptors have an effectiveness close to 0.8 SSPK. If the defense effectiveness drops to 0.6, the extent to which damage can be limited drops significantly (e.g., around 110 warheads leak through a 600-interceptor defense if the SSPK is 0.6 and approximately 20 warheads leak through a 1000-interceptor defense). Thus, effective damage limitation with modest defense deployments requires high interceptor performance.

\textsuperscript{40}The additional interceptors that would be required to defend Hawaii and Alaska have been ignored—something an analyst can do but not a politician.
To avoid the instability created by a U.S. damage-limiting first-strike option, the number of ground-based interceptors should be constrained to less than approximately 200 interceptors, leaving Russia with around 300 effective weapons with which to retaliate (regardless of interceptor SSPK). Note that an important assumption here is that theater BMD interceptors do not add appreciably to the strategic BMD capability—an assumption one may worry about if thousands of theater interceptors are deployed (especially in light of past debates about the strategic BMD capability of high-performance surface-to-air missiles. An ABM Treaty–constrained deployment of 100 ground-based interceptors has minimal effect on the Russian retaliatory capability. From Figure 15 one observes that 100 interceptors still provide the United States with an effective accidental launch protection system, provided the interceptor SSPK is above 0.7 and likely threats contain no more than around 50 objects.

CONCLUDING OBSERVATIONS

The United States and Russia cannot arbitrarily reduce the size of their strategic nuclear arsenals and the alert rate associated with these forces, and then deploy limited nationwide ballistic missile defenses, without eventually undermining deterrence. If U.S.-Russian relations continue to be cooperative in the future, then a robust deterrent posture between these two states may be of less concern. However, one cannot predict with certainty that crises between these former adversaries will never occur. Therefore, it behooves U.S. and Russian leaders to avoid creating, in cooperative times, a strategic nuclear deterrent posture that, in a crisis, proves to be weak. The likelihood of conflict with Russia may be small in the future, but with events this fateful, leaders should err on the side of caution.

The perspective advanced in this analysis is not that a weak day-to-day deterrent posture leads to premeditated nuclear war in peacetime, but rather that it forces the strategically disadvantaged side to react in ways to redress its perceived vulnerability. This reaction may appear provocative in a crisis and, hence, could increase the chance of miscalculation, thereby making crises more difficult to control and increasing the likelihood of inadvertent nuclear war.

With offensive forces constrained by START II, and assuming no nationwide ballistic missile defenses on either side, neither the United States nor Russia can significantly limit damage by striking first with an all-out counterforce attack. The mutual hostage relationship that characterizes mutual deterrence will remain. In particular, on generated alert—the most likely posture for U.S. and Russian forces if a crisis develops—both countries will have around 2200–2500 strategic nuclear weapons that can survive attack and pene-
trate the opponent’s strategic air defenses to arrive on target in the opponent’s homeland. This retaliatory capability is sufficient to hold at risk a wide range of military, leadership, and economic targets even after weapons have been subtracted for a secure reserve force.

At lower alert rates, the retaliatory capability drops. For example, on day-to-day alert the United States will retain the capability to deliver up to approximately 900 weapons after absorbing a Russian counterforce first strike, assuming the United States does not reduce its SLBM alert rate. Projected Russian force postures are more vulnerable at low alert rates because relatively few Russian SLBMs are assumed to be at sea and few mobile ICBMs are assumed to be out of their garrisons when a U.S. counterforce attack arrives. Even so, approximately 450 weapons could be delivered against the United States in retaliation on day-to-day alert, unless a significant number of weapons are withheld for a Russian secure reserve force.

If one assumes that deterrence rests on the ability to inflict casualties on the opponent’s population, then deterrence under START II is more robust. On day-to-day alert, the relatively small (by Cold War standards) surviving arsenals can still do tremendous damage. For example, 900 arriving U.S. weapons would be able to threaten approximately 60 million Russian fatalities (approximately 60 percent of the Russian urban population), and 450 arriving Russian weapons can threaten around 50 million U.S. fatalities (approximately 30 percent of the U.S. urban population).

Based on these observations, one might conclude that there is room for further offensive force reductions beyond the levels set by the START II Treaty, or for further reductions in U.S. or Russian peacetime alert rates. The question of further cuts depends on whether 2200–2500 effective weapons is more than is necessary to deter all plausible resurgent Russian threats (including a revanchist Russian state led by Vladimir Zhirinovsky). If one believes deterrence rests on the ability to credibly inflict societal destruction, the answer will likely be that further cuts are acceptable (desirable), possibly down to as few as 1000 strategic nuclear weapons in each side’s arsenal. However, if one believes that deterrence of a resurgent Russia requires the ability to credibly threaten Russian military as well as economic assets, then reductions below START II levels are probably not possible, unless they are minor.

With respect to peacetime alert rates, again one might conclude that further reductions are possible. Certainly the U.S. peacetime alert rate seems more than adequate under START II. However, the crucial point to bear in mind is that reductions in the peacetime alert rate increase one’s dependence on force generation to provide an adequate deterrent in times of U.S.-Russian tension. In principle, strategic force generation is a stabilizing act because it increases the effectiveness of each side’s deterrent. Politically, however, force generation, especially if it occurs rapidly at the onset of a crisis, is frequently viewed as a
provocative act. Therefore, "mobilization instabilities" may become a central feature of the future U.S.-Russian strategic nuclear interaction. Managing the dynamics associated with U.S. and Russian strategic nuclear force mobilization will become crucial for preventing future crises from spiraling out of control. This mobilization spiral does not necessarily lead to war, as some have argued was the case prior to World War I. However, it will make crises more difficult to control. Greater transparency with respect to U.S. and Russian force generation procedures might help reduce the chance for misperception. These observations are predicated on the assumption that neither the United States nor Russia deploys an appreciable limited nationwide ballistic missile defense. If such a defense is deployed, this would have a significant impact on the above observations.

Nationwide ballistic missile defenses have two competing effects. They strengthen deterrence by complicating the opponent's first strike, and they weaken the opponent's deterrent by undermining his secure second strike. Of these two effects, the second dominates at modest levels of defense. In particular, with 600 ground-based interceptors deployed on both sides (space-based defenses are not considered in this analysis), the Russian retaliation on day-to-day alert could be reduced to as few as 20 arriving weapons after a U.S. counterforce attack. Such a damage-limiting capability was never achieved by either side during the Cold War. Russia's deterrent is the first to become vulnerable because it has fewer weapons deployed in survivable basing modes on day-to-day alert. If one considers offensive force reductions below START II, obviously less defense would be required to limit damage to the same degree.

On generated alert, the U.S. and Russian START II deterrent forces are quite robust, even at defense levels of 600 ground-based interceptors. Therefore, if one believes that U.S. and Russian forces will always be on generated alert in a crisis, then deterrence will be robust even in the presence of limited nationwide ballistic missile defenses. However, it is the transition from peacetime to generated alert that concerns us. Since deterrence at low alert rates is compromised, the Russians have a strong incentive to improve their de-

---

41 This same critique was leveled against nationwide ballistic missile defenses in the mid-1980s. However, at that time the level of defense needed to upset the strategic balance would have been quite large. Under START II, relatively low levels of defense can undermine deterrence.

42 This assumes highly effective ground-based interceptors, i.e., interceptors with a single-shot kill probability of 0.8 and sufficiently accurate and timely track information to allow for a shoot-look-shoot defense firing doctrine.

43 The defense required to seriously interfere with either side's deterrent on generated alert is around 3500 ground-based interceptors—deployments that will not likely be realized in the near future.
terrent posture. Again, the point here is not that a U.S. President might be tempted to launch a surprise attack out of the blue, but rather that the Russians, realizing their strategic vulnerability, will be forced to improve their secure second-strike capability. This is the main strategic problem created in an environment characterized by low force levels (i.e., START II levels), low alert rates, and limited nationwide ballistic missile defense deployments on the order of 600 ground-based interceptors or more.

Several solutions to this problem are possible. First, Russia could simply maintain a higher fraction of its submarines at sea or mobile ICBMs out of garrison in peacetime. This is not likely to occur for reasons of cost. Second, Russia could threaten to launch its vulnerable forces out from under a U.S. counterforce attack. This is unattractive because it increases the likelihood of an accidental or unauthorized Russian missile launch in the midst of a crisis—particularly in light of the relatively poor condition of Russia’s ballistic missile early-warning network in the wake of the collapse of the Soviet Union (though obviously this system may be improved by the time START II is finally implemented). This, in turn, increases the chances for inadvertent nuclear war, a situation neither country wants.

Third, the Russians could increase the alert rate associated with their strategic nuclear forces early in a crisis. From a strategic perspective, increasing the alert rate is a defensive act because it increases the fraction of the force that can survive a counterforce first strike and, hence, strengthens deterrence. However, it is also consistent with offensive intent. Force generation thus is likely to be politically provocative because it forces the United States to alert its forces fairly quickly to match the Russians. Again, one is faced with the prospect of mobilization spirals, as discussed above.

Finally, the best approach would be to limit the size of the ballistic missile defenses on both sides. If nationwide ballistic missile defenses are limited under START II to fewer than 200 ground-based interceptors deployed nationwide, deterrence would be minimally affected because the Russians would still have around 300 effective weapons with which to retaliate. A defense limited to 100 ground-based interceptors deployed nationwide, as required by the ABM Treaty, clearly would not upset the strategic nuclear balance. Note that an important assumption here is that theater ballistic missile defenses do not add appreciably to the strategic ballistic missile defense capability—an assumption one may question if highly capable theater defenses are deployed.

In addition, 100–200 highly effective ground-based interceptors should be adequate to protect against a wide range of accidental, unauthorized, or deliberate third-country ballistic missile attacks, unless one assumes that a large number of decoys cannot be discriminated by the defense. Thus, it should be possible to have a reasonably effective accidental launch protection system without undermining either the U.S. or Russian strategic nuclear deterrent—a
requirement written into U.S. law by the 1991 congressional Missile Defense Act. Such a defense system may violate some elements of the ABM Treaty.\footnote{Even if one assumes only 100 interceptors are deployed, the ABM Treaty would still have to be modified to allow upgrades to each country's ballistic missile early-warning network (possibly including space-based components), and it would have to be modified to allow interceptors to be deployed at more than one site.} However, this is not necessarily inconsistent with U.S. and Russian post–Cold War security objectives.

In general, as U.S. and Russian nuclear arsenals shrink, stable deterrence will increasingly depend on maintaining a significant fraction of forces on alert, particularly if limited nationwide defenses are deployed. Reducing the number of weapons in each side's arsenal or “de-alerting” elements of the strategic nuclear force as a symbol of “nuclear disengagement” may improve political relations in peacetime; however, it can create strategic vulnerabilities that may make future crises more difficult to control. Though deterrence of a hypothetical Russian attack is not foremost in the minds of U.S. policymakers, it still behooves them to carefully weigh the implications of further reductions below START II levels, further cuts in the peacetime alert rate, and the deployment of limited ballistic missile defenses beyond several hundred ground-based interceptors before taking the next steps to reshape the U.S.-Russian strategic nuclear landscape.

**BIBLIOGRAPHY**


Jervis, Robert (1976), Perceptions and Misperceptions in International Politics, Princeton University Press, Princeton, NJ.

Kent, Glenn, and David Thayer (1989), First-Strike Stability: A Methodology for Evaluating Strategic Forces, RAND, Santa Monica, CA.


Wilkening, Dean (1994), The Impact of START II and Limited Nationwide Ballistic Missile Defenses on Deterrence and Crisis Stability, RAND, Santa Monica, CA.