CENTRAL REGION STABILITY IN A DEEP-CUTS REGIME

Paul K. Davis

December 1989

P-7610

The RAND Corporation

Papers are issued by The RAND Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of RAND's contracts or grants. Views expressed in a Paper are the author's own and are not necessarily shared by RAND or its research sponsors.

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

PREFACE

This paper was prepared for a symposium entitled "Conventional Stability in Europe: Prerequisites and Analytic Requirements" held at the University of the Bundeswehr in Munich, Germany during the period October 10-13, 1989. I updated the paper slightly in December 1989 to reflect recent events in Eastern Europe. The paper, less the preface and summary, will appear as a chapter in a book stemming from the symposium edited by Dr. Reiner Huber and will be published by NOMOS in Baden-Baden in the spring of 1990.

SUMMARY

There is widespread interest in the possibility of deep cuts in conventional force in Europe, cuts going well beyond those that will be necessary under the Conventional Armed Forces in Europe (CFE) treaty now being negotiated. It is not uncommon to hear suggestions about final levels approximately 50 percent of NATO's current levels. However, there has been considerable concern expressed by Western analysts about the potential *destabilizing* aspects of such deep cuts and, as of 1989, NATO governments had been prudently conservative on the matter.

This paper critically reviews the static-analysis arguments underlying concerns about low force levels and presents the results of both simple analytic modeling and more complex simulation modeling exploring the deep-cuts regime (Davis, Howe, Kugler and Wild, 1989). It concludes that *from a theoretical perspective there is nothing inherently destabilizing about a deep-cuts regime—so long as the sides have approximately equal force-generation capabilities (force levels versus time).* The analysis indicates that there is no clear-cut "operational minimum" for the defender, although defensive maneuvering would become increasingly critical as the level of forces employed in the Central Region drops below a fuzzy threshold of about 20 to 30 "equivalent divisions," roughly 25 percent below the level of in-place and reinforcing forces nominally associated with the Central Region after enactment of the current NATO CFE proposal. Even below this fuzzy threshold, military stability (i.e., force-posture stability) could be high with respect to classic large-scale invasions and a contingent version of forward defense could still be a viable strategy. At low enough levels—below what might be called the "offensive minimum"—the attacker would be unable to achieve and maintain a strategic victory even if he won the initial battles.

On the negative side, the paper concludes that a scaled-down version of NATO's current layer-cake force posture would be highly inappropriate in a deep-cuts regime. At low force levels it would be especially important for the defender to have many attributes not characteristic of NATO today: unity of command, commonality of doctrine and logistics, flexibility in operational strategy, and the capability and willingness to conduct maneuver warfare from the outset, including both counterattacks and fallback movements—first to deter the attacker from stripping forces from some regions of the front to concentrate forces elsewhere, second to counterconcentrate quickly in response to early intelligence indicators of likely main-thrust attack sectors, third to avoid decisive

battles under adverse conditions (e.g., by withdrawing to more defensible lines if necessary), fourth to reinforce main-thrust sectors quickly with forces maneuvered over considerable distances, and, lastly, to counterattack and restore lost territory. Serious rethinking and restructuring would be necessary.

The maneuver capability needed would depend on the opponent's posture, which could be greatly affected by "stabilizing measures" ("operational arms control") negotiated in the CFE process. Of particular value would be the substantial or total withdrawal of Soviet forces and infrastructure from Eastern Europe and other constraints making it difficult to achieve operational surprise, which is particularly dangerous at low force-to-space ratios. Stability at low force levels would also be enhanced by force structures assuring that the defender always has a mobility advantage in his own rear areas. This may imply encouraging personnel carriers, air cavalry, air-mobile infantry forces, and capability to rapidly deploy obstacles, while constraining mobile air defenses that could move with an attacker.

Moving from theoretical parity to realism, there are more problems. Even if the sides had static parity in the Atlantic-to-the-Urals (ATTU) aggregate, this would not assure *dynamic parity* in a Central Region conflict. The D-Day force ratio in the Central Region could be adverse because of faster Pact force generation or the use by the Pact of out-of-region forces such as those in the ATTU but not nominally associated with the Central Region, or even from out-of-ATTU forces from areas such as Central Asia and the Far East. The significance of an adverse D-Day theater force ratio (e.g., 1.5:1) would be substantially greater at force levels below the "operational minimum" than it would be today (except in the limit of force levels so low as to preclude a permanent strategic victory by the attacker even if he were successful in early battles). With such a theater-force-ratio advantage, the attacker could concentrate of 3 or 4 main sectors and hope to accomplish large-scale encirclements. *This threat of out-of-area forces appears to be the most fundamental obstacle to achieving Central-Region stability at low force levels.* Although the United States has additional forces that could be used in the Central Region, the Soviet Union would have an advantage in deployment times. Also, light forces such as those of Spain and Turkey are unsuitable for Central Region conflict, while the bulk of "out-of-area" Soviet forces are mechanized.

As a practical matter, the principal NATO mechanisms for dealing with the "out-of-area" threat to Central Region stability in a low force-level regime appear to be: (a) limiting through arms control the number of out-of-area ground forces and imposing stabilizing measures (e.g., constraints on readiness, large-scale movements and exercises)

to increase warning time and the likelihood that response would be prompt and cohesive (Davis, 1988); (b) developing the capability for early large-scale interdiction efforts in the event threatening redeployments begin; and (c) reducing the assessed threat as the nations and armies of Eastern Europe become increasingly independent.

Given the stakes involved, it is important for NATO to proceed carefully if deep cuts are to be implemented in a CFE II. It seems unlikely that the Soviet Union will soon be able to absorb cuts going beyond those required of her by CFE I, so time is available for NATO to make the needed changes in forces, doctrine, and strategy, and for the sides to extend negotiations to handle the out-of-area problem and cope with the changes occurring in Eastern Europe, which are making obsolete the focus on the intra-German border (IGB) and images of monolithic-Pact versus monolithic-NATO conflicts.

CONTENTS

PREFACE iii
SUMMARY v
CONTENTS viii
Introduction 1
Defining Terms of Debate: What is "Stability?" 2
Static Analysis of the "Operational Minimum" 3
Understanding the Significance of Theater-Force Ratio 10
Simulation of Defense at Low Force Levels 15
When Parity Doesn't Mean Parity 22
Summing Up 24
REFERENCES 26

INTRODUCTION

Both NATO and the Warsaw Treaty Organization (WTO) have proposed negotiated reductions to parity in ground forces at levels significantly lower than NATO's current levels. It is therefore appropriate to ask how low those force levels might be made. It may seem natural to assume "the lower the better," but many military and civilian figures in the West have argued that reductions below a threshold sometimes called the "operational minimum" might actually *decrease* stability. As of early 1989 it was widely believed within the NATO community that this threshold might be only 10 percent or so below NATO's current force levels. The analysis described in this paper argues that no clear-cut operational minimum in the sense of a sharp threshold exists. A regime of much lower force levels (e.g., 20 divisions per side for the Central Region) could be stable or unstable, depending on details of force postures, strategy, and criteria; furthermore, at low enough force levels—below what might be called the *offensive minimum*—a would-be attacker would be unable to achieve and exploit a strategic victory even if he won the first campaign. There *are* reasons— serious reasons— for NATO to worry about stability at low force levels, but those reasons should be seen more as incentives for wise and innovative force planning than as excuses for avoiding the substantial demobilization that common sense suggests is decades overdue.[1]

The paper has five parts. The first discusses the *concept* of stability. The second uses static methods to discuss the notion of the so-called operational minimum. The third introduces a simple analytic model to explain the sensitivity of the warfighting balance to theater-force ratio and to discuss some of the consequences of the sides asymmetrically reducing to nominal parity and then accepting additional deep and symmetrical cuts. Both the static analysis and analytic model lay the groundwork for the fourth part of the paper, which describes theater-level simulation experiments to assess the feasibility of stability at and well below the static estimate of the operational minimum. Even the simulation experiments emphasize simplicity and transparency so as to clarify certain transcendent issues for the low force-level regime. The final section then summarizes conclusions relevant to the CFE negotiations and national defense programs.

---

[1]This paper is for the most part a shortened version of Davis, Howe, Kugler and Wild (1989), but it includes a more pointed summary of the author's personal conclusions and some comments reflecting the dramatic changes occurring in Eastern Europe as the paper was being edited in December 1989.

DEFINING TERMS OF DEBATE: WHAT IS "STABILITY?"

One of the more subtle problems in the debate about stability has been the definition of stability itself. Regrettably, analysts (including the author) have often added to the confusion by focusing strictly on a military-technical definition. The tightest version of this is that stability exists if and only if both sides correctly believe that an attack by *either* side would be successfully defeated by the defender conducting a forward defense strategy—either by holding at or close to the border in the first place or, in the event of unavoidable initial penetrations, by conducting counterattacks restoring the status quo ante.[2]

There is much to be said for this definition. It is concrete; it focuses on capabilities rather than intentions; it relates to NATO's operational strategy of forward defense; it recognizes the two-sided nature of the problem, including the role of perceptions and the non-zero-sum nature of the game; it recognizes that counterattacks may be necessary; and it allows stability to be measured with models and simulations (see especially Huber (1988). Many Western and Eastern analysts have recently embraced this or a closely similar definition, a development which is itself a tangible sign of remarkable East-West intellectual progress.

This said, there are also many problems with the definition. First, it pertains only to what the author calls *force-posture stability under forward defense strategy;* it has nothing to say about what might be called *international relations stability*, the related *likelihood-of-military-crisis stability, arms-race stability,* or *domestic political stability.* This creates a credibility gap between analysts and policymakers.

Further, as applied in models and simulations, the definition creates an artificial requirement for a particular version of forward defense that would be militarily irrational in some scenarios. By doing so and by ignoring distinctions among battles, theater campaigns, and wars, the definition establishes unreasonably stringent criteria for early battlefield success by the defender. This, in turn, insidiously encourages the corruption of models or model calibration so as to permit "success" under conditions where that

---

[2]For discussions of what the author calls "force-posture stability under forward defense," see especially Huber (1988) and Rohn (1989). Those interested in stability theory may also wish to examine certain strategic-nuclear studies that treat conceptually analogous issues in detail. See Kent, De Valk, and Thaler (1988) and Wilkening and Watman (1986) on force-posture stability; see Davis (1989a) for a theory of crisis stability that incorporates behavioral factors and decisionmaking.

particular kind of success would be unlikely in the real world. One result is a credibility gap between analysts and operationally oriented military officers (and historians).

Finally, it should be emphasized—however impolite this may seem in forums in which East and West are meeting and talking—that it is *not* sensible to define stability as though the geostrategic situation were symmetric. The above definition derives from a game-theory-like image in which the intra-German border (IGB) is the natural boundary and in which both sides seek to guarantee forward defense capabilities at that boundary. That image is fundamentally wrong from a strategic perspective. Its perniciousness can be seen by visualizing an outcome that would achieve this type of "stability": nearly impenetrable fortress walls along the border, backed up by high densities of infantry and the permanent occupation of Eastern Europe by large numbers of Soviet forces. That is not an image of what the author would regard as desirable stability.[3]

Having raised these issues, the remainder of this paper is nonetheless focused primarily on force-posture stability. The paper's thrust and conclusions, however, depend in part on recognizing the other dimensions of the stability issue.

STATIC ANALYSIS OF THE "OPERATIONAL MINIMUM"

As of 1988 and early 1989, many analysts spoke of an "operational minimum" of forces below which stability, by which they meant force-posture stability under forward defense strategies, would be endangered—*even at parity*. The logistic basis for this concern is easily understood: if the defender has to cover the same frontage while the forces to do so are reduced, and if we assume that the attacker has the first move—choosing and concentrating on the points of attack without an immediate defender reaction—there will come a point in reductions at which the defender's density of forces (force-to-space ratio) will be unable to prevent the attacker from penetrating before the defender can respond, because the defender's tactical reserves will have to maneuver over long distances. The fact of a penetration by no means implies success of the invasion, and many nations have successfully defended themselves over the centuries with low overall force-to-space ratios, but the strictest version of forward defense requires that no such penetrations be permitted. The question then is where this point occurs: how many divisions must the defender have on a given frontage to prevent such penetrations? Later

[3]The author's arguments on this are longstanding (see Davis (1988:49-50) and Davis (1989a), but they were controversial until the events in Eastern Europe late in 1989, which made it evident that the IGB (and simple images of the East-West military competition) are becoming historical artifacts.

in the paper we shall use dynamic analysis to demonstrate that in fact no sharp point or threshold exists. We will also consider alternatives to an unconditional forward defense strategy. However, the dynamic analysis uses a simulation model that also depends on force-to-space considerations, and the underlying parameters dominating that analysis are closely related to the so-called operational minimum. Thus, the static analysis provides background for understanding, setting parameter values in, and assessing uncertainties of the dynamic analysis.

Estimates of the operational minimum are based on taking the frontage to be defended, dividing by a rule of thumb for the maximum frontage a division can cover while holding ground, and multiplying by or adding in a factor providing for operational reserves. Sometimes the reserve factor is omitted and the calculation is performed as though all divisions are on the front from the start. In this case, a different rule of thumb should be used. Either approach sounds straightforward, and such calculations have been used for decades and reflected in high-level policy studies as well as informal work and academic papers.[4] Despite its usefulness, however, the approach is riddled with opportunities for ambiguity, confusion, and error. As a result, drastically different estimates of NATO's so-called operational minimum (from approximately 15 to 45 divisions for the Central Region) have been presented by or can be inferred from the presentations of respected organizations.

The author and colleague Robert Howe emphasized these discrepancies early in 1989 and concluded after research that they stem principally from different but usually implicit assumptions about the following (see Davis, et al., 1989 for details):

- *Geographical frontage* (from 675 km to 1000 km for the Central Region, depending on whether one considers the Danish and FRG-Austrian borders, and the extent to which the defender defends on a straight-line front rather than following the border's contour)
- *Effective military frontage* (variously estimated from 30 percent to 100 percent of geographic frontage)
- *Method of counting "divisions"* (e.g., division flags versus the use of "equivalent divisions (EDs), "armored division equivalents" (ADEs), "division equivalents in firepower" (DEFs) or "division equivalents in manpower" (DEMs))

[4]See, for example, Liddell-Hart (1960), OSD / PA&E (1979), Mako (1983), Flanagan and Hamilton (1988), and Mearsheimer (1989).

- *Stringency of defense requirement* (from nominally holding ground everywhere to holding ground operationally while permitting temporary and strategically insignificant tactical penetrations)

- *Treatment of reserves* (e.g., assuming all divisions are on the line versus distinguishing between on-line divisions and those divisions in corps, army-group, or theater-level reserves)

- *Rule-of-thumb maximum divisional frontages* (i.e., disagreements on the rules of thumb that persist even with all other factors well defined)

To obtain a high-end estimate one might, for example, use a geographic frontage of 750 km, ignore issues of terrain usability, apply a rule of thumb of 25 km/division, and then throw in a reserve factor of, say, 50 percent (for the total of corps, army-group and theater-level reserves). The result: 45 divisions. The same result is sometimes achieved using 900 km, a rule of thumb of 30 km per division, and a 50 percent reserve fraction. The author believes such high-end estimates are incorrect. The oft-cited doctrinal rule of thumb of 25 km per division applies properly to sectors of front that must be defended heavily because the terrain is suitable for large-scale offensive operations. By contrast, much lower densities should be adequate for heavily forested, urbanized, or mountainous portions of the front—*so long as* the defensive commander and forces are competent and mobile operational reserves are available to cope with unanticipated offensives through such regions (Davis, 1988:62).

Figure 1 provides an approximate, unofficial, but illustrative breakdown of required tactical densities by type of terrain.[5] It assumes a rigorous forward defense and ignores the impact of tactical air forces. If the defender had superiority in air forces or if tactical penetrations could be allowed in certain regions, the rules of thumb could be relaxed significantly, especially for rough and closed terrain. Closed terrain is extremely mountainous or, for example, heavily urbanized; rough terrain might be heavily forested and/or rather mountainous with relatively few roads; mixed terrain might have enough forested areas and hills to provide good defensive cover, but would have a higher density of roads; open terrain is essentially all usable by tactical units.

Figure 1 gives requirements in two ways: requirements for firepower as measured by equivalent divisions (EDs) (a U.S. armored division counts as 1.0 ED) and

---

[5]Figure 1 comes from Davis et al. (1989). The author personally believes that *if* operational surprise can be made unlikely through operational arms control in the CFE talks (Davis, 1988), the rules of thumb shown here will then be too high by perhaps 25 percent when measured in EDs or that they will apply even when the defender's operational reserves are rather modest, especially if some temporary penetrations can be tolerated.

requirements in division equivalents in manpower (DEMs). The concept here is that in defense-favorable terrain the defender's firepower density can be much lower, but to prevent local penetrations (e.g., by specialized infantry) that might undercut a much larger sector's defense lines, there must still be a substantial *effective* density of manpower. If one is conservative on such matters, the only remedy is a high density of manpower on the ground. If one is more optimistic about the responsiveness and effectiveness of air-mobile infantry, attack helicopter units, tactical fixed-wing aviation, tactical and operational reserves, and modern systems for reconnaissance, surveillance, targeting and long-range maneuvering of fire, then one may conclude that the DEM requirements stated in Fig. 1 are much too high.

Strong differences of opinion exist about current effectiveness and responsiveness—even among the author and his colleagues at RAND[6]—but consensus exists about the importance of achieving high effectiveness and responsiveness, especially as reductions proceed. Discussion of such matters takes place in several different terminologies. One group may emphasize the need to develop tailored divisions that can better exploit special features of terrain; these might be *hybrid divisions* intermediate between standard mechanized and infantry divisions. Another group may emphasize Reconnaissance Surveillance Tracking and Acquisition (RSTA) capabilities and maneuvering of fire. Yet another group may talk more in the language of "nonoffensive defense," emphasizing the value of obstacles, innovative use of infantry, and so on. A key concept here is the relative mobility of the defender and the attacker in the rear area of the defender (see also Huber, 1989). Stability is enhanced to the extent

[6]On the cautionary side, defenders have been surprised and routed by attackers who used rough terrain (e.g., the Ardennes forest) that *should* have been eminently defendable had they been adequately covered, but they were not. Also, the author has the impression that some Western military officers have a tendency to underestimate the usability of terrain and the potential speed of tactical maneuver forces (see, for example, Donnelly (1988:28) and Simpkin (1984)). Liddell-Hart (1960) also emphasizes that defensive adequacy depends critically on the *quality* of defending forces, and discounts by perhaps 50 percent the effectiveness of units with numerous conscripts. At the same time, it seems likely that defendable frontages have increased significantly over the last three decades for technological reasons, but the rules in Fig. 1 do not reflect that. Based on WW II experiences, Liddell-Hart estimated in 1960 that a front line of ten mobile divisions should be able tactically to hold a geographic frontage of 400 km along the northern half of the IGB. Figure 1's methodology would give a very similar result. If Liddell-Hart was correct in 1960, then the numbers in Fig. 1 might reasonably be raised by 20 to 50 percent today as the result of increased mobility, range, and firepower (as a point of comparison, a 1960 armored division's ED score was probably between 0.5 and 0.8). As noted earlier, however, the more optimistic figures may be prudent only if force postures change so as to make operational surprise extremely unlikely.

that the defender always has the mobility advantage. This may be achieved, for example, by systems such as transport helicopters and Infantry Fighting Vehicles (IFVs), which are extremely vulnerable except in one's own rear areas.

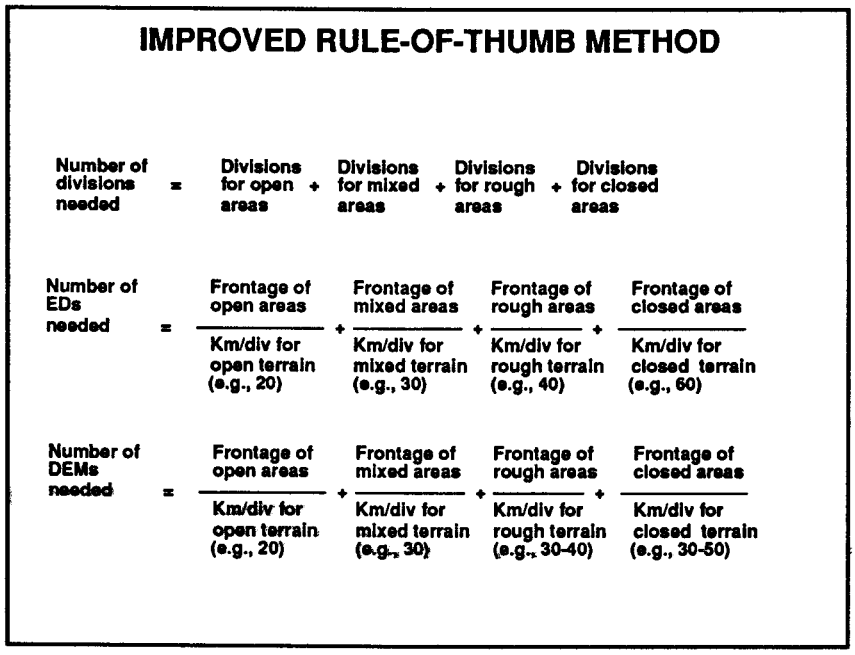


Fig. 1

For analytic work it is often useful to simplify the treatment of terrain by working in terms of the "effective military frontage," which is a fraction of the geographic frontage. In practical terms, this becomes the numerator into which one divides the rule of thumb. As an example, if a sector 100 km wide had 10 percent closed, 60 percent rough, 20 percent mixed, and 10 percent open terrain, then one could estimate the same requirement using Fig. 1 or by characterizing the sector as having an effective military frontage of 65 percent (the sum of open and mixed plus a 50 percent allowance for the rough and closed) and then using the rule of thumb for open-to-mixed terrain on that "effective" military frontage. The results would be essentially the same in the two calculations (2.8 versus 2.6 EDs).[7] More commonly, military analysts discount rough and

[7]Unfortunately, when one maps n variables into n-2 there are different ways of doing so and it is very unusual for those developing terrain data bases for analytic models and war games to document their rationale and definitions. As a result, terrain data cannot be compared readily from

closed terrain even more (e.g., using a 33 percent factor), in which case the effective military frontage in the example would be 53 percent and the requirement 2.1 EDs. Implicitly, this amounts to relaxing the rules of Fig. 1 to permit some tactical penetrations.

With that approximation, Fig. 2 shows a nomogram for computing the operational minimum using one's own judgments about: geographic frontage, effective military frontage, the rule of thumb for a defending division's maximum frontage while holding ground, and the reserve multiple. The "correct" planning factors should come from operational military commanders with intimate familiarity with the relevant terrain, but the strawman example assumes 750 km of geographic frontage (the figure most often used over the years in policy studies of Central Region defense), 60 percent of that for effective military frontage, a rule of thumb of 25 km per equivalent division, and a reserve factor of 1.50. The strawman result, then, is a statically derived strawman "requirement" for 27 EDs. This translates into approximately 33 division equivalents in firepower (DEFs) or average NATO divisions. If we assume 300 tanks per ED as a rough figure, then the "requirement" is for about 8100 tanks.[8] There is considerable uncertainty in all these figures and the author prefers an estimate about 25 percent lower.[9]

With this background on force-to-space issues and static methods of estimating force requirements for forward defense, let us now consider some dynamic considerations—first with a simple analytic model, and then with a theater-level simulation model.

---

one model to another. An additional complication is that the apparently same terrain-data items in two models may have subtly different meanings. See Davis, et al. (1989b:34-35).

[8]Under NATO's CFE proposal, the sides would have *approximately* 10,500 tanks or 35 EDs nominally associated with Central Region conflict (in-place forces plus reinforcing units). Thus, the 27-ED "requirement" suggests the ceilings could be reduced by roughly another 25 percent.

[9]The original motivation for using 750, 60 percent, and 25 km of effective military frontage per ED was that these correspond to the nominal values of certain related data items in the RAND Strategy Assessment System (RSAS) simulation used by the author and colleagues and discussed briefly in a later section of this paper. Also, these were "middle of the road" values. In retrospect, based primarily on historical data such as that in Liddell-Hart (1960), the author himself would either use a larger rule of thumb (about 31 km per ED) *or* a smaller reserve multiple (1.25). In that case, the statically derived "operational minimum" would be about 22 EDs or about 27 DEFs or divisions. High-resolution gaming accounting explicitly for weapon-on-weapon capabilities in the terrain of interest (e.g., JANUS) might be useful in shedding light on this estimate.

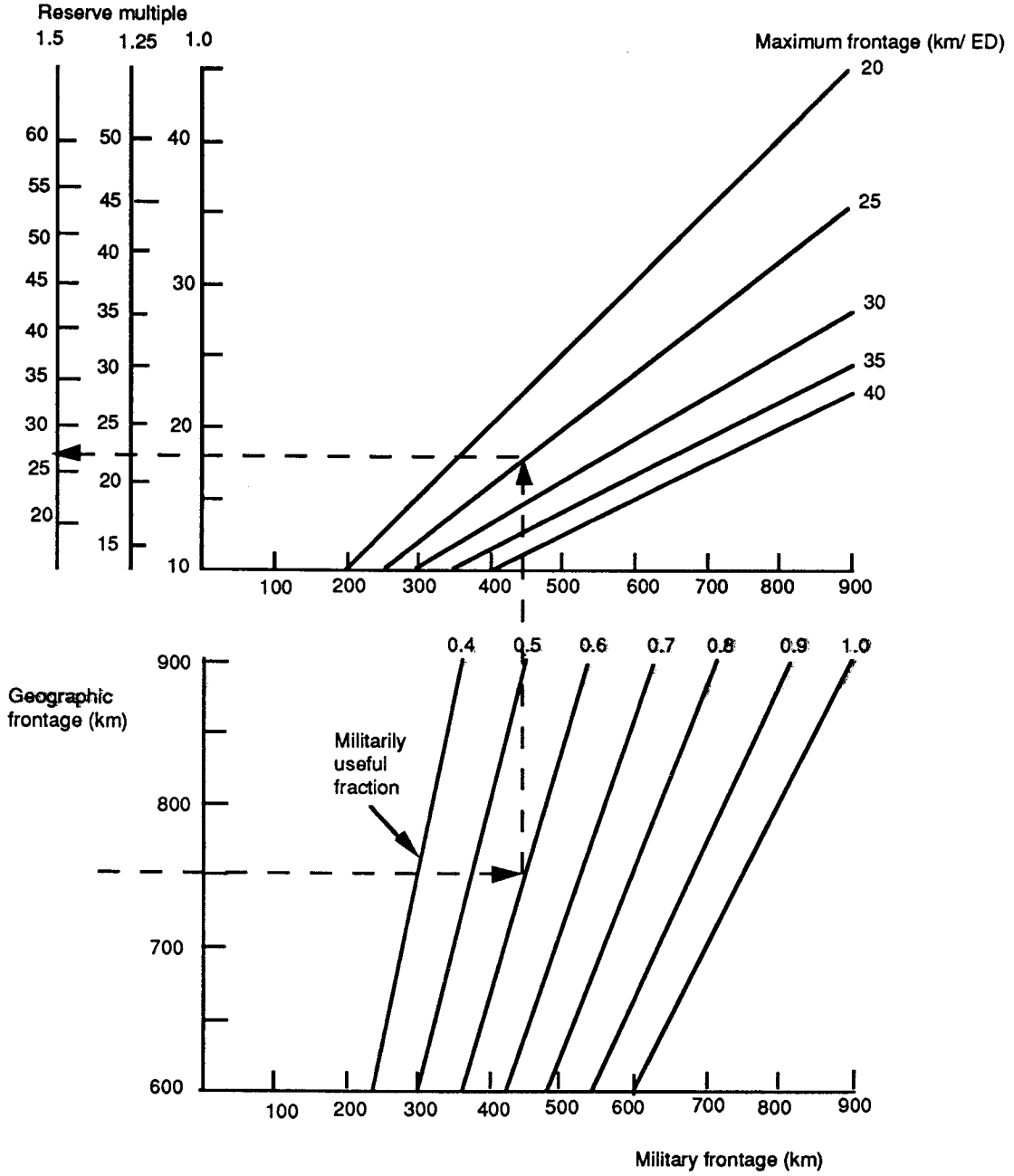


Fig. 2—Nomogram for static calculation of operational minimum

(1 + reserves/on-line forces = reserve factor)

UNDERSTANDING THE SIGNIFICANCE OF THEATER-FORCE RATIO

Before presenting some dynamic analysis of combat at low force levels it is useful to discuss the implications of theater-force ratio in determining the attacker's strategy and the defender's challenges. Primarily for the purpose of establishing certain concepts, the

author developed a simple analytic model describing the game of concentration and counterconcentration.[10] The simplest version of the model pretends that each side has only one type of division, that all divisions are identical, that airpower can be ignored, and that there are a number of identical sectors. The attacker concentrates equally on some sectors (main-thrust sectors), tolerating an adverse force ratio on the others. Both sides initially have reserves, but those are subsequently committed uniformly to the main-thrust sectors. Reinforcement of the defender's main-thrust sectors, then, are with reserves and not, in the simplest model, from adjacent sectors. The issue is to understand what kinds of sector-force ratios can be achieved as a function of the sides' choices. The variables of the model are as follows:

Inputs

A, D: number of attacker/defender divisions

(1+f): theater-force ratio $1+f = A/D$

L: frontage of theater
w: frontage of a sector
s: number of sectors $s = L/w$
t: number of main-thrust sectors
h: force ratio attacker accepts on non-main-thrust sectors
R: fraction of A initially in reserve
R': fraction of D initially in reserve

Outputs

C: ratio of forces on main-thrust sectors before reserves are committed

C': ratio of forces on main-thrust sectors after reserves are committed by both sides

It follows that

$$C = \{ s (1+f) (1-R') - h (s-t) (1-R) \} / t (1-R)$$

$$C'= \{ t (1-R) C + s R' (1+f)\} / \{t (1-R) + sR) \}$$

Or, inverting to solve for t as a function of the force ratio *desired*,

$$t = \text{integer part of } \{ s(1+f) (1-R') - hs (1-R) \}/ \{ (1-R) (C-h) \}$$

---

[10]Similar models have been derived independently over the years by Andrew Hamilton, Reiner Huber, and Wilbur Payne, although some important differences exist among them. Similar models also underlie published Soviet formulas and nomograms such as those in Hines (1989). The author's model has been extended recently by Jerome Bracken to include certain effects of aircraft and air defenses.

Figure 3 illustrates the formulas for C and C' assuming nine sectors (e.g., the number of NATO corps sectors in the Central Region, including for this purpose the Danish sector), a 2:3 force ratio h on non-main-thrust sectors, two main-thrust sectors, and reserve fractions of 1/3 and 1/6. Note how strongly C depends on theater-force ratio and how much lower C' is than C. The horizontal lines show possible rules-of-thumb requirements for an attacker and it seems clear that at parity, by contrast with today's force ratio in excess of 1.5, the attacker's prospects appear dim, especially if he is unable to prevent timely commitment of the defender's reserves.

Figure 4 shows the sensitivity of this to the force ratio on non main-thrust sectors, h. Here we see that the attacker can hope to do much better if only he can tolerate the risks associated with small values of h. Doctrinally, armies would probably prefer values of h greater than 2/3 as a minimum. However, suppose that the defender could be relied upon not to counterattack on the non-main-thrust sectors. In that case, strategy would supersede doctrine and the attacker might indeed strip as many forces as necessary from "other" sectors so as to achieve his desired force ratios on main-thrust sectors. This demonstrates in mathematical form some of the consequences of static defensive concepts.

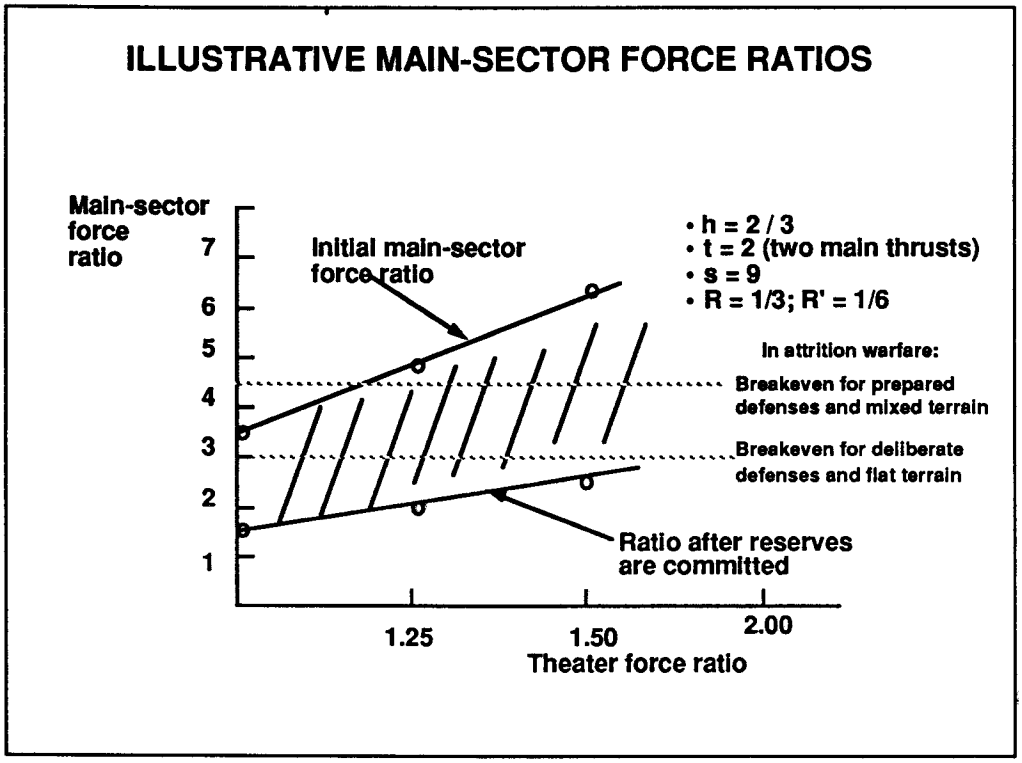

ILLUSTRATIVE MAIN-SECTOR FORCE RATIOS
Main-sector force ratio
7
6
5
4
3
2
1
Initial main-sector force ratio
• h = 2 / 3
• t = 2 (two main thrusts)
• s = 9
• R = 1/3; R' = 1/6
In attrition warfare:
Breakeven for prepared defenses and mixed terrain
Breakeven for deliberate defenses and flat terrain
Ratio after reserves are committed
1.25
1.50
2.00
Theater force ratio

Fig. 3

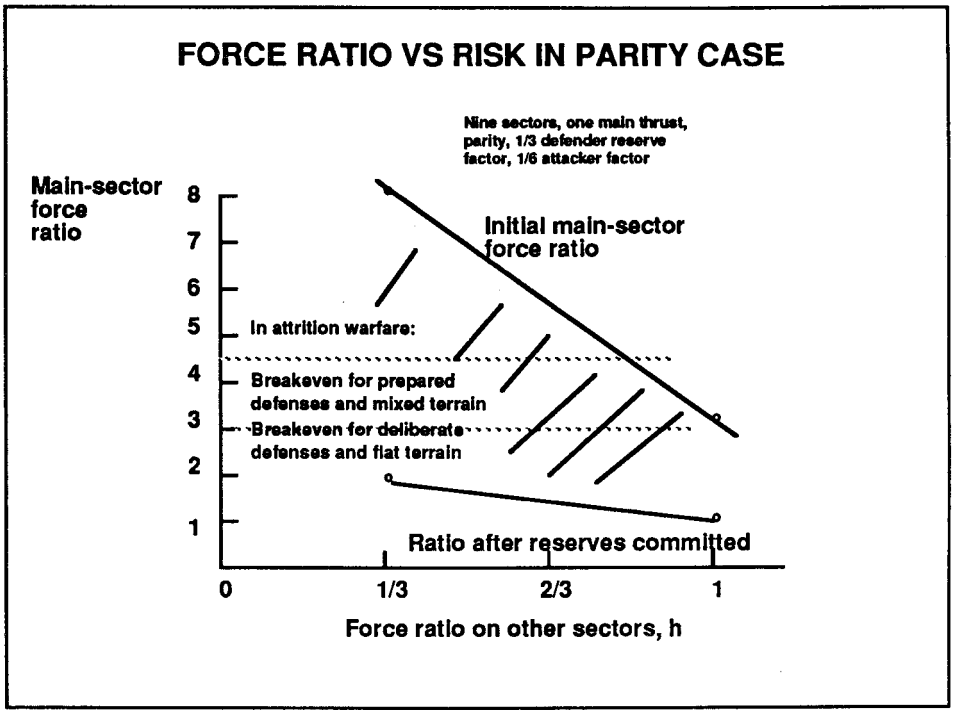


Fig. 4

Figure 5 shows the number of main thrusts possible as a function of theater-force ratio, assuming the requirement to achieve a main sector-force ratios of 4.5, with force ratios on other sectors of 1/3 or 2/3. With a theater-force ratio of 1.5 (less than today's WTO/NATO force ratio), it would be easily possible for the attacker to have three main thrusts. By contrast, at parity, the attacker could have only one corps-sized main thrust unless he reduced his main thrust-force ratio.

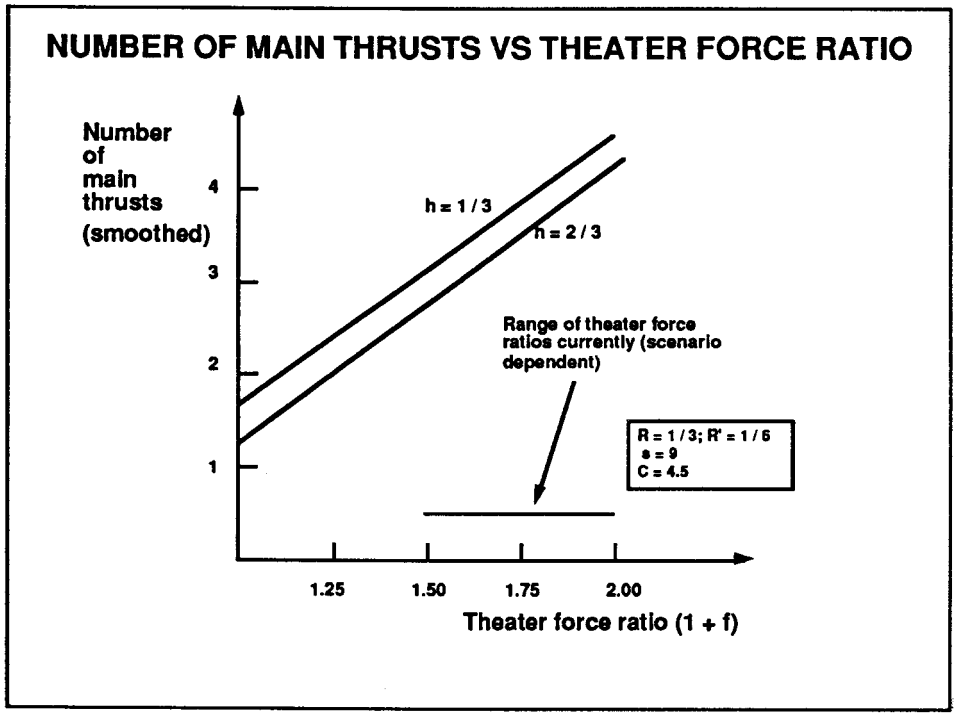


Fig. 5

Some of the conclusions one can derive from experimenting with this simple model are:

- The attacker's prospects improve quickly with increasing D-Day theater force ratio.
- The attacker's prospects at parity are poor, *unless* the defender has Achilles Heels (e.g., maldeployment, rigid command-control, and few mobile reserves—the situation describing France in 1940 when the Germans attacked through the Ardennes).
- The attacker's prospects improve if the defender—by virtue of capability, doctrine, or strategy—will not quickly counterattack, allowing the attacker to strip forces from some sectors to concentrate on others.
- The defender's prospects improve by holding more reserves, covering sectors in proportion to strategic significance, and being able and willing to counterattack on all sectors.
- The defender's prospects improve if likely main thrusts can be identified early to permit at least some anticipatory counterconcentration (the initial main-thrust force ratio will then be between C and C').

- At parity, the attacker will be unable to achieve large main-thrust sector force ratios (i.e., 3 - 4.5:1) on more than one or two corps-sized sectors.[11] If he narrows his main-thrust sectors, he can have proportionally more main thrusts, but each of those will be more vulnerable to logistical problems and flanking attacks.[12]

SIMULATION OF DEFENSE AT LOW FORCE LEVELS

With this background we are now in position to discuss and understand a series of simulation experiments conducted by the author and colleagues (see Davis, Howe, Kugler, and Wild (1989) and, for documentation and details on the simulations, Wild, Howe, and Davis (1989). At the outset of this work the entire subject of defense at low force levels and parity was so poorly understood that it seemed important to keep analysis simple, even when using simulation. The variables of most interest to us in these simulation experiments were:

- *Force levels*, assuming parity (Would the simulations show a breakdown of defense at force levels below the "operational minimum" as suggested by static analysis?)

- The defender's *operational strategy*, assuming an optimized attacker strategy (What if the defender did not insist on a strict version of forward defense in all cases?)

- The defender's *preparations and command-control efficiency* in anticipating or reacting to the attacker's concentrations (As noted earlier, we expected the results to depend on the race between the attacker's penetration efforts and the defender's commitment of reserves; the attacker's penetration rate would depend on defender preparations and command-control efficiency would be critical in determining how quickly counterconcentration occurred.)

---

[11]The author would argue that a 4.5:1 requirement is reasonable for cases in which the defender has made significant preparations and has substantial air-to-ground advantages not captured in the simple model. As noted in Liddell-Hart (1960), Allied attacks in the Normandy campaign rarely succeeded unless the attacking troops had an operational-level superiority of more than five to one, even with complete domination of the air. The Soviets conducted a number of successful Eastern Front operations with somewhat lower force ratios, but in these the Germans in defense had a dangerously low force-to-space ratio, a factor that will be considered in the next section.

[12]These problems are ignored in the nomogram rules of thumb for concentration described in the Soviet literature and summarized in Hines (1989). It should be emphasized, however, that the Soviets have long considered alternative concepts of operations based on rapid penetration of specialized independent brigades followed by corps-level exploitation forces. Should such a doctrine and associated capabilities be fielded, the traditional calculations based on large-scale concentration of forces would no longer apply. It is by no means obvious that such alternative strategies would work, especially in the absence of operational surprise. It is nonetheless predictable that such strategies will be of even greater interest in the future.

Having milked the simple analytic model of the last section for the kinds of insight one can get without recourse to geography, we were also interested in seeing how the simulation's treatment of military geography would affect results. The RSAS simulation (Bennett, Jones, Bullock, and Davis, 1988) distinguishes among NATO's corps sectors and geographic zones within those sectors; each zone is characterized by geographic and military frontages along and perpendicular to the avenues of approach, a terrain-determined defender-advantage multiplier, the presence of natural obstacles such as major rivers, and the presence and depth of any prepared defenses. The attacker may be "shoulder-space limited" (i.e., there is a parameterized minimum military frontage per attacker division). The model simulates the maneuvering of corps, army-group, and theater-level reserves. There are parameters corresponding to decision times, movement rates, and logistics- and traffic-related constraints on the rate at which reserves can be fed into a troubled sector. The model represents a number of historically rooted force-to-space considerations, including: the attacker's shoulder-space constraint (there is a parameterized minimum military frontage per attacker division) and the parameterized dependence of attrition and Forward Line of Own Troops (FLOT) movement on the defender's density (as well as FLOT force ratio, terrain, type battle, etc.). All of these parameters are uncertain, but their nominal values are consistent with the illustrative figures assumed in the earlier static estimate of the operational minimum.

A conceptually important aspect of the RSAS is its explicit prediction of explosive breakthroughs at defender densities below a parameterized threshold, the nominal value of which the author inferred from historical data and correction factors adjusting for the increase in divisional combat power over the years (e.g., roughly a factor of 2.5 in ED scores). If a defender persists in attempting to hold ground despite a low force-to-space ratio, he fares more and more poorly (at constant force ratio) as his force-to-space ratio drops until, at a threshold,[13] the attacker is considered to "break through," after which the attacker is granted very high rates of advance and favorable exchange ratios until and unless the defender is able to blunt the pursuit-phase momentum by bringing up operational reserves or tactical air forces. This non-Lanchesterian treatment, although aggregated, captures some critical features of large-scale maneuver warfare and has great

[13]With nominal parameter settings, an armored division will essentially hold ground in the RSAS at an effective military frontage of 25 km under average conditions (e.g., deliberate defenses across a main avenue of advance in mixed terrain); it will break catastrophically if attempting to hold ground with an effective military frontage of 50 km. The corresponding figures for geographic frontage are about 40 and 83 km. All such parameters can be varied interactively in sensitivity analyses, which are quite important given the uncertainties.

explanatory power in reviewing World War II campaigns. Figure 6 illustrates the results one gets from the model for an operation in which the attacker is in fact able to break through.

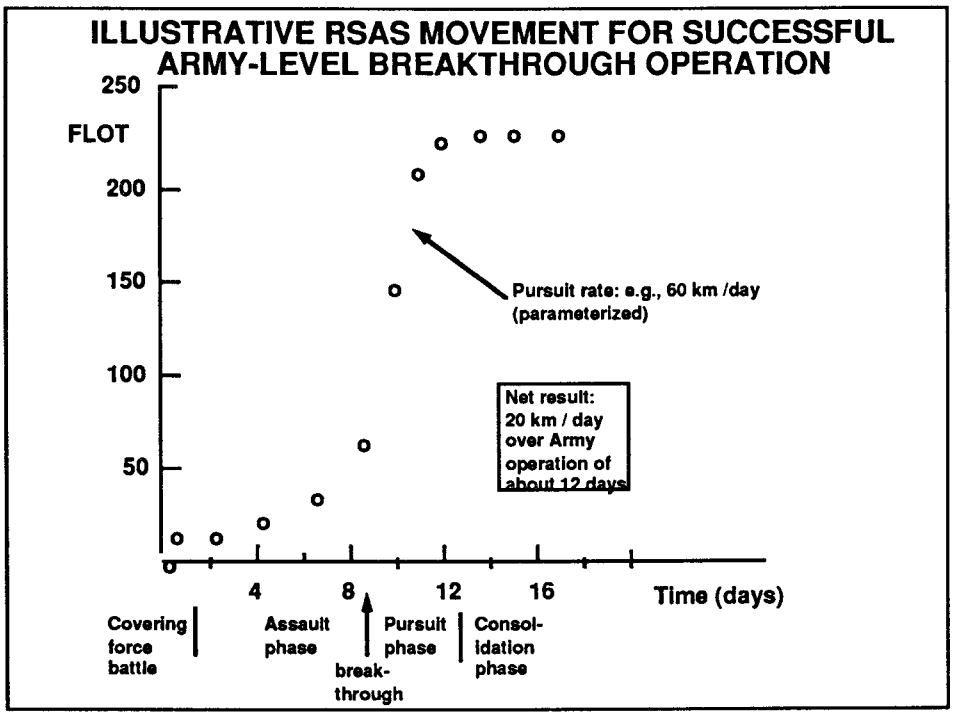


Fig. 6

With these aspects of the simulation in mind as the key points of focus, we made drastic simplifications in other respects—stripping out much of the complexity that characterizes more typical RSAS analysis. The simplifications were in many respects analogous to those made in developing and using the simple analytic model of the last section, notably: (1) generic forces rather than named units with different characteristics; (2) uniform defense of all sectors initially; (3) no air forces (neither rotary nor fixed-wing); (4) main thrusts in the Dutch and, usually, in the Ge I corps sectors because of their strategic significance; (5) relatively sluggish defender command-control as measured by the speed with which operational reserves are committed and maneuvered to reinforce main-thrust sectors; (6) a risky attacker strategy on non-main-thrust sectors (a 1:2 force ratio); and (7) all forces available in the theater on D-Day. In subsequent work, of course, we have relaxed most of these simplifications. In particular, recent work has examined the potential influence of tactical air, which may favor the attacker or defender

depending on detailed assumptions about operational surprise, air-to-ground effectiveness in disrupting maneuver as well as causing attrition, and many other factors. We have also considered a range of mobilization scenarios.

Because parity is so favorable to the defense, we chose a baseline case in which the defender had no prepared defenses on D-Day. The combination of assuming this and the absence of tactical air forces corresponded to a deliberately very conservative baseline, but one from which it was possible to conduct useful sensitivity excursions. Figure 7 shows movement rate as a function of time for the the various force levels examined. In each case, NATO followed a strict forward defense and the theater-force ratio was parity. We see that at force levels of 27, 36, and 45 EDs, the defense held at or east of the Weser River—a clear failure for the Pact. At the 18-ED level, however, Red achieved an early operational-level breakthrough, which it was able to exploit faster than Blue could commit reserves and prepare defenses at the river line. As a result, Red's momentum was only slowed, not stopped, at the Weser, and the Red offensive crossed the Rhine within a few weeks.

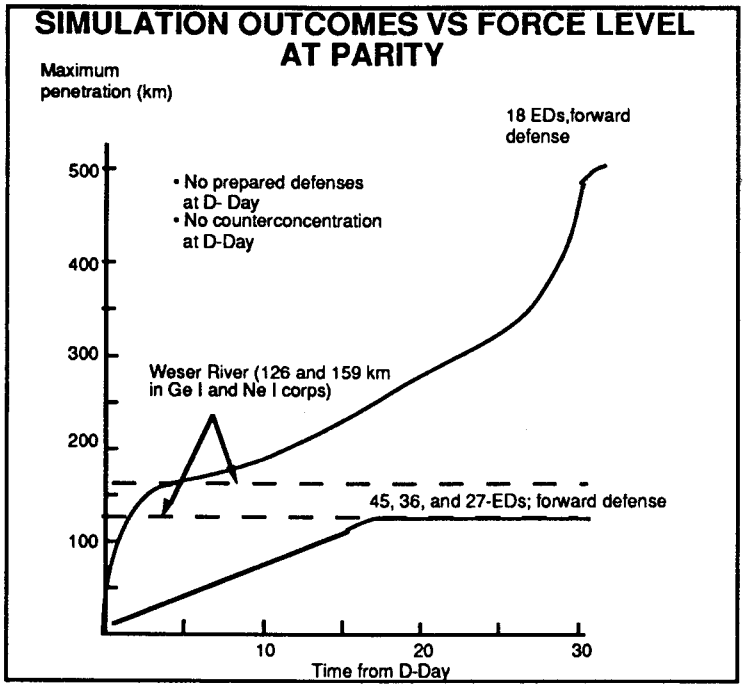


Fig. 7

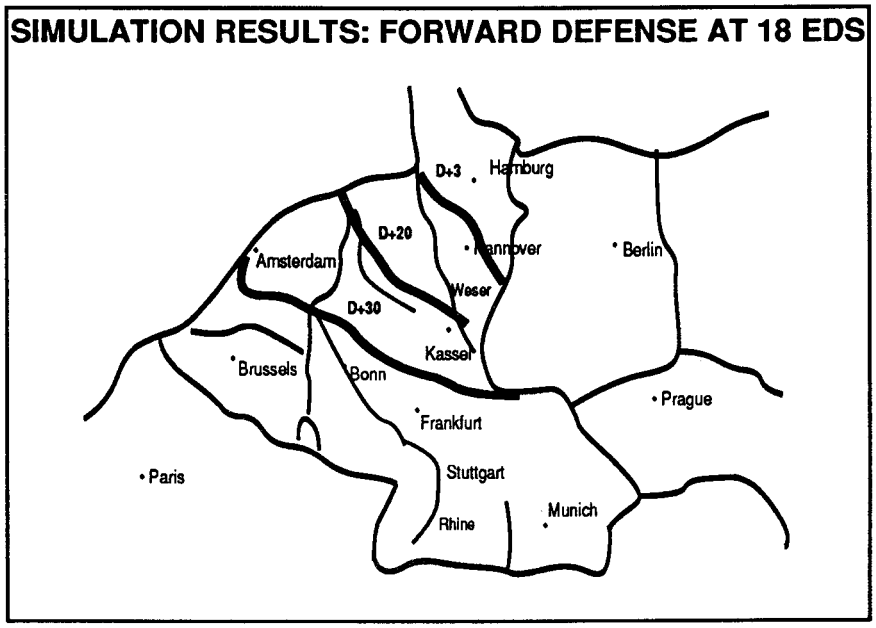


Fig. 8

The principal point of Figs. 7—8 is that the simulation results reflect the expected breakdown of defenses somewhere below 27 EDs. In a sense, this is merely "verifying" the computer program, since—as noted earlier—the sensitivity to force-to-space ratio is built in. It would seem, then, that low force levels would be destabilizing. But now let us consider the implications of some alternative assumptions. Figure 9 shows the results of varying assumptions about operational strategy, command-control efficiency, and prepared defenses. We see that if Blue conducts a delay-imposing fallback to the Weser instead of trying to maintain a forward defense with inadequate forces, Red's attack stalls out at the river line. This is because by denying Red its breakthrough, Blue is able to slow Red's movement. This in turn buys time to commit operational reserves and prepare defenses at the river line. The results, then, highlight the fact that the catastrophic collapse of defense below the alleged "operational minimum" is strongly dependent on assumptions about strategy. In fact, one can argue that it is difficult for a defender to *lose* at parity unless he insists on fighting a forward defense under bad circumstances such as when the attacker has achieved operational surprise by concentrating without the defender's counterconcentrating.

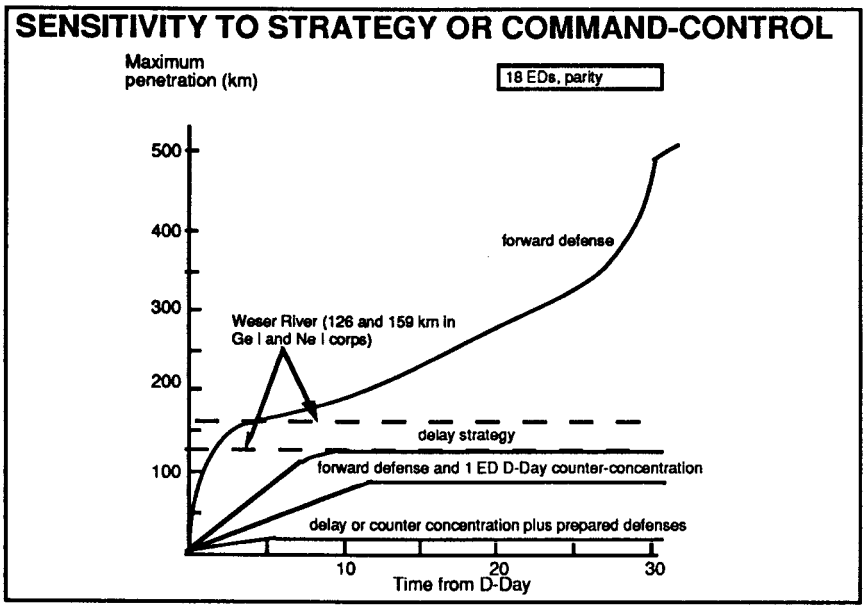


Fig. 9

Figure 9 also shows results assuming a modestly improved Blue command-control system. Here we assumed that Blue observes and diagnoses Red's concentration efforts well enough to commit *one* additional division to the main-thrust corps sector before D-Day. The effects are dramatic. This may at first appear to be an artifact of the model, but a simple back-of-the-envelope calculation reminds us that in these cases 1 ED is not a small increment, but rather a large one. It changes the defender's tactical density on D-Day from about about 1 to 2 EDs per 80 km of geographic frontage (or about 50 km of militarily effective frontage). As discussed earlier in the paper, we should *expect* a defender to break at the lesser density and to hold at the higher density.

The bottom line in Fig. 9 shows the consequences of assuming, in addition to modest counterconcentration, 20 km of prepared defenses at the IGB on D-Day. Here Blue is able to successfully conduct a forward defense by any standards. Many observers would argue that this is the most likely of the cases treated in Figs. 7—9, since one might expect that neither side would be likely to suffer the complete failure of intelligence and command-control implied by the base case.

It follows from this deliberately simple analysis that successful defense is quite possible at low force levels and parity, and that even a forward defense is possible in many cases (e.g., if the defender preferentially defends what turns out to be the main-thrust sectors, or if the defender is able to slow the attacker's initial movements long enough to counterconcentrate with ground forces). On the other hand, there are "bad cases," and one conclusion from exploring the low-force-level regime is that while best-estimate cases are probably defense favorable, the variance of results among plausible cases is larger than at higher force levels and parity (Wild, et al., 1989), just as one would expect intuitively. Because there are so many instances in history in which a defender with adequate capability and approximate parity has maldeployed his forces and lost quickly, it is important that the defender be capable, on short notice, of adopting a delay-and-withdrawal strategy. The following quote is timeless:

> [The] crucial factor in the defense of any wide front is the *time* factor. This turns not only on the relative mobility of the attacking and defending forces, but on the defender's correct appreciation of the attacker's line of advance and the degree in which the attacker's mobility is restricted by natural obstacles, fortifications, and counterthreat (Liddell-Hart, 1960, pg. 30).

If we now draw upon lessons learned from the simple analytic model as well as the simulation effort, it follows that there are certain very important attributes of forces if force-posture stability is to exist at low force levels. These attributes are:

- *Forces well suited to maneuver*, including counterattacks at the tactical and operational levels (recall from Fig. 4 the advantage accruing to the attacker if he can be confident his opponent will be static; also, in the "good cases" of Fig. 9, the defender would have to counterattack to restore the border, although that was not simulated in these runs)
- *Reliable capabilities for delay operations at the front* (the fallback strategy assumed in Fig. 9 requires especially difficult maneuver operations)
- A *large theater reserve* of maneuver forces
- A *flexible operational strategy permitting contingent forward defense* (i.e., forward defense in most cases, with alternative strategies in worst cases such as operational surprise)
- An *effective command-control system* designed for rapid diagnosis, decision, action, monitoring, and follow-up
- *Unity of command* (essential, since at low force levels there can be no bargaining about whether particular units will or will not be employed where they are needed and in the manner they are needed)

- *Political flexibility to develop and exercise contingent operational strategies* that include options for preferential defense of strategically critical sectors, giving up space for time temporarily, and counterattacking at the operational level

This says nothing about air forces, because we have not yet completed and published our analysis of how air forces affect stability, but a side wishing to hedge its ability to defend should want responsive air forces capable of slowing penetrations while ground forces can counterconcentrate. Whether air forces are on balance stabilizing or destabilizing depends significantly on their relative effectiveness over enemy or friendly territory, which in turn depends significantly on air defenses. It can be argued that mobile air defenses should be constrained quantitatively.

Unfortunately for NATO, the above list of desired attributes is quite sobering. Bluntly put, *the author would expect NATO to have grave difficulties conducting a successful defense at low force levels, even under conditions of parity, unless there were major changes in NATO's force posture and command-control system.* It would be potentially quite dangerous to implement deep cuts by merely scaling down NATO's current layer-cake defense with its independent corps sectors, maldeployment of forces, diverse and inconsistent national logistics systems, etc. The cup, then, is both half empty and half full.

WHEN PARITY DOESN'T MEAN PARITY

Even more troublesome, because it is less readily solvable, is the fact that negotiating parity in the Atlantic-to-the-Urals (ATTU) region would by no means imply parity in a Central Region war. There are two principal difficulties here:

- The attacker might be able to generate and deploy forces more quickly than the defender (by having, as of M-Day, a higher proportion of active or ready-reserve units).

- Both sides have many forces that could be employed in the Central Region even though they are nominally assigned elsewhere. The Soviets are by far the biggest problem here, because: (a) they have on the order of two hundred divisions overall, 66 of which are outside the ATTU; and (b) they are far more capable of shifting forces to the Central Region from other regions within and outside the ATTU than NATO is. The forces of NATO's Northern and Southern Regions, for example, are ill-suited for Central-Region combat.

Figure 10 illustrates the significance of these problems with a single example in which Red attacks with 27 divisions and Blue has only 18 as of D-Day. Even with a

delay defense, Blue's defenses buckle. This should not be surprising if we remember the sensitivity of sector-force ratios and the number of main-thrust sectors possible with a given force ratio to the theater-force ratio (Fig. 5). With 27 divisions against 18, Red could hope to follow classic Soviet doctrine, attack on multiple fronts, and achieve strategic-operational envelopments. Simulations confirm the seriousness of this problem.

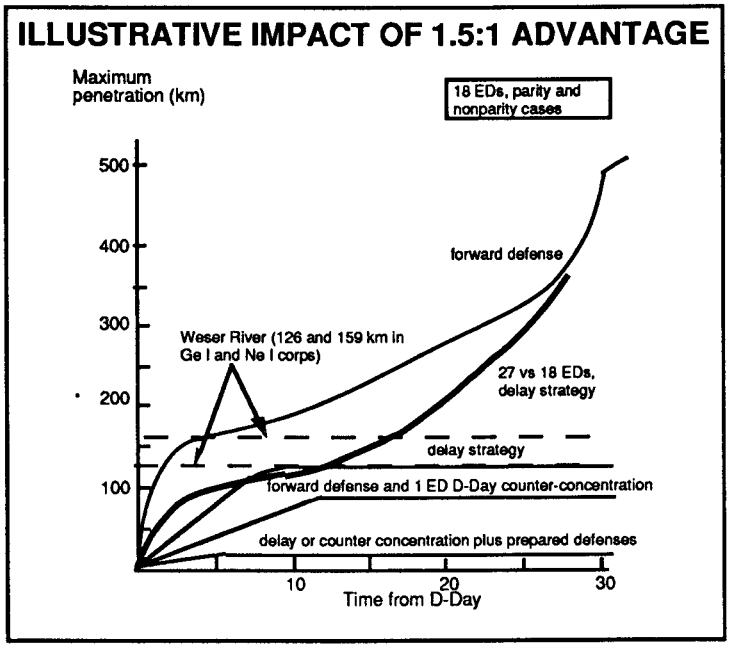


Fig. 10

The problem of "out-of-area forces" is the most fundamental of those arguing against deep cuts. The principal NATO means for dealing with the problem appear to be: (a) developing the capability for early large-scale interdiction campaigns in the event of improper redeployments, (b) negotiating limits and operational constraints on "out-of-area" ground forces, and (c) reassessing the threat in the light of the increasing independence of East European governments and armies.

SUMMING UP

The CFE-mediated process of moving toward parity will greatly improve force-posture stability in the Central Region. However, many wars have been fought at parity, and the defender has often lost. Thus, parity is no panacea. To make things worse, there

is a major difference between nominal parity as achieved in negotiations for large aggregations such as the ATTU, and *dynamic parity* (force levels versus time) in the Central Region in time of war.

At low force levels such as 50 percent of NATO's current levels (counting all forces NATO would expect to be able to employ in the first month of a Central Region war), force-posture stability might be high or low, depending on details. There is no theoretical reason to believe that a low-force-level regime should be stable or unstable, although the importance of operational-level defensive maneuver would become increasingly important at force levels below what has been called the operational minimum. That minimum should be regarded as a fuzzy threshold rather than a sharp one, but to the extent that it is meaningful analytically, its value is probably between 20 and 30 "equivalent divisions," which is much lower than the figure often quoted. In the author's judgment, the higher values are based on static analysis with obsolete planning factors, overly constraining assumptions about NATO's operational strategy and command-control rigidity, and a failure to internalize fully the implications of parity and the likelihood of a post-CFE regime in which Soviet forces will no longer be properly poised for a relatively rapid assault across the IGB. Nonetheless, there *are* judgments here, and a low-force-level regime could be highly unstable if NATO did not make fundamental changes in force posture, command-control, and operational strategy. A scaled-down version of today's layer-cake posture and related operational concepts would be highly inappropriate. Thus, force-posture stability could worsen as reductions proceed much below NATO's current levels. On the other hand, at sufficiently low levels—below what might be called the offensive minimum, an attacker would probably be unable to win a permanent strategic victory, even if he won the initial battles. This would be a strong deterrent.

Since deep cuts are unlikely in the immediate future (if for no other reason than the Soviet Union's likely inability to absorb quickly more cuts than those necessary under CFE I), there is time for NATO to make necessary adjustments in its force posture, and for the sides to negotiate ceilings and other measures that would enhance the stability of the low-force-level regime. These measures (often called operational arms control) might include the substantial or complete withdrawal of Soviet forces *and attack infrastructure* from Eastern Europe, constraints on the number of high-readiness reserve units, ceilings and constraints on out-of-area ground forces, and constraints on certain types of large-scale exercises and movements. In addition, the sides may evolve toward force postures

with a smaller ratio of attack-suitable and defense-suitable weapons and units, while retaining high mobility and flexibility. All in all, there are many reasons for optimism.

REFERENCES

Bennett, Bruce Wm., et al., *Main-Theater Warfare Modeling in the RAND Strategy Assessment System 3.0*, The RAND Corporation, N-2743-NA, September 1988.

Davis, Paul K., *Toward a Conceptual Framework for Operational Arms Control in Europe's Central Region*, The RAND Corporation, R-3704-USDP, November 1988.

Davis, Paul K., *National Security Planning in an Era of Uncertainty*, The RAND Corporation, P-7605, September 1989.

Davis, Paul K., *Studying First-strike Stability With Knowledge-Based Models of Human Decisionmaking*, The RAND Corporation, R-3689-CC, April 1989.

Davis, Paul K., et al., *Variables Affecting the Central-Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels*, The RAND Corporation, N-2976-USDP, 1989.

Donnelly, Christopher, *Red Banner: The Soviet Military System in Peace and War*, Jane's Information Group Ltd., Sentinel House, Coulsdon, Surrey (UK), 1988.

Flanagan, Stephen J. and Andrew Hamilton, "Arms Control and Stability in Europe," *Survival*, September / October 1988.

Hines, John, "Calculating War, Calculating Peace: Soviet Military Determinants of Sufficiency in Europe," to be published.

Huber, Reiner K., "On Structural Prerequisites to Strategic Stability in Europe Without Nuclear Weapons: Conclusions from the Analysis of a Model of Conventional Conflict," *Systems Research*, Vol. 5, No. 3, pp 211ff, 1988.

Huber, Reiner K., "Parity and Stability: Some Conclusions from the Geometry of Military Operations in Central Europe," 1989, a chapter in this volume.

Kent, Glenn A., Randall J. DeValk and David E. Thaler, *A Calculus of First-Strike Stability (A Criterion for Evaluating Strategic Forces)*, The RAND Corporation, N-2526-AF, June 1988.

Kokoshin, Andrei, et al., *Problems of Ensuring Stability With Radical Cuts in Armed Forces and Conventional Armaments in Europe* , Novosti Press, Moscow, 1989.

Liddell Hart, B.H. "The Ratio of Troops to Space," *Armor*, May-June, pp 24-30; available also in the April 1960 issue of *Military Review* and as a chapter in Liddell Hart's 1960 book, *Deterrent or Defense: A Fresh Look at the West's Military Position*, Praeger, New York, 1960.

Mako, William P., *U.S. Ground Forces and the Defense of Central Europe*, The Brookings Institution, Washington, D.C., 1983.

Mearsheimer, John J., "Assessing the Conventional Balance: The 3:1 Rule and Its Critics," and "Reassessing Net Assessment," *International Security,* Vol. 13, No. 4, Spring 1989.

Office of the Assistant Secretary of Defense for Program Analysis and Evaluation (OSD/PA&E), *NATO Center Region Military Balance Study: 1978-1984,* 1979.

Rohn, Lauri, *Conventional Forces in Europe: A New Approach to the Balance, Stability, and Arms Control* , The RAND Corporation R-3732-USDP/AF, 1989 (based on a Ph.D. dissertation for the RAND Graduate School of Policy Studies).

Simpkin, Richard, *Red Armour—An Examination of the Soviet Mobile Force Concept,* Brassey's, London, 1984. See also the same author's *Deep Battle,* Brassey's, 1987.

Wild, William G., Robert D. Howe, and Paul K. Davis, *Simulated Central Region Conflicts at Low Force Levels.*, The RAND Corporation, N-3006-USDP, 1989 (forthcoming).

Wilkening, Dean A., and Kenneth Watman, *Strategic Defenses and First-Strike Stability,* The RAND Corporation, R-3412-FF/RC, November 1986.