IDA PAPER P-2380

DEFENSE AT LOW FORCE LEVELS:

The Effect of Force to Space Ratios on Conventional Combat Dynamics

Stephen D. Biddle, *Project Leader*
David Gray
Stuart Kaufman
Dennis DeRiggi
D. Sean Barnett

August 1991

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

IDA Log No. HQ 90-35453

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this publication was conducted under IDA's Independent Research Program. Its publication does not imply endorsement by the Department of Defense, or any other Government agency, nor should the contents be construed as reflecting the official position of any Government agency.

This Paper has been reviewed by IDA to assure that it meets high standards of thoroughness, objectivity, and appropriate analytical methodology and that the results, conclusions and recommendations are properly supported by the material presented.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY ***(Leave blank)***

2. REPORT DATE
August 1991

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Defense at Low Force Levels: The Effect of Force to Space Ratios on Conventional Combat Dynamics

5. FUNDING NUMBERS
Independent Research

6. AUTHOR(S)
Stephen D. Biddle, David Gray, Stuart Kaufman, Dennis DeRiggi, D. Sean Barnett

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-2380

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311-1772

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 words)***

This paper develops, tests and applies a systematic theory relating force to space ratios and conventional combat outcomes, and describes a simple PC-level computer model developed to automate the calculations associated with that theory. The paper is intended in part to illuminate policy issues relating to conventional force reductions in Europe, and the development of post Cold War strategy and force structure for the NATO Alliance. More broadly, however, it is also intended to contribute to an improved understanding of the dynamics of conventional warfare at low force levels generally—and to the development of an improved body of theory for explaining the outcomes of armed conflict at the theater level.

The paper concludes that the widespread perception that there exists a minimum force to space ratio for successful defense is largely incorrect. While the force to space ratio does affect combat outcomes, and while lower force to space ratios do tend to favor attackers over defenders, this effect need not be decisive, and the relationship between force density and defense effectiveness is not independent of the size of the attacking force or the doctrine and weapons used by the two sides.

14. SUBJECT TERMS
conventional warfare, theater level, force to space ratios, force density, theory, validation, doctrine, tactics, force employment, operational level of war, arms control, Europe, armored warfare, mechanized warfare, optimization, simulation, experimentation, mathematical modeling, JANUS

15. NUMBER OF PAGES
318

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
Unclassified

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

IDA PAPER P-2380

DEFENSE AT LOW FORCE LEVELS:

The Effect of Force to Space Ratios on Conventional Combat Dynamics

Stephen D. Biddle, *Project Leader*
David Gray
Stuart Kaufman
Dennis DeRiggi
D. Sean Barnett

August 1991

INSTITUTE FOR DEFENSE ANALYSES
IDA Independent Research Program

PREFACE

This paper was produced by the Institute for Defense Analyses (IDA) under the IDA Independent Research Program. The paper develops, tests and applies a systematic theory relating force to space ratios and conventional combat outcomes, and describes a simple PC-level computer model developed to automate the calculations associated with that theory. The paper has several purposes. It is intended in part to illuminate policy issues relating to conventional force reductions in Europe, and the development of post Cold War strategy and force structure for the NATO Alliance. More broadly, however, it is also intended to contribute to an improved understanding of the dynamics of conventional warfare at low force levels generally—and to the development of an improved body of theory for explaining the outcomes of armed conflict at the theater level. There is a large and heterogeneous literature on the conduct of conventional warfare, but very little of it was prepared with the clarity required to support selection among competing hypotheses by systematic comparison with experience; a major purpose of this paper is thus to contribute to the development of a more rigorous, cumulative approach to the study of cause and effect in this crucial field of inquiry.

The fundamental implication of this theory is that the widespread perception that there exists a minimum force to space ratio for successful defense is largely incorrect. While the force to space ratio does affect combat outcomes, and while lower force to space ratios do tend to favor attackers over defenders, this effect need not be decisive, and the relationship between force density and defense effectiveness is not independent of the size of the attacking force or the doctrine and weapons used by the two sides. If the defender adapts his operational doctrine to suit the demands of a lower density battlefield, and if cuts in defensive forces are accompanied by cuts in offensive forces, then it should be possible to defend effectively even at very low ratios of force to space. Given this, there is no purely military floor on acceptable NATO force levels—as long as the Alliance negotiates appropriate limits on Soviet forces, and as long as NATO militaries make appropriate adjustments in operational doctrine.

To substantiate these conclusions, the paper is organized as a brief main report which summarizes the theory and applies it to the policy issues of NATO troop reductions and Alliance strategy. This summary is supported by a series of appendices which describe the theory and the process by which it was developed in substantially greater detail. That process began with an extensive review of existing theoretical literature to establish the current state of knowledge with respect to the effects of force density. The results of this review are described in appendices A and B. An explicit hypothesis relating force density to combat outcomes was then developed and tested. The equations constituting the resulting theory are derived, motivated, and described in appendix C. Testing was conducted by controlled experimentation using a highly detailed, disaggregate combat simulation, the Lawrence Livermore National Laboratory's JANUS model. The testing process, results, and epistemological issues relating to the use of simulations as *in vitro* experimental tools are described in appendix D. A FORTRAN code was then written to automate the calculations associated with the equations embodying the final theory. This code constitutes a simple, theater-level model of conventional combat embodying the relationships described in the theory. This VFM (for Variable Force eMployment) model is documented in appendix E. The data file used to produce the base case results is described and documented in appendix F. Sensitivity analyses are given in appendix G, and a bibliography is provided in appendix H.

This work has benefited from the contributions of many individuals. Within the IDA study team, David Gray performed the review of Western literature, and executed JANUS experiments and regression analyses. Stuart Kaufman, now of the University of Kentucky, performed the review of Soviet theoretical literature. D. Sean Barnett developed the VFM model's optimization routine, and wrote the associated section of the VFM code. Dennis DeRiggi reviewed the equations and, together with D. Sean Barnett, wrote the FORTRAN code for the VFM model. Stephen Biddle developed the theory and the strategy for testing it, designed the test procedures, conducted the analyses using the model, and directed the study as a whole. The paper was formally reviewed by Dr. Jeffrey Grotte and Mr. John Tillson of the IDA staff, Dr. Jerome Bracken of Yale University, and General Ennis Whitehead (U.S. Army, retired). The authors are also grateful for the many useful comments provided by informal reviewers, including Dr. Joshua Epstein of the Brookings Institution; Dr. Peter Feaver of Duke University and the Mershon Center at Ohio State; Col. David Glantz of the U.S. Army's Soviet Army Studies Office at Ft. Leavenworth; Professor John Mearsheimer of the University of Chicago; Dr. Ivan Oelrich of the Office of Technology Assessment; Dr. Robert Pape of the University of Michigan, and the members of the Arms Control Seminar of the

University of Michigan's Program for Peace and International Security Research; and particularly to their colleagues at IDA, especially Dr. Peter Brooks, Col. W.M. Christenson (U.S. Army, retired), Mr. Marshall Hoyler, and Dr. Victor Utgoff. Invaluable administrative and production assistance was provided by Mrs. Bernie Aylor, Ms. Cori Bradford, Ms. Eileen Doherty, and Ms. Barbara Fealy.

TABLE OF CONTENTS

PREFACE iii

A. INTRODUCTION 1

B. HISTORY OF THE ISSUE 4

C. A THEORY OF FORCE TO SPACE RATIOS 7

  1. Force to Space Ratios and the Dynamics of Theater-Level Conventional Warfare 8
  2. Intervening Variables 10
  3. Theoretical Implications 16

    a. The Effects of Force to Force Ratios 26
    b. The Effects of Barrier Defenses 27
    c. The Effect of Tactical Warning 29

D. POLICY IMPLICATIONS 31
  1. European Force Reductions 31
  2. Alliance Strategy 38

APPENDIX A: REVIEW OF WESTERN LITERATURE

A. INTRODUCTION A-1

B. MILITARY THOUGHT PRIOR TO WORLD WAR I A-2

C. MID 20th CENTURY: BASIL LIDDELL HART A-11

D. THE CONTEMPORARY DEBATE: THE CONVENTIONAL BALANCE, ARMS CONTROL, AND FORCE TO SPACE RATIOS A-17

E. CONCLUSIONS A-30

APPENDIX B: SOVIET VIEWS ON THE EFFECTS OF FORCE TO SPACE RATIOS

A. INTRODUCTION B-1

B. DENSITY: WESTERN OBSERVERS B-3

C. ALTERNATIVE WESTERN VIEWS B-8
D. ASSESSMENT OF THE DEBATE B-11
E. THE SOVIET LITERATURE B-12
F. CONCLUSIONS B-31

APPENDIX C: THEORY

A. INTRODUCTION C-1
B. KEY VARIABLES C-2
C. ANALYTIC CONTEXT: THEATER DYNAMICS C-5
D. TEMPO C-8
E. DEFENSIVE DEPTH C-17
F. DEFENSIVE RESERVES AND COUNTERATTACK C-27
G. WEAPON TECHNOLOGY C-40
H. TERRAIN C-58
I. EQUATIONS C-63

APPENDIX D: TESTS FOR VALIDITY

A. INTRODUCTION D-1
B. TEST METHODOLOGY D-1
C. EXPERIMENTAL DESIGN D-5
D. EXPERIMENTAL PROCEDURE D-8
E. RESULTS AND STATISTICAL ANALYSIS D-10

APPENDIX E: VARIABLE FORCE EMPLOYMENT (VFM) MODEL DOCUMENTATION

A. INTRODUCTION E-1
B. USER'S MANUAL E-3
C. PROGRAM LISTING E-8

APPENDIX F: BASE CASE DATA

A. INTRODUCTION F-1
B. DATA FILE LISTING F-2

APPENDIX G: SENSITIVITY ANALYSES

A. INTRODUCTION G-1
B. DISCUSSION G-1

APPENDIX H: BIBLIOGRAPHY

LIST OF ILLUSTRATIONS

Figure

I-1 The Effect of Force to Space Ratios: Theoretical Implication 17
I-1A The Effect of Force to Space Ratios: Conventional Wisdom 17
I-2 The Effect of Force Employment Constraints 20
I-3 Optimal Force Employment as a Function of Force to Space Ratio (Initial Deployment or Predeployed Depth) 22
I-4 Optimal Force Employment as a Function of Force to Space Ratio................. 22
I-5 Optimal Force Employment as a Function of Force to Space Ratio................. 23
I-6 The Effect of Force to Force Ratio 27
I-7 The Effect of Barrier Defense.. 28
I-8 The Effect of Tactical Warning 31

Table

I-1 European Force Reduction Outcomes 33

DEFENSE AT LOW FORCE LEVELS: THE EFFECT OF FORCE TO SPACE RATIOS ON CONVENTIONAL COMBAT DYNAMICS

A. INTRODUCTION

Defense at low force levels has become a central issue for conventional net assessment and force planning. In Central Europe, for example, major troop reductions are now all but inevitable given the dramatic political changes of the recent past. While some forces will remain, it is now clear that any foreseeable East-West conflict would occur at much lower force levels than those of the past four decades. Moreover, with the relaxation of East-West tensions in Central Europe, other security concerns acquire new salience. The prospect of East-*East* conflict among the emerging nations of Eastern Europe, for example, has become a significant issue. Any such conflict, however, would involve radically smaller forces—although the frontiers to be defended are potentially quite large. More broadly, an essential question for the development of any new security architecture for a multipolar Europe is the ability of small armies to defend effectively within a diverse system of potential coalitions. Nor is the issue of defense at low force levels confined to Europe. Elsewhere in the world, force levels are often much lower than has been the case for the traditional NATO-Warsaw Pact confrontation, yet the danger of armed conflict can be quite high. Pakistan, for example, defends a frontier with India twice the length of the old Inter-German border, but with only half the troops of NATO.

Little is known, however, about the effectiveness of conventional defenses at such low force levels, or about the proper design or employment of such small forces. For most of the postwar era, the attention of the defense planning community focused on warfare between large armies on the inter-German border. Until very recently, even modest reductions in those forces seemed unlikely, while the prospects for deep cuts seemed too remote to warrant significant analysis. As a result, the military consequences of low force levels have heretofore received limited attention.

Yet there is reason to believe that defense at low force levels may be a very different proposition. It has been argued, for example, that to defend a fixed frontier requires a certain minimum number of divisions—i.e., a minimum "force to space ratio." At defensive force levels above this minimum, it is argued that combat produces a slow-moving war of position, favorable to defenders on prepared terrain. If the defender falls below this minimum density, however, it has been argued that combat becomes a war of maneuver characterized by deep penetrations, encirclements and meeting engagements, fought in the depths of the defense and favoring mobile attackers over static defenders. Moreover, this minimum force density is generally argued to be independent of the size of the opposing force. To mount an effective forward defense, it would therefore be necessary to provide at least this minimum force, even if the attacking army were also small in size relative to the length of the contested frontier. Even an attacker to defender (or force to *force*) balance of parity, it is argued, could still produce defeat for the defender if the force to *space* ratio dropped below the forward defense minimum.[1]

If true, this conception of defense at low force levels has important implications. Most estimates of the minimum force to space ratio fall in the neighborhood of one division per 25 to 30 kilometers of front. For the reduced forces of the new Europe, however, this density is quite high. NATO, for example, is virtually on the threshold today; future troop cuts will thus push NATO well below this minimum. Even the Soviet Union will be hard pressed to maintain forces sufficient to defend its own borders at this troop density, while no other East European army can provide such a density today, much less

---

[1] For exemplary arguments, see James A. Thompson and Nanette C. Gantz, Conventional Arms Control Revisited: Objectives in the New Phase (Santa Monica, CA: Rand, 1987), Rand Note N-2697-AF; John J. Mearsheimer, "Numbers, Strategy, and the European Balance," International Security, Spring 1988 (Volume 12, No. 4), pp. 174-185; General John R. Galvin, "Some Thoughts on Conventional Arms Control," Survival, April, 1989, pp. 99-107; Stephen J. Flanagan and Andrew Hamilton, "Arms Control and Stability in Europe: Reductions are not Enough," Survival, September/October 1988, pp. 448-463; James W. Moore, "The Estimation of Optimum Force Size and Force Reduction Potential in Conventional Arms Reduction Negotiations," Arms Control, September 1988 (Volume 9, No. 2), pp. 116-133; Operational Minima and Force Buildup of the Warsaw Pact and NATO (Bonn, Federal Republic of Germany: Federal Ministry of Defense, 1989), unpublished manuscript; Jack Beatty, "The Exorbitant Anachronism" The Atlantic Monthly, June 1989, pp. 40-52; Leonard Sullivan, Jr., Security and Stability in Conventional Forces: Differing Perceptions of the Balance (Washington, D.C.: The Atlantic Council of the United States, 1988), pp. 8-9, 39, 60-5; Comments of General Hans Henning von Sandrart, Commander in Chief, Allied Forces Central Europe, as reported in Peter Adams, "NATO Has Little to Barter in Conventional Arms Talks, Commander Says," Defense News, November 7, 1988, p. 21; Andrew J. Goodpaster, Gorbachev and the Future of East-West Security: A Response for the Mid-Term (Washington, D.C.: The Atlantic Council of the United States, 1989), pp. 1-17, esp. p. 11; United States General Accounting Office, NATO-Warsaw Pact: Assessment of the Conventional Balance (Washington, D.C.: Government Printing Office, 1988), Main Report and Supplement, GAO/NSIAD-89-23 and 23A, pp. 13, 18, supplement pp. 42-3, 63. For a more detailed review of the public debate on this issue, see appendices A and B.

after further troop reductions. If this conception of the effects of force to space ratios is true, the consequences for military stability in the new Europe could thus be unsettling. More immediately, NATO must make near term decisions regarding specific troop cut proposals and possible revisions of Alliance strategy. Under a conception of force to space ratios such as that described above, however, it would be difficult to argue that any realistic force level could provide an actual defense of the Alliance's borders, and it would require NATO to abandon its strategy of Forward Defense.

To know whether this is so, we need a deeper understanding of the underlying dynamics of defense at low force to space ratios. Current arguments on the nature of force to space minima are useful as a point of departure, but as yet there has been no systematic description of the relationship between force density and defense effectiveness. Without such a description, however, it is difficult to know whether the minimum defensive density is above or below that achievable by any given state; whether the minimum can be altered by changes in technology or doctrine; or even whether such a "minimum" force to space ratio exists at all, independent of the force to *force* ratio between the two combatants.

The purpose of this paper is thus to develop such a description—a rigorous, carefully specified theory relating force to space ratios and conventional combat outcomes. We will then use this theoretical foundation to address some particular policy issues of significance for U.S. and Alliance decision making in the near term, specifically: how far can NATO reduce its forces and retain a credible conventional defense against some potential future Soviet attack, and would deep cuts in ground forces compel NATO to modify or abandon its declaratory strategy of Forward Defense?

In particular, we will argue that a minimum force to space ratio does not exist independent of the size of the attacking force and the doctrine and weapons used by the two sides. While the force to space ratio does affect combat outcomes, and while lower force to space ratios do tend to favor attackers over defenders, this effect need not be decisive. If the defender adapts his operational doctrine to suit the demands of a lower density battlefield, and if cuts in defensive forces are accompanied by cuts in offensive forces, then it should be possible to defend effectively even at very low ratios of force to space. Given this, there is no purely military floor on acceptable NATO force levels—as long as NATO negotiates appropriate limits on Soviet forces, and as long as NATO militaries make appropriate adjustments in operational doctrine.

To substantiate these conclusions, the balance of the main report is organized in four sections. A brief history of the force to space ratio issue is provided to establish an analytic context. An overview of the theory developed in the study is then provided, followed by a discussion of its application to the policy issues of NATO troop reductions and Alliance strategy. The main report is supported by a series of appendices which describe the theory and the process by which it was developed in substantially greater detail. That process began with an extensive review of existing theoretical literature to establish the current state of knowledge with respect to the effects of force density. The results of this review are described in appendices A and B. An explicit hypothesis relating force density to combat outcomes was then developed and tested. The equations constituting the resulting theory are derived, motivated, and described in appendix C, as are the limitations and bounds of application of that theory. Testing was conducted by controlled experimentation using a highly detailed, disaggregate combat simulation, the Lawrence Livermore National Laboratory's JANUS model. The testing process, results, and epistemological issues relating to the use of simulations as *in vitro* experimental tools are described in appendix D. A FORTRAN code was then written to automate the calculations associated with the equations embodying the final theory. This code constitutes a simple, theater-level model of conventional combat embodying the relationships described in the theory. This VFM (for Variable Force eMployment) model is documented in appendix E. The data file used to produce the base case runs described below is described and documented in appendix F. Sensitivity analyses are given in appendix G, and a bibliography is provided in appendix H.

B. HISTORY OF THE ISSUE

While the salience of force to space ratios in the public debate is a recent development, the issue itself is much older.[2] Occasional references to the effect of force density on combat results can be found as early as the 1830s.[3] The first sustained treatment, however, was by European military officers in the decades prior to the First World War. The issue arose in the context of the widespread effort to come to grips with the meaning of the new, high firepower weapons technology that had become available in the

---

[2] For a more detailed treatment of the literature on force-to-space ratios, see appendices A and B.

[3] Clausewitz, for example, observed that: "In fact, a fairly constant ratio exists between the size of a force and the area it can occupy .... it is enough to say that the relationship between the two is permanent and fundamental." Carl von Clausewitz, On War, translated and edited by Michael Howard and Peter Paret (Princeton, NJ: Princeton University Press, 1976), Book VI, Chapter 25, p. 472.

late nineteenth century. Writers such as Wilhelm Balck and Jean Colin concluded, in effect, that against machine guns and rapid-fire artillery, a direct frontal assault could no longer succeed. To advance against such weapons required that attackers find a flank, or a gap, against which an assault could be directed without trying to overpower an intact defense directly. If a defender could present a continuous front, however (i.e., one with neither flanks nor gaps), these writers concluded that any attacker would incur prohibitive losses.[4] This conclusion led to a variety of fairly elaborate calculations of the number of men per meter required to produce such a continuous front—in effect, calculations of minimum force to space requirements.

This theme largely disappeared from military writing in the immediate aftermath of the war. The issue resurfaced with Basil Liddell Hart's work in the late 1930s. In effect, Liddell Hart argued that insufficient force densities on the Polish border made the Poles vulnerable to German attack in spite of the general defense-dominance he espoused at the time, but that the more densely populated French frontier was proof against invasion.[5]

With the coming of the Second World War, the apparent failure of Liddell Hart's predictions cast a general pall over his prewar assessments.[6] Moreover, in the aftermath of Hiroshima, attention turned to the question of nuclear weapons and their implications. Thus the issue of force to space ratios again subsided from view.

Liddell Hart, however, returned to this theme in 1960, codifying his thoughts on force density in a book chapter and a corresponding article.[7] In these later writings, Liddell Hart set the basic terms of the modern debate over force density. Much like Balck and Colin, he argued that the defender's ability to create a continuous front was of

---

[4] See, for example, Jean Colin, The Transformations of War, translated by L.H.R. Pope-Hennessy (London: Hugh Rees, Ltd., 1912); and Wilhelm Balck, Tactics, translated by Walter Kruger (Ft. Leavenworth KS: U.S. Army Cavalry Association, 1915 translation of the fourth edition of 1908).

[5] See, for example, Basil H. Liddell Hart, The Defense of Britain (London: Faber and Faber, Ltd., 1939), pp. 54, 96, 107, 123. Liddell Hart subsequently sought to downplay the latter argument and stress the former. See his later treatment of these issues in The Liddell Hart Memoirs, Vol.II (New York: G.P. Putnam's Sons, 1965), pp. 138, 253.

[6] John J. Mearsheimer, Liddell Hart and the Weight of History (Ithaca and London: Cornell University Press, 1988), pp. 151-6, 178-9; see also Brian Bond, Liddell Hart: A Study of his Military Thought (New Brunswick, NJ: Rutgers University Press, 1977), pp. 112-115, 119-121.

[7] Basil H. Liddell Hart, "The Ratio of Troops to Space," Military Review, Vol.XL, April 1960, pp. 3-14; and Deterrent or Defense: A Fresh Look at the West's Military Position (New York: Praeger, 1960), pp. 97-109.

crucial importance to the success of the defense. The number of men required to create such a continuous front over a given distance he labeled the minimum ratio of "troops to space." While attacks against a continuous front would require in excess of a 3:1 force superiority to succeed, even modest attacks could succeed against defenses below the minimum troop-to-space ratio.

At this point, the issue again largely disappeared until John Mearsheimer rediscovered it in the context of the conventional balance debate of the early 1980s.[8] While significant, force to space ratios were but one of several issues addressed in that debate. With the INF Treaty in 1987 and the emergence of a serious opportunity for conventional arms control by mid-1988, however, the implications of force density for combat results assumed paramount importance. Following publication of an influential RAND study which explicitly linked force density and arms control policy,[9] the late 1980s thus brought about a dramatic expansion in the volume of literature on the effects of force to space ratios. At the same time, Alliance policy came to reflect the conclusion that a minimum force to space ratio determines a floor for NATO force reductions, and the idea became an essential underpinning of the NATO position in the CFE I negotiation.[10]

But while CFE has driven the force to space ratio issue to unusual salience in the larger public debate, the issue itself is thus much older. Concern for defense at low densities has been present in the military literature for at least the last hundred years, and has waxed and waned at regular intervals since then. These various ups and downs have not, however, produced a formal theory of force to space ratios sufficient to sustain attempted falsification, or to answer the kinds of detailed questions that emerged once the policy community discovered the issue. How strong is the force to space ratio effect? Would a fifty percent force reduction lead to the collapse of Western defenses, or to a moderate increase in an attacker's ability to take and hold ground? Can the disadvantages of a lower force to space ratio be offset by reductions in the force to force ratio, and if so, by

---

[8] See John J. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," International Security, Vol.7, No.1 (Summer 1982), pp. 3-39; also Conventional Deterrence (Ithaca and London: Cornell University Press, 1983), pp. 181-3; and the somewhat later "Numbers, Strategy and the European Balance," op. cit.

[9] James A. Thompson and Nanette Gantz, Conventional Arms Control Revisited: Objectives in the Next Phase, op. cit.

[10] See, for example, General John R. Galvin, "Some Thoughts on Conventional Arms Control," op. cit.; Peter Adams, "NATO Has Little to Barter in Conventional Arms Talks, Commander Says," op. cit.

how much? Is the minimum force to space ratio wholly independent of the weaponry or doctrine of the other side, or are there changes in the nature of the attacking army that could lower the floor on defensive force levels? Are there changes in defensive forces or doctrine that could lower the minimum? To answer these questions it is necessary to move beyond the existing literature and to develop a more systematic explanation of the relationship between density and combat outcomes.

C. A THEORY OF FORCE TO SPACE RATIOS

How, then, do force to space ratios affect combat outcomes? To answer this question we will advance and test an explicit causal theory. Of course, causation in conventional warfare is clearly very complex; it involves a host of issues other than force to space ratios per se. To develop a meaningful explanation we will therefore begin with the larger context of theater-level combat as a whole and distill from this complex process an abstraction of its underlying dynamics in terms that allow us to identify the role played by force density.

In particular, we will describe the dynamics of theater-level conventional combat in terms of a race between attacker concentration and defender counterconcentration. The effect of force density can then be explained in terms of the initial conditions it establishes for this race, and how these initial conditions influence the ultimate outcome. We will then describe some important intervening variables affecting the relationship between force density and combat outcomes as suggested by these dynamics. Given the resulting theory, we can then deduce both the relationship between force density and combat outcomes, and the degree to which that relationship is sensitive to changes in other variables.

To facilitate this process of deduction, the variable interactions which comprise our explanation of theater dynamics have been specified more precisely as a series of formal hypotheses. This more formal treatment facilitates testing and allows us to interconnect our hypotheses in an explicit mathematical model. This model, which thus embodies the causal explanation developed in the theory as a whole, enables us to derive the relationship between force density and combat outcomes by observing changes in the model's output as we vary input force density.

We will not, however, attempt to specify these formal hypotheses (or the resulting mathematical model) here. Detailed derivations of the hypotheses, the model, and the validity testing conducted to evaluate those hypotheses are provided in appendices C, E, and D, respectively. Instead, our immediate goal is to outline the general logic of cause and effect underlying the more detailed formulation in the appendices, and in so doing, to motivate the relationship between force to space ratios and combat outcomes deduced from that formulation.

1. Force to Space Ratios and the Dynamics of Theater-Level Conventional Warfare

Let us begin by assuming that the theater attacker ("red") chooses a point of attack and concentrates a large fraction of his forces opposite that chosen point, defending elsewhere with the remainder. We will further assume that, initially, the location of this point is unknown to the theater defender ("blue"). Prior to discovering this point of attack, blue distributes his forward forces across the length of the frontier. Forward forces' mobility is limited by their proximity to the enemy; once they are committed, they are difficult to disengage. Defensive reserves are more mobile. Once the defender locates the point of attack, withheld reserves thus move to that point, while engaged forward forces defend in place. Upon arrival, reserves assigned to passive reinforcement dig in astride red's axis of advance. Reserves assigned to counterattack concentrate against a chosen point on the flank of the red penetration and launch a smaller scale equivalent of the red theater offensive in an attempt to cut off the red spearhead.

Prior to the arrival of those reserves, however, red's local concentration provides a high attacker:defender force to force ratio at the point of attack. Red attempts to exploit this local advantage by overwhelming the initially outnumbered forward defenders and breaking through into blue's vulnerable rear area before sufficient reserves arrive as to make further advance impossible. If red is able to break through, continued defense in a theater as shallow as Central Europe would be extremely difficult. As a point of departure, we will assume that successful breakthrough is tantamount to the catastrophic failure of the defense. If red fails to break through, his net territorial gain amounts to the ground taken prior to being halted by the arrival of blue's reserves, less any territory retaken by reserves assigned to counterattack. In either case, however, red's strategic objective is assumed to be to take and hold as much blue territory as possible—ideally by breakthrough and annihilation of the opposing army or, alternatively, by continuously opposed advance.

Given this race between red concentration-penetration and blue counterconcentration, what role does force density play? In effect, force density establishes the starting points for the race. Given our assumptions as to blue's initial state of knowledge, the largest initial ground defense that blue could mount at the point of attack is simply that fraction of the entire blue theater force that would occupy the red attack frontage if all blue forces were allocated forward (since, prior to discovering the location of red's main effort, blue must defend the entire frontier). Blue could deploy a smaller initial defense by withholding some of that theater force as a mobile reserve, but the upper limit on the size of blue's initial defense is determined by force size—that is, by the force to space ratio.

For red, on the other hand, the largest initial ground assault that can be mounted is determined not so much by force size as by the terrain at the point of attack. Red can only concentrate a finite ground force on a finite front, regardless of the size of the force available to red in the theater as a whole. Attackers in excess of the terrain's carrying capacity can eventually be directed against the point of attack, but they must initially occupy follow-on echelons to the rear of the assault wave until space is opened for their commitment. While these follow-on forces are of substantial value to red, they cannot play a direct role in the initial assault.

Thus, as force levels increase—on both sides—the potential size of blue's initial defense increases, but the size of a given red assault wave is terrain-limited and (for a given frontage) cannot increase with overall force size. Additional red forces beyond this limit are of value as follow-on units, but they cannot participate directly in the initial assault. Likewise, if force levels decrease, blue's maximum initial forward defense decreases in size, but red's initial assault wave again remains the same. Fewer follow-on forces will be available behind this initial wave, but the number of forces simultaneously engaging blue at the point of attack can be maintained at the carrying capacity limit until red has too few troops in the theater to reach the limit on the given front (while maintaining adequate security forces away from the point of attack).

This implies, however, that, *ceteris paribus*, the lower the force to *space* ratio, the higher the initial red:blue force to *force* ratio at the point of attack. If the size of red's initial assault wave is a terrain-determined constant, while the potential size of blue's initial forward defense is proportional to blue's theater force level, then a smaller theater force for both sides means a smaller defense against a constant attack and thus a higher initial force to force ratio at the key point. Fewer follow-on forces will be available to back up the initial assault, but the initial assault itself will take place at more favorable odds for the attacker, and thus the process of concentration, penetration and counterconcentration will begin with a bigger head start for red than if force levels were higher. Other things being equal, lower force to space ratios thus tend to favor the attacker.

Conversely, higher force to space ratios imply a larger potential initial defense for blue against a constant red initial attack, and thus a lower initial force to force ratio at the point of attack. This in turn implies less early success for red, and a smaller head start in penetrating the blue defense prior to blue counterconcentration. Other things being equal, higher force to space ratios thus tend to favor the defender.[11]

These effects pertain to both theater-level attackers and to smaller scale *counter*attackers. That is, red must not only take ground from blue by penetration but must also hold that ground against counterattack. But as low force to space ratios make it more difficult for blue to prevent red's theater attack from penetrating, so low force to space ratios also make it more difficult for red to prevent blue's counterattack from penetrating. Thus force levels affect both the initial theater attack and the eventual blue counterattack—and in both cases lower force levels tend to make penetration easier, again, other things being equal.

2. Intervening Variables

Other things, however, are not necessarily equal. A variety of other important variables affect the dynamics described above in ways that impinge on the role of force density, including:

- Theaterwide Force to Force Ratios, that is, the ratio of red to blue forces in the theater of war;

---

[11] Some writers, especially Liddell Hart and pre-World War I European theorists, also stress the importance of a "continuous front" (see appendix A). They argue that at low force densities, it becomes impossible for defenders to maintain an unbroken wall of fire across a long frontier. Gaps thus appear through which attackers can maneuver against the defender's flanks and rear. Taken literally, however, it is not clear that this is a sound description of the late twentieth century battlefield. Few military writers today anticipate a strictly continuous, linear front at *any* force density. Rather, defenses in depth and mechanized attackers are generally expected to produce a granular zone of contact (see, for example, Headquarters, Department of the Army, FM 100-5, Operations (Washington, D.C.: USGPO, May 1986 edition), pp. 2-3) in which a "continuous" or a "discontinuous" front is a less useful distinction than simply the ratio of the forces engaged across the attacker's assault frontage at the point of attack. Thus we emphasize the latter here. Even at very low densities, where it is clear that some approach routes will not be covered by significant fire, it is assumed that blue can at least outpost such routes so as to provide intelligence on enemy movements for the purpose of directing blue counterattack or reinforcement (which constitutes the bulk of blue's combat activity at such densities; see figure I-4 and accompanying text below). On the necessity of outposting at low force densities, see Major William Bentson, The Problem of Width: Division Tactics in the Defense of an Extended Front (Fort Leavenworth, Kansas: U.S. Army School of Advanced Military Studies, 1987).

- Weapon Mix, that is, the balance of armor, infantry, artillery, air, and advanced conventional munition or "ACM" support available to each side;
- Terrain (especially "man-made" terrain in the form of barrier defenses); and
- Force Employment, that is, the operational concepts or military doctrine by which the available forces are used in battle.

Of these variables, force employment is perhaps the most challenging to describe analytically. Any operational doctrine represents a broad and often subtle collection of guidelines for employing a force—and even if the totality of these official guidelines could be pinned down, individual commanders in the field ultimately determine how those official guidelines become actual practice.[12]

Yet force employment is clearly a central issue for the dynamics of low density warfare: when the existing literature argues that a low density battlefield will produce a "war of maneuver," it is effectively suggesting that the employment of forces will change for one or both sides — and that the net result of this change will undermine the prospects for effective defense. To capture the essential effects of force to space ratios thus requires that the interaction of force density and force employment be captured analytically.

One way to do this would be to simulate (by direct imitation) the detailed movements of the two sides' forces over time as an operation unfolds. Sand-table games, field maneuvers, and map exercises, for example, all trace the stops, starts, turns and dispositions of individual units over three-dimensional terrain in ways that enable the form of any given maneuver to be recognized directly. Such techniques enable analysts to trace out the consequences of any particular sequence of movements and counter-movements, and offer the flexibility to represent a wide range of different sequences. But while such techniques can show how any given sequence might play out, it is effectively impossible to examine all potential combinations of all potential turns, starts, and stops;

---

[12] By "operational doctrine" we follow the definition given by the U.S. Military Academy at West Point: "the [officially accepted] body of ideas .... concerning the use of available military resources to attain strategic ends in a theater of war. As the link between tactics and strategy, it governs the manner in which operations are designed to meet strategic ends and the way in which campaigns are conducted." John I. Alger, Definitions and Doctrine of the Military Art (Wayne, NJ: Avery Publishing, for the Department of History, United States Military Academy, West Point, New York, 1985, pp. 7, 5). Operational doctrine is thus neither Alliance strategy (e.g., Forward Defense and Flexible Response in NATO), nor small unit tactics (e.g. assault formation or the siting of weapons for maximum engagement range). Where possible, however, we will use the more specific term "force employment" as is defined in greater detail below.

alternatively, attempts to develop "rules" by which to identify the one right or best sequence of individual movements have been unsuccessful to date. As a result, although there are many uses for such techniques, they are thus ill-suited for making general observations about the effects of density on combat outcomes.

An alternative approach would be to step back from the detailed movements of individual units and develop instead a more abstract description of the relationship between broader classes of alternative operational concepts and their effects on combat outcomes. Rather than literally walking individual units around a hypothetical battlefield, we would instead isolate a discrete set of key dimensions along which to distinguish the practical force employment alternatives and then describe how differences with respect to those key dimensions of force employment affect combat outcomes—and how their effects change as force to space ratios change.

But how are we to identify such a set of key dimensions, or key aspects, of a force employment concept? Fortunately, the military literature on force density eases our task somewhat by emphasizing the importance of a small number of essential operational issues for the effectiveness of defenses at low force density. While this is no guarantee that these aspects of doctrine are sufficient for our purposes, the experience embodied in the literature at least provides us with a sound point of departure.

These key aspects of force employment are five: depth, reserves, counterattack, "tempo," and concentration. Respectively, these are defined for our purposes as the distance from the initial line of contact to the defender's rear defense line (in kilometers);[13] the fraction of the defender's total forces withheld from contact for use as mobile reserves; the fraction of those reserves used for counterattack (as opposed to

---

[13] Defensive "depth" is further separated into *predeployed* depth (the number of prepared defensive positions initially manned by forward forces) and *rolling* depth (a function of the fraction of the defender's forces in any given forward position that are withdrawn for use in secondary positions behind the predeployed lines), as is discussed in greater detail below. These combine according to a functional relationship described in appendix C to determine the ultimate depth of the defense as a whole. The theory as developed in appendix C is also written to support variation in additional dimensions of force employment—e.g., the attacker's casualty threshold for breaking off a local assault, or the defender's attempted velocity of counterattack, although these descriptors are as yet only partially implemented in the VFM model. For a more rigorous definition of these variables, see appendix C.

passive reinforcement); the attacker's assault velocity (in kilometers per hour); and the frontage of attack (in kilometers).[14]

In these terms, the U.S. Army's 1970s doctrine of "Active Defense" can be characterized as one with a small fraction of total forces held in reserve, a small fraction of reserves used for counterattack, and a limited deployment depth for committed forces. The Army's current AirLand Battle doctrine, by contrast, is one with a higher fraction of total force in reserve, a higher fraction of those reserves used for counterattack, and a deeper forward deployment. A "blitzkrieg" offensive doctrine is distinguished by high velocity on a narrow front; a more cautious offensive would be conducted at lower attempted velocity on a broader front.[15]

Force employment affects many aspects of the theater dynamics described above. The defender's allocation of forces between forward deployment and reserve, for example, determines how close to the maximum will be the initial defense mounted by blue at the point of attack. It also determines the rate at which blue reserves will build up at the point of attack once that point is located and, thus, the rate at which blue counterconcentrates. (The more force blue deploys forward, the smaller the reserve and thus the slower the rate at which those reserves arrive at a randomly located point of attack;[16] the more force blue holds in reserve, the faster the build up, but the smaller the initial

---

[14] Note that neither "deep penetration," "envelopment," nor "meeting engagement" appear in this list. In effect, these forms of combat activity represent ends rather than means with respect to the issues that concern us here. To the extent that, for example, a "deep penetration" occurs, it is because the defense has materially failed in its primary task of preventing breakthrough, which is a necessary precondition for a deep penetration by the attacker. While different armies will choose different forms of maneuver for the exploitation and pursuit phases of an operation, and while these differences are not insignificant, for our purposes the effectiveness of the defense has already been seriously compromised if any of these maneuvers are available as feasible alternatives for the attacker. In effect, our focus here will be on distinguishing the preconditions under which these maneuvers can occur - and on assessing the ability of defenders to prevent them from obtaining.

[15] On the distinction between Active Defense and AirLand Battle, see for example John L. Romjue, From Active Defense to AirLand Battle: The Development of Army Doctrine, 1973-1982 (Ft. Monroe, VA: Historical Office, U.S. Army Training and Doctrine Command, 1984), esp. pp. 3-22, 51-74; and Huba Wass de Czege and L.D. Holder, "The New FM 100-5," Military Review, July 1982, pp. 53-70. The term "blitzkrieg," while commonly used, is rarely defined in such a manner as to make possible systematic comparisons with plausible alternatives. For a definition that goes beyond "winning quickly," see Mearsheimer, Conventional Deterrence, op. cit., pp. 33-43; for historical examples, see e.g., Charles Messenger, The Art of Blitzkrieg (London: Ian Allen, Ltd., 1976).

[16] Assuming that blue reserves are initially deployed in assembly areas distributed across the theater without direct knowledge of the location of red's point of attack. Thus, their arrival rate is simply the product of the density of those reserves and their average speed, which is proportional to the total number of reserves in a theater of fixed length.

defense.) The depth of the defended zone likewise influences the strength of the blue opposition to the initial red assault (since for a given forward allocation, depth can be increased only by spreading out a fixed number of defenders, putting fewer in reach of any given assault wave). Depth also affects the time available for moving reserves prior to red breakthrough and reduces the effectiveness of attackers by gradually wearing down the coherence and coordination of a given red assault wave over distance. The fraction of defensive reserves used for counterattack determines the strength of the red flank defense required to prevent blue from cutting off the red spearhead; thus it also determines the "overhead cost" of holding the ground gained by penetration.

For the attacker, the frontage of the attack determines red's overall concentration, in terms of the number of attackers ultimately to be committed against each kilometer of defended frontier (given a fixed allocation of force to the point of attack). The attack frontage also affects red's vulnerability to blue counterattack: the narrower the front, the shorter the distance blue's counterattack need advance to sever red's line of communication (and thus the stronger the flank defense that red must deploy if a given blue counterattack force is to be stopped quickly enough). Finally, the velocity of the attack affects the casualties the attacker will suffer to penetrate the defense, and the rate of that penetration. High velocity permits higher potential rates of advance but requires greater exposure and allows less preparation—and thus can be obtained only at the price of higher casualties in a given assault. Lower velocity limits the rate of advance but permits more extensive preparation and more covered approaches—and thus produces fewer casualties for a given assault.

As for the other intervening variables given above, the weapon mix, for example, affects the casualty price the attacker must pay to penetrate the defense at a given force to force ratio and assault velocity, and thus influences the ability of the attacker to exploit an initial imbalance at the point of attack.[17]

---

[17] For a detailed discussion of the effects of variation in the weapon mixes (that is, the *combined arms balances*) of the two combatants, see appendix C; in brief, however, the more infantry-heavy the defense, the higher the price the attacker must pay for a high-speed advance at a given local force to force ratio (and for a given offensive weapon mix), but the lower the price he pays at low velocity. Defensive artillery, by contrast, tends to have less effect on the attacker the higher his velocity, while defensive armor effectiveness tends to be relatively insensitive to the pace of the attack. For attackers, offensive artillery can reduce losses for a given assault, but this effect is much stronger at low velocity (where there is time for an extensive preparatory barrage) than at high velocity (where there is not). Offensive infantry is extremely vulnerable at high velocity, but can be an important asset at lower velocity (where it can be dismounted and supported with a more extensive artillery and intelligence

Terrain, particularly in the form of barriers, likewise affects the relationship between the cost of penetration and attacker velocity. An advance through a barrier system at high velocity incurs heavy casualties; at low velocity (where the attacker has more time to clear obstacles and locate dug-in defenses), barriers have smaller effects. Defensive barriers thus tend to compensate for low force levels, in that they encourage slower velocity choices by red, giving blue more time to shift reserves to the point of attack.

As noted above, terrain also limits the number of attackers that can usefully be massed on a given front. If too large an assault force is crammed into too small a space, its vulnerability to opposing artillery fire increases and it loses its ability to take evasive action under fire, to choose the least exposed path between its jump-off point and its objective, to maintain efficient formations that maximize its own firepower, or to change direction quickly to meet unexpected threats. As a result, all terrain has a "carrying capacity." Adding forces beyond the carrying capacity of the terrain produces less and less additional combat power for each unit added—and may even reduce total combat power in extreme cases.[18]

Finally, the theaterwide force to force ratio determines the number of initially unengaged follow-on forces available to the attacker at the point of attack. Although these follow-on forces cannot directly influence the local force to force ratio at the point of attack (which is determined by the terrain and the density of the defender's forward deployment), they can be of substantial value to red as replacements for spent assault echelons and as flank defenders to forestall blue counterattack. Thus, the higher the

---

preparation). As with defenders, offensive armor effectiveness is relatively insensitive to attack velocity.

[18] Difficult or impassable terrain can also reduce the militarily relevant frontage to be defended in the theater of war, thus effectively increasing the force to space ratio for a given theater and defending force by reducing "space" while holding "force" constant. See, e.g., Paul K. Davis, et. al., Variables Affecting the Central Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels, (Santa Monica, Ca.: RAND, September, 1989), pp. 15-35. Caution must be exercised, however, in assessing particular terrain as "impassable." In World War II, for example, the French, and then the Americans each underestimated the suitability of the Ardennes forest for offensive operations and were caught correspondingly ill-prepared for the German offensives of May 1940 and December 1944, respectively. See William L. Shirer, The Collapse of the Third Republic: An Inquiry into the Fall of France in 1940 (New York: Simon and Schuster, 1969), esp. pp. 609-610; Hugh M. Cole, The Ardennes: Battle of the Bulge (Washington, D.C.: Department of the Army, Office of the Chief of Military History, 1965), esp. pp. 39-40, 55-56. For additional examples, see Bentson, op. cit., pp. 9-28.

theaterwide red:blue force to force ratio, the longer red will be able to press its attack before being halted by the cumulative effects of attrition and the "overhead cost" of manning the flanks of a lengthening penetration corridor.

3. Theoretical Implications

The explanation of cause and effect outlined above implies a particular relationship between the force to space ratio and the outcome of combat at the theater level. This relationship is depicted graphically in Figure I-1.

Figure I-1 represents the output of the mathematical model (described in appendices C and E) which formalizes the description of variable interactions discussed above. It plots the combat outcome predicted by the theory as a function of the defender's force to space ratio, for a constant weapon mix and terrain, and a constant theaterwide force to force ratio of 1:1 (i.e., parity between red and blue at the theater level). Combat outcomes are assessed in terms of the theater attacker's *net territorial gain*—that is, red's maximum penetration distance (in kilometers), less any ground retaken by the defender as a result of counterattack.[19] The force to space ratio is expressed in terms of the total quantity of blue forces available in the theater for the defense of a constant theaterwide frontage. Blue forces are denominated in units of "armored fighting vehicle equivalents" (AFVEs); a frontage of 850 kilometers is assumed.[20]

---

[19] Red breakthrough is represented by an arbitrarily large net territorial gain corresponding to the catastrophic failure of the defense.

[20] An AFVE is simply a convenient, "generic" index of force size selected to facilitate summary presentation of combat results (and to facilitate comparison of highly disaggregate JANUS output and equations for the estimation of theater-level combat outcomes). A single main battle tank represents one AFVE (regardless of nationality, make or model). A single armored troop carrier with its infantry complement is also scored here as one AFVE. A carrier without its infantry is half an AFVE; the infantry without the carrier is half an AFVE. Armored antitank, air defense, command, or reconnaissance vehicles are also one-half an AFVE. Field artillery and aircraft are accounted separately in units of tubes and sorties, respectively (see appendix C), and thus are excluded from the AFVE totals per se. In these terms, NATO's post-CFE Central Region force to space ratio comes to about 37,000 AFVEs on an 850 kilometer front. For force levels, see David G. Gray, IDA Unclassified Conventional Forces Data Base (Alexandria, VA: Institute for Defense Analyses, 1989), IDA D-708. Note that AFVE scores as used here are merely a presentational convenience for describing the "size" of a heterogeneous theater force; the equations in appendix C and the associated VFM model are heterogeneous in the sense that attrition is sensitive to the specific numbers of each weapon type present, rather than merely the aggregate AFVE score as such (attrition in VFM is only partially heterogeneous, however, in that the proportional representation of each weapon type does not change over time within any given run).

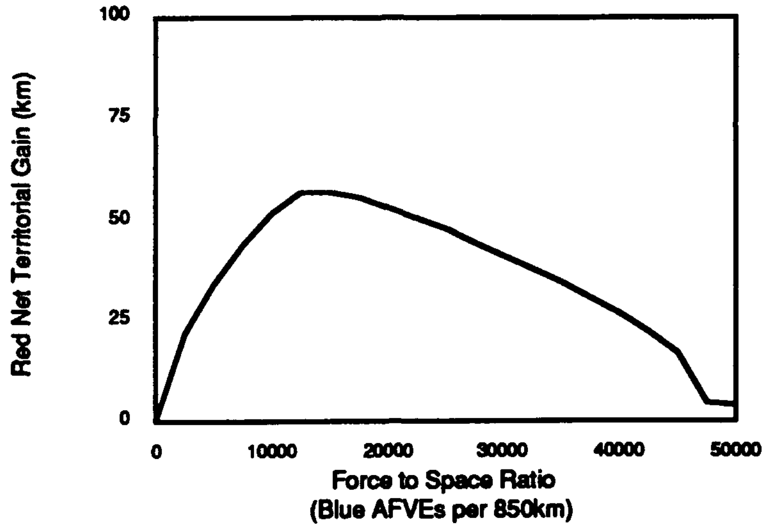

Figure I-1. The Effect of Force to Space Ratios: Theoretical Implication

To highlight the essential features of the theoretically deduced relationship, an alternative conception has been depicted in Figure I-1A. This purely notional curve corresponds roughly to the implicit understanding of this relationship that underlies much of the public debate on the question of force to space ratios. This notional "threshold" conception suggests that the defender fares reasonably well at high force to space ratios,

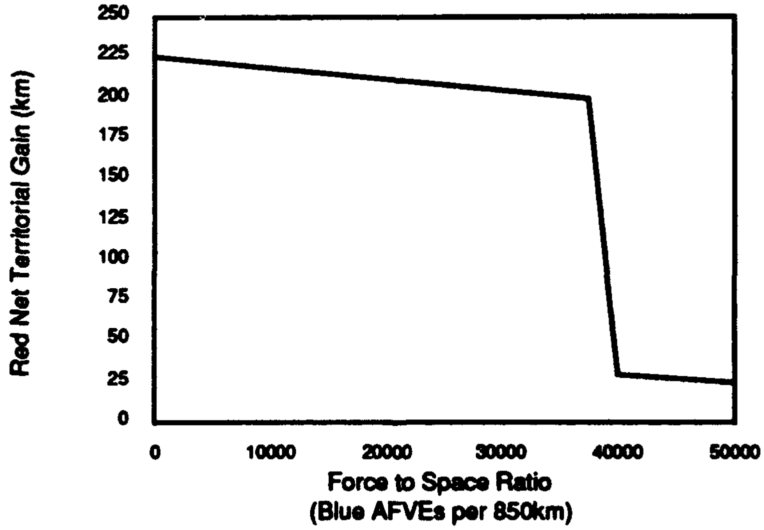

Figure I-1A. The Effect of Force to Space Ratios: Conventional Wisdom

but very badly at low density. These conditions are separated by a sharp discontinuity at a "minimum" force to space ratio where the nature of the fighting is transformed and defensive effectiveness is fundamentally undermined. Moreover, this minimum occurs at a relatively high force level corresponding to a density of roughly one forward division per 25 kilometers of front.[21]

By contrast, the theoretically deduced relationship depicted in Figure I-1 provides a very different understanding. Note, however, that there is still a force to space ratio effect in Figure I-1: net territorial gain still changes as the force to space ratio changes. Moreover, for a substantial range of force to space ratios, the direction of this effect is essentially that which the public debate suggests: for most blue force levels, decreasing the force to space ratio increases red's net territorial gain. The differences, however, are substantial. The model output constitutes a smooth, shallow, continuous curve. There is no threshold point at which the nature of combat fundamentally changes character, and there is no identifiable "minimum force to space ratio" to constitute a floor for effective conventional defense. Of course, it would still be possible to choose an arbitrary cut-off value for net territorial gains above which blue could be considered "defeated," and to label the associated force level a minimum force to space ratio. Under a continuous conception of this relationship, however, it will always be the case that for any given cut-off value, force to space ratios slightly below the associated "minimum" will produce combat outcomes only slightly worse than for force levels slightly above the minimum.[22]

Moreover, if force levels go low enough, the continuous conception suggests that outcomes will eventually improve for the defender until, at a theater force level of zero, net territorial gain likewise falls to zero. For a constant theater force to force ratio, a blue force level of zero implies a red force level of zero; thus at this extreme, red can take no territory because it has no forces with which to do so. More generally, if red's strategic objective is to take and hold blue territory, then as both sides' force levels fall, the number

---

[21] Where one division is assumed to total roughly 800 AFVEs, and where one division is assumed to be held in reserve for every two committed to the front line.

[22] Of course, inasmuch as this conclusion is based on the dynamics of concentration, counterconcentration, penetration and counterattack, we consequently cannot exclude the possibility that there may exist an administratively determined minimum based on the requirements for efficient logistical support; nor can we exclude the possibility of a minimum based on a requirement for border security or control of infiltration. What we *can* exclude, however, is the existence of a minimum based on the central issue raised in the literature on ground force density and conventional force planning: i.e., the military requirement to defend against a concentrated, high intensity ground offensive (for a more detailed discussion of the bounds of application and limitations of the theory advanced here, see appendix C).

of potential red defenders of seized territory also falls.[23] Moreover, just as the task of theater defense gets harder for blue as the blue force to space ratio decreases, so the task of red's potential counterattack defenders gets harder as their own force to space ratio decreases. Blue thus has an incentive to devote an increasing fraction of its forces to counterattack as the theater force to space ratio decreases, thereby forcing red to allocate more and more of its shrinking forces to defensive overhead, and eventually limiting red's overall advance by virtue of insufficient forces for defense of seized ground.

What is responsible for this difference between alternative conceptions of the same relationship? The answer lies with the role of force employment, and with the difference between adaptation and stasis as force to space ratios fall. Since the theory describes outcomes as a function of alternative employment choices, it enables us to observe how predicted territorial gain varies with systematic variations in the two sides' force employment. If we search over the whole range of defined employment choices, we can identify the particular combination that provides the best outcomes for the given circumstances—i.e., an optimal force employment profile. The mathematical model that implements the theory thus enables us to determine optimal force employment choices internally to the model, and consequently to re-optimize the two sides' choices to fit *changing circumstances as other* variables change. The model can of course be constrained to consider only a limited range of employment choices for either side, or even to predetermine those choices if desired. But an important property of the model is that it enables us to provide for optimal adaptation of force employment as force levels change, and to reflect the effect of such adaptation on the relationship between force to space ratios and combat results.

Figure I-2 illustrates the significance of force employment adaptation by plotting model output for two cases. The first, or base case, is identical to the curve in Figure I-1. It assumes that blue adapts its force employment optimally to suit the demands of

---

[23] It is also conceivable, especially at very low force to space ratios, that a potential invader could instead adopt what Archer Jones has termed a "raiding strategy," in which invading armies seek not to defeat the opposing army as a means of asserting control over the opponent's territory, but rather to avoid contact with the opposing ground forces while penetrating deeply enough to destroy the opponent's economic and political infrastructure for coercive purposes. See Archer Jones, The Art of War in the Western World (Chicago: University of Illinois Press, 1987), pp. 666-667 (such an objective is also similar in many ways to the aims articulated by early airpower advocates; for a concise survey, see David MacIsaac, "Voices from the Central Blue: The Air Power Theorists," in Peter Paret, ed., Makers of Modern Strategy (Princeton, NJ: Princeton University Press, 1986), pp. 624-647). As a point of departure, however, we will limit our consideration here to the more traditional objective of seizure and control of opposing territory.

decreasing force density. The second, or "Constrained Blue Employment" case, restricts the range of force employment choices available to the defender. In particular, blue is limited to a defensive depth roughly equivalent to that of blue's optimal choice at the 45,000 AFVE force level.[24] The constraint thus amounts to an assumption that blue fails to adapt as force levels fall, employing its forces in roughly the same depth regardless of the force to space ratio. Red force employment, on the other hand, is unconstrained; thus red adapts while blue (at least with respect to depth) does not.

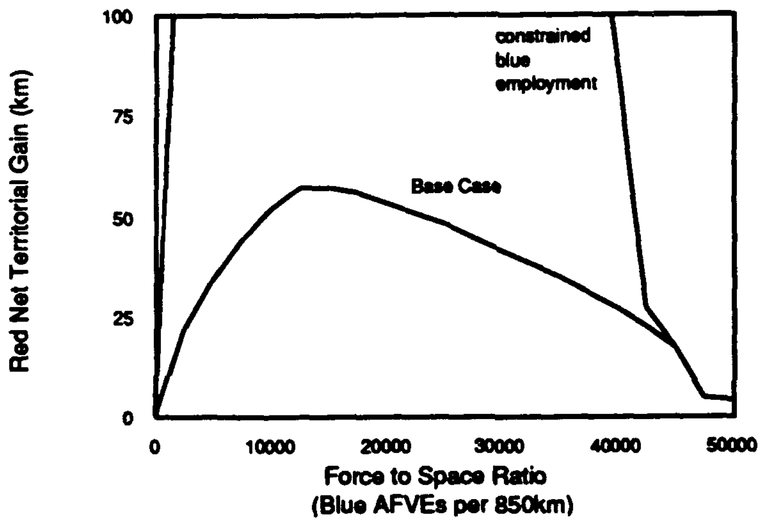

Figure I-2. The Effect of Force Employment Constraints

The result of these constraints is that net territorial gain diverges from the base case almost immediately as force levels fall. Once the force to space ratio falls below about 40,000 AFVEs in the theater, the blue defense effectively collapses, and red breaks through consistently until force levels fall to the point where neither side has meaningful forces in the theater.

---

[24] More specifically, blue's predeployed depth is capped at 10 kilometers (the optimal blue choice for a force level of 45,000 AFVEs), and blue's withdrawal fraction (which determines blue's "rolling depth," in addition to the static or "predeployed" depth) is capped at 0.78 (i.e., no more than 50 percent higher than the optimal blue choice at 45,000 AFVEs). For a more detailed discussion of depth in VFM, see appendices C and E.

This excursion result corresponds very closely to the threshold conception depicted in Figure I-1A. In effect, if we do not explicitly account for the role of changing defender force employment, then the theater dynamics described above are consistent with the public debate on the force to space ratio effect. If we consider the defender's potential to adapt doctrinally, however, we obtain very different results. In particular, we obtain a relationship without an identifiable minimum and in which the net effect of a lower force to space ratio is much less dramatic and much less problematic for the defender. Put somewhat differently, the treatment in the public debate can be seen in this light as a special case of a more general relationship—and a special case in which blue's prospects are unnecessarily grim. Defense at low force levels could thus be substantially more effective than the debate suggests if blue, and not just red, adapts its force employment to suit the demands of a lower density battlefield.

The nature of the adaptation required is depicted graphically in Figures I-3, I-4, and I-5, which show optimal force employment choices as a function of the force to space ratio as computed by the model for the base case considered above.[25]

Figure I-3 gives the defender's optimal solution for depth of initial deployment (or *predeployed* depth) as a function of the force to space ratio.[26] Optimal predeployed depth

---

[25] These optima, however, should properly be regarded as first order approximations. The VFM code does not analytically solve for true optima; rather, numerical techniques are used to approximate those optima within user-determined constraints on computer run time. As run time is increased, approximation accuracy is improved, but only in the limit is the true optimum obtained (note also that as approximation accuracy is increased, the curves depicted in figures I-3 through I-5 typically become smoother; for a more detailed treatment of approximation accuracy, see appendix F). Perhaps more importantly, the current version of the VFM code makes a number of simplifying assumptions in order to facilitate computation and reduce run time. In particular, Red's assault frontage and Blue's counterattack frontage and velocity are exogenously determined on the basis of standard planning factors or observed performance of the code rather than optimized endogenously. As a consequence, the resulting optima in some cases display more extreme variation as a function of force density than would be the case if these variables were treated endogenously. (Blue's optimal counterattack velocity, for example, falls very steeply above a force density of about 12,000 AFVEs in the theater. If blue were free to reduce counterattack assault velocity as force density increase, the slope of this curve would be less steep, and it is likely that the minimum counterattack fraction would not fall all the way to 0.01 as is depicted in figure I-4.) The importance of the results in figures I-3 through I-5 is thus not so much the specific values depicted—which are necessarily initial approximations—but more the nature and direction of change illustrated, and in the nature and importance of their effects on net territorial gain.

[26] *Total* depth includes both initial, or predeployed depth, (as shown here) and rolling depth obtained via withdrawal of those predeployed forces. For optimal withdrawal fractions, see figure I-4 and the associated discussion below; for a more detailed treatment of both rolling and predeployed depth, see appendix C.

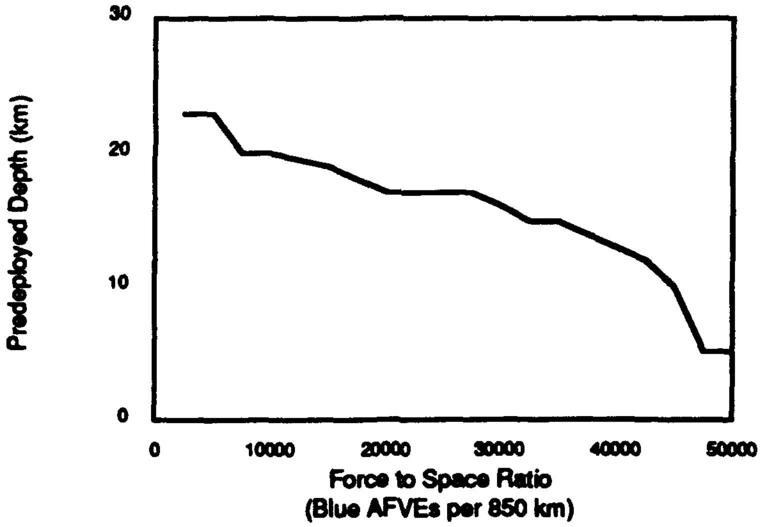

Figure I-3. Optimal Force Employment as a Function of Force to Space Ratio (Initial Deployment or Predeployed Depth)

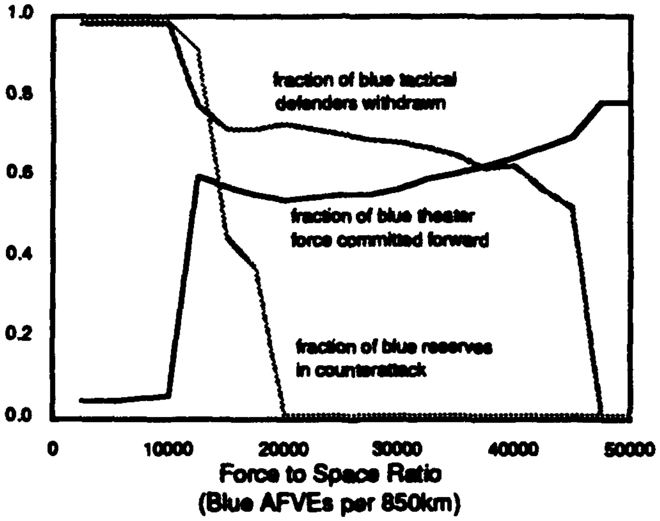

Figure I-4. Optimal Force Employment as a Function of Force to Space Ratio

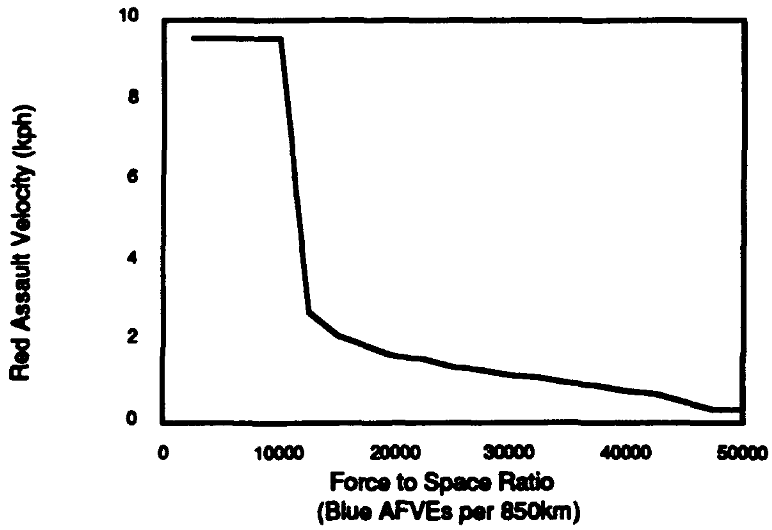

Figure I-5. Optimal Force Employment as a Function of Force to Space Ratio

is greatest at low force to space ratios, and smallest at high force to space ratios. Depth buys time for defensive reserves to arrive at the point of attack and tends to erode the efficiency of attackers that must advance over extended distances under the constant threat of fire from concealed defensive positions. Depth thus makes breakthrough harder for attackers. Depth also spreads defenders away from the frontier, however, and thus puts fewer defenders simultaneously within range of any given assault echelon. Deeper defenses thus make it harder for attackers to break through, but at the price of making it harder for defenders to halt an attacker quickly near the frontier. At low force to space ratios, the threat of attacker breakthrough is highest, and the opportunity to bring the attacker to a quick halt near the frontier is low anyway. Under these circumstances, depth is most valuable. At high force to space ratios, attackers are less likely to break through, and a less spread-out defense has an opportunity to impose an early halt on a terrain-limited attacker. Depth is least valuable under these conditions.

Figure I-4 gives the defender's optimal solutions for the fraction of total forces to be committed forward, the fraction of total reserves to be allocated to counterattack, and the residual force level (as a fraction of initial strength) at which blue tactical defenders are withdrawn from a given defensive position (and thus blue's ability to add to the defense's predeployed depth by "rolling with the punch" and giving ground). Optimal forward allocations are lowest at low force to space ratios. At low force to space ratios, blue cannot create a forward defense heavy enough to halt red quickly, even if blue allocates his entire force forward. Because blue must defend the length of the frontier, his forces are spread too thinly to meet red's concentrated attack at anything like even odds—and if he attempts to halt red at the frontier with a heavy forward allocation, blue will then lack the reserves required to counterconcentrate over time. The result of a high blue forward fraction at low force to space ratios is thus likely to be a red breakthrough. Under these circumstances, blue is better off accepting some loss of territory while amassing reserves, rather than trying to stop red too quickly and suffering catastrophic rupture of a forward defense too thin to enforce an immediate halt. A low forward fraction (with a correspondingly large reserve) thus provides a better outcome for blue at low force to space ratios than does a high forward fraction.

At high force densities, on the other hand, blue has an opportunity to halt red quickly by meeting red's terrain-constrained assault with a heavier forward defense. With a larger theater force available to him, blue can now create a sizeable defense along the entire front. Red's theater force is also larger, but terrain constraints prevent him from using it all immediately at the point of attack. Blue can thus afford to deploy more of his forces forward, and thus yield less ground prior to final counterconcentration, when theater force levels are higher. Higher force to space ratios thus encourage higher forward fractions for blue.[27]

Optimal counterattack fractions, on the other hand, are highest at low force to space ratios and lowest at high force densities. In general, the dynamics described above imply that the lower the defensive force density, the harder tactical defense becomes relative to tactical attack. While this is true for blue's theater defense, it is also true for red's flank defense. In effect, at high force levels, flank defense is relatively easy for red. Blue thus has an incentive to use his reserves for passive reinforcement rather than counterattacking an easily defended flank. At low force levels, however, flank defense is difficult for red, giving blue an incentive to devote an increasing fraction of his reserves

---

[27] Moreover, the higher the theater force-to-*force* ratio, the lower the optimum forward fraction for blue. Here, with an assumed theater force-to-force ratio of 1:1, blue's optimal forward fraction approaches 0.8 for force levels in excess of 50,000 AFVEs per 850 km. At a theater force to force ratio of 1.75:1, by contrast, blue's optimal forward fraction is under 0.4 at 50,000 AFVEs per 850 km. It should also be noted that "reserves" as discussed here include any blue forces present in the theater capable of rapid lateral redeployment (in effect, cross-corps or even cross-division movement). By this definition, all division and corps—as well as army group or theater—reserves are included in the "reserve" total (and thus excluded from the forces considered "forward committed."). For a more detailed discussion of reserves in VFM, see appendices C and E.

to counterattacking this more vulnerable flank as the theater force to space ratio decreases.

Optimal withdrawal fractions are likewise inversely related to the force to space ratio. Withdrawal adds "rolling depth" to a defense by enabling defenders who survive a given tactical engagement to fall back to rearward positions from which further resistance can be offered. Inasmuch as these rearward positions are not fully occupied at the outset, this process of withdrawal thus increases the effective depth of the defended zone beyond that of the defender's predeployed positions. The earlier the withdrawal order is given, the larger will be the number of surviving defenders available for redeployment in this way, and the greater the resulting overall depth of the defense. By the same token, however, the earlier the defender withdraws from a position, the lighter the casualty toll the attacker must pay in order to take that position. Rolling depth via withdrawal thus comes only at the price of lower attacker casualties per kilometer of ground taken. Where depth is most important for the defense, this price is most worth paying. As argued above, the lower the force to space ratio, the more essential is depth to the defender's ability to hold. Thus, optimal withdrawal fractions are highest at low force densities and lowest at high densities (where less overall depth is required).

Figure I-5 gives the attacker's optimal velocity (in kilometers per hour) as a function of the theater force to space ratio.[28] Red's optimal velocity choice is also inversely related to force density. At low force to space ratios, blue depends heavily on reserve arrivals and mounts a relatively weak forward defense (assuming optimal blue employment choices as described above). Red thus has an incentive to advance rapidly so as to penetrate as far as possible before those reserves arrive at the point of attack; moreover, the casualty price to be paid for that rapid advance is relatively low while blue's initial forward defense is weak. Conversely, at high force to space ratios, blue generally relies less on reserve arrivals over time and deploys a stronger initial forward defense. Red thus simultaneously faces a higher casualty price for rapid advance and has less incentive to advance rapidly in order to avert blue reserve arrivals. Thus the velocity choice that

---

[28] For the current version of VFM, velocity is the only attacker force employment choice optimized within the model; attacker frontage is treated exogenously as a function of theater force size by reference to standard planning factors (see appendices C and E). While the theory is structured in such a way as to make endogenous treatment of attacker frontage possible, this is not implemented as such in this initial code.

maximizes net territorial gain for red is highest at low force to space ratios and lowest at high force densities.

As a whole, then, the theory suggests a gradual transition in blue's optimal force employment from a relatively shallow, forward-oriented posture at high force to space ratios to a deeper, more counterattack-oriented posture with a higher fraction of available forces in theater reserve at low force levels. For red, the theory suggests a transition from a high velocity offensive at low force levels to a more deliberate, slower paced attack at high force densities. In more traditional terms, the theory describes, in effect, a gradual shift from a stiff blue positional defense at one extreme to an elastic delay-in-depth with reliance on a large counteroffensive reserve force at the other extreme. When blue employs available forces in this manner, optimal red employment resembles the traditional blitzkrieg only at very low force densities—and even here, the defensive combination of depth and counterattack makes such a posture unlikely to break through. If both sides are free to optimize, the net result of these adaptations is a modest increase in red's net territorial gain as force levels decline—until blue's increasing counterattack threat limits red's ability to hold seized ground, and net territorial gains fall accordingly.[29]

a. The Effects of Force to Force Ratios

As suggested above, the relationship between force density and combat outcomes is affected by a variety of intervening variables. Figure I-6 illustrates the effect of perhaps the most obvious of these, the red:blue theater force to force ratio (ffr). Whereas the ffr was held constant at a value of 1:1 in Figures I-1 through I-5, here it is varied systematically between a maximum of 2:1 and a minimum of parity.

---

[29] Of course, this representation is necessarily an abstract one; the actual implementation of doctrinal change requires answers to a wealth of detailed operational, logistical and command and control questions not directly addressed here. In a NATO context, for example, changes such as those depicted in Figures I-3 to I-5 would almost certainly require a variety of modifications in Alliance military organization if they are to be fully effective. Increased emphasis on theater-level counterattack or cross-corps reinforcement, for example, would be complicated by NATO's current "layer cake" corps structure, and by the inconsistent interoperability of NATO materiel and command procedures. The purpose of these assessments, however, is less to provide a detailed list of required changes in procedures or organizations than it is to develop an overview of the nature of the required change and the magnitude of its potential effect on outcomes. To these questions, the answers provided by Figures I-1 to I-5 are depth, reserves, and counterattack—and a moderate impact for decreased force-to-space ratios if these changes are realized.

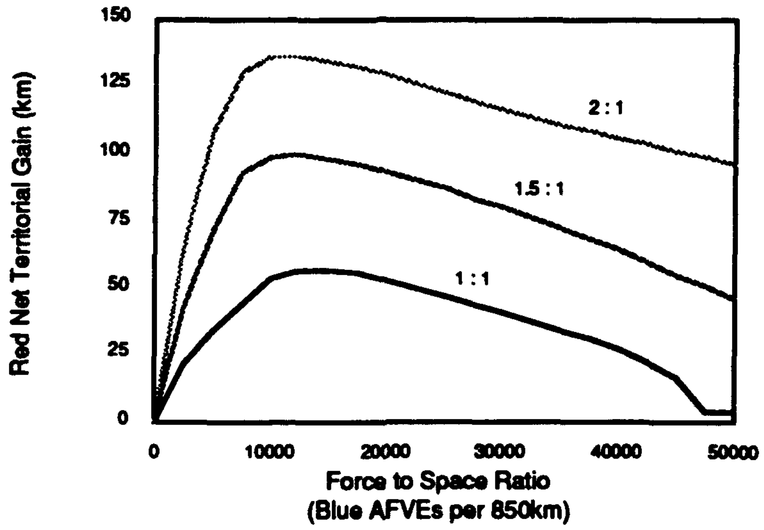

Figure I-6. The Effect of Force to Force Ratio

For each of the resulting curves, reducing force density for a constant ffr hurts the defender (at least for force levels above about 10-15,000 blue AFVEs). But there exist many combinations of changes in ffr and changes in force to space ratios that leave blue better off as a package. If, for example, blue were to move from an initial condition of a 2:1 ffr at a force level of about 50,000 AFVEs to parity at a force level 50 percent smaller than this, net territorial gain would actually decrease from 100 to under 50 kilometers.

b. The Effects of Barrier Defenses

Another intervening variable is terrain—and in particular, man-made terrain in the form of barrier defenses. The base case assumes a nominal defensive obstacle system capable of increasing attacker losses (for a given assault speed) by 50 percent relative to a defense with no barriers. Figure I-7 depicts the effects of more extensive obstacle deployments by comparing the base case results described above with two excursion cases representing the addition to the base case obstacles of two notional packages of barrier defenses of increasing magnitude.

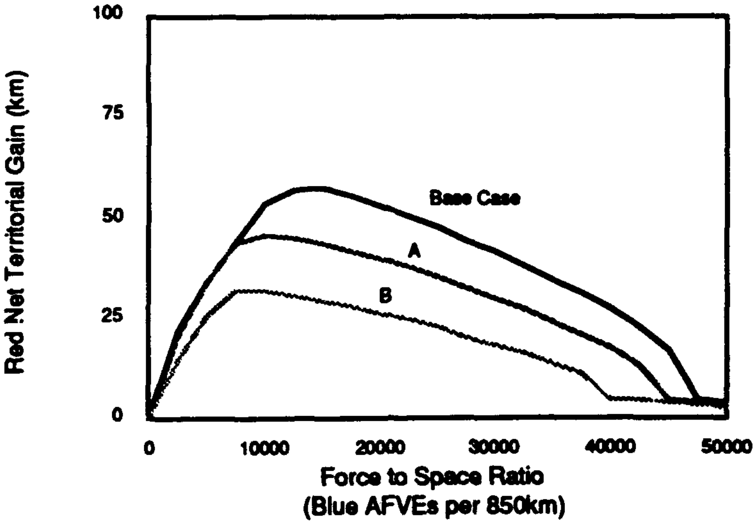

Figure I-7. The Effect of Barrier Defense

The U.S. Army has done a number of studies to estimate the potential impact of improved barrier defenses. In particular, a study done at the Army Training and Doctrine Command's Systems Analysis Activity (TRASANA) in 1978 concluded that a particular combination of protective construction, camouflage and minefields could produce over a 100 percent improvement in nominal loss exchange ratios for a typical small unit engagement at a given attack velocity and local force to force ratio.[30] Case A in Figure I-7 represents the substitution of a package corresponding to that considered by TRASANA for the less extensive barrier assumed in the base case. Case B represents the substitution of an even more extensive package corresponding to an almost 200 percent improvement in local loss exchange ratios.

Of course, more extensive barriers improve blue defensive performance: both excursion cases produce lower net territorial gain than the base case for all force to space ratios. Note, however, that improving barrier performance by about a factor of three (case B in Figure I-7) did not reduce red's net territorial gain by a factor of three at low force densities. Just as blue is able to mitigate the negative effects of decreased force density through changes in force employment, so red is able here to reduce the impact of

---

[30] Department of the Army, United States Army Training and Doctrine Command Systems Analysis Activity, Effects of Barriers in a Combat Environment (White Sands, NM: TRASANA, 1978), as referenced in U.S. Army Corps of Engineers, Engineer Studies Center, Survivability—The Effort and the Payoff, R-81-8, June 1981, pp. 21-23.

blue's barrier deployment by changing the way red's forces are used. In this case, red's optimal velocity fell from a value of 1.6 kilometers per hour at a force density of 20,000 AFVEs in the theater in the base case to about 0.9 kilometers per hour in case B. The slower rate of advance allows for a more extensive and better-prepared barrier clearance effort to support the red assault. By pausing long enough to discover barrier systems by advance reconnaissance, to deploy the combat engineers required to conduct a deliberate breach, and to support the clearance elements with a more extensive smoke and fire support preparation, red is able to reduce his losses relative to the nominal casualties estimated by TRASANA for a constant velocity attack. On balance, red is still worse off as a result of blue's barrier deployment (since red's decreased velocity permits blue to deliver more reserves to the point of attack for any given red advance distance). The degree of change, however, is substantially smaller as a result of red's ability to adapt force employment to suit changing circumstance.

c. The Effect of Tactical Warning

The basic theater dynamics described above assumed that the blue defender knew nothing of the location of red's point of attack prior to red actually crossing the border. In effect, only once the attack was under way could blue begin to counterconcentrate—and even then, reserves could not begin to move until commanders had had time to formulate plans, disseminate orders, and organize the forces to be moved in road march order from hidden positions in dispersed assembly areas. In the base case, it is therefore assumed that reserve movement toward the point of attack cannot commence until about four hours after red begins the main attack.[31] But what if the defender has some tactical warning—

---

[31] See Statement of General Fred K. Mahaffey, Director, Requirements Office of the Deputy Chief of Staff for Operations and Plans in Department of Defense Authorization for Appropriations for Fiscal Year 1981, Hearings Before the Committee on Armed Services, United States Senate, Ninety Sixth Congress Second Session, Part 5 (Washington, D.C.: U.S. Government Printing Office, 1980), p. 3030. General Mahaffey estimates three hours of command and control time required from the moment a decision is reached by higher command to begin counterconcentration to the time a brigade or larger reserve formation could be given movement orders; in addition, we assume here that one hour is required for the theater commander to process the necessary data and make that decision. Thus, if the stimulus for action is the initiation of the red attack, it follows that reserve units would receive movement orders four hours after the attack begins. Of course, this assumes that the blue theater commander has positively identified the true point of attack as of the time that attack begins. It is possible that risk-averse commanders could take more time to be certain that feints had not been mistaken for real attacks, or to discern with greater confidence the intended direction of the real attack (and thus the appropriate deployment point for the arriving reserves).

that is, intelligence on the location of the main attack sufficient to enable counterconcentration to begin prior to the onset of the attack itself?[32]

Figure I-8 answers this question by comparing the base case with three excursions representing improvements in tactical warning sufficient to enable blue to begin moving reserves 12, 24, and 36 hours prior to red's actual attack. The result of these improvements in warning is to flatten the relationship between force to space ratios and net territorial gain. In fact, if blue has enough advance warning of the location of red's attack, the force to space ratio effect disappears almost entirely for force to space ratios between about 5,000 and 35,000 blue AFVEs in the theater (and the effect of reductions below that level is to reduce, not increase, red's net territorial gain). It is far from clear that tactical warning on this order is a realistic possibility; red could, for example, redirect its lead forces onto an alternative axis of advance with less than the 36 hours of visible warning time that is assumed for the third excursion case here. Nevertheless, it is important to recognize the crucial role of warning time assumptions in any estimate of the effects of force to space ratios—and it is important to note the potential utility of improvements in warning for improving a defender's ability to cope with the demands of a lower density battlefield.

---

[32] Note that we refer here not just to indications that an attack is imminent, but to more specific battlefield intelligence as to where that attack will fall and in what relative strength (for a less restrictive definition, see Richard K. Betts, Surprise Attack (Washington, D.C.: Brookings, 1982), pp. 4-5). This is very different from *strategic* warning—meaning intelligence indications that an opponent is preparing for war. Tactical warning times are typically much shorter, and tactical warning of the sort described here can be substantially more difficult to obtain. Moreover, as noted above, advance tactical warning indications may or may not be immediately acted upon by a theater commander, since the consequences of committing reserves against a feint could well be to lose any chance of disengaging them in time to affect the outcome at the real point of attack. An additional advantage of defensive depth in this context is that it gives commanders time to distinguish feints from real attacks by observing differences in penetration distances rather than by reliance on intelligence indicators alone, and thus permits higher confidence decision making.

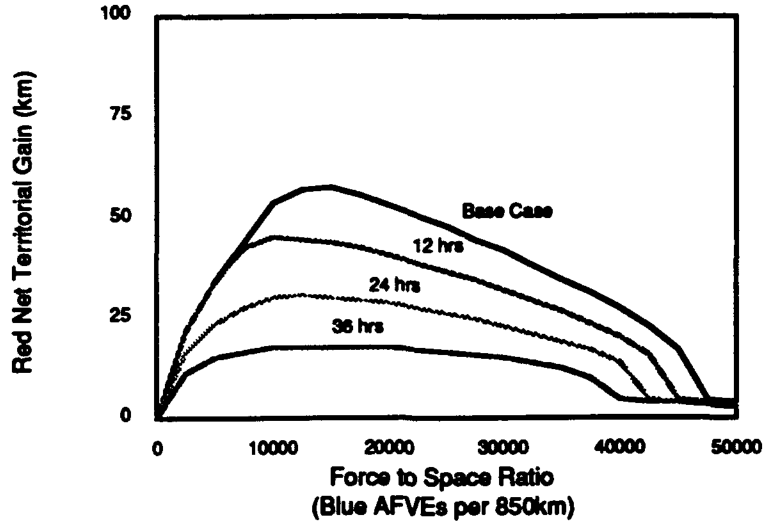

Figure I-8. The Effect of Tactical Warning

D. POLICY IMPLICATIONS

In short, then, force to space ratios do affect combat results, but theory suggests that as long as the defender adapts its force employment, this relationship is continuous, smooth, and shallow. There is no single "minimum" force to space ratio for effective defense, and the relationship as a whole is relatively weak. Moreover, the effect of force density on combat results depends on a variety of other variables, including the force to force ratio, the prevailing weapon mix, terrain, and of course, force employment. As a result, the defender can mitigate much of the potential negative effect of reduced force levels through policy initiatives designed, for example, to improve tactical warning or strengthen existing barrier defenses. Even without such initiatives, lower force levels need not prove catastrophic given appropriate adaptation of defensive force employment, but policy options are available for reducing the military impact of force reductions should one wish to do so.

If this is the nature of the underlying phenomenon, then what are its implications for policy? To answer this question, we will focus on two policy issues in particular: deep cuts in European conventional forces (whether negotiated or unilateral), and NATO strategy.

1. European Force Reductions

With respect to European force levels, the most important implication of the theory is that there is no identifiable minimum force to space ratio to provide a floor on force density for effective defense. While lower force levels do favor attackers for a substantial range of potential reductions, the magnitude of this effect need not be large. In general, then, we should expect that the military consequences of force reductions per se would be relatively modest—and may even be advantageous for certain combinations of force cuts and changes in the East-West balance or the weapons with which the forces are equipped.

More specifically, the theory's implications for particular NATO reduction options are summarized in Table I-1. As a baseline for comparison, the VFM model was used to compute a predicted outcome for a case corresponding to the NATO-Warsaw Pact balance as of late 1988, prior to Gorbachev's announcement of unilateral Soviet force reductions. Although the force levels that either side would actually bring to a conflict are inherently uncertain, it was assumed as a basis for analysis that the Warsaw Pact could mount an attack with some 70,000 AFVEs within about 2 weeks of mobilization. With comparable mobilization time, NATO was assumed to be capable of meeting this attack with about 43,000 AFVEs.[33] Given the weapon mixes available to the two sides in 1988, and optimal force employment by each side, the outcome calculated by the VFM model suggests a potential Pact ground gain of somewhat more than 70 kilometers.[34]

---

[33] For force levels, see Gray, op. cit. For a general discussion of the problem of uncertainty in assessing force levels, see Stephen D. Biddle, "The European Conventional Balance: A Reinterpretation of the Debate," Survival, March/April 1988, pp. 99-121.

[34] With the resulting force employment optima being, for NATO, a forward fraction of 0.42, a predeployed depth of 26.5 kilometers, a withdrawal fraction of 0.68, and an allocation of about 1 percent of theater reserves to counterattack; and for the Pact, an optimal assault velocity of 1.4 kilometers per hour.

For NATO, these computed optima are in fact broadly consistent with NATO's actual employment profile in 1988. Of 124 available central front brigades, for example, NATO was expected to hold about 56 in either corps or army group reserve (corresponding to a forward fraction of about 0.45; see Gray, op. cit.). Perhaps 21 of these brigades belonged to either the U.S. III Corps or the French First Army (counting each smaller French division as the equivalent of one U.S. brigade) and would thus have been considered for use in counterattack, but it is not clear what fraction of these forces would actually have been employed in this role. While the number would surely have been greater than one percent of the total theater reserve, it would probably have been less than 20 percent. Finally, the predeployed depth computed as the optimum above would certainly not have exceeded the bounds of NATO's existing plans for a "main battle area" of between 40 and 70 kilometers depth (see David C. Isby and Charles Kamps, Jr., Armies of NATO's Central Front (New York and London: Jane's, 1985), pp. 199-209, 265-175).

For the Pact, the computed optimum assault velocity of 1.4 kilometers per hour is significantly slower than that implied by the Soviets' own 1988 doctrine—which in terms of our definition of assault velocity (see appendix C) comes to between 2.3 and 4 kilometers per hour. This corresponds to the net closure rate implied by a 25 to 50 minute fire support preparation followed by a 12 kilometer per hour regimental advance in two echelons, followed by support and command elements, over an assumed

Table I-1. European Force Reduction Outcomes

Scenario	Net Territorial Gain (km)
1988 Force Levels	72
Post CFE Force Levels	31
50% of CFE Force Levels	54
25% of CFE Force Levels	52
100% of CFE Force Levels, Eastern Europe Neutral	3
25% of CFE Force Levels, Eastern Europe Neutral	8
NATO at 50% of CFE Force Level, USSR at 100% of CFE Force Level, Eastern Europe Neutral	116

While it is difficult to say whether this outcome would or would not have met Soviet war aims in an invasion of the West, it clearly does not represent either the complete collapse of the NATO defense or the annihilation of the Soviet attacker. Rather, at the high force density of the 1988 deployments, NATO's defenses perform well enough to contain the

---

distance corresponding to a five kilometer average separation between NATO defended positions in depth; see Headquarters, Department of the Army, FM 100-2-1, The Soviet Army: Operations and Tactics (Washington, D.C.: USGPO, July 1984), pp. 5-17, 5-22, 5-27. Against the computed NATO optimum employment profile, a Pact assault velocity in excess of 2.3 kilometers per hour produces excessive casualties at the point of attack without permitting a commensurate increase in the rate at which the attacker actually penetrates the defense, with resulting net territorial gains of less than 60 kilometers; increasing assault velocity to 4 kilometers per hour reduces territorial gains to only 32 kilometers. There is some evidence that the Soviets themselves were coming to similar conclusions in the late 1980s and were consequently moving in the direction of a less tank-heavy, more infantry and combined arms oriented organization and doctrine as a means of penetrating what they perceived to be an increasingly dense NATO defense (for a more detailed discussion see appendix B). In our terms, this would imply a slower average assault velocity (note that "assault velocity," or the closure rate of assault forces with defended positions at the tactical level, and "rate of advance," or the overall rate at which the invader takes ground, are not necessarily the same quantity—the discussion above pertains solely to the former. For a more detailed treatment of the distinction, see appendix C).

Soviet offensive—even at an adverse force to force ratio—but not without some loss of territory.

If we instead assume forces corresponding to the Central Front's share of the CFE treaty limits, then the Soviets' potential ground gain falls to 31 kilometers.[35] CFE I reduces NATO's force to space ratio (from 43,000 AFVEs in the theater at D-day to about 37,000), but compensates for this modest density reduction with a major reduction in the theater force to *force* ratio (from 1.65:1 to 1:1). The net result of the two changes is that the invader-favorable effect of a somewhat lower force to space ratio is more than offset by the defender-favorable effect of a much lower force to force ratio, with the cumulative consequence being a substantial reduction in the net territorial gain predicted by the theory.

Of course, the forces permitted under CFE I will not necessarily be maintained by either side in coming years. If we assume further reductions to 50, and to 25 percent of CFE I force levels on both sides, the predicted ground gain increases relative to the CFE I outcome. In neither case, however, does the reduced force density produce a Soviet breakthrough—and in neither case does the predicted outcome resemble a catastrophic failure of the NATO defense. Whereas CFE I cuts both the force to space and the force to force ratio, the further reductions considered here affect only the ratio of force to *space*—the balance between NATO and Pact forces is assumed to remain constant at 1:1 as further cuts are implemented. As we have seen, the theory suggests that lower force to space ratios can increase net territorial gains, but that mutual force reductions eventually leave attackers without sufficient forces to hold the ground they take. Thus a mutual 50 percent force cut from CFE I levels increases the Soviets' potential gains by some 23 kilometers, but a further cut to 25 percent of CFE I actually decreases net gains slightly.

With the effective dissolution of the Warsaw Pact, however, it is not clear that a NATO-WTO conflict at CFE force levels is necessarily the most important case. Many possible alternatives could be posited; for illustrative purposes, we will focus here on the possibility of a future NATO-USSR conflict in which the East European states of Poland, Czechoslovakia, Hungary, Romania and Bulgaria remain neutral, but do not contest Soviet transit of their territory prior to its initiating hostilities with the West. Two variations of this scenario have been examined: one in which NATO and the USSR each

---

[35] For corresponding force levels, see Gray, op. cit.

commit forces equivalent to their respective CFE limits, and one in which a combination of further force reductions and other contingencies prevents either side from committing more than 25 percent of its CFE ceiling to a potential conflict. In addition, we have examined an asymmetric reduction scenario in which NATO reduces its forces to a level 50 percent below its CFE ceiling, but the USSR commits a force equivalent to its ceiling under the treaty.

The results of these analyses suggest that an unassisted Soviet attack would have little success against a united NATO defense regardless of force level—that is, regardless of the force to space ratio—as long as further reductions are not too asymmetric. Even at 25 percent of the respective CFE limits, the potential Soviet ground gain is less than a dozen kilometers when both sides reduce proportionally. Without East European support, a Soviet attack at the CFE sublimit force level would be substantially outnumbered by a unified NATO defense.[36] An outnumbered attacker faces a difficult military problem against reasonably employed defensive forces at any force level.[37] If, however, NATO reduces significantly faster or further than the USSR, then the potential territorial gain in the event of a Soviet attack could increase substantially. Such circumstances combine the effects of a lower force to space ratio and a higher force to force ratio, and can produce territorial gains in excess of 100 kilometers if, for example, the Soviets commit forces equal to their CFE limit against a NATO force reduced to 50 percent of that limit.

Overall, then, the theory suggests that NATO can cope with even very deep cuts in force levels. For these results to obtain, however, NATO reductions must be accompanied by cuts in Soviet forces, and NATO force employment must suit the demands of a lower density battlefield. If NATO military doctrines remain unchanged as force levels fall, theory suggests an increasing risk of breakthrough as force levels fall below the CFE level. To minimize this risk, NATO militaries will have to deepen forward defensive

---

[36] Fighting alone against a unified NATO, with only the forces permitted it in the Atlantic-to-Urals (ATTU) region, the Soviet Union could expect only a 0.66:1 theaterwide attacker:defender force to force ratio. Of course, the Soviets could, if strategic circumstances permit, reallocate forces from outside the ATTU region in the event of a war in the West. For an illustrative example of the consequences of force imbalances at lower force levels, see discussion below. On the general problem of diplomatic and political uncertainty in the determination of theater force levels, see Biddle, op. cit., pp. 99-121.

[37] Even outnumbered attackers can succeed, however, if defenders employ their forces poorly. Even at an adverse theater force to force ratio of 0.66:1 as assumed here, for example, the Soviets could still break through the NATO defense if NATO were to fight a rigid defense in place (i.e., a withdrawal fraction of 0) at a predeployed depth of less than 20 kilometers (assuming a counterattack fraction of one percent, and a forward fraction of 0.4, as per the optimal 1988 NATO employment).

positions, hold a larger fraction of total forces in reserve, and use more of those reserves for counterattack than today.[38] Such changes may require different training and may be more demanding of the skills of commanders and their troops than current doctrine.

Alternatively, if NATO force levels fall significantly faster or further than those of the Soviet Union, then the combination of a lower force to space ratio and a higher theater force to force ratio could leave NATO militarily worse off than it was in 1988—even if NATO employs its forces optimally.[39] For deep cuts to be sustainable militarily, it is thus important that these cuts be two-sided in nature.

There seems little reason to believe, however, that these two conditions could not be met. With respect to opposing force levels, the unification of Germany and the effective dissolution of the Warsaw Pact have already produced a major reduction in the forces potentially available for any attack on NATO. In combination with ongoing unilateral cutbacks in Soviet and East European armies, it seems likely that the size of any potential attack force will continue to fall in coming years. Moreover, the ongoing arms control process provides a formal mechanism by which NATO can both establish a ceiling on future Soviet force levels, and verify those force levels to its own satisfaction.

With respect to changes in NATO force employment, it is important to recognize that, although their influence on outcomes is potentially large, the changes themselves constitute differences of degree, not of kind, relative to current practice. Any military

---

[38] In particular, were NATO to maintain the employment profile computed to be optimal for the 1988 conditions of a 1.65:1 theater force to force ratio at a NATO force level of 43,000 AFVEs, then the theory predicts that NATO could avert breakthrough for deep force reductions at theater parity, but with unnecessarily high net territorial gains in the event of attack, and very little margin for error. (In effect, the advantage of a force to force ratio reduction from 1.65:1 to 1:1 enables the force employment choices which are optimal at 43,000 AFVEs to be acceptable, if sub-optimal, at, e.g., 18,500 AFVEs. By contrast, the force employment choices which are optimal for the post-CFE conditions of parity at 37,000 AFVEs would produce breakthrough at a force level of 18,500 AFVEs—i.e., without any further offsetting improvement in the theater force to force ratio.) As a general rule, for a constant force to force ratio, lower force to space ratios cause force employment optima to shift in the direction of larger fractions of theater force in reserve, larger fractions of theater reserves in counterattack, deeper initial deployment of those forces committed forward, and a higher withdrawal fraction for forward-committed forces. Such a shift cold provide a substantially increased margin for error under conditions of low battlefield force densities.

[39] Assuming, of course, that the Red Army constitutes a disciplined, coherent force capable of posing a serious military threat, which is by no means a certainty for the post-Cold War era. If so, however, a combination of numerical imbalance at lower force levels and ill-suited force employment choices could have particularly severe military consequences in the event of an attack. If, for example, the Soviets committed forces equal to their CFE limit against a NATO force reduced to 50 percent of that limit, and NATO defended with a pre-CFE operational doctrine, the predicted result would be a clean

doctrine consists of both static and mobile, active and passive, and forward and reserve elements. All modern armies are trained to counterattack and to reinforce; all modern armies deploy forces forward and withhold mobile reserves; and all modern armies conduct delays and practice defenses-in-place. The particular balance differs from doctrine to doctrine, but the underlying elements are common to all. Thus, the adaptation described here does not represent a wholly novel approach to force employment, but rather amounts to a change in emphasis among functions already present in the operational repertoire of current NATO military organizations.[40] Such changes in emphasis are a normal part of an army's response to ongoing developments in technology, threat, and strategic circumstance; the U.S. Army, for example, has already changed its official doctrine four times since 1945,[41] and is in the process of a fifth major modification now.[42] There is every reason to believe that further doctrinal development would eventually proceed in all NATO armies as a natural consequence of an Alliance decision to reduce its forces. While this process requires conscious effort on the part of both military and civilian organizations—especially given the pace of ongoing changes in European security relationships, it is thus more reasonable to assume doctrinal change than to assume that armies would operate identically on high and low density battlefields.

Deep cuts in force levels thus need not undermine effective conventional defense. This does not necessarily mean that any given reduction proposal should be adopted, however. There are larger questions of alliance politics and national foreign policy goals that must be addressed before a sound decision can be reached on the force level that best suits U.S. and NATO interests.

---

breakthrough (as opposed to the 115 kilometer ground gain reported in Table I-1) even without the participation of the East European states in the attack.

[40] See note 29 above for a more detailed description of the indicated changes.

[41] Specifically, the introduction of the Pentomic Division in 1956, the Reorganized Armored Division (ROAD) concept in 1962, the Active Defense in 1976, and the AirLand Battle in 1982; and not including the many less sweeping modifications of doctrine between these formal revisions, or the shift in emphasis toward counterinsurgency that characterized the Vietnam era. See, for example, Robert A. Doughty, The Evolution of U.S. Army Tactical Doctrine, 1946-76 (Ft. Leavenworth, KS: Combat Studies Institute), Leavenworth Paper No.1, esp. pp. 12-25, 40-50; and Romjue, op. cit. See also Paul H. Herbert, Toward the Best Available Thought: The Writing of Field Manual 100-5, "Operations" by the United States Army, 1973-1976, Ph.d dissertation, The Ohio State University, 1985; and Wass de Czege and Holder, op. cit., pp. 53-70.

[42] See, for example, "Calling the Shots," Army Times, November 26, 1990, pp. 22, 24, 61.

What this does mean is that deep cuts in European force levels are primarily a political, rather than a military question. If we conclude that major reductions in forces are in our larger national interest, then the results of this study suggest that the military effects of those reductions can be managed. Whether they should be managed is not a question that can be answered on the basis of a military analysis.

2. Alliance Strategy

A second policy issue of immediate significance for U.S. decision making concerns the validity of NATO strategy in an era of lower force levels. Since 1967, NATO strategy has consisted of two key elements: a nuclear strategy of Flexible Response and a conventional strategy of Forward Defense. Each is now being re-evaluated in light of the diminished threat from the Soviet Union. In the case of Forward Defense, however, this re-evaluation is motivated not only by changing threat perception, but also by changing perceptions of the military capacity of a reduced Western force structure. In particular, it has been widely suggested that deep force cuts will make Forward Defense impossible—even given corresponding reductions in Soviet forces. Consequently, it is argued, the Alliance must renounce the strategy of Forward Defense, and begin deliberations to define an alternative which would be better-suited to the demands of a low density battlefield.

The theory advanced here, however, suggests that the defensive potential of a reduced NATO force structure need not be substantially lower than that of today—and that there is consequently no military need to abandon the underlying aims embodied in the strategy of Forward Defense. This is not to suggest that sweeping change in the institutional underpinnings of European security either is or is not required by the demands of the post-Cold War world, and it is certainly not to suggest that NATO's military posture can be left unchanged in response to these developments. Profound change in many aspects of NATO's posture will clearly be required.[43] Moreover, the Alliance is very likely to abandon the term "Forward Defense" as a description of its

---

[43] At a minimum, the "layer cake" arrangement of national corps sectors will almost certainly be eliminated, the level of multinational military integration will almost certainly be lowered from the current system of multinational army groups to the level of multinational corps or even divisions, and the theater c3 system for coordinating the maneuver of Alliance forces will almost certainly require substantial modification [on the latter point in particular, see Paul Stares, Command Performance: The Neglected Dimension of European Security (Washington, D.C.: Brookings, forthcoming)].

wartime aims.[44] But it is important to recognize that these changes need not undermine the substance of the political commitment embodied by the strategy of Forward Defense as the Alliance has understood it for the past two decades—and indeed, it is highly likely that this commitment can in the future be met at substantially lower levels of conventional forces.

To substantiate this contention, we must answer two questions: What is Forward Defense, and what has it meant in specific military terms? Will deep cuts in NATO's military forces compel NATO to adopt a posture fundamentally inconsistent with that meaning?

As for the first of these questions, Forward Defense is a declaratory policy describing the broad aims of Alliance strategy in the event of war—it is not a specific operational doctrine for the employment of military forces. On the contrary, in NATO, operational doctrine is a national, not an Alliance, responsibility. NATO in fact has 16 *different* operational doctrines governing the actual use of the forces assigned to it. The American doctrine of AirLand Battle, for example, is unique to the U.S. Army. The British Army of the Rhine operates according to a different, more positional doctrine; German doctrine emphasizes fluid operations from shallower deployments, and so on.[45] Forward Defense establishes the bounds within which these national doctrines may vary, but it is not in itself a blueprint for military operations.

Moreover, the bounds established by Forward Defense are quite broad. The real substance of Forward Defense is a commitment by NATO to the Federal Republic to defend as much of its territory as possible. The specific nature of this commitment, however, is deliberately ambiguous. On the one hand, Forward Defense has clearly excluded the alternatives of a voluntary withdrawal to the Rhine, or of a defense by guerilla warfare in the German interior.[46] On the other hand, however, it clearly does *not*

---

[44] See, for example, John Broder, "NATO Ministers Take Steps to Fundamentally Change Alliance," Los Angeles Times, May 24, 1990, p. 16; also "NATO to Change European Strategy," Defense News, October 8, 1990, p. 2.

[45] See, for example, Phillip A. Karber, "In Defense of Forward Defense," Armed Forces Journal International, May 1984, pp. 27-50; also Isby and Kamps, op. cit., pp. 199-209, 265-175.

[46] The former was in fact NATO policy prior to German rearmament; see, for example, Roger L. L. Facer, Conventional Forces and the NATO Strategy of Flexible Response (Santa Monica, CA: RAND, 1985), pp. 15-17; also Karber, op. cit., pp. 27-50; Isby and Kamps, op. cit., pp. 14-15. The latter has been proposed by a variety of "Alternative Defense" advocates, including, for example, Gene Sharp and

constrain NATO to deny the Soviets a single inch of Federal territory, or to deploy forces in a shallow cordon defense at the international border. On the contrary, existing plans call for a delay-in-depth across a covering force zone extending some 15 to 40 kilometers deep. Behind that is a main battle area of between 40 and 70 kilometers depth. It has been intended that a Soviet offensive would be halted, and ultimately reversed, within the latter zone—a prescription which more or less guarantees that the Soviets would initially overrun some German territory, even if the defense were wholly successful according to plan.[47] Forward Defense is thus clearly not a specific mandate for a shallow, linear tactical deployment at the international border. Rather, within the broad guidance of defending as much of the Federal Republic as possible, Forward Defense already provides substantial latitude for military organizations to operate their forces as they see fit.

Turning to the second question, are the military requirements of defense at low force levels inconsistent with the underlying aim of Forward Defense? Given the degree of latitude that has always existed for military implementation of that aim, the answer is no. The analysis above suggests that, even at very low force levels, NATO's ability to deny a Soviet attacker control of German territory need not be significantly lower than that of today. Indeed, for some force levels such a defense could on balance prove *more effective than today's, in that* lower force density would be accompanied by parity (or better) between NATO and Soviet forces. To be clearly inconsistent with a strategic aim as broadly defined as Forward Defense has always been, a military posture would have to undermine significantly NATO's ability to hold ground; it is far from clear that the consequences described here fit such a description.[48]

Perhaps the fundamental policy implication of this analysis is thus that deep cuts are primarily a political, rather than a military, question. Defense at lower force to space

---

Horst Afheldt; for a survey, see, e.g., Jonathan Dean, "Alternative Defense: Answer to NATO's Central Front Problems?" International Affairs Winter 1987/88 (Vol.64, No.1), pp. 61-82.

[47] See Isby and Kamps, op. cit., pp. 15, 269. For a more general discussion of the question of depth and elasticity in Central Europe, see Defining Stability in the European Theater, Hearings Before the Defense Policy Panel of the Committee on Armed Services, House of Representatives, One Hundredth Congress, Second Session, HASC No 100-104, 1989, pp. 121-122.

[48] As noted above, effective defense at lower force levels will also require a variety of changes in the organization of NATO's deployed military posture. The point is not that NATO can operate in the future as it does today. Rather, the point is that the necessary changes can be subsumed within even the Alliance's *current* definition of Forward Defense without doing violence to the principles upon which that strategy has operated over the preceding two decades.

ratios can be managed militarily, and without requiring wholesale abandonment of the fundamental principles underlying NATO's declaratory strategy. To do so requires that we ensure that reductions in NATO forces be matched by those of the USSR. To do so also requires that a variety of operational changes be made in the military doctrines of NATO member states, and it may require that the Alliance military organization be modified to facilitate cross-corps reinforcement and interoperability. This study has not attempted to provide a detailed or comprehensive list of the necessary modifications; this can only be done by the military organizations involved. Nor has it been intended as a road map for a CFE negotiating strategy that would best ensure that force levels remain balanced as both sides reduce their strength. What this study has attempted to provide is a conceptual structure for understanding the nature of the changes introduced by radical reductions in force density, for assessing their interaction and relative magnitude, and for relating elements of change and of continuity in a rapidly evolving international security environment. To this end a systematic theory has been developed, tested, and exercised. In the process, an approach has been developed with the potential to address a broad range of military issues—and thus to provide at least a partial contribution to the larger problem of planning for national security in this era of rapid change.

Appendix A

REVIEW OF WESTERN LITERATURE

David G. Gray

A. INTRODUCTION

This appendix is intended to provide a thorough treatment of prior thought on the issue of force to space ratios from the perspective of western military theoreticians. Soviet literature is treated separately in appendix B.

Force-to-space ratios are by no means a new issue, but the history of western thought in this area is highly episodic. Given this, the literature is treated in three sections, dealing respectively with works prior to the First World War, contributions in mid-century, and the modern debate as it has developed after about 1980. Authors whose insights are comprehensive or particularly significant are examined in some detail.

For each author or school of thought, several key questions are posed. First, how is the relationship between force density and combat results characterized, and why? Does decreased density benefit attackers or defenders? Is the relationship continuous, or discontinuous; steep or shallow? Second, what other variables or related effects, if any, do the authors cite as significant? Is the relationship between density and combat results essentially fixed, or is it susceptible to change as a result of the influence of other battlefield phenomena? Finally, how important an effect is the force to space ratio seen to be? *Is it considered to be of paramount importance, or to be of secondary concern* relative to other issues addressed by the writers?

Following the description of existing literature, an assessment is made of its strengths and weaknesses. On the basis of that assessment, implications for the further development of theory on this question are derived.

In essence, this appendix concludes that the existing literature is of substantial value as a source of insight into the range of effects and interactions to which force density is relevant on the conventional battlefield. That insight, however, is neither systematic nor conclusive. Terminology is inconsistent, classification of evidence is often ambiguous, and bounds or limits are rarely defined. The literature thus offers at best a partial description of this phenomenon. To provide an appropriate basis for policy making, a more complete and systematic understanding is necessary—and will be undertaken in appendices C, D, and E on the basis of the foundation provided by the existing literature as described here.

B. MILITARY THOUGHT PRIOR TO WORLD WAR I

The effects of force density have been a topic of discussion among military writers for over 150 years. Clausewitz, for example, wrote in the immediate aftermath of the Napoleonic Wars:

> In fact, a fairly constant ratio exists between the size of a force and the area it can occupy. This ratio cannot be expressed in numbers; besides, it is subject to change by other circumstances. Here it is enough to say that the relationship between the two is permanent and fundamental. One may be able to march to Moscow with 500,000 men, but never with 50,000—even if the invader's strength relative to the defender's were greater in the second case than in the first.[1]

The first sustained treatment of the issue, however, arose in the decades preceding World War I. The late nineteenth and early twentieth centuries were a time of great change in European military science. New technology was arriving at an unprecedented rate, and the question of its proper use and ultimate effect on military operations captured the attention of a generation of theorists and doctrine writers. In particular, the problem of attacking a defense armed with modern weaponry attracted widespread attention.[2] In the opinion of most experts, the firepower newly available to defenders (in particular, machine guns and modern rifles), when combined with easily deployed field fortifications (most notably, barbed wire), made a defensive position almost invulnerable to frontal assault.[3] Nevertheless, a "cult of the offensive" dominated pre-war military doctrines.[4] The majority of writers concentrated on describing the means by which defensive firepower could be overcome by a decisive attack. The more pragmatic of these writers focused on the firepower available to the attacker, arguing that the firepower of the attack

---

[1] Carl von Clausewitz, On War, translated and edited by Michael Howard and Peter Paret (Princeton, NJ: Princeton University Press, 1976), Book VI, Chapter 25, p.472. As will be seen, few other writers have shared Clausewitz' aversion to quantification of the effects of force to space ratios—specific "minimum force densities" are in fact characteristic of the literature as a whole. Interestingly, Clausewitz also takes particular account of the impact of force density on *attackers* rather than on defenders alone. For most writers, the force to space ratio is primarily (if not exclusively) a defensive concern.

[2] For example, see the discussion of "the cult of the offensive" in both Stephen Van Evera, "The Cult of the Offensive and the Origins of the First World War," and Michael Howard, "Men Against Fire: Expectations of War in 1914" in Steven Miller, ed., Military Strategy and the Origins of the First World War (Princeton, New Jersey: Princeton University Press, 1976).

[3] Howard, pp. 42-43.

[4] This "cult of the offensive" preached the superiority of the attack over the defense. Nonetheless, it recognized that a successful frontal assault would be both costly and difficult. For a contemporary perspective, see Freidrich von Bernhardi, On War of Today. (New York: Dodd, Mead & Company, 1914), pp. 155-160.

could overwhelm that of the defense.[5] The majority of experts, however, simply emphasized the moral superiority of the attack, reasoning that spirit could overcome technology.[6]

Within this context, however, a few writers approached the problem from a different perspective. Rather than concentrating on what an attacker might do to overcome a defender's advantage, they developed insights into what a defender had to do to gain that advantage. Most significant for this study, a number of these authorities concluded that the defender was required to maintain a minimum force-to-space ratio. In general, these writers were inspired by the combat results of the Boer War and by the Russo-Japanese War of 1905. In particular, the views of two writers are crucial in this respect: Commandant Jean Colin of the French War School, and General Wilhelm Balck of the German army.[7]

Lieutenant General Wilhelm Balck commanded an infantry division on the Western Front during World War I. Prior to the war, Balck, then a Colonel, wrote Tactics, a study of the infantry tactics and army regulations used by the major powers. The first edition of Tactics appeared in 1896. The fourth edition of the work incorporated the lessons of the Russo-Japanese war, and was issued in 1908. In France, Commandant Jean Colin of the Ecole De Guerre wrote The Transformations of War in 1912, in which he drew primarily upon the evidence of modern battles (from the German-Austrian War of 1866 through the Russo-Japanese War of 1905) to make judgements on the state of modern warfare. Although both Colin and Balck looked to contemporary battles for inspiration, they labored at different tasks. Colin produced a work of military history and analysis. Balck's Tactics, on the other hand, was a pragmatic book firmly grounded in the

---

[5] In particular, (then) Colonel Ferdinand Foch made this argument while teaching at the Ecole de Guerre in 1900. See op. cit., p. 42.

[6] See Ibid., pp. 55-57, and van Evera, op. cit., pp. 60-61.

[7] It must be stressed that even these authors—in particular, Jean Colin and Wilhelm Balck—lay firmly within the mainstream with respect to the "cult of the offensive." Although (as will be seen) Colin held that a frontal attack had relatively little chance of succeeding, he nonetheless believed firmly in the moral value of attack, and of the frontal assault in particular: "if it is the outflanking movement that is productive of victory, it is the frontal attack that reaps the moral fruits of victory, and it is by prolonging it as a direct pursuit that one obtains great results." See Jean Colin, The Transformations of War, translated by L.H.R. Pope-Hennessy (London: Hugh Rees, Ltd., 1912), pp. 70-73. Balck, although he stressed the importance of firepower to the outcome of a battle, also held that the moral effect of the bayonet made it indispensable to the infantry. He wrote, "If the infantry is deprived [of the bayonet] ... An infantry will be developed which is unsuitable for attack and which moreover lacks a most essential quality, viz. The moral power to reach the enemy's position." See Wilhelm Balck, Tactics, Volume I, Introduction and Formal Tactics of Infantry, translated by Walter Kruger (Fort Leavenworth, Kansas: U.S. Army Cavalry Association, 1915), p. 383.

infantry regulations of the day. It was a manual intended to offer combat guidelines to the prospective commander. These different objectives drove the two authors to distinct (but not contradictory) insights.[8]

In his work, Colin offers a more detailed examination of why a minimum force-to-space ratio should exist. Balck, on the other hand, concentrates on identifying what the minimum actually might be for a specific situation. As each author tends to focus on distinct issues, the work of Colin is emphasized in the first sub-section (called "Discontinuous Fronts") and that of Balck dominates the second sub-section (called "Important Variables"). First, however, we examine those ideas common to both Balck and Colin.

In particular, three common points are worthy of note. First, lower force to space ratios favor attackers over defenders, and force densities that permit successful attack are separated from densities that prohibit attack by a unique minimum or threshold value. Second, Colin, in particular, argues that this minimum exists because of the deleterious effect (on the integrity of the defense) of gaps in the front line.[9] Third, Balck (and to a lesser extent, Colin) identifies a number of independent variables as important to a determination of the minimum force-to-space ratio. These include weapon technology, the use of field fortifications, the defensibility of the natural terrain, and methods of force employment (e.g., the attacker's ability to launch flank attacks).

---

[8] On one significant issue, Balck and Colin appear to disagree. As will be seen, the belief that gaps pose a danger to the defense is central to Colin's explanation of why a minimum force-to-space ratio should exist. Balck, on the other hand, writes (p. 411)

> This [the need for a minimum force-to-space ratio] does not imply that the position must be held in equal strength all along the line; portions of the line that are very difficult to attack need only be kept under observation. Gaps in the defensive line are, as a rule, of very little value to the assailant, as the defender will frequently be able to sweep the space in front of them from a flank.

Note, however, that in discounting the danger of the gaps, Balck makes three important assumptions: 1) the gaps are in very difficult terrain; 2) the gaps are kept under observation; 3) the gap are small enough that they can be covered by fire. When viewed from the scale used by Colin (in general, armies of hundreds of thousands deployed along fronts many kilometers in length), it seems probable that gaps such as are discounted by Balck would not even appear as such.

[9] Colin, p. 158.

Balck and Colin's arguments stem from their belief that modern defenses are almost invulnerable to frontal assault. Consequently, the weakest parts of a defender's front line are the flanks.[10] With respect to this point, Balck writes

> The German Infantry Drill Regulations (par. 397) further emphasize the fact that, when well-trained infantry employs its rifles to good advantage in defense, it is very strong in front; that it can hold a position with a comparatively small force; and that, in this case, it has only one weak spot, the flank, which it must seek to protect by distribution in depth. This view is fully borne out by the recent events in South Africa and Manchuria.[11]

Colin's studies of contemporary battles lead him to a similar conclusion: frontal assaults against a prepared defense almost invariably fail.[12] After analyzing the results of the Franco-German War of 1870 and the Russo-Turkish War of 1877, Colin concludes that "fronts are inviolable".[13] Likewise, in the Boer War, the British could not penetrate the defensive lines of the Boers with a frontal assault.[14] In contrast, he argues that flank attacks often succeed. In the Boer War, the British succeeded when they resorted to flanking maneuvers and operations against opposing lines of communications rather than direct assault.[15] In the Russo-Japanese War, the Japanese won at Liao Yang and Mukden with turning maneuvers that exerted pressure on the Russian flanks.[16]

Balck and Colin, like others, attribute the power of the modern defense to three main factors: 1) the firepower of modern weapons; 2) the use of trenches to protect

---

[10] They were not alone in this belief. The elder Molke once wrote, "...little success can be expected from a mere frontal attack, but very likely a great deal of loss. We must therefore turn towards the flank of the enemy's position." As quoted in von Caemmerer, The Development of Strategical Science During the 19th Century, (London: Hugh Rees, Ltd, 1905), p. 82. Count von Schlieffen, as is widely known, disdained the frontal assault and based the Schlieffen Plan on a massive flank attack. He once wrote, "flank attack is the essence of the whole history of war." As quoted by Gerhard Ritter, The Schlieffen Plan: Critique of a Myth, (Westport, Ct.: Greenwood Press, 1958), p. 50.

[11] Balck, p. 229.

[12] Indeed, although Colin admits that frontal attacks do succeed, Colin's examination of history had convinced him that "such a success ... Is very rare, so rare that a general would be mad to seek deliberately for victory through a frontal attack. " See Colin, pp. 70-71. As noted earlier, however, Colin simultaneously believed in the moral power of the frontal attack.

[13] Ibid., p. 41.

[14] Ibid., p. 61-62.

[15] Ibid., p. 62. See also Howard, p. 46.

[16] Ibid., p. 63. See also B. H. Liddell Hart, "The Ratio of Troops to Space", Military Review, Vol.XL, April 1960, p. 4.

defending infantry; 3) the use of barbed wire to hinder attacking infantry.[17] When a defensive position combined these three advantages with those traditionally given to the defender (e.g., superior knowledge of the terrain), it became almost impossible for an attacker to take that position from the front. Of these factors, they judge firepower to be the most important.[18] If the firepower of the defense were greater than some minimum, then the attacker would not be able to launch a successful frontal assault. Consequently, they conclude that there is a minimum number of men necessary to achieve this level of defensive firepower.

Moreover, both Balck and Colin maintain that this minimum number declined during the nineteenth century. Colin bases his conclusion on evidence from the Boer War and the Russo-Japanese War of 1905. In the Boer War, the British failed to penetrate fronts that were defended by a force to space ratio of one man per two or three meters.[19] Colin writes that "the Boers astonished the world" with these results.[20] During the Russo-Japanese War, the Japanese front at the battle of Liao Yang and the Russian front at the battle of Mukden were each held with force-to-space ratios of 4 men/meter.[21] In comparison, a force-to-space ratio of 4 men/meter was approximately half of the density normal to the American Civil War and the Franco-Prussian War of 1870.[22] Balck, in a table entitled "Influence of Various Rifles on the Density of Battalions," documents the drop in force-to-space ratios over the course of the entire 19th century. For example, he estimates that 11,000 muzzle loading rifles were used per kilometer at Waterloo, while

---

[17] See Colin, op. cit., pp. 61-63 and p. 158; Balck, op. cit., pp. 227-229 and p. 241.

[18] Balck writes, "The infantry combat is decided by the combined action of long firing lines." Only those forces that could shoot at an enemy could contribute to the outcome of a battle. See Ibid., p. 233.

[19] Colin, op. cit., p. 61. Colin used the following figures for his calculations. At Modder River, the Boers held 7 kilometers with 3000 men; at Magersfontein, 10 kilometers with 5000 men; at Colenso, 12 kilometers with 4000-5000 men. These figures give force-to-space ratios of .43 men/meter, .5 men/meter, and .38 men/meter respectively.

[20] Ibid., p. 61.

[21] For the battle of Liao Yang, see Ibid., p. 147-148. For the battle of Mukden, see Ibid., p. 157.

[22] According to Balck, the French force-to-space ratio at the battle of Gravelotte in 1870 was approximately 7200 rifles per kilometer, or roughly 7 men per meter. See Balck, op. cit., p. 240. According to Liddell Hart, during the first part of the Civil War, a defensive force-to-space ratio of 12,000 men/mile, or approximately 7500 men/kilometer, was considered normal. This ratio was also used in the Franco-Prussian war. See Liddell Hart, op. cit., p. 4. Others estimated the drop in force-to-space requirements to be even more severe: "if forty years ago we counted ten men per pace of front, we could do to-day with three men or less per metre." See Von Bernhardi, op. cit., p. 156. Furthermore, as late as 1893, General Lewal, French commander of 17th Army Corps, estimated that an army of 60,000 men could not safely hold a front of more than sixty kilometers, a force-to-space ratio of 10 men per meter. See Von Caemmerer, op. cit., p. 87.

the Russians used 4000 magazine rifles per kilometer to hold the line at the battle of Mukden.[23]

Both men conclude that the reason for this decline in battlefield densities lay with new technology. As suggested by the title given to his table of historical force-to-space ratios, Balck argues that the advent of modern rifles had been the primary driver behind the decline in the defender force-to-space ratio.[24] Colin likewise attributes declining densities to the use of modern weapons, and adds that increased use of trenches and wire also played a factor in the decline of minimum force-to-space ratios.[25]

1. Colin: Discontinuous Fronts

Colin, unlike Balck, stresses the need for continuous fronts on the defense. A capable attacker would choose to attack the flanks created by the gaps and weak spots in a discontinuous line, and could thereby win a victory that could not have been achieved with a frontal attack.[26] In effect, every discontinuity in the line offers an attacker two additional flanks that could be exploited.[27] Colin concludes that a continuous line, as measured in men per meter, is the first requirement of a successful defense.[28]

Colin supports this conclusion by reference to the battle of Sha-ho during the Russo-Japanese War. The Russians occupied a series of strong but discontinuous positions, and the battle was reduced to a number of small fights for these fortified localities. The Japanese, by dint of superior morale and training in small unit tactics, defeated each position in detail and successfully pierced the Russian front. In the other battles of the Russo-Japanese War, the Russian fronts were more evenly held, and the battles were decided upon the extreme right or left.[29]

---

[23] Balck, op. cit., p. 240.

[24] Ibid., pp. 227 and 240.

[25] Colin, op. cit., pp. 61-63 and p. 158.

[26] Ibid., p. 157.

[27] By "discontinuous line", Colin refers either to the defensive tactic of relying upon a series of fortified positions or to economies of force that result in a line of alternately weakly and strongly defended zones. See Ibid., pp. 152-154.

[28] Ibid., p. 157.

[29] Ibid., pp. 151-154. The force-to-space ratio at Sha-Ho was approximately 4 men/meter. The force-to-space ratio at the battle of Mukden was also 4 men/meter. Colin attributed the success of the defensive line at Mukden to the even distribution of troop strength and the uninterrupted trench line, which combined to make the line continuous. See Ibid., p. 157.

Colin was not the only military expert to link defender force-to-space ratios and the probability of a successful penetration of the defensive line. Count Alfred von Schlieffen, Chief of the German General Staff from 1891 to 1905, raised such a possibility in his analysis of the dynamics of a German advance through Belgium:

> The aim must always be to envelop the enemy's left flank with a strong right wing... Should the enemy try to prevent the envelopment by extending his left wing, he will so weaken his front line that a break-through may well become possible. [30]

Schlieffen supported his reasoning by calculating the force-to-space ratio, in men per meter, of the French defensive line:

> The three positions Verdun-Dunkirk, Verdun-La Fere-Abbeville and Verdun-La Fere-Paris have approximately the same length. With the addition of the line Belfort-Verdun, each is about 500 kilometers long. [Assuming a French field army of one million.] This gives an average of two field army infantrymen per metre. [After Territorial troops have been added to the forces of the field army] ... An average of four men per metre may be reckoned upon. If the German right wing has been made strong, it may be hoped that the position Verdun-Dunkirk can be penetrated.[31]

Thus the French, to cover the entire length of their line, would be forced to defend with a force-to-space ratio of four men per meter. In Schlieffen's opinion, this force-to-space ratio was low enough that a successful penetration might be possible. Schlieffen thus clearly linked defensive force-to-space ratios to the chance of breakthrough by the attacker.

In fact, such calculations were widespread among European military planners on the eve of the war. Ironically, the French themselves used an argument similar to that of Schlieffen to conclude that if the Germans attempted such an attack, the *German* force-to-space ratio would fall to two men per meter. Against such a density, the French believed that their own attack would cut the German forces in half. On this basis, the French discounted the possibility of a German invasion through Belgium.[32]

---

[30] Ritter, op. cit., p. 157.

[31] Ibid., p. 157.

[32] See Barbara Tuchman, The Guns of August (New York: Macmillan Publishing Co., 1962), p. 28.

2. Balck: Important Variables

Balck, by contrast with Colin, stresses the variety of battlefield considerations that affect the density required for a defense to resist frontal attack. Balck defines the fundamental question as:

> with how weak a force may I occupy the position and still obtain the frontal strength described in the regulations, and how strong can I make the general reserve so as to bring about a decision?[33]

In reply, Balck concentrates on four factors that he argues determine the answer: 1) the weapons available to the defender; 2) the terrain of the defensive position; 3) the minimum size of the reserve force; 4) the mission the defender is tasked to accomplish.

The characteristics of the weapons available to the defender are the major factor in determining the number of men required along the actual firing line itself. Balck ascribes the decisive role in the outcome of a battle to fire effect:

> In deciding how many men are required to occupy or attack a position, the principal point to be considered is the effect of fire. The modern long range magazine rifle will, no doubt, enable us to defend a position with a smaller force than was possible in the past with the older less improved weapons.[34]

Given the technological capabilities available to the defender, the actual number of men required to hold a front would depend upon features specific to that piece of ground. Thus, Balck declares that "the number of troops which will be required to hold a given piece of ground must be determined separately in each case" [35]. Balck lists three terrain-related factors that bear on a calculation of the minimum force-to-space ratio:

> 1) the strength, natural or artificial, of the position; 2) obstacles in its front; 3) salient angles which can be easily enveloped.[36]

Physical features that determine the "strength" of a position are "favorable terrain, cover, and intrenchments (sic)". [37] Balck does not elaborate upon the "obstacles" to which he refers. Presumably, these include major terrain features, such as rivers, mountains, and

---

[33] Balck, op. cit., p. 411.

[34] Ibid., p. 227.

[35] Ibid., p. 231.

[36] Ibid., p. 232.

[37] Ibid., p.241.

swamps, and perhaps also, more localized factors such as are subsumed under the "strength" of a position. The importance of "salient angles" lies in the vulnerability of the flank of a defense (as discussed earlier).

Although Balck maintains that reserves contribute nothing to the immediate outcome of the battle, he concludes that they are necessary because they allow a commander "to exercise a constant influence on the course of action" throughout the battle.[38] However, Balck holds that the size of a reserve force is inversely related to that of the forward force:

> Distribution in depth and frontage are interdependent; the greater the frontage, the less the distribution in depth, and visa versa. [39]

As the battle is actually decided by the forces on the firing line, Balck argues that only the minimum number of forces should be retained in reserve.[40] The question is how to determine that minimum.[41] Balck argues that the size of the minimum reserve force reflected the amount of risk associated with a particular defense. The larger the risk, Balck reasons, the larger the size of the minimum reserve force.[42]

Balck identifies three more specific variables as particularly important to this determination of the required reserve size: 1) the defensibility of the flanks (a function largely of terrain); 2) the need for reinforcements along the front line (considered by Balck to be a function of mission); 3) the need for a counter-attack force (a function of

---

[38] See Ibid., pp. 222-225, and p. 232. Reserves serve four important functions: 1) to reinforce the front line; 2) to guard the flanks of the line; 3) to launch a counter-attack; 4) to guard against unexpected contingencies. Forward forces, once engaged in a battle, can only rarely be re-directed toward another task. In performing their missions, reserves offer the commander a flexibility that cannot be provided by forces along the firing line.

[39] Ibid., p. 225.

[40] Ibid., pp. 222-225. The retention of the bare minimum reserve force would be the optimal solution. Balck, however, argues that the danger of retaining too small a reserve force is greater than that of retaining too large a reserve force.

[41] Balck writes, "The result of the combat depends in many cases upon a happy answer to this question." Ibid., p. 225. Note that Balck's method of determining how to distribute one's forces implies that one should calculate the minimum required reserve force, retain that force (or perhaps a little more, to be conservative) in reserve, and deploy the remainder of one's forces on the firing line.

[42] Ibid., p. 229.

mission).[43] The first and third of these factors are the most important.[44] Balck concludes that the combined effect of these three variables produce a non-linear relationship between the size of the minimum reserve force and the level of aggregation of a unit .[45]

C. MID 20th CENTURY: BASIL LIDDELL HART

Captain Basil Henry Liddell Hart has come to be regarded as one of the twentieth century's most influential military theorists. Over a fifty year career, Liddell Hart created a body of work that touches upon almost every aspect of warfare. Of particular

---

[43] Ibid., pp. 229, 232, 233. The mission a unit is tasked to perform affects both the minimum size of the reserve and the minimum size of the forward force. Moreover, a change in mission may disproportionately change the size of one force with respect to the other. Two examples given by Balck serve to define the boundaries of this variable's effect. These examples give illustrative force sizes (expressed in number of rifles) for a delaying defense and a deliberate defense. Each defense is conducted along a one kilometer front on unspecified (and, one assumes, neutral) terrain:

Delaying Defense

front line: 300 rifles
reserve forces: 520 rifles
  flank defense: 60 rifles/flank
  reinforcements: 0 rifles
  counter-attack forces: 400 rifles
minimum force-to-space ratio: 82 rifles/meter

Deliberate Defense

front line: 1000 rifles
reserve forces: 1400 rifles
  flank defense: 200 rifles/flank
  reinforcements: 200 rifles
  counter-attack forces: 800 rifles
minimum force-to-space ratio: 2.4 rifles/meter

Two features of these examples are noteworthy. First, although the size of the reserve force is much larger in the case of the deliberate defense, the proportional size of the reserve force is slightly smaller (reflecting the greater risk associated with a delaying defense). Second, to Balck, the chief distinction between the goals of the two defenses is the length of time each was designed to resist the enemy. The delaying defense seeks to slow the overall advance of the enemy. The deliberate defense is designed to last indefinitely.

[44] The importance of counter-attack is shown by Balck's examples. The importance of the flanks is based in Balck's belief that the flanks represent the most vulnerable part of a defense. In particular, Balck believed that a unit whose flanks were secure (either due to the presence of friendly neighbors or because of terrain obstructions) could hold a significantly larger frontage than a unit required to guard its own flanks. See Ibid., pp. 230 and 235.

importance for this study, Liddell Hart developed a set of insights on the impact of force to space ratios that have had an important effect on the modern debate.[46]

Liddell Hart's first comments on force-to-space ratios appeared in late 1938 or early 1939.[47] It was not until 1960, however, that he codified his thoughts on force density in a chapter of his book, Deterrent or Defense: A Fresh Look at the West's Military Position.[48] Liddell Hart then published an expanded version of this chapter in the U.S. Army's professional journal Military Review under the title "The Ratio of Troops to Space."[49] It is on this last articulation of Liddell Hart's views on the question that we will focus here.

In "The Ratio of Troops to Space," Liddell Hart examines the force-to-force and force to space ratios characteristic of the battles of the two world wars. Based upon these studies, he concludes that force-to-*force* ratios can not accurately predict the outcome of a battle. Even apparently overwhelming force superiorities for the attacker do not suffice to achieve a quick victory. The answer, Liddell Hart argues, lies in the power of a dense

---

[45] Ibid., pp. 230, 234-235, 241. Balck concludes that "the [maximum] frontage [for a given unit] does not increase in proportion to the size of the force." See Ibid., p. 241. Balck attributes this non-linear relationship to two causes. First, flank protection is provided by the aggregate unit. The flanks of a battalion deployed along the front line are secured by its neighbors. However, the corps to which that battalion belongs must secure the flanks of the corps by deploying several battalions to guard each flank. Second, Balck holds that aggregate units, in addition to establishing their own reserves, often supply the reserve needs of their constituent parts.

[46] Although Liddell Hart was far and away the most significant commentator on force-to-space ratio issues during the mid 20th century, he was not the only writer to touch on the subject. In "Corps Defense on a Broad Front," for example, Lt. Colonel E. M. Postlethwait analyzed the problem of a corps forced to defend a front roughly twice as broad as that prescribed by doctrinal norms. He concluded that an attacker would penetrate such a defense more easily. However, a successful defense would be possible if the defense took steps to limit the significance of this penetration. He emphasizes the need for high mobility, strong intelligence-gathering, a strong central reserve (to launch counter-attacks or block penetrations), and all-around defense of positions. See Lt. Colonel E.M. Postlethwait, "Corps Defense on a Broad Front", Military Review, July, 1949, pp. 50 - 56.

[47] See John J. Mearsheimer, Liddell Hart and the Weight of History, (Ithaca and London: Cornell University Press, 1988), pp. 120-121, and note 75 in particular.

[48] Basil H., Liddell Hart, Deterrent or Defense: A Fresh Look at the West's Military Position (New York: Praeger, 1960), pp. 97-109.

[49] Basil H., Liddell Hart, "The Ratio of Troops to Space." Military Review, Vol. XL, April, 1960.

defense, as expressed in its force-to-space ratio.[50] If a defender can maintain a force-to-space ratio above some minimum value, an attacker must wage a campaign of slow, costly advances, even if the attacker significantly outnumbers the defender. Conversely, if the defender fails to achieve this minimum force-to-space ratio, the attacker is granted the opportunity to achieve a decisive, speedy victory, even if the force-to-*force* ratio does not significantly favor the attacker.

To Liddell Hart, the two world wars demonstrated that a defender possessed a considerable tactical advantage over an attacker. During 1915, the Germans, defending on the Western front with a force-to-space ratio of 6000 men/mile, held the line against Allied assaults striking with a local force superiority of approximately five to one.[51] During the fall of 1918, the Allies, with an overall force superiority of three to one, did not succeed in breaking through the network of German defensive lines.[52] During World War II in North Africa, the British at Tobruk and the Germans at El Alamein successfully defended against attackers with force superiorities of three to one or greater.[53] In the last stages of the war, German defenders held out for significant periods of time against superiorities of five, or even seven to one.[54] Liddell Hart concludes that a defender, even if heavily outnumbered, can maintain a viable defensive line for a considerable time, but only if some minimum force-to-space ratio is maintained along the front. At force-to-space ratios above this minimum, Liddell Hart, like Balck and Colin, argues that the strength of a continuous defensive line is too great for an attacker to overcome.

Liddell Hart defines this minimum force-to-space ratio to be "the extent of space that troops armed with modern weapons, other than nuclear ones, can cover with a closely

---

[50] It is important to note that Liddell Hart's standard for a successful defense did not actually require the defender to win a battle or campaign. For example, the Germans (the defenders in the majority of Liddell Hart's cases) lost the majority of the battles cited by Liddell Hart. Liddell Hart's standard for a successful defense is based upon the length of time that defense held its line against a superior attacker. Thus, the German defense in Normandy in 1944, faced with a much superior attacker, did not lose quickly and is adjudged to be successful. In May, 1940, the Allies lost quickly to an inferior attacker, and thus failed to defend successfully.

[51] Ibid., pp. 4-5.

[52] Ibid., p. 6. During the fall offensives, the Allies achieved local force superiorities of up to sixteen to one. Despite the overwhelming success of the steamroller approach used by the Allies, at the time of the Armistice, Liddell Hart notes that Allied commanders felt that "Germany is not broken in a military sense."

[53] Ibid., p. 8.

[54] Ibid., pp. 8-9. In Normandy, it was estimated that the Allies needed a superiority of five to one if an attack were to succeed. On the Eastern front, the Russians achieved local superiorities of seven (or more) to one.

interwoven network of fire".[55] Like Colin and Balck, after examination of the defensive force-to-space ratios of major historical battles, Liddell Hart concludes that the defender's force-to-space ratio had dropped significantly over time.[56] Liddell Hart also identifies a number of independent variables important to the issue of defender force-to-space ratios. These include terrain features, technological characteristics of weapons, methods of force employment (most notably, the relative tempo of the attack and the defense, the defender's distribution of forces, and defender's employment of his reserve), and troop quality. In particular, Liddell Hart stresses the importance of (what he called) the "time factor" in providing a link between tactical force-to-space ratios and theater-level outcomes.

Liddell Hart draws a distinction between the tactical force-to-space minimum and the strategic minimum. The tactical minimum applied to the battlefield and is the force-to-space ratio that determines the outcome of breakthrough battles. However, a defender, in all likelihood, would not have to guard all sectors of his front at once. Therefore, the strategic minimum represented the smallest in-theater force that could defend the length of that theater. Liddell Hart's examination of historical battles leads him to conclude that the strategic minimum had historically been much lower than the tactical minimum, but that the distinction was lost in World War I. Armies chose to cover the worst-case scenario by maintaining the tactical minimum along the entirety of a front.[57] Liddell Hart argues that the key factor in determining the difference between the two minima is in the

---

[55] Ibid., p. 11.

[56] Liddell Hart based his conclusion upon the following estimates of defensive force-to-space ratios: For the Napoleonic Wars, 20,000 men/mile; for the American Civil War, early period, 12,000 men/mile; for the Civil War, late period, 5,000 men/mile; for the Boer War, 600-800 men/mile; for the Russo-Japanese War, 8000 men/mile; in World War I, the average force-to-space ratio (including reserves) in active sectors was 1 division per 3 miles (6,000 men per mile); the force-to-space ratio of divisions actually deployed along the front line (not including reserves) was one division per 4-6 miles (4,500 - 3,000 men/mile); for World War II, the Battle of France, 3.5 miles/division; for World War II, El Alamein, 8 miles/division; for WWII, Normandy, 10 miles/division; for WWII, Eastern Front, 20 miles/division. See Ibid., pp. 4-9.

It must be noted that Liddell Hart follows no consistent pattern in calculating these force-to-space ratios. Some of these force-to-space ratios are calculated at the theater level (Russo-Japanese War, Battle of France, Normandy, WW II Eastern Front) while others are calculated at the tactical level (Napoleonic Wars, El Alamein). Some included reserves (the first WW I number) while others do not (the second WW I number).

[57] Ibid., pp. 12-13. Liddell Hart calculates strategic force-to-space ratios by dividing the length of a defender's theater front by the forces in that theater. For example, Napoleon held the frontier of France with a strategic force-to-space ratio of 180 men per mile. The Confederate Army defended its frontier with a strategic force-to-space ratio of 250 men per mile. Both of these figures are an order of magnitude smaller than the tactical force-to-space ratios calculated by Liddell Hart (and noted earlier). In contrast, in World War I, the strategic force-to-space ratio for the Germans in 1915 was one division per five miles. The tactical force-to-space ratio was one division per three miles. See Ibid., p. 4.

time required for a defender to react to an attacker's decisive thrusts.[58] If this period of time were too short for an attacker to damage the defense irreparably, the defender could afford to defend his frontier more thinly. The key variables in determining this period of time are the relative mobility of the attacker and the defender, and the defender's ability to apprehend correctly the attacker's routes of advance.[59]

Using an argument similar to the one used to tie tactical and strategic force-to-space minima, Liddell Hart addresses the strategic consequences of breakthrough in terms of the force to space ratio.[60] The issue is the interaction between the tactical force-to-space ratio at the point of attack and the relative operational mobility of the defender's reserve. If the local force-to-space ratio drops low enough, the attacker can achieve a breakthrough. The significance of this breakthrough depends primarily on the relative operational mobility of the defender's reserves (a product primarily of the tempo of the attack, the reserve's rate of movement and the defender's initial force distribution); these reserve forces are the only forces that the defense can readily use to stop the breakthrough once it occurs. Therefore, the combination of high tactical force-to-space ratios and high relative reserve mobility results in a situation highly favorable to the defender. Breakthroughs are very unlikely to occur. If they do, they can contained quickly. This environment characterized World War I.[61] The combination of low force-to-space ratios and low defender mobility relative to the speed of the attacker results in a situation highly advantageous for the attacker. Breakthroughs are likely to occur, and if they do occur, they often result in a quick, strategic victory for the attacker. This condition characterized France in May, 1940.

Given this interaction, one can define the minimum tactical force-to-space ratio as the lowest force-to-space ratio that gives the defender time to concentrate reserves and deny the attacker a strategic breakthrough. At force-to-space ratios above this minimum, the attacker may still achieve localized, tactical breakthroughs. However, these

---

[58] This is the "time factor" referred to earlier.

[59] Ibid., p. 13.

[60] The discussion above is drawn primarily from Liddell Hart's analysis of the fall of France in May, 1940. To Liddell Hart, this battle represented the archetypical instance of low tactical force-to-space ratio and low reserve mobility. The Allies deployed their forces poorly. They maintained a small, immobile central reserve, and left the defense of the Ardennes "perilously weak." Once the Germans (who relied upon a high tempo of operations) penetrated this region of low defensive force-to-space ratio, the Allies possessed no forces that could be used to stop the advance. See Ibid., pp. 6-7.

[61] If the rate of advance of the attacker is very slow (as was the case in World War I), the mobility of the defender's reserves could be relatively low (in modern terms), and high *relative* mobility could still be the result. See Ibid., p. 6.

breakthroughs will not lead to strategic gains for the attacker.[62] In addition, Liddell Hart identifies a number of other independent variables that have a secondary influence on this interaction. These are: 1) the defender's ability to apprehend correctly the attacker's routes of advance (as defined above); 2) the degree of surprise initially gained by the attack; 3) the defender's reaction time; 4) the attacker's reaction time.

Liddell Hart explains the twentieth century's drop in force-to-space ratios by the simultaneous improvement in the technologies employed by armies.[63] Modern equipment lowers the force-to-space ratio by enabling a smaller force to create that "closely interwoven network of fire" over a larger area. Like Balck and Colin, Liddell Hart argues that greater weapon range and more powerful weapons allow units to concentrate more firepower at greater distances.[64] Liddell Hart also attributes this decline to the advantages conveyed by the combination of improved tactical troop mobility (due to the advent of ground and air mechanization) and vastly superior surveillance and communication methods (caused largely by the widespread use of radio and the reconnaissance potential of aircraft). Enemy advances can be detected along a much wider front, and troops can move more quickly and from more distant locations to confront the enemy.[65] Liddell Hart concludes that, if anything, the doctrinal force-to-space guidelines (of his era) were too high. If the Boers in 1900, relying primarily upon magazine rifles for firepower and exclusively upon horses for mobility, could hold their line while heavily outnumbered with a force-to-space ratio of 800 men/mile, then, Liddell Hart argues, modern armies should be able to deal with an even lower force-to-space ratio.[66]

Two further variables figure prominently in his discussion of the tactical force-to-space minimum: terrain, and troop quality. Terrain that acts to impede the mobility of an attacker lowers the minimum force-to-space ratio. Examples of such terrain include natural obstacles such as mountains, rivers, and forest, and artificial obstacles such as

---

[62] Consider, for example, the German breakthrough attempts in March and April of 1918. The Germans achieved tactical breakthroughs, but the relative operational mobility of the Allied reserve forces was high, and these breakthroughs were stopped before strategic gains were achieved. See Ibid., p. 5.

[63] Ibid., pp. 11-12.

[64] Ibid., pp. 9-11.

[65] Ibid., pp. 9-11.

[66] Ibid., p. 11.

fortifications and minefields.[67] Higher troop quality acts to lower the force-to-space minimum by rendering all military operations more efficient, allowing fewer troops to produce an equivalent net output of military strength.[68]

In calculating historical force-to-space ratios (both tactical and strategic), Liddell Hart expresses his results in units of men/mile or miles/division. He himself admits that these units, while easily used in calculations, did not express other factors important to a calculation of military strength: "equipment, terrain, area, communications, training, tactical methods, leadership, and morale".[69]

D. THE CONTEMPORARY DEBATE: THE CONVENTIONAL BALANCE, ARMS CONTROL, AND FORCE-TO-SPACE RATIOS

Recent discussion of the impact of defender force-to-space ratios has been driven by the policy debates over the conventional balance and conventional arms control. While these debates have produced a large literature, much of which at least touches upon

---

[67] For a discussion of the obstacles that might be encountered by a Soviet invasion of West Germany, see Ibid., pp. 13-14. Mines were used by Rommel at El Alamein, and fortifications by all sides in World War I. Note that terrain must be occupied to be of benefit to the defender. The French in May, 1940 considered the Ardennes to be impassable, left it largely unoccupied, and paid for their mistake. See Ibid., p. 7.

[68] Ibid., p. 10. In Liddell Hart's opinion, the Germans in World War II were able to operate effectively with a lower force-to-space ratio than NATO could safely risk. He attributes this to NATO's "mixture of nationalities, different training systems, and other handicaps." In particular, Liddell Hart disparaged the possibility that short-term conscripts could defend with a low force-to-space ratio.

[69] Liddell Hart admitted that proper consideration of certain of these variables was difficult to incorporate into numerical calculations. However, he encouraged readers to try.

the question of force density minima,[70] the central theoretical contributions of this work can be appreciated by a detailed review of the work of three authors in particular: Professor John J. Mearsheimer of the University of Chicago, the RAND conventional arms control policy group, and Professor Archer Jones of North Dakota State University.

Before turning to a more detailed review of these sources, however, there are several broad conclusions common to the modern literature as a whole. First, it is widely held that force density is an important contributor to battlefield outcomes, and that lower densities tend to favor attackers over defenders.[71] Moreover, the relationship between

---

[70] For individual references, see the following: William Mako, U.S. Ground Forces and the Defense of Central Europe, (Washington, D.C.: The Brookings Institution, 1983); Stephen Flanagan and Andrew Hamilton, "Arms Control and Stability in Europe: Reductions Are Not Enough", Survival, Vol. XXX, no. 5 (September/October, 1988), pp. 448-463; Barry Posen, "Measuring the European Conventional Balance" in Steven Miller, ed., Conventional Forces and American Defense Policy (Princeton, N.J.: Princeton University Press), pp. 79-120; Phillip A. Karber, "In Defense of Forward Defense", Armed Forces Journal International, May 1984, pp. 27-50; Jack Snyder, "Limiting Offensive Conventional Forces", International Security, Vol. 12, No. 4 (Spring, 1988), pp. 48-77; Leonard Sullivan, Security and Stability in Conventional Forces: Differing Perceptions of the Balance, (Washington, D.C.: The Atlantic Council, 1988); Lynn Whittaker, Report of a Conference on "U.S. Conventional Forces: Current Commitments, Future Needs", (Cambridge, Ma.: Center for Science and International Affairs No. 88-2, 1988); Klaus Wittman, Adelphi Paper 239: Challenges of Conventional Arms Control, (London: Brassey's, 1989); James Moore, "The Estimation of Optimum Force Size and Force Reduction Potential in Conventional Arms Reduction Negotiations", Arms Control, Volume 9, No. 2 (September, 1988), pp. 116-133; Federal Republic of Germany Ministry of Defense, Operational Minima and Force Buildup of the Warsaw Pact and NATO; John Galvin, "Some Thoughts on Conventional Arms Control", Survival, April, 1989, pp. 99-107; "The General's New Order for Europe: Andrew Goodpaster and the Eisenhower Vision", Arms Control Today, May, 1989, pp. 3-8; Andrew Goodpaster, Gorbachev and the Future of East-West Security: A Response for the Mid-Term, (Washington, D.C.: The Atlantic Council, 1989); John J. Mearsheimer, "Assessing the Conventional Balance: The 3:1 Rule and its Critics", International Security, Vol. 13, No.4 (Spring, 1989), pp. 54-89; John J. Mearsheimer, Conventional Deterrence, (Ithaca, N.Y.: Cornell University Press, 1983); John J. Mearsheimer, "Numbers, Strategy, and the European Balance", International Security, Volume 12, No. 4 (Spring, 1988), pp. 174-185; John J. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe", International Security, Volume 7, No. 1 (Summer, 1982), pp. 3-39; James A. Thompson and Nanette C. Gantz, Conventional Arms Control Revisited: Objectives in the New Phase, (Santa Monica, Ca.: RAND, 1987); Bruce W. Bennett, et. al., Main Theater Warfare Modeling in the RAND Strategy Assessment System (3.0). (Santa Monica, Ca.: RAND, 1988); Paul K. Davis, et. al., Variables Affecting the Central Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels, (Santa Monica, Ca.: RAND, September, 1989).

[71] For example, Phillip Karber writes, "... The outcome of combat is less dependent on the ratio of forces than it is to the density of the defense - that is, the ratio of force to space." Karber, op. cit., p. 36. William Mako identifies the minimum force-to-space ratio as one of three factors that should be used to determine the ground forces required to deal with a given contingency (the other two being the force-to-force ratio and the minimum size of the operational reserve). He goes on to quote former Secretary of Defense Schlesinger as stating, "Whether we are talking about Central Europe or Korea, if a front is to be held along its length with a reasonable degree of confidence, there must be a minimum density of manpower along that front, with no significant gaps between units." Mako, op. cit., p. 35-36. Jack Snyder, writing on a potential arms control agreement, offers a more picturesque metaphor describing the same phenomenon:

force to space ratios and combat outcomes is widely held to be discontinuous—meaning that a minimum or threshold force to space ratio exists at which the nature of the campaign changes character and the relative fortunes of defenders and attackers change dramatically.[72] Above the minimum, combat is dominated by attrition effects and the pace of battle is slow.[73] Below the minimum, the outcome is decided by a faster-moving

---

> A defender's force needs depend as much on the width of the front to be defended as on the size of the attacking force. At Thermopylae, a few good Greeks held off the Persian multitudes by clogging the narrow pass so that the Persians could attack them with only a few men at a time. (Snyder, op. cit., p. 66.)

This consensus view has not been unchallenged, however. See, for example, Joshua Epstein, "The 3:1 Rule, Adaptive Modeling, and the Future of Security Studies", International Security, Vol. 13, No. 4 (Spring, 1989), esp. pp. 123-4, footnote 84.

[72] For example, NATO's Supreme Allied Commander General John Galvin writes:

> Furthermore, force-to-space ratios and the dictates of terrain mean there are certain force levels below which the West cannot reduce...Western defensive doctrine allows for divisional frontages of 40-60 km in the defence... *Reductions in this force could not cut very deeply before the considerations of terrain and force-to-space ratios would become a dominant factor* [emphasis added]. In order to cover the front and carry out the defensive mission, Allied Command Europe would be forced to conduct more mobile operations... (Galvin, op. cit., p. 103.)

As discussed in greater detail below, the RAND combat model CAMPAIGN bases its determination of the success of an attempted breakthrough on a discontinuous relationship between a defender's force to space ratio and the character of the ensuing battle. See Bennett, et. al., op. cit., p. 55. Jack Snyder argues that if a defender had just enough forces to establish a defensive line with this minimum force-to-space ratio, it would be in an attacker's interest to trade a relatively large reduction of his own forces for even a small reduction of the defender's forces. See Snyder, op. cit., p. 67.

[73] William Mako, drawing explicitly upon Liddell Hart's ideas on force to space ratios, writes:

> ...once the ratio of defenders to the amount of space being defended reaches some threshold, an attacking force—despite a numerical superiority of 5:1 or even higher—can well find it impossible to move rapidly through that defense. The defense may eventually be worn down and pushed back; but, in the meantime, it can perhaps hold up the offense long enough for defender reinforcements to arrive or for counterattacks to get under way. (Mako, op. cit., p. 36.)

war of maneuver.[74] The minimum density at which this transition occurs is generally held to lie in the range of 25-30 kilometers per division.[75]

A second point of broad consensus is that force-to-space ratios primarily affect breakthrough battles, in which an attacker attempts to penetrate a prepared defense.[76] In particular, the force-to-space ratio plays a role in determining the length of time required for an attacker to penetrate a defense.[77]

Finally, many authors agree that the relationship between force density and combat outcomes is affected by variables such as terrain, and some have suggested that it is also affected by variations in weapon types.[78] Moreover, military authors in particular have emphasized the importance of force employment for outcomes on a low density battlefield. It is often asserted that lower force to space ratios would force NATO member states to change their national military doctrines; it is sometimes argued that to do so would force NATO to modify or abandon its declaratory strategy of forward defense.[79]

---

[74] William Mako, for example, writes:

> For planning purposes, analysts in the Office of the Secretary of Defense posit twenty-five kilometers as a standard frontage for a U.S. Heavy division in Central Europe. Divisions might, if necessary, hold wider fronts...At some point, the defense would be stretched so thin that a breakthrough could occur. (Ibid., pp. 35-36.)

[75] Analysts who cite a number in this range include William Mako, Barry Posen, James Thomson and Nanette Gantz, Leonard Sullivan, John Mearsheimer, Stephen Flanagan and Andrew Hamilton, Klaus Wittman, the German Ministry of Defense, and the analysts who produced the RAND "Operational Minimum" study. For specific references, see the bibliographic footnote given earlier in this section.

[76] See, for example, Mearsheimer, "Numbers", op. cit., pp. 177-179. See also Posen, op. cit., pp. 106-110.

[77] For example, Phillip Karber writes

> Thus, for example, a NATO mechanized brigade screening a 30-kilometer sector of open terrain would be hard pressed to delay a Soviet force at 2:1 odds for more than several hours, whereas the same unit in well-prepared positions on a 10-kilometer constricted movement corridor could stop a force at 6:1 odds for several days. (See Karber, op. cit., p. 36.)

[78] Mako, for example, estimates that an infantry division could not defend as wide a front as an armored division. See Mako, op. cit., p. 37. In general, the literature argues that the minimum force-to-space ratio is lower in forests or urban areas. Mako, however, maintains that the minimum is actually higher in those situations. See Ibid.; Sullivan, op. cit., p. 39, Karber, op. cit., pp. 33-36, and especially the second RAND study.

[79] SACEUR General John Galvin and retired SACEUR General Andrew Goodpaster, in particular, have emphasized the importance of force employment in this regard. As noted above, General Galvin argues that a reduction in NATO's force level, thereby lowering NATO's force to space ratio, could compel a change in NATO's plan of operations. General Goodpaster argues that:

1. John Mearsheimer

Drawing upon Liddell Hart's work, John J. Mearsheimer has developed a description of the process by which force-to-space ratios influence combat outcomes. In particular, Mearsheimer concludes that the defender's force-to-space ratio plays a key role in determining an attacker's ability to initiate blitzkrieg warfare. In turn, the ability to wage a blitzkrieg plays a central role in Mearsheimer's assessments of the conventional balance in Central Europe.[80]

Mearsheimer's description is based upon two key insights: 1) terrain constraints limit an attacker's ability to concentrate; 2) the defender possesses a considerable tactical advantage in combats determined by attrition. Starting from these insights, Mearsheimer concludes that force-to-space ratios play the deciding factor in determining the character of a battle: whether the conflict will be determined by a slow campaign of attrition or by a fast campaign of blitzkrieg and maneuver. If the defender's force-to-space ratio is above some minimum value, the attacker has effectively no chance of achieving the quick breakthrough necessary to initiate a blitzkrieg.[81] Below this value, a high probability exists that an attacker will penetrate the defensive line quickly and will thereby be able to attempt a blitzkrieg.[82]

> If we go down the road [to alliance parity at force levels of 50% current NATO levels], it will be necessary to devise new doctrine and new plans and new modes of employment and command...

See Goodpaster interview, p. 5. (It is important to note, however, that General Goodpaster is more optimistic regarding NATO's prospects in the event of such change than are most analysts in the current debate. In the above reference, for example, Goodpaster advocates that the reductions described be seriously considered by the Alliance. See also Goodpaster, Gorbachev and the Future of East-West Security, op. cit.). For other observations by U.S. Military officers on the importance of force employment on a low density battlefield, see, for example, Bentson, op. cit., and Postlethwait, op. cit.

[80] Mearsheimer first wrote on the subject in his article "Why the Soviets Can't Win Quickly in Central Europe." Subsequently, force-to-space ratios played a part in Mearsheimer's theory of conventional deterrence, as presented in Conventional Deterrence. The most complete description of force-to-space constraints appears in "Numbers, Strategy, and the Conventional Balance."

[81] It is important to note that Mearsheimer uses force-to-space ratios to determine the immediate likelihood of a successful breakthrough battle. If the breakthrough does succeed, the eventual outcome of the battle will be decided by other factors. The most notable of these are the defender's ability to stop the breakthrough and the defender's strategic depth. The first is largely a matter of the strength of a defender's mobile reserves and the competence shown by commanders in employing those reserves. The second is a matter of geography. For a more complete discussion of blitzkrieg warfare, see Mearsheimer, Conventional Deterrence, op. cit., pp. 35-52.

[82] In general, the defender does not automatically lose if the minimum force-to-space ratio cannot be maintained or achieved. In effect, the attacker achieves a breakthrough and must then successfully exploit this advantage. As indicated above, measures can be taken to neutralize the penetration. However, in the specific case of NATO's forward defense of Central Europe, a breakthrough of any sort could well be disastrous. See Mearsheimer, Conventional Deterrence, op. cit., pp. 48-49.

Mearsheimer's description is based on the interaction of attacker concentration and defender counter-concentration. In general, a defender possesses a tactical advantage over the attacker.[83] The attacker seeks to overcome this advantage by concentrating at the point of attack and achieving local force superiorities. Mearsheimer argues that an attacker's ability to concentrate is limited by the local transportation network and the local geography. Any forces in excess of this limit would have to be placed in echelons behind the forward forces, and would have no immediate impact on a battle (the "crossing the T" phenomenon).[84] This constraint on concentration proves crucial because it severely limits the potential advantage of large attacker force superiorities. If the defender can maintain some minimum force-to-space ratio along the front line, the attacker cannot achieve the immediate force superiorities required to break through the defensive line.[85]

The minimum force-to-space ratio required to hold a particular length of front is determined by the particular terrain features of the front and by the capabilities of the specific units tasked to hold that terrain (e.g., what force imbalance are they able to handle). Obstacles, such as "rivers, mountains, forests, swamps, urban sprawl, man-made defensive positions," limit an attacker's avenues of advance, thereby effectively reducing the length of front to be defended.[86]

2. RAND

The RAND Corporation has released a number of studies in which force-to-space ratios play a part in determining the stability of the conventional balance in Central

---

[83] The advantages traditionally assigned to the defense include the ability to fight from protected positions, better firing positions, familiarity with the terrain, and the ability to use obstacles (minefields, tank traps, ditches) to slow or kill the attacker. See Mearsheimer,"Assessing the Conventional Balance ...", op. cit., p. 57. Mearsheimer expresses these advantages in the 3:1 rule of thumb. This rule states that an attacker requires a force ratio of 3:1 at the point of attack to achieve a breakthrough. See Mearsheimer, "Numbers ...", op. cit., p. 177. At the present time, the 3:1 rule is the subject of intense debate in the community. See International Security, Vol. 13, No. 4 (Spring, 1989), and especially the articles of Epstein and Mearsheimer.

[84] Mearsheimer, "Numbers," op. cit., p. 178. These "stacked up" forces would be crucial in the latter stages of a breakthrough attempt. Presumably, however, by the time these forces came into play, the defenders would also have called upon reserves. The breakthrough attempt would therefore become a battle of attrition dependent upon rates of reinforcement. See Ibid., p. 179.

[85] The minimum force-to-space ratio would depend upon the force inferiority one is prepared to accept. Using the 3:1 rule, and setting the maximum attacker concentration at 1 brigade per 7 kilometers, Mearsheimer feels that a good rule of thumb for the minimum force-to-space ratio is 1 brigade per 15 kilometers. See Ibid., p. 178, note 7.

[86] Ibid., p. 177. See also the discussion of a Soviet advance through Central Europe in Mearsheimer, "Why the Soviets...", op. cit., pp. 23-29.

Europe. Two of these will be reviewed in detail: Conventional Arms Control Revisited: Objectives in the New Phase, and Variables Affecting the Central Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels. Both studies are based on the output from the CAMPAIGN model, a theater-level combat simulation developed at Rand.

CAMPAIGN is the land warfare component of the Rand Strategy Assessment System.[87] CAMPAIGN may be used either interactively or as a closed simulation (no human interaction). When used non-interactively, all force-employment decision-making is performed by a number of "expert systems" decision models.[88] In the interactive mode, these decision models make only some of the force employment decisions.[89] In either mode, a series of combat decision rules are important to the estimation of outcomes.

In structure, CAMPAIGN is a modified "piston model" of ground combat.[90] Attackers move along axes of advance that run across a terrain grid of Central Europe.[91] Combat interactions are assessed between attackers and defenders on the same axis; thus each axis is a semi-independent "piston" with a separate FLOT (forward line of troops) which advances at a rate determined by the balance of forces on the given piston. Models of this structure cannot explicitly model operational maneuver (e.g., breakthroughs, encirclements, flank attacks, and so on) as such. Instead, CAMPAIGN represents the effects of maneuver indirectly by modifying the nominal strengths and advance rates of the forces on a given axis to reflect the impact of the maneuver phenomena presumed to take place within the piston.[92] The nature of the maneuver presumed to occur is estimated by the model as a function of the local combat environment—and in particular, as a function of the force-to-space ratio.

---

[87] For a more complete description of the RSAS environment of which CAMPAIGN is a part, see Bennett, et al.

[88] Ibid., p. 4.

[89] The simulation automatically makes many of the brigade-level decisions. For example, on a given attack axis, a decision model known as the "axis commander" determines which forces (of those allocated to that axis of advance) fight in the first echelon and which are held in reserve. As attrition occurs, the "axis commander" rotates brigades between reserve and forward deployments. Ibid., p. 53.

[90] Ibid., p. 11.

[91] A given cell of the grid might be 60 km deep by 80 km wide. Ibid., p. 10.

[92] Ibid., p. 47.

To do this, CAMPAIGN breaks a battle into cyclical phases: preparation, assault, breakthrough or stalemate, and (given breakthrough) exploitation and pursuit.[93] Attrition and attacker rate of advance are calculated differently for each phase of battle.[94] The highest defender attrition and attacker rates of advance occur during the breakthrough and exploitation stages.[95] Therefore, the attacker's immediate tactical objective is to achieve and exploit a breakthrough along the major axes of advance. The defender, in turn seeks to prevent a breakthrough. Failing that, the defender attempts to end the exploitation phase of battle by re-establishing a stable defensive line.[96] If a new defensive line is established, the cycle of battle begins anew.

Within CAMPAIGN, the defender's force-to-space ratio plays the key role in governing the transition between the assault phase and the breakthrough (or stalemate) phase. If the force-to-space ratio falls below some threshold value, a breakthrough is assumed to occur, and the battle automatically shifts to the breakthrough and exploitation phase.[97] If the force-to-space ratio remains above this breakthrough threshold for the entire combat, stalemate eventually ensues.[98] The exploitation phase ends if and when the defender force-to-space ratio rises above the breakthrough threshold. This threshold force-to-space ratio is determined independently of the attacking force, and varies with the type of terrain and with the degree of terrain preparation.[99]

CAMPAIGN defines four ranges of the defender's force-to-space ratio: maximum density, hold density, minimum density, and breakthrough density. These threshold

---

[93] Ibid., pp. 54-55.

[94] Ibid., p. 47.

[95] For defender attrition rates, see Ibid., p. 59. For rates of advance, see Ibid., p. 62.

[96] *In general, a new defensive line is established by the deployment of operational reserves.*

[97] Ibid., p. 55. The force-to-space ratio is measured in divisions along the FLOT (forward line of troops) divided by the length of FLOT held. Divisional reserves are counted (i.e., if one brigade of the division were held in the rear as a reserve, this brigade would still be counted) but divisions stationed behind the FLOT as operational reserves would not be counted. RAND measures divisional strength in equivalent divisions (ED), in which one U.S. Armored division is equal to one ED. An illustrative breakthrough minimum for Central Europe might be 60 km/ED. See Ibid., p. 18.

[98] Within CAMPAIGN, the force-to-force ratio of attacker to defender must exceed some minimum in order for an attacker to continue to attack. See Ibid., p. 44. If the force-to-space ratio remains above the minimum for a sufficient period of time, attrition will this force-to-force ratio to fall below the minimum required to attack, and stalemate will result. See Ibid., p. 55.

values are user-inputs.[100] The breakthrough threshold is explained above. The other thresholds reflect varying levels of defensive effectiveness.[101] In addition to the role described above, these thresholds are secondary factors influencing an attacker's rate of advance.[102] Above the hold density, the attacker's nominal rate of advance is multiplied by a factor less than one. Between the hold and the minimum densities, the nominal rate of advance is not changed by force-to-space ratio considerations. Below the minimum density, the nominal rate of advance is multiplied by a factor greater than one. Below the breakthrough level, the breakthrough velocity is used.

In Conventional Arms Control Revisited: Objectives in the New Phase, James A. Thomson and Nanette C. Gantz of RAND use this model to analyze the impact on NATO stability of an arms control agreement authorizing significant force cuts.[103] They paint a stark picture of the task confronting NATO arms control negotiators:

> To have some effect, proposed reductions should be substantially asymmetric, probably at least at a Pact/NATO ratio of 5:1 in overall combat capability...Smaller asymmetries are likely to leave the balance more precarious from NATO's standpoint.[104]

This conclusion is based on the role of low force to space ratios in promoting breakthrough conditions, as described above. Because NATO fares badly once local force densities fall too low, NATO in effect requires a certain minimum force to man the front thickly enough to keep the battle from shifting into breakthrough conditions. The rule of thumb used in the CAMPAIGN runs performed for the Thomson and Gantz study held that the minimum force-to-space ratio required to maintain a coherent defense is 25

---

[99] Ibid., p. 55. Terrain modification of the minimum force-to-space ratio is based upon the degree to which the terrain facilitates defense (presumably by providing cover, advantageous lines of sight, and other benefits). See Ibid., p. 56. Four levels of defense preparation are possible: Hasty, deliberate, prepared, fortified. Hasty defenses require the highest force density to prevent breakthrough, fortified defenses the lowest density. See Ibid., p.49.

[100] Sample values for these thresholds are given for Central Europe. These are, respectively, 15 kilometers per standard division, 25 kilometers/ std. Div, 40 km/std. Div., and 60 km/ std. Div. Ibid., p. 18. The nature of a standard division is determined using the WEI/WUV III methodology. See Ibid., p. 23.

[101] The maximum density represents the strongest possible front-line defense. The minimum density represents the weakest front-line defense that has any chance of withstanding an attack. Ibid., p. 18.

[102] Ibid., pp. 61-63. The primary factors determining the attacker's rate of advance are the type of engagement (determined by the phase of the battle and the degree of defensive preparation) and the force ratio. This nominal rate of advance is then modified by terrain considerations and by the defender's force-to-space ratio.

[103] Thomson and Gantz, op. cit., p.6

[104] Ibid., p. 16.

kilometers per equivalent division (ED).[105] Similarly, the Pact requires at least a certain forward force density away from the point of attack. Forces in excess of these minima constitute operational reserves, to be fed forward to replace losses at the point of attack.[106] The ratio of these residuals (i.e., the balance between the two sides' reserves) then largely determines the outcome of the battle of attrition that results. This ratio of excess forces is thus, as Thompson and Gantz put it, "a good measure of the balance [in Central Europe]."[107] At present, the ratio of excess forces predicted by Thomson and Gantz is "roughly 4:1" in favor of the Warsaw Pact.[108] If arms control reductions are to improve the conventional balance in Central Europe, the agreements must therefore authorize asymmetric cuts of (at least) approximately 4:1 (Pact forces for NATO forces). This critical ratio of acceptable reductions is thus driven largely by the force to space minimum required to forestall the transition to breakthrough. The lower the force to space minimum, the larger the excess forces on each side; given that NATO has fewer forces than the Pact, lowering a common density floor will reduce the ratio of Pact to NATO reserves, thus lowering the degree of asymmetry required for a given arms control agreement to be in NATO's interest.[109]

In Variables Affecting the Central Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels, Paul K. Davis, Robert D. Howe, Richard L. Kugler, and William G. Wild reach two conclusions. First, they estimate the minimum force (including reserves) required to defend the Central Region to be 27 equivalent divisions (ED).[110] Second, at force levels below this "operational minimum," they suggest that the stability of NATO's defense is problematic, but that a successful defense

---

[105] Ibid., p. 12. This corresponds to the "hold density" used in CAMPAIGN. See Bennet, et al, op. cit., p. 18.

[106] Ibid., p. 12.

[107] Ibid., p. 12.

[108] Ibid., p. 12.

[109] Consider the following example. At this force level, NATO would require thirty divisions to hold a 750 kilometer inter-German border. If NATO is assumed to have 40 total divisions, NATO's excess forces are 10 divisions. If the Pact is assumed to have a total force of 80 divisions, the ratio of excess forces is 5:1 (if the Pact is assumed to match NATO's distribution along the line). Now consider an instance where the minimum force to space ratio was 40 km/division. NATO would now have roughly 19 divisions on the front line, and 21 divisions of excess forces. The ratio of excess forces would now be 61:21, or slightly less than 3:1.

[110] Davis, et. al., op. cit., p. 31. The front is question is 750 kilometers in length. It includes the Danish sector but excludes the Austrian border. This gives a force-to-space ratio of roughly 28 kilometers per ED.

is nevertheless not, in theory, impossible.[111] In the process they provide additional insight into two issues in particular: 1) the effect of terrain on minimum force-to-space ratios; 2) the potential effect of defender force employment at low force-to-space ratios.

The authors divide Central European terrain into four classes: closed, rough, mixed, and open.[112] Associated with each terrain type is a rule of thumb that determines the minimum tactical force-to-space ratio with which that terrain type may be defended.[113] These rules of thumb reflect the extent to which the terrain impedes and/or channels the advance of the attacker, and the extent to which the terrain differentially aids the defender (e.g., by providing cover, advantageous lines of sight, ambush positions, or easily fortified rear positions such as rivers).[114] Of the four terrain classes described, closed is assessed to be the most defense-favorable, followed in order by rough, mixed, and open.[115]

Starting from these terrain-based force-to-space minima, the authors calculate NATO's minimum force size to be 27 EDs.[116] To arrive at this figure, the authors estimate that 60 percent of NATO's 750 km front is "militarily usable"; the remainder of

---

[111] Ibid., p. Vii.

[112] "Closed" terrain refers to densely urbanized areas or very mountainous areas. " Rough" terrain refers to forested areas penetrated by only narrow roads or to moderately mountainous or hilly areas penetrated by few roads. "Mixed" terrain refers to a mix of rough and open terrain. "Open" terrain refers to relatively flat and easily trafficable regions such as plains or croplands.

[113] These rules of thumb are: for closed terrain, 60 km/ED; for rough terrain, 40 km/ED; for mixed terrain, 30 km/ED; for open terrain, 20 km/ED. Ibid., p. 24. These values represent authors' estimates (see Ibid., pp. 15-6, and p. 25).

[114] In addition, the following non-natural characteristics of a piece of terrain figure prominently in the determination of the minimum tactical densities appropriate to that piece of terrain: The available road network, the depth of the terrain, the position of the defense line with respect to the terrain (forward, within, rear), the defender's mission, possible defender force tailoring (specialized light infantry for forests, mountains, or urban areas), the presence of sensors and obstacles, attacker force tailoring (specialized assault troops for forests, mountains, or urban areas). Ibid., p. 22. In general, terrain that impedes the attacker's movement and/or channels the attacker into certain approach routes also differentially aids the defender in other ways. However, while assuming as a point of departure that more difficult terrain benefits the defender, the authors take care to note that this is not always the case. The same terrain characteristics that prohibit or limit major avenues of advance can also shorten lines of sight, to the detriment of the defender who must be able to detect the advance of the attacker. Dense forests characterized by many minor routes of advance provide the best example of this. A defense in a dense forest may require less firepower than a defense on an open plain but more manpower (i.e., by dividing the defense's firepower into many discrete units, the front may be held more completely, if less strongly). See Ibid., pp. 23-4; c.f. Mako, op. cit., p.37.

[115] Davis, et. al., op. cit., p. 24.

the front need not be held with significant strength. The nature of this usable terrain is open-to-mixed. The minimum force-to-space ratio for open-to-mixed terrain is 25 km/ED. Thus, NATO requires at least 18 EDs forward. In addition, the authors estimate that NATO's operational reserve should be 50 percent of NATO's forward force. Therefore, NATO's operational minimum is 27 EDs overall.[117]

With respect to defenses conducted below the operational minimum, the authors conclude that success is problematic, but possible in certain situations. The authors used CAMPAIGN to examine the outcome of a war in Central Europe in which NATO's force level was 18 EDs. Under conservative assumptions in which NATO attempts to hold at the inter-German border, CAMPAIGN predicted that the Warsaw Pact would penetrate to a depth of about 300 kilometers in approximately three weeks. (In comparison, under the same conditions, a defense conducted with 27 EDs would halt the Pact at the Weser River).[118] The authors then tested the effectiveness of the defense in two instances in which the original conditions were changed. In the first variation, NATO was assumed to fall back to the Weser in order to buy time for counterconcentration, and construction of fortifications on the Weser. In the second variation, NATO was assumed to begin moving reserve forces to the main-attack sector prior to the actual assault.[119] In both instances, NATO conducted a successful defense with 18 EDs.[120]

3. Archer Jones

Archer Jones, in <u>The Art of War in the Western World</u>, treats force-to-space ratio issues more broadly than the majority of the modern literature. Of particular interest for

---

[116] Ibid., p. 11 and p. 31. This value represents the density of forces per kilometer of frontage as viewed from the army group perspective, and is assessed to be "the minimum operational density of forces needed to maintain a cohesive line and hold ground for perhaps a week or so with reasonable confidence."

[117] Ibid., p. 31. The authors stress that this figure represents a "strawman" estimate. For the authors' comments on their choice of 50 percent as the fraction of reserve forces, see Ibid., p 17.

[118] Ibid., p. 54. For the specific assumptions, see Ibid., p. 52. In general, the authors assumed parity, a NATO attempt to hold at the inter-German border, no differential concentration on the part of NATO, sluggish defender command and control, and no prepared defenses on D-day.

[119] Presumably, this advantage reflects enhanced sensor and intelligence capabilities on the part of NATO.

[120] Ibid., p. 55. The authors did not model counter-attack, so the possible outcome of an attempt to regain the status quo ante is undetermined. However, the Pact's unsuccessful attempts to penetrate the Weser line do cause the force ratio to shift decisively in NATO's favor (Pact/NATO = .76). See Ibid., p. 55-56.

our purposes is Jones' argument that the impact of density on combat outcomes is a function of *both* the attacker and the defender force-to-space ratios.[121]

In eras when the force-to-space ratios of both the attacker and the defender have been high, Jones concludes that warfare is characterized by slow battles of attrition. Flanks are extended to become unassailable. Attacks are instead made frontally against continuous fronts that cannot be penetrated quickly. Overall defensive effectiveness in an era characterized by high force-to-space ratios is therefore high.[122]

Moreover, Jones argues that defensive effectiveness has also been quite high in eras when the force-to-space ratios of both the attacker and the defender have been low. This is because the attacker is simply not numerous enough to conquer and hold territory.[123] Even if the attacker wins a decisive victory over the defender's armed forces, he is not powerful enough to occupy the country.[124] Moreover, such an attacker may not even be able to defend his own borders in the event that the defender opts to counter-invade rather than to intercept the attack.[125] Jones, in effect, predicts that, if attacker and defender force levels are comparable and vary together, an attacker does best at mid-range force-to-space ratios, and worst at high and low force-to-space ratios.[126]

Jones limits these conclusions, however, to cases in which both attackers and defenders employ weapons with broadly similar capabilities. In particular, Jones stresses

---

[121] The following discussion is based on Archer Jones, The Art of War in the Western World, (Chicago: University of Illinois Press, 1987), esp. pp. 666-667. In addition, Jones (like Balck, Colin, and Liddell Hart) concludes that the prepared defender is effectively invulnerable to direct frontal assault. Jones argues that this conclusion is a constant throughout the history of war. Ibid., p. 691.

[122] Jones cites World War I as the best example of this phenomenon. See Ibid., pp. 440-441.

[123] When the military strength of both attacker and defender is roughly equivalent, the attacker, in an era of low force to space ratios, suffers from an inability to force decisive battle upon an opponent. For example, see Jones' description of the Thirty Years War, especially pp. 227-232. However, even if an attacker's strength were such as to dominate the defender militarily, the attacker would not be able to hold the defender's territory. For example, this situation obtained during the French operations in Russia and Spain during the Napoleonic wars, and during the English operations in France during the Hundred Years War. See Ibid., pp. 168, 355, 366-67, 371, 440-41, and 666-67.

[124] For example, Hannibal's victory at Cannae did not allow him to end the war on satisfactory terms. See Ibid., pp. 65-69.

[125] Jones suggests that in situations where neither side has sufficient force to hold territory, the best option for a potential combatant will often be a "raiding strategy," in which the freedom of maneuver resulting from low force densities is used to pillage the opposing countryside while avoiding contact with the opposing army. See Ibid., pp. 666-667. This situation obtained during much of the Thirty Years War, when Gustavus, Tilly, and Wallenstein conducted many logistics raids, but fought relatively few frontal battles. See Ibid., pp. 223-243.

[126] Ibid., pp. 666-667.

the potential impact of differing weapon mobility for the attacker and the defender.[127] A highly mobile defender (e.g, tank forces) faced with a slow-moving attacker (e.g., foot infantry) may be able to defend his country successfully regardless of his force-to-space ratio.

E. CONCLUSIONS

The literature on force to space ratios is thus fairly extensive and extremely heterogeneous in nature. It has taken shape intermittently over a span of more than a century, and includes contributions from theorists, practitioners, civilian analysts and serving military officers. Interest in the issue was high in the decades prior to the First World War; fell in the postwar years; was revived by Basil Liddell Hart prior to the Second World War and again in the early 1960s; and fell again until rediscovered by John Mearsheimer in the early 1980s. The public debate over the conventional balance and especially conventional arms control has driven the issue of force density to an unusual salience in the last years of the decade, but it is thus not a new issue.

Notwithstanding the heterogeneity of this literature, several conclusions are nevertheless matters of broad consensus. First, it is generally believed that defenders require a minimum *force for defense of a fixed front.* Defenders without at least this minimum force density lose the advantages of position and concealment normally associated with the tactical defense, and thus risk breakthrough when confronted with a concentrated attack. Since attackers can only concentrate a finite force on a finite front, however, a more densely defended line can present an attacker with a difficult target, slowing offensive advance rates long enough for defenders to counter offensive concentration by moving reserves to the point of attack.

Second, while there is some disagreement among sources, most argue that this defensive minimum is largely independent of the size of the opposing force. Thus, small forces in large theaters are widely regarded as promoting the offensive—even if the two sides are of roughly equal strength.[128] As a consequence, it is widely argued in the modern literature that the minimum force to space ratio constitutes an effective floor for conventional arms control more or less regardless of Soviet concessions—or alternatively, that Soviet concessions must be extraordinarily asymmetric in nature for NATO to be able to accept significant reductions in its own forces.

---

[127] Ibid., pp. 666-667.

[128] For an exception, see Jones, op. cit., pp.666-667.

Finally, while different authors cite different lists of secondary (or intervening) effects that influence the minimum force to space ratio, some of these effects recur frequently enough to be treated as points of consensus. In particular, three classes of such effects can be identified: terrain, weapons technology, and force employment.

With respect to force employment, three particular aspects are most frequently addressed: the "tempo" of the campaign, the defender's capacity to counterattack a penetrating attacker, and the distribution of the defender's forces (both in depth, and most often, with respect to the maintenance of reserves).[129] As noted earlier, "tempo" determines the severity of a potential breakthrough. If the attacker is moving faster than the defender can react, then the defender requires a greater density of forces forward than if the defense can keep up with the attacker. The ability to counterattack determines the defender's ability to contain the attacker if and when he penetrates into the defense.[130] Thus, limited counter-attack capability implies a higher minimum force-to-space ratio. The third force employment variable, the distribution of forces, concerns the depth of the defense and the proportion of total force that is held in reserve. [131] Deep defenses deny the attacker the ability to penetrate the defense by winning a single battle, thereby slowing the attacker's rate of advance.[132] Deeper defenses imply sufficiency for lower force-to-space ratios. The size and deployment of reserve forces determine a defender's ability to counter-concentrate. Small, poorly placed reserve forces deny the defender any chance of counter-concentrating at the point of attack; this implies a need for higher force-to-space ratios.

Weapon technology is often described as a primary driver in observed changes in force to space minima over time. Indeed, the introduction of new, high firepower weaponry in the late nineteenth century was the initial inspiration for the first period of active consideration of density effects in the decades preceding World War I. Many

---

[129] The modern literature offers relatively few specific guidelines with respect to the interactions of force employment and force-to-space ratios. It therefore cannot be said that any true consensus has formed with respect to the specific nature of these interactions. Nonetheless, among those authors who comment upon force employment (e.g., Galvin, Goodpaster, the RAND analysts), there is relatively little disagreement.

[130] In May, 1940, the initial deployment of the Allies left them with no forces with which to counterattack the German spearhead. This nullified any chance the Allies might have had to stop the German advance short of the Channel. See Liddell Hart, op. cit., pp. 6-7.

[131] Depth as used here is not related to the size of the reserve. Therefore, one can increase the size of the reserve force without increasing the depth, and visa versa.

[132] See, for example, Mearsheimer, Conventional Deterrence, op. cit., pp. 49-50.

authors point to advancing weapons technology as a driver for a general reduction in density minima over the course of the twentieth century.[133]

Terrain effects are addressed in two broad categories: natural features and artificial terrain features and/or enhancements. Natural features include rivers, hills, mountains, plains, forests, and swamps. Artificial terrain features include roads, bridges, and urban areas. Man-made military enhancements include trenches, the use of wire, minefields, and other attempts at fortification. The literature stresses two aspects of terrain: trafficability and advantageous combat features (such as cover, good lines of sight, ambush locations, etc.). Higher trafficability (such as that provided by plains or extensive road networks) implies higher combat tempo and, therefore, a need for higher force-to-space ratios. Likewise, lower trafficability (characteristic of mountainous regions or mine-fields) implies a need for lower force-to-space ratios. Terrain characterized by many advantageous combat features (such as forests or fortified positions) allows the local defender to fight more effectively. This implies that lower force-to-space ratios would suffice.

Force to space ratios have thus been the subject of considerable if sporadic attention for an extended period of time, and this consideration has produced a substantial degree of consensus on certain broad aspects of the issue. Corporately, the resulting literature thus represents a major source of insight into the nature of the issue, and in particular, into the range of effects and considerations that must be taken into account when evaluating the consequences of a given force density.

Yet few writers on the subject have accorded it extensive or primary treatment; most often it is addressed instrumentally in the course of an investigation to which it relates, but is not the direct subject. This has contributed to an unstructured and largely non-cumulative literature. As a consequence, the insight provided by this body of thought, while rich, is neither systematic nor unambiguous. Terms are inconsistently defined; consequences are rarely specified in other than a very general manner; and

---

[133] Interestingly, however, the possibility that changes in weapon technology could reduce the floor on NATO force levels for arms control purposes has not been widely explored. For a more detailed treatment of this issue, see appendix C.

preconditions are often left unstated.[134] Author's estimates and rules of thumb are most often the underlying foundation upon which minimum force estimates are based. While these are not necessarily without merit—and are substantially superior to the absence of assessment—they are best regarded as expedients until a more rigorously based understanding can be developed.[135] In the meantime, results are best regarded as provisional hypotheses. Thus, while this body of existing thought is of considerable value as a point of departure, it does not as yet constitute a complete or wholly sufficient description of the impact of force to space ratios for the purposes of the defense planning community. For this, further exploration of these issues is required.

---

[134] The term "force to space ratio" itself has been variously defined as theater forces divided by length of theater front; forward forces divided by length of theater front; theater or forward forces divided by area of deployment; or forces at the point of attack divided by the attack frontage (see, e.g., Liddell Hart, op. cit., pp. 4-9; Davis, op. cit., pp. 11-13; Mearsheimer, "Numbers," op. cit., pp. 177-179). Alternatively, the consequences of a low force to space ratio have sometimes been described as offensive breakthrough even at "low" force to force ratios, and sometimes as a style of combat favoring attackers; the consequences of a force to space ratio above the minimum are sometimes described as "successful" defense, other times as delayed offensive breakthrough, and other times as a style of combat favoring defenders (see, e.g., Mako, op. cit., pp. 35-36; Karber, op. cit., p. 36; Mearsheimer, Conventional Deterrence, op. cit., pp. 48-49; Galvin, op. cit., p. 103; Liddell Hart, op. cit., pp. 6-7). Bounds of applicability are typically left unstated, but the nature of the treatment often either directly implies or at least indirectly suggests very broad applicability. Examples, for example, are frequently drawn from both world wars and widely divergent theaters of war within those conflicts (see esp. Liddell Hart, op. cit., pp. 4-13). It is thus difficult to infer whether these relationships are meant to be applicable, for example, to modern warfare in the Middle East, Africa, or the Indian subcontinent; whether some new weapon technology would undermine the relationship; or whether the nature of the relationship is specific to some particular military doctrine or national strategy.

[135] On the epistemology of judgmentally based estimates and rules of thumb, see John Mearsheimer, "Assessing the Conventional Balance", International Security, Volume 13, No. 4 (Spring, 1989).

Appendix B

SOVIET VIEWS ON THE EFFECTS OF FORCE TO SPACE RATIOS

Stuart Kaufman

A. INTRODUCTION

A complete review of existing thought on the relationship between force-to-space ratios and combat outcomes requires a careful survey of Soviet, as well as Western, theoretical literature. Soviet military writing is extensive and often very systematic in nature. Moreover, it has been suggested by a number of prominent Western analysts that the particular question of force-to-space ratios has played an important role in Soviet writing—particularly with respect to the problem of pre-emptive attack as a means of circumventing the mobilization of a dense NATO defense. The Soviet treatment of this issue, however, takes place in the context of a very different analytic tradition from that of the Western literature. As such, to do justice to Soviet thought on this issue requires separate treatment.

The purpose of this appendix thus is to summarize and evaluate the state of knowledge in the Soviet military literature with respect to the issue of defensive force density, and to assess the arguments by Western analysts with respect to the role of force to space ratios in that Soviet literature. The central conclusion of this review is that little evidence has been discovered to suggest that force to space ratios are a central or overriding concern for the Soviets. There are no publicly available Soviet studies specifically about force to space ratios, at least for the past two decades. Density is one among many variables in Soviet calculations, and it is not notably more prominent than others. Discussion of the issue is often brief or indirect. General Staff planning methodologies, for example, demonstrate a force to space effect, but this property is implicit in models that focus primarily on the "correlation of forces" at the point of attack.[1] The Soviets do believe that, other things being equal, lower densities favor attackers over defenders. But all other things are not necessarily equal, and the Soviets focus more on the other things—especially the force to force ratio, defensive depth, terrain preparation, weapons effectiveness, and troop quality. Moreover, when the Soviets address the issue of dense defenses, it is most often in the context of prescriptions for overcoming them (most often through the use of offensive artillery).

---

[1] See John Hines, "The Operational Calculations for Equal Security Under Army Control," Conference Paper presented at "International Symposium on Conventional Stability in Europe: Prerequisites and Analysis Requirements," 10-13 October 1989, German Armed Forces University, Munich; Col A. Gaponov, "Correlation of Forces and Rate of Advance," Voennaya Mysl', Nov. 10, 1971; Professor Fritz Stoeckli, "Soviet Operational Planning: Superiority Ratios and Casualties in Soviet Front and Army Operations,: RUSI Journal, Spring 1989.

An important caveat about this, or indeed any other issue of Soviet military thinking, is that Soviet military theory is in an extraordinary degree of flux in the late 1980s. A number of Soviet sources imply that a new period of military development began for the Soviet Union around 1985-86.[2] This seems to reflect a growing recognition in the Soviet military, spurred by the work of Marshal Ogarkov, the former Chief of the General Staff, of the implications of new advanced and precision-guided conventional weapons technologies.[3]

In May 1987 another major shift was announced: the Warsaw Pact's announcement of a move toward a defensive posture of "reasonable sufficiency," with a primary emphasis on avoiding war. Yet another major change was Soviet President Mikhail Gorbachev's December, 1988 announcement of deep unilateral cuts in the Soviet Ground and Air Forces. Thus at the end of the 1980s Soviet military thinking is being altered simultaneously by new fundamental assumptions about Soviet goals in a possible war, by radical cuts in Soviet military capabilities, and by a belief that recent changes in military technology are "revolutionary in their character."[4]

These changes have a number of implications for this review. The fact that Soviet ideas are changing means that older statements (even five years old or less) may not reflect current Soviet views. Further, since the changes are both so basic and still continuing, even the most recent Soviet statements are likely to reflect either tentative or unofficial views. Finally, the shift to a proclaimed defensive strategy means that views about Soviet offensive success are likely to be particularly subject to change. To the extent that Soviet statements represent general theories about military cause and effect, they are of value as theory even as circumstances (and doctrine) change. But given current changes, any projections about the future directions Soviet views may take must be highly tentative.

These caveats in mind, this appendix will continue with a discussion of the school of Western analysts mentioned above, followed by a discussion of other Western analysts

---

[2] The Soviet sources are Lt. Gen. V. Reznichenko, "Sovetskie Vooruzhennye Sily v poslevoennyy period," Kommunist Vooruzhennykh Sil No. 1, January 1988; the book he is reviewing is by A.A. Babakov and V.V. Larionov, Evolutsia voennogo iskusstva: etapy, tendentsy printsipy (Moscow: Voenizdat, 1987). C.F. Colonel David M. Glantz, "Soviet Operational Art and Tactics in an Era of Reform," Soviet Army Studies office, Fort Leavenworth, KS, April 1989, p. 25 footnote.

[3] See Colonel General Makhmut Akhmetovich Gareev, M.V. Frunze, Military Theorist (London: Pergamon-Brassey's, 1988), p. 216. C.F. Mary C. Fitzgerald, "Marshal Ogarkov on the Modern Theater Operation," Naval War College Review, Autumn 1986.

[4] Reznichenko, "Sovetskie Vooruzhenye Sily," op.cit., p. 87.

with differing views. With a consensus elusive, we will turn to the Soviet literature over the past two decades, ending this section with some inferences about current Soviet views. Finally, we will conclude with some summary observations about Soviet views on the issue of force-to-space ratios.

B. DENSITY: WESTERN OBSERVERS

Most Western discussions of Soviet military literature on defensive force- to-space ratios take place in the context of a larger debate about the strength and prospects of NATO's defenses. One theme in this larger debate has particular significance for this review, i.e., the argument that NATO's defense, if fully deployed, would be so dense as to be virtually impenetrable to Soviet attack. Because the Soviets fear attacking such a dense defense, they are argued to believe that the success of their attack would depend on pre-empting the deployment of the defense. In works in this vein, the discussions of density are usually subordinate to and closely bound up with this issue of pre-emption (and, in particular, how Soviet pre-emption is driven by NATO density) but if true, these arguments are clearly of importance for the general question of the relationship between force-to-space ratios and combat results. We shall therefore discuss the issues together.

The first major Western discussion of this type was in a article published by Phillip Karber in 1976. Karber wrote:

> Soviet writers have long held that density—the ratio of force to space—is the key variable influencing rate of advance. The greater the quantity of force in a given area the slower the movement, and conversely with low force-to-space ratios the battlefield becomes granular rather than linear, fluid instead of static. Instead of nuclear weapons to disperse the defence, the [Soviet] armour advocates call for pre-emptive manoeuvre—attacking the defence before it mobilizes and deploys a dense anti-tank defence. Soviet writers note that surprise attacks with conventional weapons offer the same opportunities as nuclear strikes for low force densities: fluidity of manoeuvre, and a high initial rate of advance.[5]

A similar line of argument was taken up a few years later by Christopher Donnelly, and still later by Charles Dick, both of the Soviet Studies Centre at the Royal Military Academy, Sandhurst. Donnelly writes:

> During conventional battle, however, due to the high density of anti-tank weapons in NATO armies, and due to the resilience of a strong defense to Soviet air or artillery bombardment, an attack on a prepared defensive position will normally require the troops to dismount and attack on foot, in

---

[5] Phillip A. Karber, "The Soviet Anti-Tank Debate," Survival, 1976, p. 111.

> close co-operation with accompanying armour and under cover of well coordinated artillery fire.[6]
>
> From the very serious attention given to NATO defense, and the great strength imputed to it, it is certain that in future wars NATO defensive strength could easily be sufficient to compel the Soviet Ground Forces to engage in a massive concentration of effort in order to maintain the tempo of their offensive, and hence win the war very quickly. Put another way, this Soviet realization of the potential strength of a modern prepared defense must make pre-emptive surprise attack ever more attractive to every Soviet soldier from corporal to Commander-in-Chief. To quote the most common 'cry' voiced by contributors to the last debate,"you forestall—you win all."[7]

Donnelly, like Karber, argues here only that pre-empting is attractive for the Soviets, not that it is necessary.[8]

Later, Donnelly and especially Dick go much further. They begin from the assumption that the Soviets believe it is essential that, should a war occur, they be able to win a quick victory. Donnelly argues, "If the war drags on, there is a high risk that it will develop into a catastrophic nuclear exchange and/or that the strains of war will destroy the Soviet bloc from the inside."[9] Dick adds that Soviet economic weakness relative to the West would by itself doom the Soviet Union to defeat in a long war.[10]

The need to win a war quickly creates the demand for a blitzkrieg strategy. Donnelly and Dick argue, however, that the Soviets are dubious that they could win a blitzkrieg war against a fully deployed NATO defense. The reason? The density of NATO's defense makes achievement of a breakthrough too difficult. Dick argues:

> Of crucial importance is the defender's ratio of force to space. A defense which is overstretched can be penetrated and destabilized by an enemy

---

[6] C.N. Donnelly, "Tactical Problems Facing the Soviet Army," International Defense Review (henceforth IDR), No. 9, 1978, p. 1406.

[7] Ibid, p. 1412.

[8] In another article written a year later, Donnelly explicitly argues that the Soviets see victory as possible without preemption: "The sheer weight of the massed tank formations plus artillery and air support would be such that they would overwhelm the defense on a narrow front." (C.N. Donnelly, "Soviet Tactics for Overcoming NATO Anti-Tank Defenses," IDR No. 7, 1979, p. 1106.) Some other analysts make the argument for Soviet emphasis on quick victory and surprise without mentioning NATO defense density as a significant factor in Soviet eyes (see James H. Hansen, "Countering NATO's New Weapons: Soviet Concepts for War in Europe," IDR No. 11, 1984, p. 1621. See also C.N. Donnelly, "The Soviet Operational Maneuver Group: A New Challenge for NATO," Military Review, March 1983, p. 44, for a similar argument.)

[9] Donnelly, Military Review, March 1983, p. 44.

[10] C.J. Dick, "Catching NATO Unawares: Soviet Army Surprise and Deception Techniques," IDR No. 1, 1986, p. 21. C.F. Vigor, op.cit.

> with mere parity in strength. But, with the advantage of the initiative, the enemy will be able to concentrate all of his efforts on the chosen sector of attack.
>
> The traditional massing of men and material on a breakthrough sector can no longer be accomplished. In the face of sophisticated modern reconnaissance means and nuclear weapons, such a course could be suicidal. Even if NATO was unlikely to use nuclear weapons (for instance, in the early days of the war), some contemporary weapons can be almost as devastating to massive troop concentrations. Given the effectiveness of modern defense, a traditional breakthrough operation must be uncertain of success—it is certainly incompatible with the demand for speed.
>
> The worst case facing the Soviets would be the need to conduct a conventional breakthrough against a prepared enemy under the nuclear shadow. While the Soviets would simply not initiate hostilities in this situation, no matter what diplomatic humiliation resulted, they could be forced to mount such an operation against enemy strategic reserves.[11]

For Dick, then, the Soviets would under no circumstances attack a fully deployed (and therefore dense) NATO defense because they have no confidence in their ability to break through it. Elsewhere, Dick explains that it is the quality of modern anti-tank weapons which makes a dense defense so formidable:

> NATO deploys anti-tank weapons with a range, accuracy and destructive power *unknown during the* Great Patriotic War. Modern tanks and ATGWs, when dug-in, are seen as having a 5:1 advantage over tanks advancing in the open. Combined with rapid minelaying techniques and a plethora of hand-held anti-tank weapons, they pose a formidable barrier to mechanized assault.
>
> It is no longer sufficient to mass more men and tanks per kilometre of front to overwhelm the enemy. There is simply not enough room to bring enough mass to bear. The contemporary problem is one of the ratio of force to space. To illustrate this point with an example, it is not possible to cram more than 40 tanks per kilometre into an assault wave. Faced with a density of 15 heavy and medium anti-armour systems per kilometre (a typical NATO deployment), the attacking tanks will, according to Soviet calculations, take 65% casualties—more than enough to bring the attack to a halt.[12]

Donnelly's discussion of the issue essentially agrees with Dick's:

> Current Soviet calculations show that NATO, when fully deployed, could establish a defense that will resist attempts at breakthrough with conventional weapons alone. This is because the growth in effectiveness of anti-tank weaponry (including tanks and mines) in recent years has

---

[11] Charles J. Dick, "Soviet Operational Concepts, Part I," Military Review, September 1985, pp. 35-36.

[12] C.J. Dick, "Soviet Operational Art, Part I: The Fruits of Experience," IDR No. 7, 1988, pp. 759-760.

> made density (the ratio of force to space) as important as the correlation of forces (ratio of force to force) in establishing a strong defense.[13]

Colonel David Glantz of the U.S. Army's Soviet Army Studies Office agrees that the Soviets view fully deployed enemy defenses to be extremely difficult to penetrate. He defines a prepared enemy defense to be one fully occupied by enemy troops (an unprepared defense, in contrast, is occupied only by an enemy covering force, and is therefore, by definition, less densely manned).[14] After outlining Soviet operational practice against prepared enemy defenses, Glantz concludes, "The Soviets strongly believe requisite offensive success can be achieved only against an unprepared or partially prepared defense."[15]

Density alone, however, is widely acknowledged to be but one of several important issues for the success of an attack. In particular, the role of supporting artillery in suppressing a dense defense is clearly pivotal. Dick and Donnelly have argued, however, that the Soviets do not think attacks on such defenses would be very successful even with good artillery preparation and support.

Donnelly makes the point rather less strongly. Examining a series of articles on the use of artillery in the Soviet Ground Forces journal, Donnelly lists a number of problems the Soviets had identified. Those relevant to breakthrough operations against dense defenses include: "the high proportion of moving armoured targets which are difficult to locate, hit and damage;...the high and also fluctuating speeds of the assault which the artillery is supporting; the extreme effectiveness of enemy counter-bombardment, especially with advanced projectiles; the difficulty of locating enemy batteries in defensive positions; the need to locate and destroy individual weapons capable of delivering nuclear warheads...."[16] Donnelly's conclusion is cautious: "it is clear that these doubts assail the Soviets themselves, but it is also clear that they have given much thought to their resolution."[17]

Dick is less equivocal. Using Soviet norms, he calculates that a breakthrough attempt on a front of 5-6 kilometers might demand 30,000-40,000 rounds for artillery preparation. He argues that this would require improbable logistical capabilities,

---

[13] C.N. Donnelly, "Future Soviet Military Policy, Part 2: Where and How," IDR No. 2, 1989, p. 141.

[14] David M. Glantz, "Operational Art and Tactics," Military Review, December 1988, pp. 37-38.

[15] Ibid, p. 39.

[16] C.N. Donnelly, "The Wind of Change in Soviet Artillery," IDR No. 6, 1982, p. 737.

[17] Ibid, p. 744.

excessively dangerous concentration of forces, and excessive expenditure of time. His conclusion: "Not surprisingly, the Soviets regard the time, ammunition and casualties involved in gnawing through prepared defenses as unacceptable, even were they to be more certain of success than they actually are. This tactical problem adds yet more weight to the conclusions which they have drawn from strategic considerations. The Soviets must fragment NATO's defenses before they have been formed...."[18]

The conclusion of Dick and Donnelly is that, to the Soviets, "some considerable degree of surprise is essential" for Soviet success—i.e., that the Soviets believe they can succeed only against a defense which has not yet achieved maximum density.[19] Dick lists a number of advantages the Soviets would hope to seize by surprising the enemy in this way: pre-empting NATO's reinforcement plans, acting as a force multiplier to avoid the need for breakthrough battles and ease the insertion of operational maneuver groups, lessening casualties and logistical burdens by guaranteeing mobile operations, and acting before Warsaw Pact allies can find a way to avoid involvement in battle.[20] These advantages of successful surprise are substantial, and Dick shows good evidence that the Soviets stress the importance of surprise for most of those reasons.[21]

Glantz makes a similar case. His point is this: "Drawing heavily from research done on the theme "the initial period of war" or, specifically, what a nation's army must do to win rapid victory or avoid precipitous defeat, the Soviets have concluded that the principal prerequisite for victory is the surprise conduct of rapid operations by forces concentrated well forward."[22] How is this achieved? The Soviets, he says, envision their "forces operating in a nuclear-scared configuration employing operational and tactical maneuver in the critical initial period of war to pre-empt and quickly overcome enemy defenses, to paralyze the enemy's ability to react and to win rapid victory within carefully defined political limits."[23] The purpose is precisely to overcome the dangers otherwise attendant in an attempt to assault a dense defense.[24]

---

[18] C.J. Dick, IDR, No. 7, 1988, p. 760.

[19] Donnelly, Military Review, March 1983, p. 44.

[20] C.J. Dick, IDR No. 1, 1986, pp. 21-22.

[21] For obvious reasons, one of the issues the Soviets do not discuss is the need to surprise their own allies.

[22] Glantz, Military Review, December 1988, p. 33.

[23] Ibid, p. 34.

[24] The assumption of preemption also explains why the Soviets expect a future war to be characterized by "great maneuverability...and the absence of a continuous front line in defense and attack." V.G. Reznichenko (ed.), Tactics, 1984, Translation Bureau, Secretary of State Department, Ottawa, Canada,

C. ALTERNATIVE WESTERN VIEWS

There is thus a substantial school of thought in the Western literature which argues that NATO's high force to space ratio (once fully deployed) compels the Soviets to pre-empt or fail. An alternative view is proposed by John Hines and Phillip A. Petersen, analysts for the Rand Corporation and the Defense Department, respectively. Hines and Petersen's view of Soviet thinking is that the Soviets see defensive density as a problem that can be handled, and tactical surprise (not pre-emption) as sufficient for offensive success. They do note the density problem, writing:

> At the bottom of the tactical hierarchy, the Soviet platoon leader must penetrate a NATO defense dense with antitank guided missiles (ATGMs). NATO ATGMs, from protected positions, can destroy all of his armored fighting vehicles and tanks while still at a distance of several kilometers and while the NATO weapons themselves are still well out of range of the tanks' main guns.[25]

Petersen and Hines argue, however, that such a defense can be neutralized. "Under the concept of 'integrated fire destruction'," they say, "the volume of fire to be used against sectors of NATO defenses selected for penetration is likely to be overwhelming." They estimate a Soviet first-echelon division on the main axis of attack would be supported by almost 300 artillery and mortar tubes, about 80 fighter-bombers, and probably 18 multiple-rocket launchers, a battalion of surface-to-surface missiles, and a regiment of attack helicopters as well. They conclude:

> The concentration of such a mass of fire throughout the enemy's tactical depths in a sector sometimes narrower than 6 kilometers is likely to achieve the 60 percent destruction norm required by Soviet doctrine. Because potential antitank guided missile (ATGM), tank, and artillery positions are to receive special attention within the sector, the combat effectiveness of unprotected ATGM and artillery crews is likely to be drastically reduced if not completely destroyed.[26]

Hines and Petersen are basing their argument on different kinds of calculations from the ones made by Donnelly and especially Dick. Hines and Petersen argue that Soviet norms involve concentrating so many artillery tubes and other firepower that the

---

(Washington: USGPO, 1987), p. 36, (emphasis in original). The Soviets, according to this view, plan to attack before the enemy can set up a continuous line.

[25] Phillip A. Petersen and John G. Hines, "The Conventional Offensive in Soviet Theater Strategy," Orbis, Fall 1983, p. 706.

[26] Ibid, pp. 714-715.

Soviet bombardment would *a fortiori* be successful. Dick counters that such a concentration would be dangerous—it would become too tempting a target for nuclear attack or attack with modern conventional artillery. He also doubts the Soviet ability to supply the ammunition needed to meet Soviet norms—about 100 rounds per tube just for the artillery preparation (not counting ammunition for counterbattery and support fire).

These considerations are mostly at the tactical level. In another work, Hines explains the criteria the Soviet General Staff has proposed for use in operational-level planning. These are embodied in a mathematical model in which force-to-space ratios per se do not appear—the model is mostly about the effect of the "correlation of forces," or force ratio between offense and defense, on the attacker's rate of advance.[27] Force densities do affect outcomes in this model, but implicitly rather than explicitly.

The equation, it is worth noting, can be easily derived by making a few simple assumptions. The model assumes that the defender deploys all his forces evenly across the front, and that the attacker deploys only a minimal covering force across most of the front, concentrating everything else in one narrow attack sector. One then derives the force ratio in the attacker sector (Ca) to be:

$$Ca \approx (Cg - Cmin)\,(Wg/Wa) + Cmin$$

where Cg is the overall force ratio, Cmin is the force ratio maintained by the covering forces, Wg is the width of the front, and Wa is the width of the attack sector.[28] One can see that in this equation the expected correlation of forces in the attack sector (i.e., the attacker's force advantage there) rises as that attack sector becomes a smaller proportion of the entire front (as Wg increases or Wa decreases). For example: a Soviet front (i.e., army group) can have a main sector of attack as narrow as 20 km.[29] If the overall frontage is only 200 km, the attacker can expect to face about 1/10 of the enemy force in the attack sector. But if the same forces face each other across 400 km of front, then the attacker can muster nearly the same attack force on that 20 km sector—but he will face only 1/20 of the enemy force. Thus, the Soviets conclude, "the opportunities for the intensive use of existing men and equipment are greater as the spatial scope increases."[30]

---

[27] Hines, "The Operational Calculations," Figure 1, p. 50.

[28] Ibid, pp. 41-42.

[29] Ibid, p. 43.

[30] Colonel A.G. Tarekhan (Ret.), quoted in ibid, p. 5.

Because this model applies specifically to front-level operations, it also implies that mutually smaller forces on a constant-size front line also make attack easier. If the number of Soviet formations in the theater diminishes (and the enemy shrinks proportionally), then each remaining front (army group) holds a longer frontage, so both sides have more room for maneuver and more opportunity for offensive concentration. Thus, the prospects for at least local offensive success improve as theater-wide force densities on both sides go down.

Regarding surprise, Petersen and Hines minimize not the importance of achieving it but rather the degree of surprise the Soviets need and expect:

> Strategic surprise is virtually ruled out since the Soviets anticipate that a major international crisis would precede hostilities. Attainment of even operational surprise would be a challenge because, as the Soviets have noted, modern technical reconnaissance would make it extremely difficult to deny information about one's activity to the enemy.
>
> Whatever the means, Warsaw Pact leaders would probably hope at most to deny NATO knowledge of precisely when rather than whether the Pact would attack. In addition, they might hope to mislead us as to the main directions of their advance. Given the character and timing of the strategic offensive operation itself, however, surprise measured in days or even hours might be adequate to give the Pact the initiative and momentum they would need for success.[31]

Petersen's overall assessment is that the Soviets are confident that they have the capability to launch a successful offensive should the necessity arise. According to one Soviet source he quotes, "a well-proportioned military organization has been created, permitting the accomplishment of missions of any scale under any conditions."[32] In short, then, Petersen and Hines disagree with half of the argument formulated by Donnelly and Dick. They agree that the Soviets desire a short war, and that overcoming dense NATO defenses presents them with a serious challenge. But they argue that artillery and air support would probably be sufficient to suppress those dense defenses, and therefore conclude that a degree of tactical surprise regarding the time and place of attack is sufficient, in Soviet eyes, to make success possible.

---

[31] Petersen and Hines, p. 734.

[32] Ibid, p. 733.

D. ASSESSMENT OF THE DEBATE

What are we to make of these arguments? One group of analysts argues that defensive density plays a central role in Soviet thinking, driving a Soviet conclusion that they must pre-empt or fail in case of war. John Hines, on the other hand, argues that while defensive density plays an important, though implicit, role in Soviet calculations, it is not a central issue. Hines and Petersen do not detect an exclusive Soviet reliance on pre-emption.

Unfortunately, a complete assessment is complicated by the scarcity of documentation in some arguments. Dick and Donnelly, for example, provide as formal documentation only one published source which is directly relevant to this issue.[33] The source is a 1978 article by a Lieutenant General of Artillery Yu. Kardashevskiy.[34] Kardashevskiy offers a table showing the probability of offensive success as a function of the density of attacking tanks and defending anti-tank weapons. The table shows that a 3:1 superiority of tanks results in a 50% probability of success for the attackers, and that at high densities of anti-tank weapons (more than 20 per kilometer), no density of attacking tanks has a significant chance of success.

The article, however, is entitled "Creatively Plan the Fire Destruction of Targets," and it merely uses the table as an object lesson in what happens to an attack which is launched without effective artillery support. Most of the article is concerned with how artillerists can calculate the amount of artillery support necessary to allow offensive success against various defenses. Kardashevskiy explains:

> In defense, hand-held and stand-supported antitank grenade launchers and ATGM weapons are employed for the struggle with attacking tanks...And as research shows, a breakthrough of [a defense so equipped] is possible only in the case of the reliable suppression of the whole defense and first of all the systems of anti-tank fire.
>
> Success also depends on the correlation of antitank means and attacking tanks, as is shown in table 1.
>
> The data presented in the table show that reliable success of the attack can be achieved with a correlation of 5:1 and more in favor of the attacker.

---

[33] While Professor Donnelly apparently has not cited a source for this claim in any published work, he has used the table from the Kardashevskiy article in briefings on this subject. For example, in an IDA briefing, Professor Donnelly has told this author that he has relied on the work done for the British Ministry of Defence by Professor H.F. Stoeckli for this argument. Unfortunately, the papers in question by Professor Stoeckli are not publicly available.

[34] Dick, IDR No. 7, 1988, p. 761. (Incidentally, Dick erroneously attributes the article to 1979; it was published in 1978.)

> From here the conclusion can follow that in the planning of fire it follows to proceed from the concrete level of fire destruction of enemy targets, the result of which will be the achievement of the necessary correlation of forces and means.[35]

In other words, the thrust of Kardashevskiy's argument is not that a dense defense is insuperable, but simply that it must be suppressed by an adequate degree of artillery bombardment. He suggests no reason to believe that such suppression cannot be achieved. His article, therefore, appears insufficient in itself to substantiate the claims of Donnelly and Dick.

E. THE SOVIET LITERATURE

The more important question than Kardashevskiy's view per se is the extent to which his is a typical Soviet view. Is there further support for Dick's position elsewhere in the Soviet military press? To answer these question it is necessary to do a broader survey of Soviet military literature over the past two decades. For most purposes, the single most authoritative source is the Soviet General Staff's journal Military Thought, which is available in the West for the years up to 1973. Similarly authoritative are the Military Encyclopedic Dictionary and the Soviet Military Encyclopedia, both edited by either the Minister of Defense or Chief of the General Staff at the time they were published.

The next most authoritative source on Soviet military art is Voenno-Istoricheskiy Zhurnal (Military-Historical Journal), the articles in which often explicitly claim to be directly applicable to current conditions. For discussions of Ground Forces tactics and some operational issues, the Ground Forces journal Voenniy Vestnik (Military Herald) is also useful. Zarubezhnoye Voennoye Obozrenie (Foreign Military Review) focuses on technical issues more than operational ones, but it is occasionally relevant. Soviet Military Review is the least authoritative source, as it is published in English and may therefore have some propaganda purposes; it has nevertheless been consulted to some degree.

Selected articles from these journals have been the main sources for this review. Military Thought was considered only from 1968-73 (earlier years would have been particularly obsolete, focused almost entirely on nuclear war; later years are not available). The other journals were sampled selectively for the years 1974-89. A few

---

[35] Lieutenant General of Artillery Yu. Kardashevskiy, "Tvorcheski Planirovat' Ognevoe Porazhenie Tseley," Voennyy Vestnik, July 1978, p. 64, Author's translation.

particularly authoritative books (such as the text called Tactics) or books written by authoritative authors (such as A.I. Radzievskiy, then-head of the Frunze Military Academy), have also been considered. Also considered have been a few particularly relevant, if not necessarily authoritative books.

One major finding of the survey of the Soviet literature is that Kardashevskiy's theme is a recurring, if not extremely common one.[36] High density of an enemy defense, especially in anti-tank weapons, is from time to time mentioned as a factor making a successful attack more challenging. However, in keeping with the offensive flavor of most Soviet military writing before the late 1980s, most of those discussions focused primarily on how to overcome such dense defenses. Recommended methods usually involve effective reconnaissance followed by strong artillery and air preparation and support, high tempos of advance following the artillery preparation,[37] and appropriate concentration of attacking forces to achieve superiority in the attack sector.

When Soviets are discussing their own defenses, density is sometimes mentioned, and occasionally discussed in some detail, as one of the factors making for a strong or stable defense.

However, defense density is almost always listed as one among many such factors, and there is rarely any implication that it is the most important factor. Furthermore, defensive density is rarely described as a factor important in and of itself; rather, it is generally considered to be a measure of the defender's success in massing forces in the decisive sector. In other words, density is usually considered primarily as an indicator of the force to force ratio, rather than as being important in itself.

The article on "Density of forces and means" in the authoritative Military Encyclopedic Dictionary, edited by then-Chief of the General Staff N.V. Ogarkov, illustrates this Soviet conception of density. The full definition is:

---

[36] The survey was far from comprehensive. We were interested primarily in the evidence for the contentions of Profs. Donnelly and Dick and Col. Glantz. We therefore focused our efforts particularly on material cited by those scholars (and by Mr. Petersen), as well as on sources suggested verbally by Prof. Donnelly, Col. Glantz, Prof. H.F. Stoeckli of the University of Neuchatel, and Dr. Eugene Rumor of The Rand Corporation.

[37] In principle, there is a tradeoff between the length of artillery bombardment and the rate of advance. The Soviets in effect finesse this issue by defining the advance as beginning only when the forces leave their final jumping-off point—just before the artillery preparation has ceased.

The specific meaning of Soviet prescriptions like those listed tend to change over time. For example, Donnelly's article on Soviet artillery (IDR No. 6, 1982) discusses changes in Soviet views on the best way to implement artillery preparation and support for an offensive. In the general sense, however, these prescriptions are commonly repeated and stable over time.

> Density of forces and means—level of saturation of a region of combat actions with forces and military equipment, calculated as the average quantity of forces and means per kilometer of front. Is an indicator of the level of massing of forces and means, and also a calculated indicator in the planning of operations (battles). The necessary density of forces and means is defined by the staff on the basis of a calculation of the correlation of forces and means in the entire area (sector) of combat actions and in the directions of blows (concentration of main forces). Distinguish density of infantry (motorized infantry), density of artillery, density of tanks and density of obstacles, and by scale—operational and tactical density.[38]

Thus in this definition, density is not in itself considered a factor in defensive success; it is merely "an indicator of the level of massing of forces and means." There is no indication here that the appropriate defensive density is affected by such issues as the nature of the terrain. Instead, it is determined solely by the requirements imposed by enemy force levels.

1. Soviet Views, 1968-73

For the 1968-73 period, we can concentrate on examining Soviet views from the General Staff journal Military Thought. In a 1968 Military Thought article entitled "Artillery in Modern Combat Operations of the Ground Forces," a Col. Shkarubskiy writes:

> In order to oppose mass tank strikes, it is necessary to have a deeply echeloned antitank defense. The chief efforts here should be concentrated in tactical depth in order from the very beginning of an enemy attack to inflict decisive destruction against his tank groupings and not to permit them to break into the disposition of friendly troops.
>
> To ensure insurmountable defense at the most important zones, considerable densities of antitank means are required. Such densities can be created by deploying a certain number of these means directly in the combat formations of the defending troops and broadly maneuvering them from the depth and from secondary zones.
>
> In the Great Patriotic War, density of anti-tank means was up to 25 units per kilometer in front of probable lanes of tank approach in tactical depth, and up to 30 units in operational depth. At the present time, in connection with the presence of qualitatively new, more effective antitank means (in mind, above all, is the antitank guided missile), it has become possible to decrease the above densities.[39]

---

[38] "Plotnost' sil i sredstv," in Voennyy Entsiklopedicheskiy Slovar', N.V. Ogarkov (ed.), (Moscow: Voenizdat, 1983), Author's translation.

[39] Col. P. Shkarubskiy, "Artillery in Modern Combat Operations of the Ground Forces," Voennaya Mysl' No. 6, 1968, tr. by CIA in "Selected Translations from Military Thought" (henceforward VM), p. 65.

This article is worth quoting in detail because it displays a number of important Soviet tendencies concerning defensive force density. One is the general tendency of Soviet military writers to refuse to admit the existence of tradeoffs, in this case the tradeoff between density (measured as weapons per square kilometer) and depth (square kilometers of defended zone). For a given force size, any increase in depth will reduce density, a consequence Shkarubskiy ignores.

A book by Herbert Goldhamer entitled The Soviet Soldier devotes two sections of a chapter to this phenomenon: "Having the Best of Both Worlds," and "Everything Is Equally Important."[40] If one is confronted with a choice between, in this case, defensive density and depth, solutions which involve sacrificing one value to enhance the other are discouraged—both are too important. Thus in the above article Shkarubskiy is simultaneously recommending that the defender position his antitank weapons in depth, and that he maneuver those weapons up from the depths in order to achieve high densities near the forward edge of the defense.

An article, a few years later, suggests avoiding this tradeoff by appealing to "the concentration of the main efforts on the decisive axes," which the author identifies as "one of the most important principles of the art of warfare." If the defender achieves such concentration, it is pointed out, he will have enough forces in those key areas to achieve both density and depth.[41]

Another useful bit of information in Shkarubskiy's article is the definition of a "dense" antitank defense: 25-30 antitank weapons per kilometer of front for World War II-era weapons, and somewhat less for an ATGM-equipped defense.[42] These numbers will become relevant in the discussion below.

The third important aspect of the Shkarubskiy article is that it is primarily about the use of artillery, as the title indicates. The author devotes most of his energy to discussing the need to achieve a high density of artillery on the attacking side, both to launch heavy artillery preparation before the attack begins and to offer artillery support after it begins. He argues that the old World War II concept of an "artillery offensive" ought to

---

[40] Herbert Goldhamer, The Soviet Soldier: Soviet Military Management at the Troop Level. (New York: Crane, Russak, 1975).

[41] Col K. Kushch-Zharko, "Principles of the Art of Warfare in Defense," VM No. 9, 1973, p. 29.

[42] A somewhat later article expands on the point that increasing effectiveness of antitank weapons is an important factor to be considered in evaluating the capability of a defense. See Col I. Andrushkevich, "Combat Against Tanks in Modern Operations," VM No. 4, 1969. Attention to this point has also persisted in more recent years; c.f. V.G.Reznichenko, (ed.), Tactics, op. cit., p. 99.

be revived for use in non-nuclear operations. Thus, regardless of the characterization of a dense defense as "insurmountable," the thrust of the article is that proper use of artillery makes it surmountable. A rare example of a discussion of a dense defense without discussion of how to overcome it is an article on "Defense in the Past and Present" in a 1971 issue of the same journal. The authors, two colonels, begin by noting, "the threat of nuclear attack by the advancing force compels the defending force to avoid establishing dense formations in areas where the main attack effort is concentrated, as was practiced in the last war."[43] As will be noted below, this is not an outdated concern: Soviet discussions of conventional operations the 1980s take into account the nuclear shadow.

Nevertheless, the authors recommend later in the same paragraph "establishing...a denser fire system and system of obstacles," regardless of the danger of over-concentration. They also note, "In spite of a twofold and threefold increase in width of defense zones and sectors, the density of antitank weapons and tanks has increased greatly in comparison with what it was in the Great Patriotic War," especially due to the introduction of ATGMs. "Today the antitank defense system has merged with the overall defense system, becoming its foundation."[44] This theme, that anti-tank defense is the basis of the entire defense, has grown in importance since this article was written.

The authors reserve the greatest respect not for antitank weapons but for defending armored vehicles, noting that a U.S. or West German defensive area may have "an average density of 30-50 units [armored vehicles] per kilometer of frontage." Furthermore, if these are dug-in, "it is extremely difficult to push past such 'armored' positions using only conventional weapons...[they] stand up well under artillery fire from indirect-fire positions as well as direct fire."[45]

Perhaps to offset the impression left by two upstart colonels that dense antitank defenses are difficult to penetrate, Lt. Gen. of Artillery M. Makarychev answered a few months later in the same journal with an article entitled "Artillery in Overcoming an Anti-Tank Defense in an Offensive."[46] Makarychev characterizes enemy defenses as being dense with anti-tank weapons allowing solid fields of fire. His main message, however, is clear: "in any situation,...artillery can successfully hit antitank defense targets...Success

---

[43] Col G. Ionin and Col K. Kushch-Zharko, "Defense in the Past and Present," VM No. 7, 1971, p. 68.

[44] Ibid, p. 70.

[45] Ibid, p. 72.

[46] Lt. Gen. of Artillery M. Makarychev, "Artillery in Overcoming an Anti-Tank Defense in an Offensive," VM No. 1, 1972.

in the effort against antitank weapons in turn depends in large measure on troop preparedness to accomplish this mission."[47]

These articles, and a few other similar ones, make clear that in the late 1960s and early 1970s, Soviet military scientists were concerned with the question of defensive density. To put this concern in context, however, a number of other issues related to the requirements of offensive success received as much or more attention. The most important of these is the ratio of attacking and defending forces. In an article entitled "Correlation of Forces and Rate of Advance," for example, Col. A. Gaponov is concerned with calculating the likely rate of advance of units on the offensive.[48] His main variables are the number of forces and the effective rate of fire on both sides. Neither Col. Gaponov nor those who criticize him in later issues of the journal mentions the density of the defense per se. Gaponov's model is, in some ways a clear forerunner to the model Hines discussed (detailed above). It lacks, however, even the implicit effect of force-to-space ratios which the later model reflects.

Offensive force density, a prominent cousin to the force ratio issue, is another factor heavily emphasized in Soviet literature on the determinants of offensive success. For example, Lt. Gen. V. Reznichenko, prominent editor of standard Soviet texts on tactics, writes of World War II tactics: "massing of men and weapons created conditions for dynamic development of the attack and ensured penetration of the enemy's defense to full tactical depth normally on the first day of the offensive operation."[49]

Reznichenko mentions that densities of artillery were especially increased during the war. Other articles explain why. One of the critiques of Gaponov's "Correlation of Forces and Rate of Advance" article made the argument that all forces do not correlate equally, as Gaponov assumed. For example, these critics said, a 5:1 advantage in artillery and 3:1 in infantry would lead to a higher rate of advance than 3:1 in artillery and 5:1 in infantry. Indeed, they claim, inadequate artillery density "constituted the primary reason for a low rate of penetration of the enemy's defenses" in some unsuccessful Soviet offensives early in World War II.[50] Another article argues that the quantity of artillery needed "to neutralize the enemy's defense in the breakthrough areas" depends solely on the

---

[47] Ibid, p. 85.

[48] Col A. Gaponov, "Correlation of Forces and Rate of Advance," VM No. 10, 1971.

[49] Lt. Gen. V. Reznichenko, "Characteristic Features and Methods of Conducting an Offensive," VM No. 1, 1972, p. 68.

[50] Lt. Col L. Veselov and Lt. Col. V. Selyavin, "The Question of the Correlation of Forces and the Rate of Advance," VM No. 5, 1972, p. 73.

number of enemy battalions to be neutralized: a denser defense requires denser artillery; more defenders require more artillery.[51]

There was one book published in the early 1970s, Antitank Warfare by Major General G. Biryukov and Colonel G. Melnikov, which heavily emphasizes the importance of defensive density. The authors list a number of principles learned in World War II "which have retained their significance to this day," including "massing and distribution of antitank weapons in depth in the most important defence sectors and carrying out large-scale manoeuvres with these weapons."[52] They also pay some attention to the value of prepared defensive positions.[53] On density and depth, the authors elaborate:

> One of the reasons for our success [at Kursk] was that the antitank efforts were distributed unevenly, the greatest densities of antitank weapons being created in vital areas which ensured the stability of the defences.
>
> By August 1941 the Soviet Army had discarded the linear organisation of antitank defences and had begun to distribute them in depth...the antitank defenses thus organised proved much more stable.[54]

It is worth noting that, like Shkarubskiy, the authors are emphasizing the importance of concentrating forces in "vital areas" in order to achieve high densities specifically in those threatened spots.[55]

Unusually in a Soviet publication, the authors of this work grapple explicitly with difficult tradeoffs, such as that between density and anti-nuclear dispersion. They write:

> The constant nuclear threat calls for dispersed battle formations, including the dispersal of antitank weapons. The extent of their dispersal, however, must be such as to provide for the fire density, especially anti-tank fire, needed to repel the enemy attacking in denser battle formations. Battalion defence areas must therefore be arranged as compactly as possible, by decreasing the intervals between the company strong points.[56]

Thus, the authors are suggesting trading off some anti-nuclear security in order to increase fire density. They note, however, that one need not go too far in this direction:

---

[51] Col J. Kaczmarek, "Concerning the Density of Artillery," VM No. 12, 1971, reprinted from Mysl Wojskowa (Warsaw) No. 4, 1971.

[52] Major General G. Biryukov and Colonel G. Melnikov, Antitank Warfare (Moscow: Progress Publishers, 1972), p. 60.

[53] Ibid, p. 100.

[54] Ibid, p. 58.

[55] Ibid, p. 55.

[56] Ibid, pp. 112-113.

"Present-day weapons are not likely to require concentration in such densities as were characteristic of the final period of the Great Patriotic War because the range of their direct fire (guided fire) had doubled or even quadrupled and their effectiveness has multiplied several times over." [57]

The authors also argue for increasing density in forward positions even if at the cost of weakening positions further back. At Kursk, they point out:

> The greater part of all antitank weapons was emplaced beforehand in the forward position of the defence zone and only about 35 percent of these weapons remained in reserve and in the second echelons of regiments and divisions. Thus concentration was achieved by providing in advance the highest possible densities of antitank weapons and by increasing these densities in the course of fighting through maneuvering with the reserves and through counterattacks. [58]

The authors note that larger antitank reserves were maintained later in the war when more such forces were available. But today, they argue, "In non-nuclear war the role of the first defence position is enhanced, which means that it has to be packed with antitank weapons to such a degree as to ensure repulsion of the enemy tank attacks."[59]

Another point worth noting is that the authors give some clues about how necessary force densities should be calculated. They say, "In distributing the total number of antitank weapons required to repel the enemy attacks (determined on the basis of antitank weapons-to-tanks ratio) the defender will also have to place some of the weapons in other positions and assign some to ambush and reserves."[60] Thus, in accord with the later Military Encyclopedic Dictionary definition, they say density should be determined on the basis of the overall force ratios.

What, specifically, should the antitank-to-tank ratio be? They point out that for a defense in a prepared position, "its chances to repel the enemy almost double. Thus for defence of a battalion we may assume the mean ratio of all its antitank weapons (without

---

[57] Ibid, p. 107.

[58] Ibid, p. 61.

[59] Ibid, p. 113.

[60] Ibid, p. 114.

light grenade launchers) to the enemy tanks to be 1:1.5-2 (sometimes 1:3)."[61] This estimate accords with Kardashevskiy's later report, which shows defenders with a 1:2 ratio defeating the attacker 90% of the time.[62]

Significantly, the authors do not stop here. They also point out that the necessary ratio depends as well on the degree of enemy "fire superiority...the nature of the terrain and its organization, the forces and weapons of the defenders and their morale."[63]

What can we conclude about this book? Clearly, it shows some significant Soviet attention to the question of defensive densities. It also puts such concerns in the usual context of attention to overall force ratios (the "correlation of forces") and the concentration of forces at the decisive place and time.

However, the authors show no evidence of particular pessimism about the possibilities of overcoming enemy defenses in general, noting only that, for example, "an inadequate softening-up of the enemy defences, especially the antitank defences, may result in breakdown of the whole offensive as was the case in a number of operations and battles of the Second World War."[64]

2. Soviet Views: Mid-1970s to Mid-1980s

For the years after 1973, Military Thought—our most preferred source—is unfortunately not openly available in the West. The open Soviet literature since 1973 seems to show, if anything, less attention to the issue of defensive force density than did Military Thought earlier. Occasional articles in the Ground Forces journal Military Herald are among the few discussions found. The Kardashevskiy article discussed above was found in this journal.

Most of these articles, however, are in the "rockets and artillery" section, and as one might expect they devote their primary focus to the use of artillery, in this case for overcoming dense defenses. The actual discussions of defensive density per se are

---

[61] Ibid, p. 100.

[62] Kardashevskiy, p. 64.

[63] Antitank Warfare, p. 100.

[64] Conclusions about the mood of pessimism or optimism should be drawn from this work only with great care, however. It was not published, as is usual for such works, by the Military Publishing House (Voenizdat). Rather, it was published in English by Progress Publishers. Both the language of publication and the choice of publisher serve as cautions that the work may have some propagandistic purposes—more so than most military publications, which are clearly designed more for internal audiences. While the general arguments in the book accord with usual Soviet ideas, the mood or tone of the work is particularly susceptible to manipulation, and so should not be given much weight.

usually limited to a few sentences, for example: "As we can see, the density of antitank means in the sector of [the enemy's] defense can reach up to 50 units per km of front. And for successful struggle with them it is necessary to know well their fire and maneuvering capabilities, tactics of action, strong and weak points."[65] The title of one of these articles—"Overcoming Defenses Saturated with Antitank Means"—clearly illustrates their focus.[66]

Another of these articles, "The Reliable Fire Destruction of the Enemy—the Basis of High Tempos of Offensives," is by Major-General of Artillery G. Biryukov—one of the authors of the book discussed above. In the article he expands on one of the points mentioned in the book: the reduced standards for achieving "high" densities since World War II. He notes that a modern battalion has, for example, over twice as many tanks and over four times as many heavy anti-tank weapons as did a German battalion during World War II. Furthermore, he notes, these weapons have much greater accuracy and range, and therefore effectiveness, than did the earlier weapons. He concludes:

> notwithstanding the increase in a battalion's defensive front, its fire, and especially antitank capabilities have grown several times in the context of a stable density of attacking tanks of 20-30 machines per kilometer of front. The fire action on attacking tanks has also been increased by artillery firing from the depth of the defense and by blows from fire-support helicopters.[67]

Biryukov's discussion explains a more general tendency: Soviet writers tend to focus more on the capabilities of weapons than on their density because their density is not particularly high by historical standards—while the density of their lethal fire is very high, because of their range and accuracy. Hence it is range and accuracy of weapons, not their density, which the Soviets discuss more. In 1976, the first volume of the authoritative Soviet Military Encyclopedia appeared, written under the editorship of then-Minister of Defense Marshal Grechko. It contains an article on the "Army Defensive Operation," which includes this discussion:

> The contemporary army defensive operation is characterized by such features as the deep echeloning of forces and means dispersed in their

---

[65] Lieutenant General of Artillery A. M. Sapozhnikov, "Deystviya Diviziona pri Prorive Sil'noy Protivotankovoy Oborony," Voenniy Vestnik (Military Herald, henceforward VV), No. 8, 1980, p. 58, Author's translation.

[66] Colonel P. Konoplya, "Preodolenie Oborony, Nasyshchennoy protivotankovymi sredstvami," VV No. 6, 1980, Author's translation.

[67] Major-General of Artillery G. Biryukov, "Nadezhnoe Ognevoe Porazhenie Protivnika—Osnova Vysokikh Tempov Nastuplenia," VV No. 5, 1977, p. 79, Author's translation.

> disposition; variety in the means employed in conducting the defense; a combination of firm positions with broad maneuver of fire, obstacles and forces; the simultaneous conduct of combat actions in several directions at various depths with sharp and frequent changes in the situation; and activeness and fierce struggle for seizing the initiative.[68]

This article illustrates another current in Soviet thinking about defensive operations: a heavy emphasis on the nuclear threat. The accent here is on depth and (anti-nuclear) dispersion of defensive deployments. The article does discuss organizing the defensive system of fire and terrain preparation.[69] But troop densities which are high by historical standards—which to the Soviets means World War II standards—are simply out of the question in a nuclear-threatened environment.

The article on "Defense" in Volume 5 of the Encyclopedia (1978, edited by Chief of the General Staff Ogarkov) explains this historical shift even more clearly. The article traces the historical development of the defensive, primarily in Russia and the Soviet Union, and highlights the Soviets' favorite defensive battle—Kursk. The discussion of the Battle of Kursk focuses on the depth and the degree of preparation of the Soviet defenses, and concludes by mentioning the operational densities. When the postwar period is considered, density is not an issue: "the deeper echeloning of forces and means and the dispersal of their deployment became characteristic."[70]

In articles in Military-Historical Journal around this period, there was a peculiar ambivalence to discussions of defensive density, presumably because of the issue of the nuclear threat. For example, one article, "The Development of Tactics for Defensive Battle," begins with the fairly common assertion that it is considering issues "which, in our view, have actual significance also in contemporary conditions."[71] The key statement of conclusions appears to be this:

> The experience of the war showed that success of defensive battle depended on the correct choice of the region and area of the zone of defense, the skillful deployment of combat units, the massing of forces and means in the probable direction of the enemy's main blow, the engineering equipping of the locale, correct organization of the system of fire, active-

---

[68] K. L. Kushch-Zharko, "Armeyskaya Oboronitel'naya Operatsia," in Marshal of the Soviet Union A. A. Grechko, (ed.), Sovetskaya Voennaya Entsiklopedia, Vol. 1 (Moscow, 1976), p. 246, Author's translation.

[69] Ibid, p. 247.

[70] K. L. Kushch-Zharko, "Oborona," in Marshal of the Soviet Union N.V. Ogarkov (ed.), Sovetskaya Voennaya Entsiklopedia, Vol. 5 (Moscow, 1978), pp. 661-2.

[71] Major-General V. Chernyaev, "Razvitie Taktiki Oboronitel'nogo Boya," Voenno-Istoricheskiy Zhurnal (henceforth VIZh), No. 6, 1976, p. 20.

> ness of the forces, commanding them precisely, cooperation of types of forces, combat and material security, etc.[72]

It is worth noting that defensive density does not even make this list. Curiously, though, the author does not undervalue issues of density elsewhere in the article. About the Battle of Moscow, he writes:

> Low densities of antitank artillery (3-5 guns per km of front), dispersion of tank [and] artillery means along the front, the lack of artillery-antitank reserves [and of] reliable fire connections between antitank strong points led to the fact that strong enemy tank groups often overcame the antitank defense of our forces.[73]

It seems unclear why the author would omit the factor of density from his summary list when he seems to rate it (or at least its absence) as so important. We can only hypothesize that he wished to make the point that density was important at the time of the war, but it should not be included among factors "significant in contemporary conditions."

Articles about the offensive in the same journal in the early 80's showed little more interest in the density of the opposing defense. One example is "The Employment of Tank Subunits and Units in Breaking Through the Enemy Defense," by Colonel N. Kireyev, published in 1982. Writing about post-World War II developments, Kireyev seems to find defensive density to be a relevant concern only for the 1945-53 (i.e., pre-tactical nuclear) period.[74] When he discusses enemy defenses for more recent periods, he states that the main challenge for the attacker in a conventional war would be "breaking through a well-prepared enemy defense."[75] He briefly mentions increases in tactical depth and the use of obstacles, and discusses the improved capabilities of the "qualitatively new antitank means" of the "probable enemy." He does not explicitly mention density in this connection at all.

Indeed, even a 1981 Military Herald article entitled "Contemporary Defense" essentially ignores the factor of the density of the defense. The author, Colonel G. Ionin, mentions that "the stability of the defense is most closely connected with its activeness." He devotes a great deal of attention to the need to set up a single, integrated system of fire

---

[72] Ibid, pp. 21-2, Author's translation.

[73] Ibid, p. 28, Author's translation.

[74] Colonel N. Kireev, "Primenenie Tankovykh Podrazdeleniy i Chastey pri Proryve Oborony Protivnika," VIZh No.2, 1982, pp. 33-4, Author's translation.

[75] Ibid., p. 38.

which covers the entire front, including any gaps in deployment.[76] He explains that even at the battalion level it is suggested that units deploy in two echelons (i.e., in some depth). But he does not mention explicitly that defensive density might be desirable. To the contrary, he writes, "The experience of wars and troop exercises provides evidence that a dispersed deployment of defending forces reduces their vulnerability and thereby secures their combat capability."[77] Thus anti-nuclear dispersion is important, while increasing density is not mentioned.

What about books on the offensive published during this period? How did they treat the issue of defense density? One such book was Proryv (The Breakthrough), by Gen. A. I. Radzievskiy, published in 1979—that is, around the same time as the Military Herald articles discussed above. A book about breakthroughs, even if ostensibly focused on the experience of World War II, should be relevant for our purposes.

It turns out that there is virtually no attention to defensive force density in Radzievskiy's book. There are only a few brief mentions, for example: "In offensive operations of the Great Patriotic War the main blow was inflicted most often on a weaker, vulnerable place in the enemy defense. Such places were usually sectors with low densities of forces and means, with insufficiently developed systems of engineering equipping of the sector, occupied by forces with weak preparation and low moral-combat qualities...Seams and flanks are always considered more vulnerable spots."[78]

However, when Radzievskiy discusses strong defenses, he only rarely mentions density. For example: "The experience of breaking through a positioned, deeply echeloned defense showed that such a defense occupied by steadfast forces harbors enormous forces of resistance. A breakthrough is connected with great losses in forces and means and often surpasses the boundary of the possible."[79] Again and again, the emphasis is not on density but on action taken to prepare the terrain—how well troops are dug in, construction of pillboxes, deployment of wire and mines—and on defensive depth.[80] Density, overall, seems to be secondary; force ratios, especially in artillery, are seen as crucial.

---

[76] Colonel G. Ionin, "Sovremennaya Oborona," VV No. 4, 1981, pp. 15, 17, Author's translation.

[77] Ibid., p. 14, Author's translation.

[78] A.I. Radzievskiy, Proryv (Moscow: Voenizdat, 1979), p. 167, Author's translation.

[79] Ibid., p. 168.

[80] Ibid., pp. 23-23; p. 188.

What Radzievskiy does discuss, in great detail, is offensive force densities. Sometimes artillery densities are mentioned first, along with densities of infantry and tanks.[81] Other discussions focus on the density of artillery alone.[82] In these cases, offensive densities and the ratios of offensive and defensive forces are often given—so defensive force densities can be calculated—but defensive densities are not mentioned explicitly.[83]

A more authoritative text, Lt. Gen. V. G. Reznichenko's 1984 Tactics, pays significantly more attention to defensive force density, but in the "mainstream" vein of most earlier Military Thought and Military Herald articles—i.e., with an eye toward how dense defenses can be overcome. Reznichenko writes:

> In essence, the antitank defensive system now constitutes the basis of the defense. The density of antitank weapons has increased drastically. In the Great Patriotic War, antitank weapon density on the main axes came to about 20-25 weapons per kilometer of front, while now, according to the experience of NATO exercises, this has doubled or tripled.
>
> Moreover, the combat capabilities of antitank weapons, i.e., their firing range and accuracy, and the power of their projectiles (missiles), have increased substantially. To disrupt modern enemy antitank defense systems, it is necessary to destroy or neutralize a considerable portion of the antitank weapons (70-80 percent of their total, according to the experience of local wars) while still conducting the fire preparation for the attack. Also the attacking subunits must immediately exploit the effects of the fire strike. The swifter and more unexpected the attack, the fewer casualties the attacking troops will suffer and the quicker they will be able to cross the zone of dense, overlapping enemy antitank fire.[84]

Reznichenko, then, in 1984 and again in his 1987 edition, is repeating the "mainstream" Soviet view of defensive force density: a dense antitank defense is formidable, but can be overcome with adequate artillery support and proper tactics. It should be noted, however, that in some places Reznichenko suggests that the Soviet expectation of a maneuver war is based on an assessment that defensive densities will be low. He writes:

> Favorable conditions for maneuver were limited in the past because of the presence of continuous defensive zones. The first-echelon formations and

---

[81] Ibid., pp. 171-173.

[82] Ibid., pp. 176-178.

[83] Ibid., p. 37, p. 46.

> units, operating in narrow zones, were initially forced to carry out a frontal attack and to break through continuous enemy defenses, i.e., to create a breach in his formation for carrying out a close or deep envelopment. Now the defense is formed up with considerable gaps between defended areas and strong points. Besides, employment of nuclear weapons, or even powerful conventional weapons alone, makes it possible to inflict heavy losses on the enemy and create breaches in his battle formation in the shortest periods of time. At the same time, the great mobility of troops makes it possible to swiftly exploit the effects of nuclear and fire strikes.[85]

Thus there seems to be a contradiction in Reznichenko's analysis: on the one hand, he professes concern about the density of NATO antitank weapons, as we noted above; on the other hand, he notes the opportunities for maneuver presented by gaps between enemy strong points. This is no isolated reference; elsewhere he notes, "During World War II an infantry division would usually occupy a defensive zone 8-10 kilometers wide and 5-8 kilometers deep, but today the dimensions of a defensive zone have increased to 30-40 kilometers in frontage and 20-25 kilometers in depth."[86]

Reznichenko appears to be reinforcing a point mentioned earlier: while NATO troop densities are now low by historical standards, the density of lethal fire is much greater. Thus a defensive zone has fewer troops but more anti-tank weapons, and more lethal ones. It also has a much greater density of anti-personnel ordnance.[87] This situation simultaneously affords more opportunity for maneuver (if that fire can be suppressed) as well as danger (if the suppression of enemy fire is inadequate).

As in other discussions, however, density is not the main factor affecting the strength or "stability" of the defense in Reznichenko's book. A section on "The Battle Formation" for defense emphasizes depth and dispersion (to reduce vulnerability to nuclear attack), with some mention of camouflage and deception—but none of trying to increase density.[88] There follow sections on preparation of positions, "The Fire

---

[84] Reznichenko, Tactics, op. cit., (fn. 24), pp. 99-100. A new edition of this volume was published by the Soviets in 1987, expanding the section under discussion, though the quoted portions are unchanged. See V.G. Reznichenko, Taktika (Moscow: Voenizdat, 1987), pp. 241-242.

[85] Ibid., p. 37.

[86] Ibid. p. 65.

[87] Engineer-Lt Col. R. Balabolkin, "The Future of the Infantry of the Main Capitalist Countries." VM No. 12, 1973, p. 113.

[88] Ibid., pp. 155-158.

Plan,"and "Hitting the Enemy on the Approaches to the Defense," among which the issue of density appears only briefly in the last, and not at all otherwise.[89]

In 1985-86, a number of articles discussing defense-related topics appeared in Military-Historical Journal. Most devoted very little attention to defensive densities. One such article, on tactics between the World Wars, stated:

> The stability of a defense was to be achieved by altering the width and depth of the areas of responsibility of the defending formations, increasing the tactical densities, making skillful use of the terrain and its engineering equipment, developing the fire system, the antitank and air defenses, and improving the battle formations and operating procedures of the troops.[90]

Of the issues listed in that statement, depth of defenses received only very brief additional treatment (two sentences) while the rest were discussed in one or more additional paragraphs. All, that is, except the issue of density, which received no further explicit discussion at all.

Density receives slightly more attention in another article a few months later. This article, "Development of the Fire Plan in Defensive Combat," makes a few statements to the effect that "it is essential to mass the weapons on the most important sectors in the aim of achieving maximum densities here."[91] The article also has a paragraph and part of a table which show the increase in tactical densities on the Soviet side during World War II. The main focus of the piece, however, is on the layout of defensive fire plans—fire support for flanks, adjustments to cover obstacles, depth and number of lines, etc.

3. Current Soviet Views

As discussed earlier, Soviet military thinking has been undergoing a period of change since the mid-1980s. This change is being driven by two separate factors, either of which alone would be sufficient to drive basic rethinking by the Soviets. The first cause of the shift is growing Soviet belief that the increased precision and destructiveness of conventional weapons has revolutionary implications for warfare. The second cause of change is the shift since 1987 in Soviet political-military doctrine toward greater defen-

---

[89] Ibid., 158-172.

[90] Colonel R.A. Savushkin and Colonel N.M. Ramanichev, "Development of Combined-Arms Tactics Between Civil and Great Patriotic Wars," VIZh No. 11, 1985, tr. JPRS, p. 18.

[91] Colonel A.A. Pastukhov, "Development of Fire Plan in Defensive Combat," VIZh No. 2, 1986, tr. JPRS-UMA-86-046, p. 13.

siveness, "reasonable sufficiency," and an emphasis on the avoidance of war, rather than the fighting of one.

What effect have these changes had on Soviet views of defensive force density? To compare with the 1979 Radzievskiy book, one might examine I. M. Ananyev's 1988 Tank Armies in the Offensive.[92] Like Radzievskiy a decade earlier, Ananyev is little interested in the density of enemy defenses per se as an explanation of attack success or failure. Like Radzievskiy, when he discusses the defense at all, he focuses more on the depth of the defense[93] or on the degree of preparation.[94] And like Radzievskiy, he is more interested in the density of offensive forces than in any characteristics of the defense.[95] Thus Ananyev's views about defensive force density show little change from Radzievskiy's a decade before.

One notable change in the Soviet literature in general is that since early 1987, there have been relatively many articles about the defensive in Soviet military publications. While there are still some "how-to" discussions of attack, for example in Military Herald, there are more such discussions of defense than previously. But these articles generally do not include consideration of density as a factor governing defensive success. For example, the lead-off article in a series in Military Herald about defense discussed such questions as the depth of the defense, the terrain, degree of fortification or preparation, mobility and maneuver, and the preparation of systems of fire, but did not mention density at all. Interestingly, the article was apparently written by one of the men, a Col. G. Ionin, who wrote the 1971 article in Military Thought discussing defensive density (and not how to overcome it).[96]

Other articles which might logically have considered the density issue also did not do so. A 1987 article on "The Bundeswehr Armored Division on the Defense," for example, argued:

> Division-level defense on the modern battlefield should be active, firm, echeloned in depth, prepared for armor, vertical envelopment, nuclear weapons, and massive air and artillery strikes...Its firmness is achieved

---

[92] I.M. Ananev, Tankovie Armiy v Nastupleniy (Moscow: Voenizdat, 1988).

[93] Ibid., p. 19.

[94] Ibid., p. 262.

[95] Ibid., pp. 253, 258, 263. Unlike Radzievskiy, Ananev focuses primarily on the density of close-support tanks in the offensive, rather than on the density of attacking artillery.

[96] Colonel G. Ionin, "Foundations of Modern Defensive Battle," Voyennyy Vestnik, No. 3, 1988, tr. JPRS-UMA-88-014.

> mainly through the proper combat formations for the situation, skillful use of the terrain, use of coordinated barriers, antitank fire, and the obstinacy of the forces conducting the defense.[97]

Again, density is not mentioned here. There is a brief mention of the usual frontages for Bundeswehr units, but the implications for density are not drawn out. The goal of the Bundeswehr defense is said to be the "greatest volume of fire" possible—not the greatest density of fires. Other journals seem to be following a similar pattern. For example, a 1988 Military Herald article entitled "Motorized Infantry Battalions of the Bundeswehr in Battle" discussed the frontage such a unit would hold on the defensive, but did not mention the density of anti-tank or other weapons that would result, let alone discuss the implications of such a density.[98] Similarly, a Military-Historical Journal article on "Breaching Enemy Defenses" discussed offensive densities in detail, but did not mention defensive densities—and indeed barely mentioned opposing defenses at all.[99] The article's conclusion makes clear it is intended that the principles identified in the article be applied to current problems.

The most recent article identified which mentions defensive density explicitly is from a 1987 issue of Soviet Military Review. The discussion is as follows:

> The density of fire and power of modern anti-tank weapons have increased multifold...That is why the attack momentum acquires in contemporary conditions primary importance. If the attack is a success and the advancing tanks burst into the enemy forward area, the attack will develop successfully and the mission will be fulfilled. If the attack fails (tanks have been stopped and infantry forced to hit the ground) the commander will have to start everything anew.[100]

This discussion is slightly different from the "mainstream" one, emphasizing attack tempo rather than artillery support. Nevertheless, like the "mainstream" view, the accent is on practical methods of overcoming dense defenses, and the implication is that such defenses can be broken through.

---

[97] Colonel A. Egorov, "Bundeswehr Armored Division on the Defense," Zarubezhnoe Voennoe Obozrenie [Foreign Military Review] No. 1, 1987, tr. JPRS UFM-87-003, p. 29.

[98] N. Nikitin, "Motopekhotnye Batal'ony Bundesvera v boyu," VV No. 9, 1988.

[99] Major General (res) A.P. Maryshev, "Breaching Enemy Defenses," VV No. 3, 1988, tr. JPRS-UMJ-88-009.

[100] Major-General Ivan Skorodumov, "Attack Momentum," Soviet Military Review No. 8, 1987, p. 14.

One recent rather suggestive article argues that "stable defense is attained in large measure through the construction of an elaborate system of fortifications"[101] Unfortunately, the article is in Soviet Military Review (which is published primarily for export). Its focus fits far too neatly into the propaganda goal of convincing the West of the essential defensiveness of the new Soviet military thinking for it to be considered reliable as an indication of a trend. Nevertheless, there is also a recent Soviet book on fortifications. It is possible that Soviet interest in the subject is growing.

So what can we infer about current Soviet views on defensive density? Reznichenko's attention to it in both editions of Tactics is too important to be ignored. Nevertheless, the apparent trend away from discussing the issue in the periodical literature may be significant. One thing that this might reflect is that to the extent density matters at all, it is the density of lethal fire, rather than of men, weapons, or anything else, which signifies. This was noticed even two decades ago, in the recognition that more lethal antitank weapons, like ATGMs, can be considered "dense" even if there are fewer of them per kilometer than there were anti-tank weapons of World War II vintage.

In the context of even more destructive modern systems—what the Soviets call "reconnaissance-strike complexes" and conventional weapons "approaching low-yield nuclear weapons in destructive power"—this becomes more true than ever. Indeed, as the range, especially, of new and projected NATO systems grows, and thereby the ability of these systems to shift their fire across broad sectors of front, the density of fire against attackers in the main breakthrough sector becomes far greater than that generated by defending forces actually deployed on that sector. It is therefore the density of fire which attracts attention rather than the density of shooters. In this sense, diminishing Soviet interest in the density of defensive weapons systems is logical.

But what about the more immediate future, in which armored targets are still mostly destroyed by direct fire from at most a few kilometers away? Here, the Soviet ideas of the early and mid-80s would still apply—including their ideas about the capabilities and densities of currently deployed NATO ATGMs, as discussed by Reznichenko and others. And as these Soviets make clear, they have thought deeply about necessary steps, such as an "air operation" and strong fire support, to overcome NATO defenses,

---

[101] Lieutenant Colonel Georgy Stepanenko, "In the Interests of Stable Defence," Soviet Military Review No. 3, 1989, p. 13.

even if prepared and dense.[102] The tone of Soviet discussions of these issues is not a pessimistic one.

Another relevant factor, of course, is the restructuring of the Soviet military. As the Gorbachev troop cuts begin to bite, and as reorganizations to make the remaining troops more "defensive" proceed, Soviet offensive capabilities are likely to decline, perhaps significantly. But as of this writing, the changes are not yet significant enough to affect the assessment.

F. CONCLUSIONS

One major conclusion of this study concerns the limited scope of the Soviet literature about defensive force-to-space ratios. There is no mature, systematic Soviet theory about the effect of defensive density on the likelihood of offensive success, and limited attention is paid to this issue in the Soviet literature. There are no books or articles which are focused primarily on that issue; when it is discussed, the discussions are rarely in depth, and are often brief, passing remarks. There is therefore little evidence to support a contention that defensive force-to-space ratios play a central role in Soviet thinking.

Soviet writings do reflect the view that, all else being equal, higher force-to-space ratios favor the defender. General Staff mathematical models also demonstrate this behavior. To this extent, Soviet views of the subject agree with most Western views. However, the emphasis of Soviet writings is on ways to overcome dense defenses, with the implication that a well-planned attack using sufficient artillery will generally succeed. Thus the Soviets clearly do not view defensive density as the dominant feature of the battlefield.

To the Soviets, there are a number of factors which interact with force-to-space ratios to affect what they call the "stability of the defense." The most important of these is the force-to-force ratio, with an emphasis on counter-concentration to maximize the force ratio at the decisive point. Other important issues the Soviets consider in their analyses include defensive depth, the degree of defensive fortification, organization of the fire plan, and leadership and morale. The Soviets also note that modern weapons technology, especially nuclear weapons and "reconnaissance-strike complexes," decreases the minimum force-to-space ratio needed for defensive success by increasing the lethality of individual weapons.

---

[102] For a good discussion of these ideas, see Petersen and Hines, op. cit., Orbis, Fall 1983.

Concerning the debate in the Western literature on Soviet attitudes toward defensive density, it is clear, as all the analysts discussed here note, that the Soviets are concerned to some degree about force-to-space ratios, and believe that less dense defenses are easier to penetrate than more-dense defenses. Ultimately, the disagreement concerns the questions of salience and severity: do the Soviets see penetrating dense NATO defenses as their central problem, and a hopeless one; or do they see it as a less important and more tractable problem? There is little evidence to support the former view. Indeed, the most emphatic case to this effect in the West—Dick's—is based on Dick's own, first-principles calculation that the Soviets' proposed solutions are infeasible. While these calculations may be correct,[103] there is little information in the open literature to suggest that the *Soviets* think so.

---

[103] Some analysts disagree with this assessment. Richard E. Simpkin argues (p. 199), "...even if NATO put everything it had in the shop window...the weights and intensities of fire and troop densities which the Soviets are in a position to employ will crack the line somewhere, probably sooner rather than later." Richard E. Simpkin, Red Armour (Oxford: Brassey's 1984), p. 199.

Appendix C
THEORY

Stephen D. Biddle

A. INTRODUCTION

The literature reviews in appendices A and B suggest that neither Western nor Soviet writers have yet produced a systematic theory to explain the impact of force-to-space ratios. For the purposes of this study, our primary task must therefore be to develop such a theory and to subject it to initial testing. This appendix is intended to accomplish the first part of this task—to provide a causal theory relating force-to-space ratios to conventional defense effectiveness.[1] appendix D will provide the second of these tasks by describing the results of initial quasi-empirical testing using the JANUS combat simulation.

To develop this theory, we will begin by defining a dependent variable, or outcome to be predicted by the theory, and identifying a set of independent variables by which to explain that outcome. We will then describe the overall dynamics of theater-level conventional warfare in broad terms that will serve to establish an analytic context within which to understand the interactions of the independent variables. Given that context, we will go on to describe in some detail the behavior of each of the identified variables, and the military logic underlying that behavior. Finally, a set of equations will be presented which will describe those behaviors in quantitative terms and enable us to deduce the relationship between force-to-space ratios and combat outcomes from the sometimes complex interactions of the identified variables. To describe the military logic

---

[1] By *causal theory* we mean an explanation of a phenomenon that permits prediction and control, and which holds over an identifiable span of time and space. Such a theory consists of a system of *law-like statements*, *assumptions*, and a *causal logic* by which these elements are harnessed to explain the phenomena of interest.

Given this, a *model* represents (or, in Kenneth Waltz' words, "depicts") an underlying theory in an analytical structure which facilitates the deduction of implications from the theory itself: See Kenneth N. Waltz, Theory of International Politics (New York: Random House, 1979), p. 7. Like theories, models can be univariate or multivariate, quantitative or qualitative, simple or complex (for examples of qualitative models in the study of international relations, see, e.g., Paul Diesing, Patterns of Discovery in the Social Sciences (Chicago and New York: Aldine-Atherton, 1971), pp. 108-114; also Waltz, op. cit., pp. 7, 59, 93-5). Models are thus not alternatives to theories—and in fact cannot be constructed without at least an implicit theoretical foundation. Rather, models are merely convenient representations of the relationships explained by the theory, and especially, representations which facilitate the process of deductive inference. In particular, mathematical models can be essential tools for deducing the consequences of theories whose if-then propositions are quantitative and complex. The usefulness of the resulting model, however, is only as great as that of the underlying theory—and the construction and use of such a model is an integral element of rigorous theoretical investigation for such questions. On the general distinctions embodied in this terminology, see also Arthur Danto and Sidney Morgenbesser, "Laws and Theories" in Arthur Danto and Sidney Morgenbesser, eds., Philosophy of Science (New York: Meridian, 1960), pp. 177-181.

of these interactions, however, we will pursue a highly idealized, notional development in our discussion of the behaviors of the independent variables; this idealized treatment is intended as motivation for the equations that follow, rather than as a substitute for them. Finally, it should be noted that we will attempt neither to formally prove the theorems in this appendix nor to provide a fully axiomatized theoretical system. While desirable, such a degree of mathematical formality is beyond the scope of the present inquiry, which is necessarily focused primarily on the military and policy issues associated with defense at low force densities.

B. KEY VARIABLES

The phenomena of primary interest for this study involve conventional combat between more or less sophisticated opponents at the theater level in Central Europe. While it is possible that the resulting theory may have useful application beyond these bounds (e.g., to theaters other than Europe), as a point of departure we will restrict our scope here to the prediction of military outcomes for European, theater-level offensive ground operations.[2]

---

[2] Thus low intensity conflict, naval or amphibious warfare, strategic (or other independent) air, or nuclear warfare are excluded from consideration here. An "operation" may be defined as an interconnected series of military actions (or "engagements") of a duration corresponding to the planning horizon of the initiating combatant's theater commander. Typical examples might include Operation Cobra (the American breakout from the Normandy perimeter, 24-27 July 1944), Operation Goodwood (Montgomery's offensive east of Caen, 18-21 July 1944), Operation Citadelle (the Battle of Kursk, 4-17 July 1943), or Operation Fall Gelb (the German invasion of France, 10 May to 5 June, 1940). Warfare of longer duration or larger scope is clearly of interest, but can most readily be addressed through the analysis of its component operations. The Battle of El Alamein in 1942, for example, consisted of two distinct operations: Lightfoot (23-26 October) and Supercharge (27 October to 4 November), of which the former stalled, while the latter successfully broke through the German defense.

Our focus on the operation as the primary level of analysis is intended largely as a matter of convenience with respect to more extensive historical validation efforts which may follow the present study. Unambiguous classification of historical battles is crucial for such validation, and for this it is necessary to be very clear with respect to the boundaries of the class of phenomena being described. For this purpose, the operation is easier to delimit than, say, the campaign. As an example, the combat activity on the Eastern Front in July 1943 is particularly instructive. Is the German offensive of Operation Citadelle to be classified as a separate case from the subsequent Soviet counteroffensive east of Orel? If the level of analysis is the operation, this combat activity would be classified cleanly as two operations, and thus two historical data points—a German offensive operation toward Kursk which stalled prior to breakthrough, followed by a Soviet operation immediately to the north which broke through the German line. If the level of analysis were the campaign (an interconnected series of operations), classification would be more ambiguous. It would be arguable that this episode would constitute a single data point: a German offensive toward Kursk which ultimately resulted in a Soviet breakthrough at Orel [since the Soviet offensive was intimately—and consciously—connected to the preceding German attack, which the Soviets expected to provide a weaker target for their own attack; for a description of the offensives at Kursk and Orel, see, for example, John Erickson, The Road to Berlin (Boulder, CO: Westview, 1983), The specific outcome of interest with respect to these phenomena is the ability of one combatant to invade and subjugate another by force of arms. To do this, an aggressor must assert military control over the territory of its opponent. While the destruction of opposing military forces is likely an indispensable means to this end, the ultimate end itself is the control of opposing territory and the economic means contained in it, whether in its entirety or in part (as might be the case, for example, in an attack with limited aims). Given this, we will define the dependent variable for the theory (that is, the specific outcome to be predicted) as the *net territorial gain* an aggressor could expect in the event of an invasion. In particular, we will estimate the maximum penetration distance (in kilometers) that a putative offensive could take and hold against defensive counterattack.

The primary independent, or explanatory variable for this study is of course the defender's theater force-to-space ratio. We will define the force-to-space ratio as the ratio of defender forces present in the theater over the length of the theater frontier, or more specifically, as the number of "Blue" armored fighting vehicle equivalents (AFVEs) defending a constant 850-kilometer border at the time hostilities are initiated.[3]

Should we consider additional independent variables? A theory should be as parsimonious as possible—that is, the number of independent variables should be as few as possible.[4] Parsimonious theories are easier to test, easier to understand, and easier to apply. Indeed, the whole function of theory is to order the complex events of the real world in terms of simpler relationships that can more readily be understood and manipulated. The more parsimonious the theory, the more completely this ultimate aim of

---

inherent in the operation but not necessarily available at the campaign level thus offers a convenient delimiter and is therefore used here.

[3] An AFVE is simply a convenient, "generic" measure of force size selected to facilitate the comparison of highly disaggregate JANUS results and equations for the estimation of theater-level combat outcomes. A single main battle tank represents one AFVE (regardless of nationality, make or model). A single armored troop carrier with its infantry complement is also scored as one AFVE. A carrier without its infantry is half an AFVE; the infantry without the carrier is half an AFVE. Armored antitank, air defense, command, or reconnaissance vehicles are also one-half an AFVE. Field artillery and aircraft are accounted separately in units of tubes and sorties, respectively (see appendix C), and thus are excluded from the AFVE totals per se.

[4] Note, however, that a very parsimonious underlying theory can give rise to very intricate deductive models of phenomena. Lanchester theory, for example, is quite parsimonious, but the interactions which can be deduced from that theory can be very complex.

theory-building is met. But while a theory should be as parsimonious as possible, parsimony must be balanced against requirements for validity and utility. We must include sufficient independent variables to explain the observed variation in outcomes. Moreover, for a theory intended to inform the development of public policy, we must include a sufficient range of explanatory variables to account for the policy maker's range of available options. Of course, no single theory will ever be able to incorporate the entire scope of nuance available to the policy maker for influencing outcomes. At the same time, however, a theory will be discarded as irrelevant unless it at least addresses most of the primary options for action.

For the case of force-to-space ratios, many potential options are available by which policy makers could affect the relationship between density and combat outcomes. The balance of opposing forces could be altered through arms control, as could the nature of the weapons or equipment with which potential combatants are armed. Weapons technology could be altered to rely more heavily on tactical aircraft or long range missiles and artillery, barrier defenses could be expanded, warning and intelligence systems could be improved, or military doctrine could be adapted.

Moreover, the literature suggests strongly that the primary relationship between force-to-space ratios and combat outcomes is strongly influenced by a variety of important intervening variables. To ignore these variables would risk biasing the resulting theory and undermining its validity as an explanation of theater combat results.

Given this, we will consider a multivariate formulation. The particular variables to be included are to be determined ultimately by the result of testing (see appendix D); as a point of departure, however, the literature was used as a heuristic device to suggest candidate variables for more exhaustive consideration and test. While the existing literature is not theoretically rigorous or systematic, it does constitute the collective observations of many experienced soldiers and analysts over a substantial period of time. It is thus rich in insight. As such, it offers a degree of insurance against the danger of overlooking effects with important consequences for outcomes and provides a sound point of departure.

Three broad categories of these intervening independent variables emerged from the literature reviews in appendices A and B and survived the process of testing described in appendix D. These are weapon mix, terrain, and force employment. These categories can be broken down into more discrete (and thus more readily operationalized) concerns.

With respect to weapon mix, for example, the literature deals both with distinctions between weapon *types* (e.g., tanks as opposed to small arms) and between more and less *effective* versions of the same weapon type (e.g., muskets as opposed to machine guns; or short range, as opposed to long range, artillery). With respect to geography, we can distinguish two sub-issues: the effects of natural, and of "man-made" topography—that is, the impact of forested vs. open natural terrain, and the impact of mined or otherwise prepared ground vs. unprepared. With respect to force employment, the literature identifies as important issues the pace or tempo of the attack and its frontage; the depth of the defense; and the availability and use of operational reserves.

A fourth class of potential intervening variable is the theater force-to-force ratio—or the balance of attacker to defender forces in the theater of war. Force to force ratios are clearly important. For our purposes, however, it will prove most convenient to address their effects implicitly, via their influence on the other three classes of explanatory variables. Our consideration of force-to-space ratios will thus focus explicitly on the effects of weapon types, weapon effectiveness, terrain and its preparation, tempo, defensive depth and operational reserves—and their interaction with theater force-to-force ratios, and of course, the force-to-space ratio itself.

C. ANALYTIC CONTEXT: THEATER DYNAMICS

A theater-level operation consists of many discrete military engagements occurring over a broad geographic area and a significant period of time. Many of the independent variables discussed in the theoretical literature, however, are much more localized in their immediate effects—and our analytic treatment of these variables will similarly be grounded in their effect on the local engagement. To reach conclusions as to theater-level outcomes on the basis of such analysis, it is thus necessary to posit some relationship between the engagements that make up an operation, and to describe how the results of those engagements combine to affect the operation as a whole. In effect, it is necessary to provide a description of the theater-level dynamics of conventional combat.

To do this, we will rely heavily on the understanding of theater dynamics provided by the existing literature. This understanding is broadly consistent. Similar descriptions can be found as early as the mid-nineteenth century in the work of Carl von Clausewitz and especially Antoine-Henri Jomini; in the mid-twentieth century in the writings of Basil Liddell Hart; as recently as the late 1980s in the work of analysts such as Richard Betts; or even in the fifth century B.C. in The Art of War by the Chinese soldier-philosopher Sun Tsu.[5]

While they differ in detail, all tend to focus on the process of attacker concentration and defender counterconcentration; on the changes these produce in local force-to-force ratios; and on the consequences of those changes for the attacker's ability to break through. Most authors, for example, credit attackers with some ability to choose the time and place of their attack. This in turn enables the theater attacker ("Red") to concentrate a large fraction of his force on a narrow sector opposite that chosen point (or points), while defending elsewhere with the remainder. Initially, the location of this point is unknown to the theater defender ("Blue"). Prior to discovering this point of attack, Blue is generally assumed to be more uniformly deployed along the frontier than is Red.[6]

Once Blue locates the point of attack, he attempts to counter-concentrate his forces to match those of the attacker. This process takes time, however (especially if

---

[5] For Clausewitz, see especially the discussion of "the decisive point" and "the culminating point of the attack," in Carl von Clausewitz, On War, edited and translated by Michael Howard and Peter Paret (Princeton, NJ: Princeton University Press, 1976), Book III, Chapter 8, pp. 194-7, Chapter 11, p. 204; and Book VII, Chapter 5, pp. 528-9. For Jomini, see Antoine Henri Jomini, A Summary of the Art of War, translated and edited by J.D. Hittle (Harrisburg, PA: *The Military Service Publishing* Co., 1947), especially pp. 67-70. For Liddell Hart, see, for example, his treatment of counterconcentration in Basil H. Liddell Hart, Europe in Arms (London: Faber and Faber Limited, 1937), pp. 83, 334. For Betts, see for example, Richard Betts, "Conventional Deterrence: Predictive Uncertainty and Policy Confidence" World Politics, January 1985, pp. 153-79. For Sun Tsu, see Sun Tzu, The Art of War, trans. Samuel B. Griffith (London: Oxford University Press, 1963), esp. p. 98, verses 13 and 14.

[6] Of course, no actual deployment is ever truly uniform. The nature of the terrain and the value of the military objectives that lie behind particular stretches of front differ; hence, no sensible defender will deploy a constant force density across the entire frontier. Instead, the defender typically leaves less well-defended those sectors he thinks less valuable or less likely to be attacked and deploys stronger forces on sectors he judges most valuable and most likely to be attacked, while retaining reserves with which to react to the opponent's actual choice once observed (more on this below). In effect, the defender distributes his forces so as to minimize his expected loss of value, given a certain expectation (or in Bayesian terms, a certain subjective prior probability distribution) with respect to the attacker's choice of main attack sector.

Note, however, that a property of an optimal defender allocation across sectors (i.e., one that minimizes expected loss of value) is that the marginal utility (i.e., the marginal increase in expected total defensive value after attack) of additional force deployment will be equal across sectors. This in turn implies that the attacker will face equal expected gain of value across sectors. In effect, then, the actual dynamics of concentration and counterconcentration will mirror the simple, heuristic description given above in important respects: while the distribution of men on the ground may not be uniform, the distribution of military attractiveness for attack will be uniform. Moreover, for the special case of an undifferentiated theater in which terrain and military value are identical across sectors, and in which the defender has no knowledge of attacker plans as of some initial reference time (which need not be the time of attack), the optimal allocation of defensive troops per se will in fact be uniform. As a heuristic, and as a first order abstraction of the more complex allocation problem, an assumption of a uniform initial force distribution in an undifferentiated theater is thus an appropriate point of departure, and will be employed here.

Blue must disengage and displace forward units in contact with the enemy, rather than simply dispatching operational reserves from rearward assembly areas). Prior to the arrival of those reserves, Red's local concentration provides a high attacker:defender force-to-force ratio at the point of attack. Red attempts to exploit this local advantage by overwhelming the initially outnumbered local defender and breaking through into Blue's rear area before sufficient reserves arrive and make further advance impossible.

If Red does break through, the viability of the Blue defense is seriously undermined. Armies depend on an elaborate supply, command, air defense and transportation infrastructure. This infrastructure is extremely vulnerable to direct attack. An attacker who has broken through the defender's forward positions and gained access to the rear can thus do grave damage by destroying that infrastructure. Even a modest exploitation force surviving at the point of attack can thus pose a serious threat to the defense in the event of breakthrough (provided that exploitation force has the freedom of maneuver to range widely once in the rear).

If the attacker fails to break through, however, his maximum advance will occur at what Clausewitz called the "culminating point" of the operation—that is, the moment at which some combination of attacker losses and defender counterconcentration produces a local force-to-force ratio too small for continued advance.[7] This culminating point constitutes both the attacker's high water mark and potentially his point of greatest vulnerability, in that he will have had limited time to prepare defensive positions or redistribute forces for effective defense. A prudent attacker will thus employ his forces in such a manner as to avoid reaching such a culminating point in a condition which would allow an aggressive defender to break through the attacker's lines at this point of maximum weakness.

In effect, our task is thus to distinguish between circumstances that produce breakthrough and those which produce culmination short of breakthrough (and in the case of the latter, to estimate the attacker's advance prior to reaching his culminating point). Given the dynamics described above, this amounts to an evaluation of the crucial engagements at the point of attack in the context of a race between attacker concentration and attempted breakthrough, and defender counterconcentration and possible counterattack.

---

[7] See Clausewitz, op. cit., Book VII, Chapter 5, pp. 528-9.

D. TEMPO

What do we mean by "tempo," and how are we to operationalize it as a variable? In the theater dynamics described above, Red penetration and Blue counterconcentration constitute a race. While Blue moves reserves, Red is taking ground at the point of attack—if Red advances far enough quickly enough, he may break through before Blue has sufficiently reinforced the threatened sector. The key issue is thus one of relative speeds, or, alternatively, of *tempo*: can the attacker prosecute a local advantage quickly enough to break through before the defender can react?[8]

We can thus address the issue of "tempo" as identified by the literature in terms of two, more concrete quantities: the attacker's rate of advance, and the defender's rate of counterconcentration. Each of these quantities involves a substantial element of choice. For the attacker, this choice requires a tradeoff between advance rates and casualties; for the defender, the tradeoff involves allocation of forces between forward and reserve roles. For the time being, we will concentrate on the first of these; we will turn to the question of the defender's balance of forward and reserve elements (and its effect on the rate of counterconcentration) in the section on defensive reserves and counterattack, below.

1. ***Military Effects of Tempo:* The Trade-off Between Velocity and Casualties**

How, then, do advance rates and and casualties trade-off? Let us begin by considering the problem at the tactical level, in the form of an assault on a defended position. Let us further posit a range of conceivable closure rates for this assault of between zero and about 60 kilometers per hour—i.e., the maximum speed of modern armored vehicles. To reach this upper bound, attackers must advance in vehicles on paved roads in open country, in column formation with sufficient inter-vehicle spacing to prevent collisions in the event of an unexpected halt, and without delaying for extensive intelligence or artillery preparation. Such tactics would require extreme exposure to enemy fire, while minimizing the forward firepower of the advancing units and thus limiting their ability to

---

[8] This usage is consistent with the underlying concerns raised in the literature on this issue. When Basil Liddell Hart, for example, emphasizes the importance of "tempo" for the German invasion of France in 1940, he in effect observes that the Germans were capable of accomplishing their goals--of breaking through and exploiting that breakthrough--more quickly than the French could respond. Alternatively, when John Mearsheimer writes of the interdependence of "blitzkrieg" and the force to space ratio, he is effectively explaining the relationship in terms of the relative rate of offensive penetration and defensive reaction--and of the sensitivity of those rates to changes in force density. For a systematic treatment of the term see Richard Simpkin, *Race to the Swift: Thoughts on Twenty-First Century Warfare* (New York: Brassey's, 1985), pp. 106-112; for a more detailed description of the literature on this point, see appendix A.

reduce losses by silencing defenders with return fire. Extreme speeds can thus produce extremely high casualties.[9]

Of course, a variety of means are available for reducing losses. At a minimum, attackers can leave the road and deploy into a line-abreast formation better suited for returning fire. Cross-country movement, however, is substantially slower than road movement—especially if line-abreast formation must be maintained over broken country. Preparatory artillery fire can destroy, suppress or obscure defenders, but to do so requires that the assault be delayed long enough to allow the artillery to deliver the necessary volume of fire. Tactical reconnaissance can locate defenders and thus improve the effectiveness of offensive fire (or permit less-exposed advance routes to be identified), but it requires that the assault and the artillery preparation be delayed for scouting. Dismounted infantry can clear dug-in defensive positions on rough terrain with fewer casualties than mounted units, but in the process the assault speed is reduced to that of a walking rifleman (or less, if defensive fire drives the attackers to ground). Techniques that would reduce attacker casualties thus also reduce the attacker's net rate of closure with the enemy; to reduce casualties in a tactical assault thus requires that the attacker reduce speed.[10]

2. Graphical Analysis

These observations can be generalized to produce the hypothesis depicted in Figure C-1, in which attacker casualties in a given tactical assault (C) are related to the attacker's attempted assault velocity (V).[11] As hypothesized in Figure C-1, casualties are highest at high velocity, where attackers have the fewest opportunities to reduce their

---

[9] Exceptions do exist—e.g., where an attacker is faced with a weak direct fire defense but extremely heavy defensive artillery fire (see Figure C-18 and accompanying discussion below); or where the attacker has broken through the defended zone and is engaged in exploitation operations in the enemy rear (see the discussion under "Military Effects of Depth," below).

[10] For a more detailed treatment of the mechanics and consequences of these techniques, see the discussion under "Weapon Technology," below.

[11] "Velocity" is defined as the elapsed time between initiation of offensive action (including artillery or ground reconnaissance preparation) against the position in question and the arrival of assault elements on the objective, divided by the distance between jump-off and objective. "Casualties" are total losses (AFVEs destroyed or disabled) to the *attacker* incurred over the duration of a single engagement (i.e., incurred between the initiation of offensive action and the arrival of assault elements on the objective). In some circumstances, losses to the defender would be inversely related to these, although this need not necessarily be the case (e.g., where mounted attackers with a substantial numerical advantage could overrun a defense too quickly for the defenders to withdraw, defensive casualties would be higher for high assault velocities than for low V, where the slower rate of closure would permit surviving defenders to escape).

exposure or bring their own firepower to bear.[12] The slower the velocity, the wider the range of tactical options available to the attacker, and hence the greater the opportunity for selecting a combination that enables the defending position to be taken with fewer losses. A modest reduction in speed permits, for example, off-road deployment but excludes artillery preparation or advance reconnaissance—and thus enables a modest reduction in losses. A more substantial reduction in casualties can be obtained by combining off-road deployment and artillery-delivered smoke and suppressive fire against suspected defensive positions, but this increases the time required to complete the assault and thus can be obtained only at the cost of a further reduction in speed. To obtain greater reductions in casualties would require that the above measures be augmented by techniques such as careful pre-assault scouting of the defensive position, or support by dismounted infantry to clear difficult terrain, but to exercise these options would further reduce the velocity of the assault.

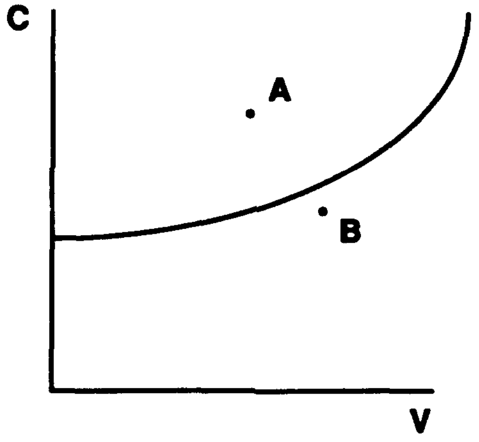

Figure C-1. The Casualty-Velocity Trade-off

The casualty-velocity trade-off hypothesized in Figure C-1 thus represents an *efficient frontier*. That is, it would certainly be possible for an attacker to produce an assault that lingers needlessly and thus suffers unnecessarily heavy losses at slow velocities (point A in Figure C-1)—but it would not be possible for an attacker to attain a given

---

[12] The graphical analyses presented in appendix C constitute theoretical postulates motivated by the discussion of military phenomena associated with the given variables. The validity of these postulated curves in quasi-empirical testing is addressed in appendix D.

velocity without suffering at least the casualties associated with that *V* (e.g., point B in Figure C-1 is infeasible). In other words, for no point on the frontier is it possible to reduce casualties without simultaneously reducing velocity.

We will posit that the curve which represents this frontier is strictly convex. This is because techniques for casualty reduction can be rank ordered from the most to the least marginally effective—that is, from the technique that offers the largest casualty reduction per unit of velocity reduction to that which offers the smallest. A rational attacker would logically choose the most marginally effective technique first, turning to successively less effective options only as necessary to reduce casualties further. Thus, the lower the velocity (and hence the more options for casualty reduction that have already been exploited), the smaller the marginal casualty reduction that can be obtained for a given further reduction in velocity, with the result that the slope of the frontier is greatest at high velocity and smallest at low velocity. As long as the marginal effectiveness of each technique is not identical, the frontier itself is thus convex.[13]

The particular curve shown here is, of course, specific to a given set of circumstances. In the following sections of this appendix we will discuss a variety of variables which determine in part the position and slope of this trade-off. For now, we will consider in detail only one of these, the local force-to-force ratio. Figure C-2 depicts the hypothesized effect of variations in force-to-force ratio (given as the attacker:defender ratio of maneuver forces in contact at the point of attack) on the basic casualty-velocity trade-off shown in Figure C-1. As shown, the higher the local force-to-force ratio, the fewer casualties an attacker will suffer in an assault at a given velocity. Alternatively, the higher the local force-to-force ratio, the higher the velocity an attacker can maintain for a given level of casualties. In effect, an increase in the local force-to-force ratio thus shifts the efficient trade-off frontier to the right; a decrease in the local force-to-force ratio shifts the frontier to the left.

---

[13] For the purposes of the discussion here we will assume, but not formally prove, differentiability and strict convexity in addition to the simple convexity argument given above. By way of motivation for this assumption, note that while any given casualty reduction technique may be discrete in nature (thus producing a piecewise linear, rather than strictly convex, frontier in combination with other such techniques), over many engagements under varying local conditions the velocity bounds and specific casualty reductions associated with any given technique will vary. Thus, as the number of individual engagements gets large, the frontier would become increasingly curve-like. It should also be noted that the empirical investigation described in appendix D supports the claim of simple convexity advanced above.

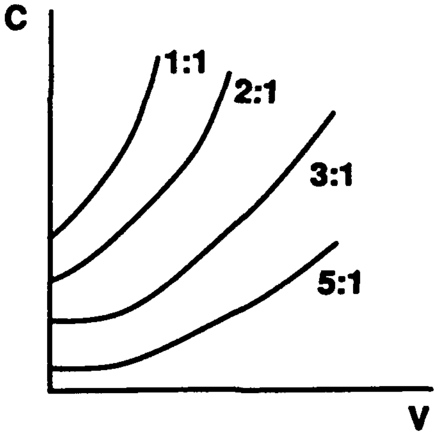

Figure C-2. The Effect of Force-to-Force Ratio on the Casualty-Velocity Trade-off

For any given force-to-force ratio, the casualty-velocity trade-off frontier thus represents the set of possible efficient choices for the attacker in conducting an assault on a defended position. While each of these points is *efficient*, in the sense that they offer the greatest speed available without an increase in casualties, they are not equally *desirable*. That is, it may be possible to conduct an assault at a velocity of 30 KPH that leads to the fewest casualties possible without slowing the speed down to 25 KPH, but that does not necessarily make higher casualties at 30 KPH preferable to lower casualties at 25 KPH. How are we to determine which of these alternatives constitutes the most desirable choice?

To answer this question, we must return to the overall dynamics described earlier. The attacker, in effect, is engaged in a race with the defender. He seeks to advance far enough to break through the defender's line before the defender can shift enough reserves to the point of attack to bring the attacker to a halt.[14] Speed is thus a virtue for the attacker. In this race, however, the efficient attacker can increase his speed only by increasing his losses. Yet a weakened attacker can be halted by a smaller quantity of defensive reserves. For a constant rate of defensive reserve arrivals, higher velocity will

---

[14] Since arriving reinforcements (as distinct from counterattackers; see discussion under "reserves," below) are not ordinarily moved directly into an ongoing firefight (to do so would be to risk their destruction in the open before they have any opportunity to dig themselves in), we will assume here that the attacker's casualties in any given assault are influenced only by the size of the defending force in place when the attacker initiates offensive action against the position in question. Thus, by slowing down, the attacker does not increase the number of defenders immediately opposing him; rather, he increases the number of defenders which *will* oppose him in his assault on the *next* defensive position.

thus allow the attacker to advance further in a given time but will also reduce the time required for the defender to move a sufficient force to the point of attack (since higher attacker velocity means higher casualties and thus fewer surviving attackers, and since a smaller reserve force will suffice to halt a smaller attacker). Alternatively, a slower attacker will cover less ground in a given time but will have more time in which to advance before the defender can deliver enough reserves to halt the larger force of surviving attackers.

We can therefore define a set of curves relating attacker velocity-casualty choices that yield the same advance distance prior to successful defensive counterconcentration (for a constant rate of defensive reserve arrival). A representative set of such "*iso-ground-gain*" curves is posited in Figure C-3. The direction of increasing ground gain is to the lower right of the plot—i.e., the direction of increasing velocity and decreasing casualties. Thus, curve IG_5 represents choices that yield a higher total ground gain than choices that lie on curve IG_4, and so on. A given iso-ground map is specific to a given rate of defensive reserve arrival. The effect of varying arrival rates is posited in Figure C-4, wherein IG_a depicts an iso-ground curve for the same ground gain, but a higher arrival rate than that of IG_b. In effect, the higher the rate at which reserves are arriving, the fewer the casualties an attacker can afford to suffer at a given velocity and still gain the same amount of ground; hence IG_a is below and to the right of IG_b.

For a given defensive reserve arrival rate, however, we will posit that iso-ground gain curves for non-zero ground gains are strictly concave, non-intersecting, and monotonic.[15] By way of motivation, let us assume an attacker velocity choice V_1, and associated casualty level C_1, that enables the attacker to advance a distance G in time t_1 before being halted by the arrival of defensive reserves. If we now consider a higher casualty level C_2, how much higher must be the new velocity choice V_2 if the new point (V_2, C_2) is to lie on the same iso-ground gain curve as the initial point (V_1, C_1)? Higher attacker velocity enables the attacker to cover the distance G in less time, i.e., at some $(t_1 - \Delta t)$. Fewer defensive reserves will have arrived by $(t_1 - \Delta t)$ than by t_1; hence the attacker can afford to incur higher casualties during the advance and still cover the same

---

[15] The special case of zero ground gain produces a vertical line at $V = 0$. We will exclude the possibility of negative ground gains. While a catastrophic assault could lead to ground loss for an attacker, for this to be an optimal outcome requires an assault (or the plausible threat of an assault) by the defender—a circumstance which we will accommodate by reversing the identities of the attacker and defender and applying the logic developed here for non-zero, positive ground gains (see the discussion of "counterattack," below).

distance G before being halted by this smaller defensive force. If the defensive reserves arrive at a constant rate, then an attacker that arrives at G two hours sooner than t_1 averts the arrival of twice as many defender reserves as an attacker that arrives at G one hour sooner. The attacker can therefore incur twice as many additional casualties during that advance and still reach G before being halted. Similarly, by arriving three hours sooner than t_1, the attacker averts the arrival of three times as many reserves as would have been averted one hour earlier, and he can therefore incur three times as many additional casualties as he could have had he arrived one hour earlier. Thus, the permissible casualty increase associated with a given Δt increases at a constant rate as Δt increases. To arrive Δt hours sooner, however, requires larger and larger increases in velocity the larger the Δt (for $\Delta t = t_1$, for example, the velocity increase required would be infinite; in effect, the attacker would have to arrive at G the same moment he sets out). Thus, to offset a constant increase in casualties ($C_2 - C_1 = C_3 - C_2$) via an increase in velocity requires an ever larger increase in velocity ($V_3 - V_2 > V_2 - V_1$). As this is true for any arbitrarily small difference in casualties ($C_i - C_j$) and associated difference in velocity ($V_i - Vj$), the result is the strictly concave shape of the curve given in Figure C-3a. Finally, since, ceteris paribus, a given attacker casualty level and velocity produces a unique ground gain, iso-ground gain curves thus cannot intersect. And since any increase in velocity will enable an attacker to cover a given distance in less time and thus avert the arrival of at least some defenders, any increase in velocity will permit a corresponding increase in casualties for the same ground gain. Thus iso-ground gain curves are monotonic in V and C.

By combining an iso-ground gain map with a given casualty-velocity trade-off frontier, we can determine an optimal choice of attacker velocity. For the purposes of this study, we have assumed that the attacker's objective is to take and hold opposing territory. This assumption enables us to define the iso-ground map as an iso-preference map for the attacker (or alternatively, an *indifference map* for the attacker). The attacker prefers any point on iso-ground curve IG_5 from Figure C-3 to any point on iso-ground curve IG_4, since IG_5 by definition represents points which produce higher net territorial gains than those lying on IG_4. If the casualty-velocity trade-off frontier represents the set of possible efficient choices, the attacker will thus maximize his net territorial gain by choosing the point on that frontier which intersects with the highest-value iso-ground gain

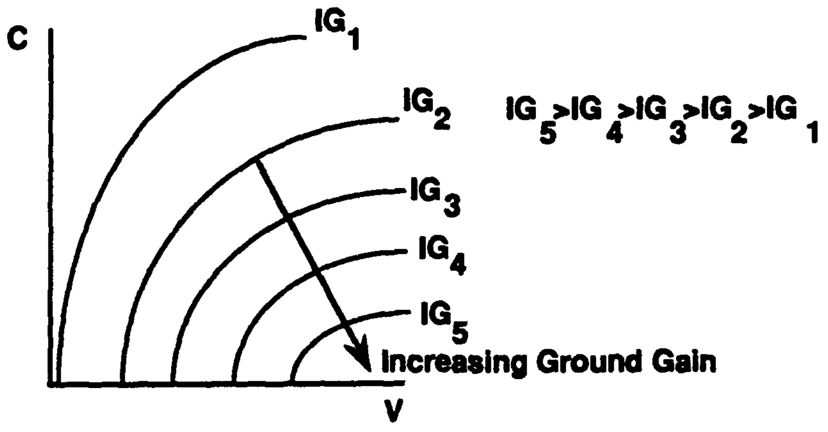

Figure C-3. Iso-Ground-Gain Curves

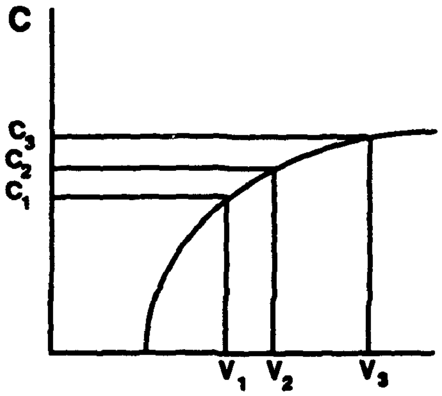

Figure C-3a. Iso-Ground-Gain Curve

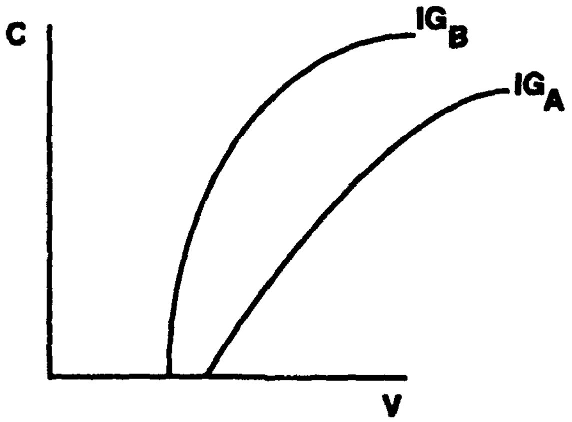

Figure C-4. The Effect of Increased Defensive Reserve Arrival Rate on the Iso-Ground-Gain Curve

curve. This will necessarily be at a point of tangency, producing an optimal choice as at point A in Figure C-5.[16]

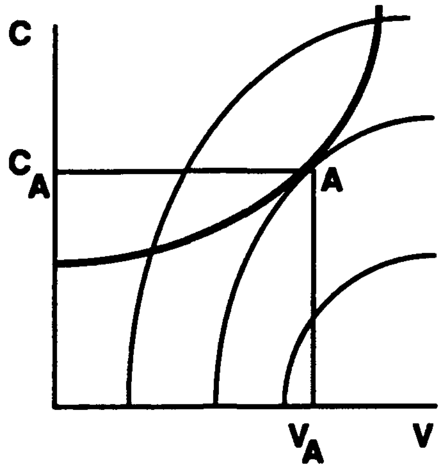

Figure C-5. Ground-Maximizing Velocity Choice

This approach of seeking points of tangency on an efficient frontier will serve as a useful heuristic for understanding and motivating the velocity treatment in the equations to follow, and for visualizing the impact of changes in weapon mixes and geography (as will be discussed in more detail below). We will find it computationally more convenient to treat the iso-ground map implicitly in the equations that follow, and the computer code developed to automate those computations conducts a search for an optimal solution

---

[16] This follows from our assumptions as to the strict convexity of the casualty-velocity frontier, the strict concavity of the iso-ground map, and the monotonicity and non-intersection of iso-ground curves. It further follows that there will be one and only one such point of tangency—and that zero velocity (and thus zero ground gain) will never be an optimal choice for red as long as there are forces available for use in attack. Of course, there may be many circumstances in which a potential attacker would be better advised to forgo attack and stand on the defensive rather than accept offensive losses which might make him vulnerable to a crushing counterblow. We will accommodate this condition, however, by constraining red to withhold reserve forces sufficient to defeat a blue counterattack. Only surplus forces beyond this self-defensive reserve can be allocated to attack. Circumstances under which red is better off standing on the theaterwide defensive will thus yield an offensive surplus of zero; any situation in which this surplus is non-zero is thus one in which red's ground-maximizing choice (for a single offensive operation) is to conduct an assault at the velocity given by the point of tangency between the casualty-velocity frontier corresponding to the local force to force ratio (and other conditions; see discussion below) and the highest iso-ground curve. Note, however, that this optimal choice is sensitive to the assumption that the time horizon for the decision is that of a single operation; if further such operations can be foreseen, the class of circumstances under which zero velocity constitutes red's optimal choice will be larger.

without explicit calculation of an iso-ground map per se. The functional relationship between velocity (or tempo) and net territorial gain described here, however, is ultimately that implemented in the equations and the associated code. That is, there is a trade-off between attacker casualties and velocity, and the ground-maximizing velocity choice will be that which best balances the conflicting demands of speed and conservation of force in light of the reaction of an active defender.[17]

E. DEFENSIVE DEPTH

What do we mean by "defensive depth?" Defenders rarely leave their entire force within firing range of the front line. A sizable fraction of that force is normally removed from direct contact with the enemy. These withheld forces can be used as mobile reserves to perform the counterconcentration function described above. Some, however, are typically employed in prepared defensive positions a modest distance behind the front line itself and are assigned the task of engaging (from those prepared defensive positions) attackers that penetrate the front line in their sector. This practice of distributing prepared defenses over some distance behind the front line constitutes a *defense-in-depth*, the area within which the attacker will encounter these prepared defenses can be defined as the *defended zone,* and the distance between the rear of this zone and the front line can therefore be defined as the *depth* of the defense.[18]

Depth, however, embodies a trade-off with mass. A commonplace among military theoreticians and doctrine writers is that forces should be concentrated at a decisive point—an injunction sometimes referred to as the "principle of mass."[19] Yet ceteris paribus, mass and depth are inversely related. For a defensive force of a given size (and given reserve withhold), depth can be obtained only by reducing the number of troops

---

[17] Of course, the treatment above does not constitute a formal proof that the relationships described above in two-space necessarily hold in n-space as well. We will not, however, attempt to provide such a proof here. As a result, we will not attempt a rigorous definition of the complete envelope of applicability for these relationships in n-space.

[18] In the literature, the distinction between depth for this purpose of prepared defenses (which lie behind the front line itself but are intended to deal with threats-in-sector) and "depth" in the sense of withheld reserves intended for employment across sectors is not always clearly differentiated (see appendix A). For our purposes, only the first of these will addressed be here; the issue of mobile reserves will be treated in detail under "Reserves and Counterattack," below.

[19] See, for example, Headquarters, Department of the Army, FM 100-5, Operations (Washington, D.C.: USGPO, 1982), p.B-2.

available to engage the enemy at any given point. Hence the only way to obtain depth is to reduce mass relative to that which could be achieved in the absence of depth.[20]

While the nature of the trade-off between depth and mass is thus straightforward—i.e., the two are inversely related—the nature of the optimal choice between the two is not. To understand this choice, and thereby to posit a functional form for relating the defender's depth choice to net territorial gain, we must begin by looking at the military effects of deep defenses on attackers.

1. Military Effects of Depth

Those effects are several fold. First, the loss of mass associated with the dispersion of defenders into depth can itself be of value in reducing the defender's vulnerability to offensive area fire weapons such as nuclear or conventional artillery. Although target planners strive to locate specific defensive targets as precisely as possible, artillery effectiveness is nonetheless quite sensitive to the overall density of defenders in the target area. Other things being equal, the greater the density of targets, the higher the number of kills per volley fired, and thus the more effective the preparatory barrage. The larger the threat from opposing artillery, the more important dispersion becomes for survivability. When the artillery threat is very great, this effect may outweigh the associated reduction in the number of direct fire weapons immediately available for opposing the attack.[21]

A second effect of depth is to reduce what might be termed the *coherence* of an attack. The effectiveness of an assault is strongly dependent on the attacker's ability to get the most out of the resources available to him. Artillery support has the *potential* to reduce offensive casualties—but only if it is properly coordinated with the local commander's maneuver scheme, properly directed against positions likely to hold defending troops, and in position itself and ready to fire at the required time. Suppressive fires must be maintained until the last possible moment, but lifted in time to permit assault elements to enter the objective without losses from friendly fire. Assault formations must

---

[20] We will consider below an exception to this principle in the form of withdrawal; as we shall see, however, the exception is only a partial one.

[21] This function of depth as a means of reducing the defender's vulnerability to artillery was especially significant in the latter years of World War I. See, for example, Timothy Lupfer, The Dynamics of Doctrine: Changes in German Tactical Doctrine During the First World War (Ft. Leavenworth KS: US Army Combat Studies Institute, 1981), Leavenworth Paper No.4, pp. 7-21; also G.C. Wynne, If Germany Attacks: The Battle in Depth in the West (London: Faber and Faber, Ltd., 1940; reprinted

be maintained in order to provide maximum firepower and prevent isolated attack elements from being overwhelmed by unexpected enemy action. Infantry and armor must be kept within mutual support distance in the face of enemy fire. Air defense systems must be maintained in positions capable of covering maneuver forces throughout their period of potential exposure.[22]

The attacker's ability to orchestrate (or, as the U.S. Army puts it, to *synchronize*)[23] these diverse elements is at its peak at the jump off point of an attack, when the benefits of advance planning and careful staging and positioning of components are closest at hand. As the attackers advance through the defense, however, it becomes increasingly difficult to maintain tight coordination of all these elements in the face of changing demands imposed by terrain, enemy forces, and the cumulative burden of unexpected events. Formations gradually spread out. Assault units lose contact with neighbors and with supporting artillery and air defense elements. Suppressive and preparatory fires must be extemporaneously arranged rather than pre-planned. Leaders and subordinates may lose communications, or subordinates may find themselves too busy for adequate reporting on rapidly unfolding events. As a result, the tight order and coordination characteristic of the early stages of an attack break down as the attack advances into depth.[24]

This property of a deep defense in promoting the gradual loss of order in attacking units might be termed the *entropic effect* of defense-in-depth: the further the attacker presses the attack, the less ordered that attack becomes, and as a consequence, the less effective the attacking forces become—*above and beyond* the effect of any casualties

---

by Greenwood Press, 1976), pp. 110-130; and Hew Strachan, European Armies and the Conduct of War (London: Allen and Unwin, 1983), p. 140.

[22] On the importance of coordination and close cooperation in the two World Wars, see Shelford Bidwell and Dominick Graham, Firepower: British Army Weapons and Theories of War, 1904-1945 (London: Allen and Unwin, 1985), esp. pp. 1-4. For a treatment extending through the present day, see Jonathan M. House, Toward Combined Arms Warfare: A Survey of Twentieth Century Tactics, Doctrine, and Organization (Ft. Leavenworth KS: US Army Combat Studies Institute, 1984), esp. pp. 188-90.

[23] See Headquarters, US Department of the Army, FM 100-5, Operations (Washington, D.C.: USGPO, 1986 edition), pp. 17-18.

[24] The World War I German doctrine of "elastic defense"—which laid the foundations for German defensive doctrine in both world wars—was particularly oriented to the exploitation of this effect (especially with respect to the separation of attacking armor and infantry). See Lupfer, op. cit., pp. 12, 13-16; Wynne, op. cit., e.g. pp. 142, 155-6; and Timothy A. Wray, Standing Fast: German Defensive Doctrine on the Russian Front During World War II (Ft. Leavenworth KS: US Army Combat Studies Institute, 1986), pp. 1-21. See also Paddy Griffith, Forward into Battle (London: Anthony Bird, 1981), pp. 79-80. For a British example, see Bidwell and Graham, op. cit., pp. 254-6.

taken by the attackers in the course of the advance. By extending the distance the attacker must cover in order to penetrate the defense, depth thus reduces (ceteris paribus) the coherence of the attack and, in the process, reduces its effectiveness.

Finally, a deeper defense buys time and provides a shield for the movement of defensive reserves to the threatened point. While within a defended zone, attackers are bound by the casualty-velocity trade-off described above. Thus speed comes only at a high price in increased casualties. Moreover, in a defended zone, opposed movement can be achieved only by massing offensive forces sufficient to defeat prepared defenders (i.e., to improve the local force-to-force ratio at least to the point where positive velocities are available as choices). This limits the attacker to advance along a few discrete axes on which these forces can be massed. If the attacker can break clear of the defended zone, however, these constraints on movement are greatly eased. Attackers can now accelerate without paying an immediate price in casualties—indeed, at this point in an operation higher velocity may well *reduce* casualties by making successful interception of the attacking spearheads harder for any surviving defensive reserves.[25] In addition to higher speeds, attackers now enjoy much greater freedom of direction. Without the requirement to mass forces in order to overcome prepared resistance (indeed, in this phase attackers typically bypass rather than engage any prepared positions they do encounter), attackers can advance along multiple axes simultaneously. In the process, they pose a serious threat to the command, logistical, air defense and transportation network upon which the remaining defensive forces in contact depend for continued resistance. This infrastructure is essential to the functioning of modern armies; once behind the defended zone, however, the attacker destroys infrastructure almost by the very act of moving through it.[26]

---

[25] J. F. C. Fuller and Basil Liddell Hart, for example, were particularly impressed with the potential of mechanized forces to reduce their losses through speed in exploitation. Fuller and Liddell Hart, however, overgeneralized this exploitation phase relationship between speed and casualties, leading to the argument that attackers should always opt for speed over preparation or firepower. The experience of World War II seriously undermined this hypothesis: see, e.g., Brian Holden Reid, "J. F. C. Fuller's Theory of Mechanized Warfare," Journal of Strategic Studies, December 1978, pp. 295-312; and J. F. C. Fuller: Military Thinker (New York: St. Martin's Press, 1987), p. 155.

[26] On the importance of the rear to the functioning of an effective defense, see, for example, J. F. C. Fuller, Lectures on F. S. R. III (London: Sifton Praed and Co., Ltd., 1932), pp. 85, 116. Perhaps the best known recent example of this phenomenon is Ariel Sharon's use of a token mechanized force to destroy the Egyptian SAM network on the West Bank of the Suez in the exploitation of the Israeli crossing of the canal in 1973. See Chaim Herzog, The War of Atonement (Boston: Little Brown and Company, 1975), pp. 231-251.

Early breakthrough is thus an important characteristic of a decisive attack. By forcing the attacker to advance further through defended territory prior to breaking through, a deep defense can thus help buy the time and protect the infrastructure needed to move reserves to the threatened point. By the same token, however, a deep defense necessarily reduces the force-to-force ratio at that threatened point in the meantime—thus allowing a relatively higher rate of advance over that distance than might otherwise have been the case. Moreover, the entropic effects of a deep defense can be countered by an attacker willing to replace spent echelons with fresh formations at frequent intervals, or who is simply willing to halt assault formations periodically to regroup. Clearly, depth is not an unmitigated virtue; how, then are we to determine the nature of the defender's optimal choice for the depth of his deployment?

2. Graphical Analysis

At this point, it will be useful to return to the casualty-velocity trade-off analysis described in the previous section. We will hypothesize that defensive depth affects this trade-off in two ways. By reducing the number of defenders within direct fire range of the attackers at any given time, depth increases the attacker:defender force-to-force ratio, and thus shifts the efficient frontier to the right. This phenomenon, termed the *mass effect* of depth, is illustrated in Figure C-6, in which CVF$_2$ represents the hypothesized effect of increased depth relative to that of CVF$_1$.

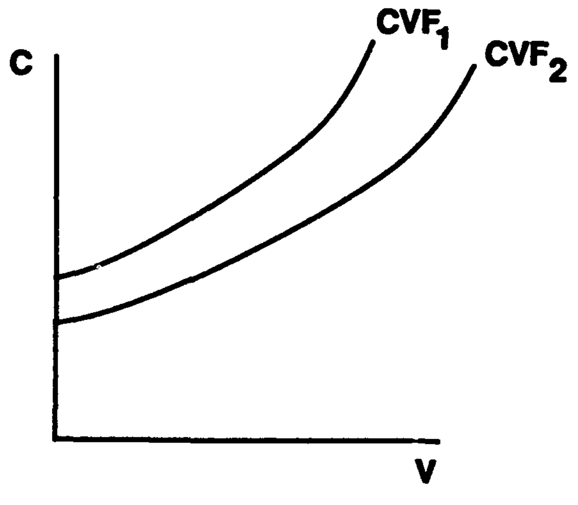

Figure C-6. The Mass Effect of Depth

To get onto the efficient frontier, however, the attacker must coordinate his available resources effectively. Depth reduces the attacker's ability to maintain the coordination, or coherence, of his forces the further he advances into that depth. The attacker can reduce this gradual loss of coherence, but only if he halts his formations periodically to reorganize or replaces spent formations with fresh units from the rear. Either of these counteractions requires delays in the attacker's advance. In effect, an attacker who wishes to maintain a high velocity under a condition of increased defensive depth must therefore accept a less coherent attack and thus a higher level of casualties in a given assault (relative to an attack against a defense of zero depth but equal local mass). If the attacker wishes to maintain coherence as he advances into depth, he must accept a slower average velocity. Depth should thus increase the slope of the average casualty-velocity trade-off frontier for engagements conducted during penetration of a defense-in-depth. This phenomenon, termed the *entropy effect* of depth, is illustrated in Figure C-7, in which CVF_B represents the hypothesized effect of increased depth relative to that of CVF_A.

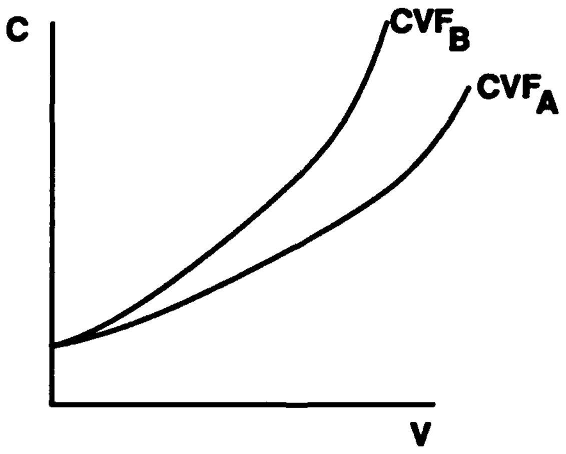

Figure C-7. The Entropy Effect of Depth

The net impact of an increase in defensive depth on the nature of the trade-off between casualties and velocity ought thus to be the sum of the mass effect and the entropy effect of the change. This net impact is illustrated in Figure C-8, in which CVF_2 represents the hypothesized effect of increased depth relative to that of CVF_1. At low velocities, the attacker is able to mitigate the effect of entropy, while enjoying the benefit of a higher local force-to-force ratio as a result of the reduced mass associated with an increase in defensive depth. The result is a reduction in the casualties associated with a given assault velocity for low values of *V*. At high velocities, the attacker still receives the benefit of a higher local force-to-force ratio, but is more and more exposed to the entropy effects of depth as *V* increases. As a result, the higher the velocity, the more the mass effect is counterbalanced by the entropy effect. At some *V'*, the two curves intersect.[27] Consequently, for velocity choices greater than *V'*, increased depth increases attacker casualties; while for velocity choices less than *V'*, the same increase in depth decreases attacker casualties.

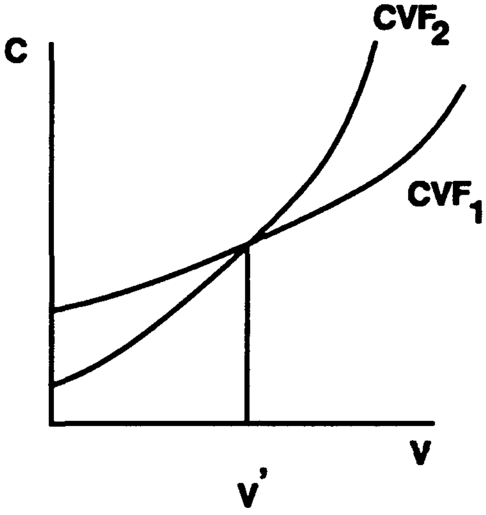

Figure C-8. The Net Impact of Increasing Depth

Figure C-9 completes the analysis by illustrating the result of these changes in the casualty-velocity trade-off on the attacker's optimal choice of velocity and consequent net territorial gain. The increased depth associated with CVF_2 relative to CVF_1 results in a decrease in optimal velocity from V_1 to V_2, a corresponding decrease in casualties from C_1 to C_2, but a net reduction in ground taken from IG_2 to IG_1. Note, however, the result of the attacker's freedom to choose his assault velocity: if the attacker is given the same

---

[27] Given the conditions posited here, there will be one, and only one, point of intersection *V'*. This follows from four properties of the posited relationship between velocity, casualties and depth: first, that the mass effect and the entropy effect of depth are additive; second, that the slope of the casualty-velocity frontier will be greater for defenses of greater depth (as implied by the entropy effect); third, that the zero-velocity casualty level will be lower for defenses of greater depth (as implied by the mass effect); and fourth, that the casualty-velocity tradeoff frontier is strictly convex, regardless of depth.

velocity V_1, but defensive depth increases to that associated with CVF_2, the loss in net territorial gain is much more substantial (i.e., he is now on IG_0 rather than IG_1).[28]

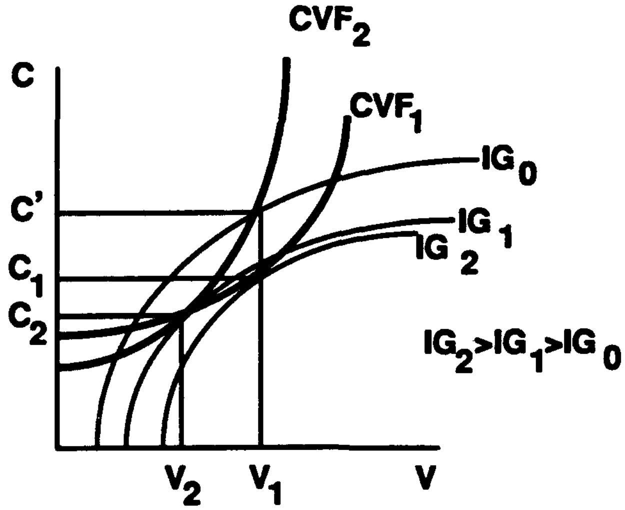

Figure C-9. Effect of Increased Depth on Optimal V and G

3. Static Depth, Rolling Depth and Withdrawal

A final point must be made regarding different forms of depth. The depth of a defense can be increased in two ways: by extending static pre-deployment of forces further into the rear, or by "rolling" depth via withdrawal. In rolling depth, defenders in a given engagement do not merely stand and fight in their given positions, but rather disengage from the attacker at some point and withdraw to the rear. They then either prepare new defenses immediately behind the initially defended zone, reinforce pre-deployed forces to the rear of the positions from which they were withdrawn, or occupy empty pre-existing positions to the rear of the ongoing engagement. Voluntary withdrawal thus enables a defender to overcome to some degree the mass trade-off inherent in establishing depth by static predeployment. Since the same forces occupy successive positions in depth, withdrawal permits a greater ultimate depth than exists at the beginning of the

---

[28] Note also that the relationship between the tangency points and corresponding optima given here are specific to the iso-ground map and casualty-velocity frontier depicted here, and are thus intended to be illustrative rather than definitive in nature.

attack, and enables a given force to establish greater depth for the same mass than could be achieved through static pre-deployment alone.[29]

Defensive withdrawal, however, trades off against reduced attacker casualties in a given engagement. By comparison with an in-place fight to the finish, withdrawal frees some defenders to fight again further to the rear, but it also spares some attackers who would otherwise have been killed by those defenders had they stayed to fight.[30]

This trade-off is hypothesized graphically in Figure C-10, in which normalized attacker casualties, C (i.e., attacker casualties given the defender's withdrawal, divided by attacker casualties assuming no withdrawal) are plotted against the fraction of the initial defending force withdrawn, W. For a fight to the finish in which no defenders are withdrawn, $W = 0$ and attacker casualties are maximized (of course, so are defender casualties). At the opposite extreme, since killed defenders cannot be withdrawn to add depth to the defense, the only way to withdraw all the initial defenders (i.e., to obtain $W = 1$) is to withdraw prior to combat,[31] in which case $C = 0$. As for points in between, if the loss exchange ratio (LER = attacker casualties divided by defender casualties) were constant throughout the engagement, the result would be a linear relationship bounded by the points (1,0) and (0,1). Ordinarily, however, the tactical LER is highest early in an engagement, when the range between attackers and defenders is longest, and lowest at the end of the engagement, when attackers have closed the range.[32] If this is true, then early withdrawal—prior to extensive defensive casualties (i.e., a high value of W)—will take

---

[29] Most historical defenses in depth relied on a combination of predeployed and rolling depth. For the paradigmatic example of the form, see the descriptions of the German elastic defense in Lupfer, op. cit., pp. 13-16; Wynne, op. cit., pp. 150-158; Wray, op. cit., pp. 3-5. In fact, the appropriate degree of reliance on withdrawal for this purpose—and the best balance between, in effect, predeployed and rolling depth—was an issue of considerable debate within the German General Staff in the First World War. See for example, Wynne, op. cit., pp. 158-163; Lupfer, op. cit., pp. 3-4, 21-22. On the Eastern front in the Second World War, Hitler's "stand fast" (i.e. no withdrawal) directive was of course extremely unpopular with the officer corps: see, e.g., the exchange described in Earl F. Ziemke, Stalingrad to Berlin: The German Defeat in the East (Washington, D.C.: US Army Center of Military History, 1987), p. 180; Wray, op. cit., pp. 118-123.

[30] As noted, for example, in Joshua Epstein, The Calculus of Conventional War: Dynamic Analysis Without Lanchester Theory (Washington, D.C.: Brookings, 1985), pp. 4-6, 16-18.

[31] While it may often be possible for the defender to pick off some attackers at long range without suffering loss, then withdraw intact, it will never be possible to guarantee this. Over many such engagements, some defending casualties will occur, with the result that while the statistical expectation value for defensive casualties may be quite small for such circumstances, it will never be identically zero unless the defender declines to offer battle and withdraws prior to firing upon the attacker.

[32] See, for example, General William E. DePuy, "Technology and Tactics in Defense of Europe" Army, April 1979, pp. 15-23.

place at a high ratio of attackers to defenders killed. Hence the 10 percent defender casualties represented by $W = .9$ will produce more than 10 percent of the attacker's total casualties, with the result that $W = .9$ implies $C > .1$. Since the ratio of attacker to defender losses decreases monotonically as the engagement proceeds,[33] the result is that normalized attacker casualties will always be greater than normalized defender casualties $(1 - W)$ until termination ($W = 0$), and the shape of the curve will be as hypothesized in Figure C-10.

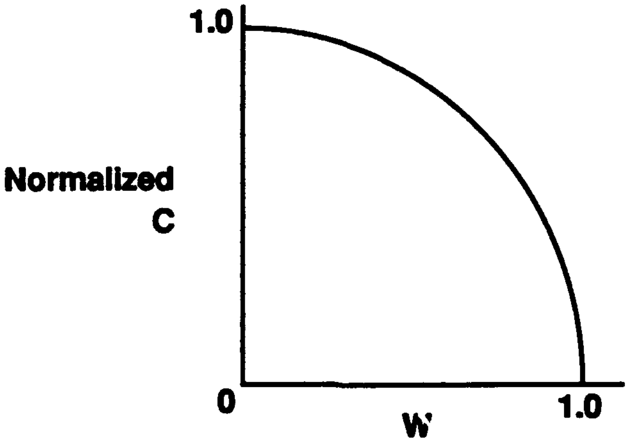

Figure C-10. Trade-off Between Defensive Withdrawal and Attacker Casualties

Voluntary withdrawal and static predeployment together determine the ultimate depth of the defended zone and, hence, the nature of the optimal attacker choice between velocity and casualties, as described above. Each represents an opportunity for defender choice, and together offer a range of possibilities varying from a shallow initial deployment with a very high fraction withdrawn (approximating a traditional delaying action) to an extremely deep but static defense (resembling many proposals for non-provocative, or non-offensive defense), to a combination of moderate depth with provision for gradual withdrawal under pressure (in effect, the German doctrine of elastic defense).

---

[33] Ibid., see esp. Figure 1, p. 19

F. DEFENSIVE RESERVES AND COUNTERATTACK

1. The Defender Reserve Fraction

What constitutes a "defensive reserve," and how is such a reserve used? With respect to the first of these questions, as we observed above, defenders rarely deploy their entire force within immediate firing range of the enemy. However, not all forces beyond direct fire range constitute "reserves" in the sense that the literature on force-to-space ratios uses the term. In the literature, a *theater reserve* is a force that is capable of lateral movement (i.e., parallel to the line of battle) over significant distances and that, hence, is well suited to the counterconcentration mission discussed above. Theater reserves are thus distinct from pre-deployed defenders-in-depth in three respects. First, reserves are typically located further to the rear[34] and held in *assembly areas* rather than being deployed in prepared defensive positions per se.[35] Second, reserves are organizationally

---

[34] Forces held too close to the front risk exposure to enemy fire and consequent restriction of safe movement. Indeed, the principal reason for retaining a "subtracted reserve" has traditionally been the difficulty of disengaging forces within artillery (or worse, direct fire) range of the enemy (see, e.g., Archer Jones, "The New FM 100-5: A View From the Ivory Tower," Military Review, Vol.58, No.2, February 1978, pp. 27-36). Defenders ordinarily seek protection from enemy fire by digging themselves in or locating behind cover. Movement often requires leaving this cover in the face of a superior enemy who may be pressing forward at the time of the attempted move. Such movement—whether for the purpose of lateral counterconcentration or merely in order to withdraw to the immediate rear—is thus very risky. While disengagement can of course be accomplished under the right circumstances, it is not a trivial operation.

[35] Forces deployed in prepared defenses are often ill-suited to immediate long distance movement—even if they are beyond immediate enemy fire—since they are distributed over (and into) the ground in accordance with the defensive potential of the terrain rather than to facilitate rapid organization of an orderly march column. The task of assembling a unit for a long road march can be a substantial undertaking; to organize a brigade-size unit into movement formation might require as much as two hours even if that unit begins in an assembly area designed specifically to facilitate this task, and even if we exclude the planning and communications time required (see Statement of General Fred K. Mahaffey in Department of Defense Authorization for Appropriations for Fiscal Year 1981, Hearings Before the Committee on Armed Services, United States Senate, Ninety Sixth Congress Second Session, Part 5 (Washington, D.C.: USGPO, 1980), pp. 3023-3039; on the complexity of the planning effort involved in cross-corps, counterconcentration-type movement, see Colonel Ted A. Cimral, "Moving the Heavy Corps," Military Review Vol.68, No.7, July 1988, pp. 28-34). This delay would be much greater for units deployed in prepared defensive positions. Reserves are therefore deployed in assembly areas well back from the immediate fighting and organized with a transition to road march in mind, rather than the defense of their position per se. (Like any military force, reserves in assembly must be concerned with security and are therefore positioned to provide for defense against unexpected attack; they are not, however, ordinarily dug into mutually supporting positions in depth while in assembly areas.) For more detailed discussions of the purposes and organization of assembly areas, see Headquarters, Department of the Army, FM 7-20, The Infantry Battalion (Washington, D.C.: USGPO, 28 December 1984), pp.D19-21; also, Headquarters, Department of the Army, FM 71-2, The Tank and Mechanized Infantry Battalion Task Force (Washington, D.C.: USGPO, 30 June 1977), pp.H8-9.

distinct from forward, "engaged" units with a defense-in-depth mission and are normally held at higher echelons of command. Finally, theater reserves are used primarily for lateral, long distance movement, whereas defenders-in-depth typically fight where they are initially deployed.[36]

As with defensive depth, the size of the defender's reserve is ultimately a choice made by the defending commander. In effect, all defenders must partition their available forces between forward elements and reserves. A given defender could conceivably choose to hold in reserve anything between zero and 99 percent of whatever force he has (with the balance deployed forward). How are we to evaluate this choice?

We will begin by formulating the choice in terms of a tradeoff between mass and response time. The first defenders an attacker will encounter will be those which are forward deployed. Only a fraction of these forward deployed defenders will be located opposite the attacker's main effort, however. The remainder will face only pinning (or "fixing") attacks designed to complicate disengagement of those forces (and to mask the location of the attacker's main effort). If the attacker advances on a narrow front relative to the length of the theater, the great majority of the defender's forward forces will therefore be occupied away from the attacker's main effort.

Defensive reserves, by contrast, must move from their assembly areas to the point of attack before they can be committed to action; thus, they are not immediately available. Reserves, however, retain freedom of maneuver by virtue of their distance from the front lines. As a result, they are not susceptible to pinning operations in the way forward forces are, and they can direct their entire combat power against the critical sector.

Forward defenders thus offer immediate availability, but at the price of directing only a fraction of their total mass against the threatened point. Reserves throw their entire mass against the threatened point but reach that point only after a possibly extended movement delay. In effect, the defender must trade time of availability against mass for a given overall force size.

---

[36] Of course, units at any level of command can be held out of combat as a "mobile reserve" by the local commander. The difference between these local assets and a theater "reserve" in the sense addressed by the force to space ratio literature lies in the ability of the withheld units to cross higher formation boundaries in lateral movement for the purpose of counterconcentration. Thus, a single company withheld by an engaged battalion's commander does not contribute to the defender's theater reserve in this sense, since it is unlikely that this company would be split away from its parent battalion and moved perhaps hundreds of kilometers to counterconcentrate against a distant point of attack. A division withheld by a corps commander, by contrast, *could* be employed in this way.

2. Graphical Analysis

The posited trade-off is depicted graphically in Figure C-11, which plots defender force size at the point of attack, *B(t)*, against time, *t*, as a function of the fraction of total defensive forces that are held in reserve, f_r. The analysis in Figure C-11 assumes that the arrival rate of defensive reserves at the point of attack is proportional to their quantity, that only a fixed fraction of total forward forces are located at the point of attack, that all reserve forces ultimately reach that point of attack, and that all defender forces are either "forward" or "reserve." The implication of this analysis is that the defender can increase the ultimate mass of forces at the point of attack by holding a larger reserve (i.e., by increasing f_r), but that to do so he must accept a lower initial force mass.[37]

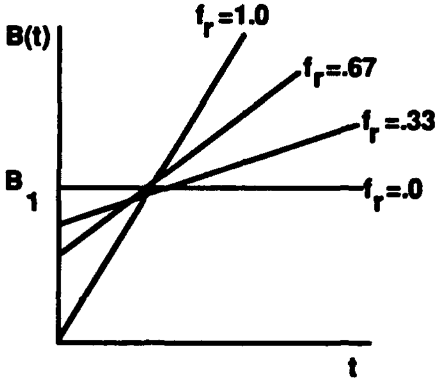

Figure C-11. Trade-off Between Defensive Mass and Availability as a Function of Reserve Fraction (f_r)

The implications of this trade-off for the attacker's choice of velocity and the resulting net territorial gain are given by Figure C-12. The initial decrease in the size of the defensive force at the point of attack associated with an increase in the defender's reserve fraction will produce a corresponding increase in the local attacker:defender

---

[37] While Figure C-11 portrays the relationship between B(t) and t as linear, this need not necessarily be the case, and this condition is not strictly necessary for the analysis that follows. One could imagine, for example, that non-uniform distributions of reserve assembly areas with respect to the location of the red point of attack could produce blue buildup rates at that point of attack that would be faster (or slower) for low values of t than for high values—or that imperfect communications could result in disproportionate delays in the arrival of the last few blue units. While the particular formulation posited for reserve arrivals in the equations to follow assumes linear arrivals, it could readily be modified to reflect alternative assumptions such as these, and the depiction given in figure C-11 is not intended to rule out such possibilities.

force-to-force ratio. We have hypothesized above that such an increase will shift the casualty-velocity trade-off frontier to the right, from cvt_1 to cvt_2 in Figure C-12. An increased reserve fraction will simultaneously increase the defender's arrival rate (recall the analysis in Figure C-11). An increased arrival rate, however, will shift and flatten the associated iso-ground curve since for a given attacker velocity—and thus a constant elapsed time to reach a given advance distance—the defender will now have delivered more forces to the point of attack. Only if these additional defenders are balanced by a reduction in the attacker's casualties for that velocity can the attacker expect to achieve the same total advance. The slower the attacker's chosen velocity, the larger the required reduction in casualties to obtain a constant advance distance, since the total time to reach that distance is longer for slower velocities, producing a correspondingly larger increase in defensive arrivals (since defenders arrive at a constant rate). Consequently, we may posit that an increase in the defender's reserve fraction will shift the iso-ground curve for a given net territorial gain from IG_1 to IG_2, producing an increase in the optimal attacker velocity from V_1 to V_2, and (in this case) an increase in optimal attacker casualties from C_1 to C_2 for the same total advance distance.

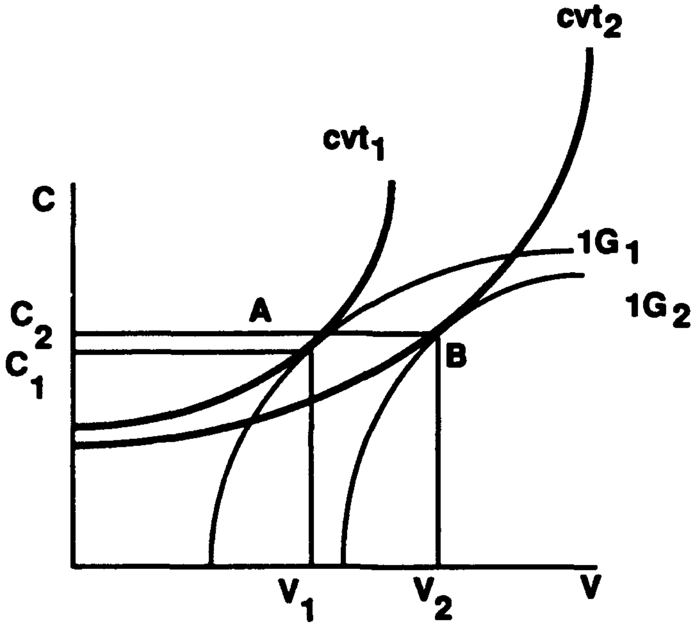

Figure C-12. Casualty-Velocity Implications of Increase in Defender Reserve Fraction

Of course, net territorial gain need not necessarily be constant with respect to change in the defender's reserve fraction (if it were always so, the issue of reserve size would be irrelevant to theater combat outcomes). Whether a given increase in defender reserve fraction will produce a net increase or decrease in attacker advance distance will depend on external circumstances (such as the ratio of the attacker's assault frontage to the length of the theater, or the speed with which the defender can prepare his reserves for movement). For any given set of such circumstances, however, we can determine an optimal defender reserve fraction by recomputing the equations describing the curves in Figure C-11 and then applying the resulting trade-off to the attacker casualty-velocity trade-off as shown in Figure C-12 for various reserve fractions. The defender will choose the reserve fraction which offers the lowest attacker-optimal ground gain solution (see the equations below for a more detailed implementation).

2. Counterattack

For a given rate of reserve arrivals, however, how should those arrivals be used? Two broad alternatives exist. Arriving reserves can be used (in much the same way as withdrawn forward defenders) to reinforce prepared defenses or to extend the depth of the existing defended zone. The impact of such employment is essentially equivalent to that of an increase over time in the strength of the forward defenses at the point of attack. Reserve arrivals can also be used, however, to counterattack.[38] The allocation of arriving reserves between reinforcement and counterattack roles thus constitutes a second defender choice with respect to the general issue of reserve employment; how are we to evaluate this choice?

We must begin by considering the basic dynamics of defensive counterattack. In particular, we can think of counterattack as a special case of the attack in general. Recall that in the standard offensive, the theater defender, Blue, partitions his forces between forward and reserve and initially distributes the forward forces uniformly across the front he judges to be susceptible to attack. The theater attacker, Red, concentrates against a fraction of that front and attempts to break through while Blue redistributes his forces by moving reserves to the point of attack. Note that as Red advances, however, he creates

---

[38] In this sense we refer to large scale, deliberate counterattacks (what the Germans called the *gegenangriff*). Smaller scale or hasty counterattacks (*gegenstoss* in German usage) may be executed by any echelon of command and do not necessarily require the massing of forces over large distances. We will focus here primarily on the former. For a more detailed treatment of the distinction and its significance, see, for example, Wynne, op. cit., pp. 97, 152-58; Wray, op. cit., pp. 87-8, 167-171.

for himself a *defensive* front along his flanks—where the theater defender may choose to counterattack the theater attacker.[39] Red's problem of flank defense, however, is directly analogous to Blue's problem of initial defense albeit on a smaller scale. Red does not know where Blue's counterattack will fall and, hence, must deploy more or less uniformly along his flank until Blue's point of counterattack is known. Blue concentrates his counterattack forces on a narrow front to achieve a high local force-to-force ratio against a fraction of Red's flank defenders, and attempts to break through that flank defense before Red can move enough reserves to the point of counterattack to thwart Blue's advance.[40]

While the basic dynamics of attack and counterattack are thus quite similar, the special case of counterattack is different in several crucial respects. In particular, Red faces a number of constraints in flank defense that Blue does not face in initial defense. Blue, for example, can choose whatever defensive depth and withdrawal fraction best suits his goal of minimizing attacker net territorial gain. Red, on the other hand, is defending the flanks of a narrow penetration corridor. Red's flank defenders are thus limited to an ultimate defensive depth no greater than the width of his attack sector. In fact, the *available* depth is even less because Red must maintain a clear channel through the center of this penetration corridor wide enough to support a high volume of supply and troop movements if he is to keep his assault spearhead moving forward.[41]

---

[39] Although the flank is ordinarily the attacker's most vulnerable point, this need not necessarily be so; for a more detailed discussion of the issue of selecting a target point for counterattack in the context of operations on the Eastern Front in World War II, see Department of the Army Historical Study No. 20-233, German Defense Tactics Against Russian Break-Throughs (Washington, D.C.: US Army Center of Military History, 1984 reprint of 1951 orig.), pp. 3-14. As a point of departure, however, we will assume here that counterattacks can best be directed against the theater attacker's flank.

[40] On the historical and theoretical importance of counterattack and, conversely, of the invader's capacity to defend against counterattack, see, e.g., Clausewitz, op. cit., e.g., Book VI, Chapter 1, pp. 357, 358, Book VI, Chapter 5, p. 370, Book VI, Chapter 8, p. 380, also Book VI, Chapter 9, p. 392; Jomini, op. cit., e.g., pp. 103, 104; Ritter Wilhelm von Leeb, Defense, edited and translated by Stefan T. Possony and Daniel Vilfroy (Harrisburg, PA: The Military Service Publishing Company, 1943 edition of 1938 original), pp. 41-3, 49-50, 54-5, 99, 111, 116-8, 121; J. F. C. Fuller, Lectures on F.S.R.III op. cit., e.g., p. 117; also "What is an Aggressive Weapon?" English Review, June 1932, pp. 601-5; Basil H. Liddell Hart, The Defence of Britain op. cit., p. 121; also The Liddell Hart Memoirs, Vol.I, 1895-1938 op. cit., pp. 166, 221, 243; Department of the Army Historical Study No. 20-233, op. cit., pp. 3-14; Wray, op. cit., e.g., pp. 3-6, 10-16, 18, 25-33, 39-48, 86-89, 93, 117-8, 138-9, 146-50, 156-161, 167-172, 175-6; House, op. cit., e.g., pp. 26-7, 98-9, 102, 127; Lupfer, op. cit., e.g., pp. 15-21, 55-56; Wynne, op. cit., e.g., pp. 147-58, 191ff., 291ff.; Headquarters, Department of the Army, FM 100-5, op. cit., e.g. pp. 134-6, 139-141.

[41] On the implications of penetration frontage for flank defense, see, for example, Department of the Army Historical Study No. 20-233, op. cit., p. 13; also Wray, op. cit., pp. 148-51 and 152.

The geometry of the penetration corridor also constrains Red's ability to move reserves to the threatened point. Whereas Blue reserves can approach the attack sector from many directions and can thus exploit a wide range of different routes to reach the same point, Red reserves can reach the point of counterattack only by moving along the narrow channel down the center of Red's penetration corridor. Road availability is thus constrained for Red (and the available roads will already be in demand for resupply of the assault spearhead), with the result that Red's reserve arrival rate is likely to be both lower than Blue's and more severely constrained by road capacity than by reserve availability per se.

Finally, the Red flank defender faces preparation time constraints not faced by the Blue initial defender. Blue owns in peacetime the territory that he will have to defend. Although few nations will exploit this advantage to its theoretical limit, in theory, Blue thus has a near-infinite amount of time to site and fortify weapon positions, clear fields of fire, emplace barriers and obstacles, and familiarize defending troops with the ground on which they will fight and the potential approach routes over which they would be attacked. Red, by contrast, can only begin to prepare his positions once they have been taken from Blue by assault. As a result, Red must also prepare these positions under fire (at least for those positions within range of Blue artillery), whereas Blue enjoys the opportunity to employ vulnerable engineering equipment and exposed laborers in peacetime without the danger of enemy fire. Thus, on average, Red's flank defenders will have to prepare in less time and under more difficult circumstances than will Blue's initial defenders.

The consequences of a successful Blue counterattack, moreover, are potentially severe for Red. If Blue breaks through with sufficient force to seal off the Red spearhead from resupply or reinforcement, Red is left in an extremely vulnerable position. Much of Red's combat power in the main attack sector is typically concentrated forward. This concentration of combat power is now cut off from resupply of munitions, fuel or food; surrounded by hostile forces and thus required to spread its resources to cover the possibility of attack from any direction; and denied a safe retreat route to the rear. An isolated, immobilized Red spearhead surrounded by hostile forces is hardly in a strong position to continue its advance. More important, it grows weaker over time simply by virtue of its isolation from resupply and thus becomes increasingly vulnerable to further pressure from Blue air or ground forces. A successful counterattack thus threatens Red with the annihilation of the assault force he had concentrated for the initial attack; at a minimum, a counterattack that breaks through compels Red to halt his offensive while spending valuable time dealing with the threat to his rear.[42]

Other things being equal, then, counterattack is thus easier than attack, since flank defense is harder than initial defense. But if these are the basic dynamics of counterattack, then what does this tell us about the defender's choice between counterattack and reinforcing roles for his arriving reserves? And how should we address the related question of the theater attacker's decision as to the quantity of his force to devote to flank defense?

With respect to the theater attacker, flank defenders are an overhead cost. The lighter the flank defense, the more force he can devote to the assault spearhead which actually takes ground and advances his offensive. On the other hand, the consequences of being cut off are so grim that it can never be an optimal choice for Red to leave his flanks so thinly guarded that a Blue counterattack breaks through. The optimal choice for Red will therefore always be to allocate just enough force to flank defense to prevent Blue from breaking through. Allocating more means accepting unnecessary overhead costs; allocating less means accepting the isolation and possibly the annihilation of his assault spearhead.

What is this minimum flank defense for the prevention of a Blue breakthrough? This of course depends on the size of the Blue counterattack force, which brings us back to the issue of the defender's allocation of arriving reserves between counterattack and reinforcement roles.

We can now define Blue's choice as an optimal allocation problem in which the defender has a fixed force available as a result of reserve arrivals, and must allocate that force between counterattack and reinforcement so as to maximize return to an objective function. Moreover, the defender's objective function can now be simplified from the broader goal of minimizing the theater attacker's net territorial gain to the narrower goal of using a fixed force *to remove the largest number of attackers from availability for continued assault*. Since force size and arrival rate are now fixed, and since the location of the threatened sector is now known with certainty, the defender's choice boils down to

---

[42] For a detailed treatment of the difficulties imposed by encirclement, and of the requirements for effective break-out of encircled forces, see Department of the Army Historical Study No.20-234, Operations of Encircled Forces (Washington, D.C.: USGPO, 1952), esp. pp. 65-6; on the latter point, see also Headquarters, Department of the Army, FM 71-3, op. cit., pp. 5-7 to 5-8.

choosing the allocation that most reduces the size of Red's assault force. Counterattack and reinforcement reduce that assault force in different ways, but each accomplishes this same end. Counterattack accomplishes this end by forcing Red to divert potential assault forces to flank duty. Reinforcement accomplishes this end by killing Red assault forces as they attack the reinforced position.

3. Graphical Analysis

Which allocation, then, removes the largest number of attackers from the assault? Figure C-13 illustrates the nature of the optimal allocation in terms of the changing marginal value of counterattack and reinforcement as a function of the fraction of available forces devoted to each role. For each option, marginal value, *MV*, is defined as the decrease in Red assault force size for an arbitrarily small increase in f_{ca}, the fraction of defending reserves allocated to counterattack.[43] Since the fraction allocated to counterattack and the fraction allocated to reinforcement sum to one, Figure C-13 gives *MV* as a function of allocation in terms of f_{ca}; for a given f_{ca}, the fraction allocated to reinforcement is simply the complement of f_{ca}.

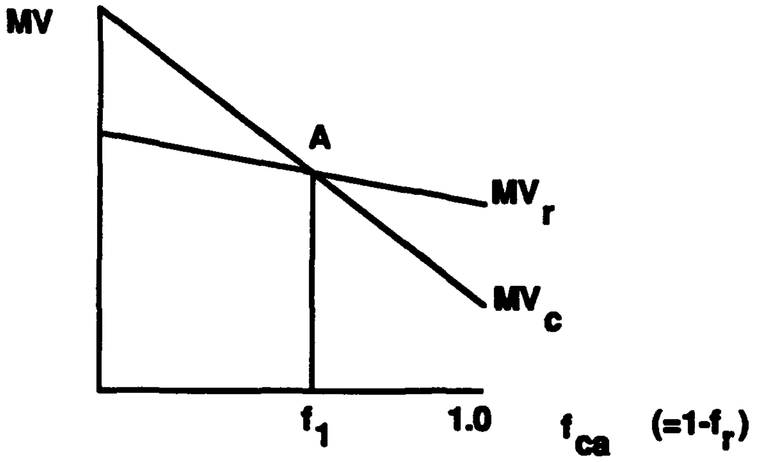

Figure C-13. Optimal Allocation of Defender Reserves Between Counterattack and Reinforcement

---

[43] Or, more formally, we may posit $MV = -dR/df_{CA}$, where R = Red assault force size, and *f* denotes the fraction of Blue reserves allocated to the given role.

For reinforcement, MV as defined here amounts to the attacker:defender loss-exchange ratio. This exchange ratio will typically increase as the defender deploys larger forces against a constant attacker; hence we would expect MV_r, the marginal value associated with an arbitrarily small change in f_{ca}, to increase as f_{ca} decreases.[44] We will therefore posit that MV_r is highest at low values of f_{ca} and lowest at high values of f_{ca}.

With respect to counterattack, MV as defined here is the number of troops Red must divert to flank defense per Blue counterattacker per kilometer of flank, times the distance the attacker has advanced (and hence the number of kilometers of flank to be defended). Since Red does not know where the counterattack will strike, an increase in the size of Blue's counterattack threat must be met by an increase in the density of flank defenders along the entire length of the flank. Consequently, the overall impact of a given increase in Blue's counterattack force is more severe the longer the Red flank it threatens. Diversions of Red forces from assault to flank defense decrease Red's ability to take ground, however, and thus decrease the length of Red's flank. Thus the larger the Blue counterattack force, the shorter the flank Red must defend, and the smaller the marginal value of further increases in counterattack size for Blue. As a result, we will posit that MV_c, the marginal value of an arbitrarily small increase in counterattack size, is highest for low values of f_{ca}, and decreases as f_{ca} increases.

Given these marginal value curves, we can find the optimal allocation between the two roles by observing that the optimal solution will generally be one for which the marginal value of the two alternatives is equal[45]—which implies that the optimal point is

---

[44] Basic lanchester theory, for example, suggests that for all but pure linear law conditions, the attacker:defender loss-exchange ratio will be inversely related to the local attacker:defender force-to-force ratio. See, e.g., Alan F. Karr, "Lanchester Attrition Processes and Theater-Level Combat Models" in Martin Shubik, (ed.), The Mathematics of Conflict (New York: Elsevier, 1983), pp. 89-126; also James G. Taylor, Lanchester Models of Warfare, 2 Vols., (Arlington VA: Operations Research Society of America, 1983), esp. Vol I, pp. 159-66. Lanchester theory, of course, has serious limitations—among them being that it displays no diminishing marginal return to very high force concentrations (see Joshua M. Epstein, The Calculus of Conventional War: Dynamic Analysis Without Lanchester Theory (Washington, D.C.: Brookings, 1985), pp. 11-12). If we assume, however, increasing marginal returns to defensive force size for low defensive force concentrations and diminishing marginal returns for high defensive force concentrations, then the conclusions given below continue to hold—the only difference being that the MV_r curve in Figures C-13 to C-15 will turn upward after some given value of f_{ca} (which of course implies different specific values for the optimal allocation, although the nature of the optimum and its behavior as force levels change will be the same).

[45] For examples from optimal allocation issues that arise in microeconomic theory, see, e.g., Walter Nicholson, Micro Economic Theory (Hinsdale, IL: Holt, Rinehart, and Winston/Dryden Press, 1978), pp. 74-76, 528-29. For exceptions, see below.

described by the intersection of MV_c and MV_r, at point A in Figure C-13 (producing a fractional allocation to counterattack of f_1). For the marginal value curves given in Figure C-13, any allocation point for which the marginal values are not equal would produce lower total value. Total value for the optimal allocation f_1 is given by the shaded area $CA + R$ in Figure C-14A. Region CA under the MV_c curve between 0 and f_1 represents the total value derived from counterattack; the region R under the MV_r curve between f_1 and 1 represents the total value from reinforcement. If Blue underallocates to counterattack by choosing point f_2 in Figure C-14B, his total value would be $CA_2 + R_2$; this area is smaller than that for f_1 in Figure C-14A by the area of the unshaded triangle L_2. L Blue overallocates to counterattack by choosing point f_3 in Figure C-14C, his total value is again lower than that of Figure C-14A by the area of the unshaded triangle L_3.

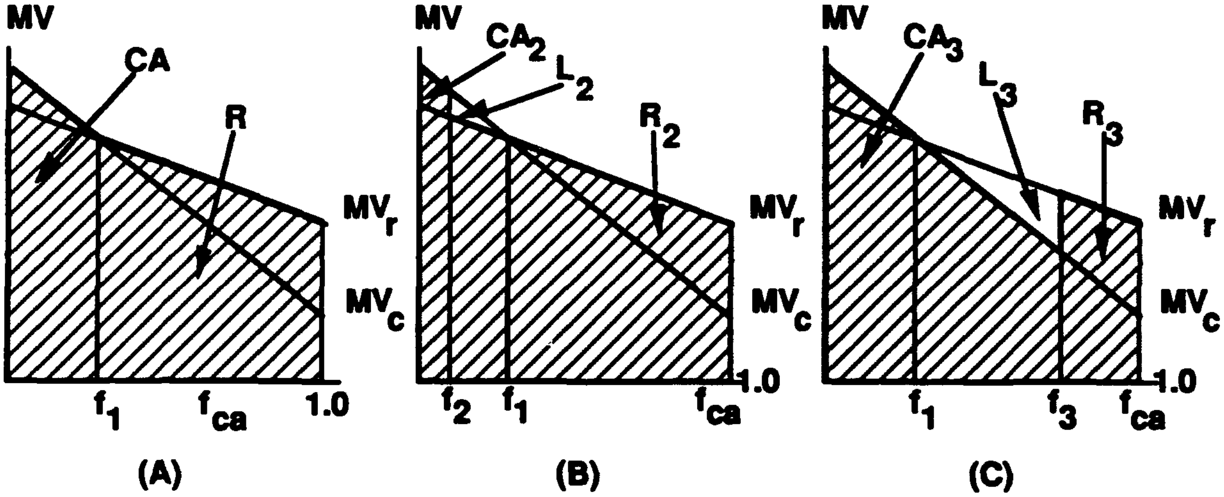

Figure C-14. Effect of Suboptimal Allocation on Total Value

This optimal allocation point is sensitive to changes in conditions. Since MV_c is a function of Red's total advance distance, conditions that would tend to increase total Red advance would tend to cause Blue to compensate by allocating a larger fraction of his reserves to counterattack. If, for example, a lower theater force-to-space ratio caused the initial force-to-force ratio at the point of attack to increase, Red would gain ground faster and the length of Red's flank at any given reference time (prior to culmination) would be longer. This would increase the total diversion of Red troops required to counter a given increase in Blue counterattack forces and, thus, would shift MV_c to the right, as shown in

Figure C-15. This shift in the marginal utility of counterattack from MV_{c1} to MV_{c2} would in turn shift the intersection point *A* to *B*, with a corresponding increase in the fraction of reserves allocated to counterattack from f_1 to f_2. As a rule, then, we would expect that lower theater force-to-space ratios would encourage the defender to adopt a more counterattack-oriented mode of employment for reserves.

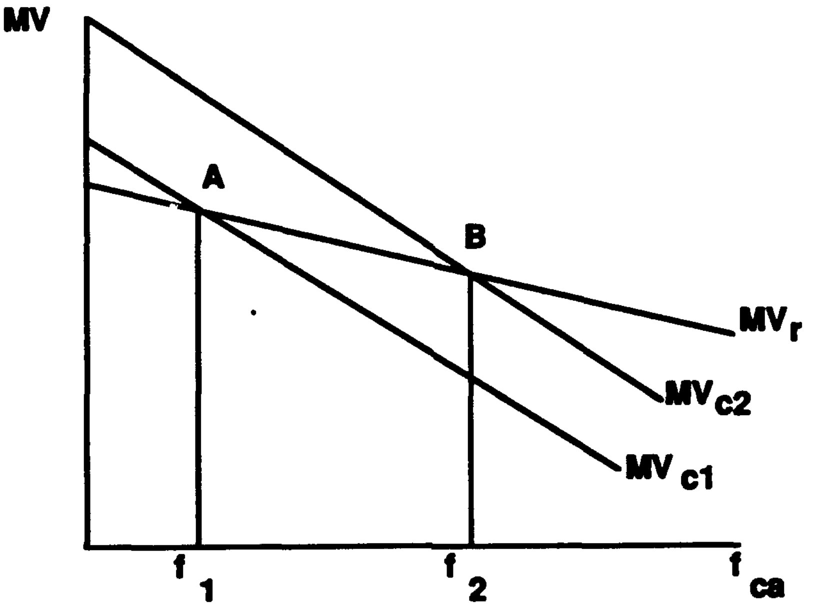

Figure C-15. Effect of Increased Local Force-to-Force Ratio on Optimal Allocation

At an extreme, changing conditions such as these could produce a corner solution of f_{ca} = 1 or 0. An extremely high initial force-to-force ratio at the point of attack, for example, could shift MV_c rightward to the point where there is no intersection point within the range of f_{ca} (i.e., 0 to 1). In this case, the optimal allocation would be entirely to counterattack (since the marginal value of counterattack is now higher at all points). Alternatively, an extremely low initial force-to-force ratio at the point of attack could shift MV_c sufficiently to the left as to deny an intersection point within the range of f_{ca}; thus a more defender-favorable force-to-force ratio at the point of attack (as we would expect at higher theater force-to-space ratios) could produce an optimal allocation of f_{ca} = 0 (since the marginal value of counterattack is now lower at all points). More generally, whenever either option dominates the other (i.e., offers a higher associated marginal value curve over the entire feasible range), the optimal solution is, of course, to allocate all reserves to that role.

There are two classes of exceptions, however. First, it is possible that under some circumstances an intersection point could constitute a total value *minimum* rather than *maximum*. If MV_r is steeper than MV_C, and if the marginal value of reinforcement is greater than that of counterattack at $f_{ca} = 0$, as depicted in Figure C-16A, then any interior choice of f_{ca} will produce lower total value than a corner solution of either $f_{ca} = 0$ or $f_{ca} = 1$, and the point of minimum total value will be given by the intersection point f_1. Under these conditions, allocating all Blue reserves to reinforcement would provide a total value greater than that of allocation f_1 by an amount equal to the area of triangle L_r in Figure C-16A. Allocating all Blue reserves to counterattack would provide an increase in total value relative to that of allocation f_1 equal to the area of triangle L_{CA}. Conditions such as these would require that: (a) the loss-exchange advantages of increased numbers of defensive shooters be very great; (b) these advantages increase without bound over all possible allocations of reserves to defensive reinforcement; and (c) the impact of Red's force diversion to flank defense have little effect on Red's total advance (i.e., that the slope of Blue's diminishing returns to increased counterattack be modest). Depending on the relative positions of MV_C and MV_r, these conditions would produce an optimal allocation of either all-counterattack (if MV_C were shallow but high) or all-reinforcement (if MV_C were shallow but low).

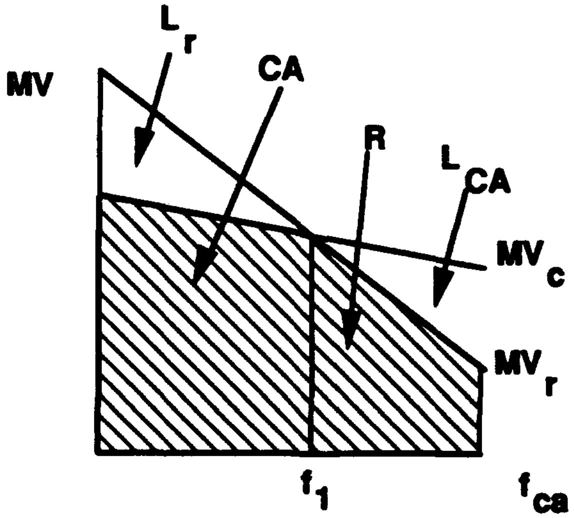

Figure C-16A. An Exception: Total Value Minimum at $MV_C = MV_r$

Second, there could be multiple points of intersection, as in Figure C-16B. Given non-linear marginal value functions, the optimal allocation will be described either by the intersection point closest to the origin (point f1 in Figure C-16b) if the area of Region P is greater than that of Region Q, or by the point $f_{ca^*} = 1.0$ if P < Q.[46]

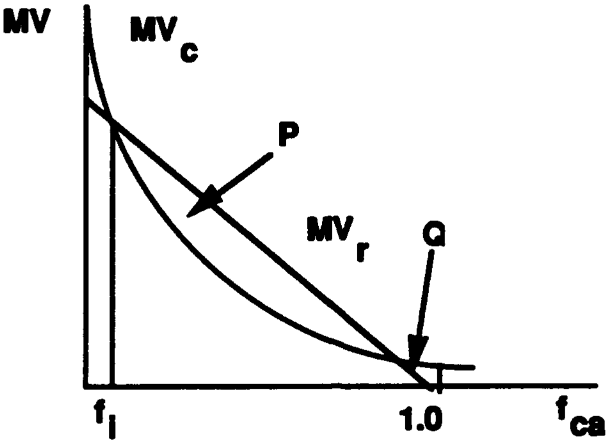

Figure C-16B. An Exception: Multiple Points of Intersection

G. WEAPON TECHNOLOGY

Unlike tempo, depth, or reserve employment, weaponry is essentially an external given for the battlefield commander. If his forces are armed with M60 tanks and M113 armored personnel carriers, the commander cannot simply choose to have them be M1s and M2s instead—he must fight with what he is given. Weapon mix, then, will not in itself be treated as a choice variable like those we have discussed above. For our purposes, technology is thus an *exogenous* variable: we do not seek to explain why technology is what it is at a given time or place. Rather, we seek to explain how an externally given set of technologies affects the *endogenous* independent variables of force employment, and of course, the dependent variable of net territorial gain. Our concern in this section is thus to describe how weapon mix affects force employment choices, and thereby to evaluate its effect on combat outcomes.

---

[46] If P = Q, then either allocation yields equal value. Alternatively, if MVc and MVr are collinear, then all allocations yield equal value and any allocation is therefore equally preferable.

"Weapon mix," as we have noted above, embodies two distinct issues as encountered in the theoretical literature: class and quality. Weapons of different *class* (e.g. tanks as opposed to infantry or artillery) affect combat outcomes differently, as do weapons of different *quality* within a given class (e.g., T55 tanks as opposed to T62s or T72s). Although both are important, we will focus here mainly on the former—i.e., the impact of different weapon classes—as this is arguably the more fundamental of the two, and in any case is a necessary precondition for an adequate understanding of the effects of quality within a class. While we thus will not directly address the impact of marginal improvements in quality, we will discuss in somewhat greater detail one particularly important (and potentially revolutionary) improvement in weapon quality—the replacement of traditional artillery with terminally guided Advanced Conventional Munitions (ACM).

To do this, we will first describe the strengths and weaknesses of the major weapon classes for theater warfare (infantry, armor, artillery and aircraft) in terms of four key characteristics: mobility, firepower, hardness, and visibility. We will then go on to address the special case of ACM and to describe the differing impact of these classes of technologies on force employment choices.

1. Mobility

"Mobility" as used here refers not just to the maximum speed of the weapon system, but also to the range of terrain types over which that speed can be maintained. Tracked armored vehicles such as tanks, armored troop carriers (ATCs), or self-propelled artillery (SPA) have high maximum speeds but are immobilized by steep slopes or heavily wooded terrain. They are also slowed (but not stopped) when moving cross-country rather than on roads. Dismounted infantry have slower maximum speeds (perhaps 5 kilometers per hour vs 65 for an M1 tank), but maintain that speed over a much wider range of terrain types. Aircraft, of course, offer the highest maximum speeds, and these speeds are effectively independent of the terrain. Fixed-wing aircraft, however, are constrained by a high ***minimum*** speed that can make target acquisition problematic, especially against concealed targets in difficult terrain.

2. Firepower

"Firepower" encompasses mass, range, and accuracy. Artillery, for example, delivers the greatest mass of munitions per unit time and can do so over long ranges, but with only limited single-round accuracy.[47] Tank guns are of smaller caliber, and typically have access to smaller supplies of on-hand ammunition. Moreover, tank guns are designed for flat trajectory *direct fire*[48] and thus are limited to much shorter engagement ranges. The high arc of howitzer and especially mortar trajectories enables artillery to engage targets located behind a terrain mask using *indirect fire*.[49] Tanks using direct fire must also accept that a certain fraction of the terrain they overwatch will be masked *dead space* against which they cannot bring fire to bear and in which opponents can therefore shelter; artillery faces fewer such constraints. Tanks thus provide less effective mass and less range than artillery. Tanks, however, fire much more accurately. Direct fire pits tanks against targets they can see directly and engage with individually aimed, high velocity, low dispersion rounds.[50]

Infantry weapons vary considerably. They include heavy, long range antitank guided missiles (ATGMs) such as the U.S. TOW or the Soviet Sagger; light, very short range disposable rockets like the U.S. LAW (Light Antitank Weapon) or the Soviet RPG series; hand grenades, and small arms. Heavy ATGMs offer longer range and better potential accuracy than tank guns but suffer from slow rates of fire and vulnerability to countermeasures.[51] Lighter weapons must be used in quantity at short range to be

---

[47] For general surveys of artillery technology and employment, see J.B.A. Bailey, Field Artillery and Firepower (Oxford: The Military Press, 1989); and Headquarters Department of the Army, FM 6-20, Fire Support in Combined Arms Operations (Washington, D.C.: USGPO, 1977), esp. appendix B, "The Field Artillery System." See also Bidwell and Graham, op. cit.; House, op. cit. For detailed treatments of Soviet artillery practices, see Chris Bellamy, Red God of War: Soviet Artillery and Rocket Forces (New York: Brassey's, 1986); and David C. Isby, Weapons and Tactics of the Soviet Army (New York: Jane's, 1988), 223-49.

[48] In which shooter and target are intervisible.

[49] In which shooter and target are not intervisible, and for which some form of spotter, or other target acquisition means remote from the firing platform must therefore be employed to direct the fire.

[50] On the tradeoffs between armor and artillery with respect to firepower, see Shelford Bidwell, Modern Warfare: A Survey of Men, Weapons, and Theories (London: Allen Lane, 1973), pp. 162-3, 53-9; also Bidwell and Graham, op. cit., p. 214. For general surveys of tank technology and tactics, see Richard Simpkin, Tank Warfare (New York: Crane Russack, 1979); Richard Ogorkiewicz, Armoured Forces (New York, Arco Publishing Company, 1970); Kenneth Macksey, Tank Warfare: A History of Tanks in Battle (New York: Stein and Day, 1972).

[51] For general surveys, see Richard Simpkin, Antitank (New York, Brassey's, 1982); R.G. Lee et. al., Guided Weapons (Oxford: Brassey's, 1983); and Seymour J. Deitchman, Military Power and the Advance of Technology (Boulder, CO: Westview, 1983), pp. 67-85. The Soviets have additionally retained a number of large caliber towed antitank guns for infantry support. See Isby, op. cit., pp. 215-20.

effective. Given trained operators in sufficient numbers, however, light infantry antitank rockets have proven highly effective historically.[52]

Aircraft armed with precision guided munitions offer potentially the same accuracy as tanks and ATGM against targets in the open. Their mass of fire per sortie is low—a single F16, for example, might carry a combat load of six Maverick air-to-surface missiles. Aircraft compensate for this limited mass per sortie, however, with an ability to concentrate many sorties at a threatened point. Their long range permits aircraft to be massed from great distances, while high speed permits that massing to be carried out quickly. Long range also permits aircraft to overfly the front lines and direct deep strikes with aimed, accurate fire against targets beyond the line of sight of ground weapons. The high minimum speed of fixed-wing aircraft, however, restricts them largely to easily located targets—typically either fixed installations or moving vehicles in the open. Helicopters, on the other hand, have no minimum speed, but their range is shorter and they are highly vulnerable when overflying hostile territory. Traditional "tube" artillery can reach beyond the ground forces' line of sight but cannot do so with the accuracy of aircraft-delivered fire. Emerging terminally-guided artillery and surface-to-surface missiles offer the potential accuracy of air-delivered munitions at comparable ranges, but may also be of limited effectiveness against stationary or concealed targets.[53]

3. Hardness

"Hardness" refers to the ability of a weapon system to survive fire directed against it. The effective hardness of a weapon system is a function of its organic armor protection, and the availability of *cover*—the interposition of earth, masonry, sandbags or other projectile-resistant substance between the target system and the weapon firing at it.

Tanks provide the maximum in mobile organic armor protection. While no armor can ensure survival against all types of attack at all ranges or from all directions, modern tanks provide frontal arc protection sufficient to require large caliber weapons for successful penetration at long range. Lighter weapons require some combination of shorter ranges or firing opportunities against more lightly armored flank or rear surfaces to be effective against tanks. Traditional artillery is of limited value against heavy

---

[52] See, for example, John Weeks, Men Against Tanks (New York: Mason/Charter, 1975), esp. pp. 68-73, and 100-104.

[53] For an overview of air to ground combat dynamics, see Deitchman, op. cit., pp. 31-65.

armor—a direct hit is required to penetrate, but the inaccuracy of standard artillery makes this improbable without prolonged barrage by large numbers of guns.[54] Tanks, however, are unlikely to remain in position under such a barrage long enough for its full effect to be felt. Their armor protection enables tanks to move out from under artillery fire; unless the barrage area is extremely wide, this will ordinarily enable tanks to escape the worst effects of opposing artillery.

Tanks are also able to exploit the advantages of natural cover by operating in *defilade*. A *hull-defilade* position is one in which the hull of the vehicle is masked, typically by the crest of a hill or the edge of a man-made entrenchment, leaving only the turret exposed.[55] The tank is thus able to fire, but more than half its theoretical target area is "hardened" by the addition of the earthen armor behind which the vehicle is sheltered. Defilade, however, is available mainly to defenders.[56] Attackers may exploit *covered approach routes* by interposing hillsides, streambanks, or buildings between themselves and opposing direct fire weapons (thus advancing in *dead space* masked from enemy fire), but typically this prevents either side from firing while such cover is in use.

Armored troop carriers are armored to withstand small arms fire and shrapnel from traditional artillery, but do not enjoy the same level of protection as tanks. Whereas tanks resist penetration from many types of anti-armor weapons *until the range to the* opponent closes substantially, ATCs can be penetrated by a much wider variety of weapons to a longer range. ATCs are also somewhat more vulnerable to artillery than are tanks, in that an APC can be overturned by a near miss that would not affect a heavier

---

[54] See, for example, the results of the Army's Human Engineering Laboratory testing of artillery battery-forward observer teams against moving vehicle targets (the HELBAT test series). For a description of the series, see, for example, R.B. Pengelley, "HELBAT Strikes Back," International Defense Review, May 1981, Vol.14, No.5, pp. 555-578.

[55] Other forms of defilade are possible, notably *turret defilade*, and *full defilade*, in which (respectively) the cupola, and none of the vehicle is exposed. Turret defilade enables observation, but not fire, while exposing only the tank commander to opposing fire. Full defilade (sometimes referred to as a "hide" position) is occupied prior to engagement, or in displacement between firing positions. See Headquarters, Department of the Army, FM 71-1, The Tank and Mechanized Infantry Company Team (Washington, D.C.: USGPO, 22 November 1988), pp. 6-27 to 6-32; also Headquarters, Department of the Army, FM 5-103, Survivability (Washington, D.C.: USGPO, 10 June 1985), esp. pp. 4-13 to 4-15.

[56] Attackers may employ overwatch forces in stationary positions to provide direct fire support for moving assault elements, and these overwatch forces may be placed in defilade, but while the defender is often able to position his entire force behind such cover the attacker can exploit defilade for only a fraction of his force. On overwatch techniques, see FM 71-1, op. cit., pp. 3-12 to 3-13 and 3-20 to 3-25.

tank. Like tanks, however, they can be employed in defilade on defense and (where possible) used to exploit covered approach routes in the attack.

Self-propelled artillery is likewise lightly armored. SPA, however, is protected mostly by cover in the form of its removal from the front lines. Long range indirect fire enables artillery to deploy beyond the reach of direct fire systems like tank guns or ATGMs; the primary threat to artillery comes rather by counter-battery fire from other indirect fire artillery systems. Counter-battery, however, is slow by contrast with direct fire. Target acquisition is more complex, and the range between shooter and target is typically long enough to require non-trivial flight times for counter-battery rounds to reach their target. As a consequence, SPA can usually evade return fire by so-called "shoot-and-scoot" techniques, whereby artillery pieces fire a mission, then quickly leave the position and displace to a new firing location before opposing counter-battery fire arrives. Shoot-and-scoot is inefficient, in that it requires a substantial movement delay between fire missions and thus reduces the scooting guns' effective rate of fire. But where counter-battery fire is a serious threat, shoot-and-scoot offers an effective option for reducing artillery losses.[57]

Dismounted infantry is without significant armor protection. Exposed infantrymen can be killed by almost any weapon found on the modern battlefield, and to the limit of the weapons' range. For protection against this array of threats, infantry requires cover. At the same time, however, infantry is uniquely suited to exploit cover. On the defense, infantry can dig themselves in, and when properly dug in, expose very little target area. A tank in hull defilade may nevertheless present up to half its total frontal area above the terrain cover.[58] An infantryman in a prepared fighting position need expose only his head and upper shoulders to bring effective fire to bear.[59] With proper overhead protection, dug-in infantry can be difficult to kill with artillery, requiring a direct hit or very near miss to penetrate an earthen foxhole roof. On the attack, the small size of individual

---

[57] Of course, shoot-and-scoot is not the only approach to improved survivability for SPA. Alternatives include dispersion, hardening (e.g., digging in personnel and ammunition, or placing howitzers in defilade), enhanced communications security, or other countermeasures to target acquisition such as chaff or direct attack of counter-projectile radars. Of these, however, shoot-and-scoot is perhaps the most effective, albeit at a price in effective rate of fire. For a more complete review of artillery survivability issues, see Bailey, op. cit., pp. 93-114; Isby, op. cit., pp. 246-7; also Headquarters Department of the Army, FM 6-20-1, Field Artillery Cannon Battalion (Washington, D.C.: USGPO, 27 December 1983), pp. 1-39 to 1-48.

[58] See, for example, Simpkin, Tank Warfare op. cit., pp. 139-40.

[59] See, for example, diagrams in FM 5-103, op. cit., esp. pp. 4-3 to 4-9.

infantrymen enables them to exploit smaller terrain features for cover. This increases the fraction of dead space in front of a defensive position, and increases the likelihood of finding a useable covered approach. Infantry's mobility in difficult terrain also enables them to utilize covered routes such as forested draws or swampy riverbeds, which would be closed to armored vehicles.

Infantry's dependence on cover creates a number of vulnerabilities, however. For example, infantry are particularly susceptible to *suppressive fire*. Suppression does not kill directly; rather, suppressive fire forces the target to take evasive action that reduces its effectiveness and thus enables other weapons to kill the suppressed target. Even perfect cover provides only partial protection. To fire, all weapon classes must risk some degree of exposure. If hostile fire is threatening enough, a weapon system can often reduce its exposure by more fully exploiting available cover, but only at the cost of ceasing its own fire. A defending tank, for example, must expose its turret to fire from the cover of hull defilade. If receiving heavy fire, it can retreat to turret defilade but cannot then return the fire. The tank may move to a new position and resume fire, but in the meantime it is effectively suppressed, enabling other attackers to close the range or to reach a flanking position from which the tank can be destroyed in spite of defilade.[60]

To force a tank into full cover in this way can be difficult since the tank's heavy frontal armor enables it to remain partially exposed under all but very heavy, aimed fire. Dug-in infantry, on the other hand, presents a different sort of target. Its exposed area is very small, making effective aimed fire of this sort very difficult. Infantry overhead cover can only be penetrated by a direct hit from a large caliber weapon. But the small area an infantryman must expose to fire is unarmored, so small projectiles are sufficient to kill or disable if the small target can be hit. Artillery and automatic small arms can efficiently spread shrapnel and small caliber bullets over a wide area, forcing infantry to pull back into full protection and, in the proces, suppressing their fire without killing them directly.[61]

Vulnerability to suppression creates a second vulnerability for infantry—it can be pinned in place by artillery fire. Many defensive positions offer withdrawal routes covered from opposing direct fire, but few such routes provide overhead cover against

---

[60] Alternatively, tanks may be suppressed to a degree by being forced to close hatches under artillery fire (that is, to "button up,") and, thus, to reduce their ability to see and hear.

[61] On infantry and suppression, see e.g., Bidwell, Modern Warfare, op. cit., pp. 156-7.

indirect fire. For most infantrymen, to leave a prepared position under an artillery barrage is to risk complete exposure in the open. By contrast, tanks and ATCs can simply leave a barrage zone and move to a secondary position. Neither armored vehicles nor infantry positions can be destroyed by artillery without a near-direct hit; the odds that randomly placed individual rounds within a barrage will strike a tank in the short time before the tank clears the area are thus very small. Infantry, however, cannot leave safely. If the barrage is maintained long enough, the odds of hitting a stationary target gradually increase. Sustained artillery fire can thus eventually destroy defending infantry outright.[62]

Aircraft pose different problems. Like infantry, they are without significant armor protection and therefore must rely on cover. Unlike infantry, however, aircraft must maintain their mobility under fire; consequently, they take "cover" by flying low to exploit terrain-masking rather than by occupying prepared positions. And unlike infantry, aircraft are capable of great speed and, hence, limited exposure times during dashes between covered points. For aircraft, then, "hardness" is largely a matter of limiting exposure time.[63] Fixed-wing aircraft accomplish this through high speed at low altitude. Helicopters do so through the use of "pop-up" techniques, in which they hover behind a terrain mask until a target is identified (typically by a second, scout helicopter) and then climb to clear the mask, engage the target, and descend behind the mask again to thwart antiaircraft fire. Either of these techniques leaves the flight crew little time to acquire and engage targets, but each is important to aircraft survivability.

4. Visibility

"Visibility" has two component parts: the ability of opponents to see the weapon in question, and the ability of the weapon crew to see the opponents (that is, *target signature* and *target acquisition*). Tanks suffer in both respects. Tanks are large, clumsy, and loud. In the attack, their size and mobility restrictions limit the availability of concealment, and in any case their advance is typically audible over long distances. In dry weather, a moving tank column can raise a dust cloud visible for miles. In the

---

[62] See, for example, John A. English, On Infantry (New York: Praeger, 1984), p. 205.

[63] Given the nature of guidance systems and target acquisition means for antiaircraft weapons, aircraft also rely on a much broader range of electronic warfare techniques to increase survivability. While important to the broader issue of aircraft survivability per se, these techniques are of only indirect relevance for our topic; we will thus not address them in detail here.

defense, tanks can be concealed much more effectively, especially against long range observation. Even here, however, their size—and the size of their armament—makes them difficult to hide completely. The muzzle flash and report from a tank gun firing can be dramatic; just the shock wave from the firing of an M1's 120-mm gun creates a noticeable rumbling of the earth for hundreds of yards. Even well-camouflaged tanks can therefore expect to be seen after the firing of a relatively few rounds, necessitating displacement to a secondary firing position if they are to regain concealment.

Tanks are also difficult observation platforms from which to spot the enemy. Especially when "buttoned up" for protection against artillery and small arms, tanks offer a limited field of view. Against long range targets, the vibration, pitching and bouncing typical of cross-country movement in tracked vehicles, together with the limited peripheral vision available through the vehicle's vision blocks, complicates target acquisition. At short range, tanks suffer a *blind zone*, an area within a distance of some 5 to 50 meters from the vehicle (depending on aspect) that contains a total of some 3000 square meters over which no crew member can see the ground.[64] Even were the platform better suited to observation, tank units are light on manpower and as a result simply have few pairs of eyes with which to search. A U.S. tank platoon consists of four tanks and a total of 16 men; a mechanized infantry platoon with four ATCs has almost three times the manpower and, thus, almost three times the number of potential observers.[65]

Infantry, by contrast, are both difficult to see and highly capable as observers. The small size that enables infantry to exploit available cover also enables them to exploit concealment so as to avoid being seen. On the defense, properly dug in and camouflaged infantry can be extremely difficult to spot, especially from long range and in forested or urban terrain. As a result, a prepared infantry position can often remain unseen by attacking armor even as the tanks pass over the foxholes. In World War II, German defensive doctrine directed well-concealed infantrymen to allow attacking armor to pass through the positions before opening fire on accompanying offensive infantry.[66] More generally, infantry at short range constitute a major threat to armor—in large part because

---

[64] In addition, there is a larger *dead zone* in which none of the vehicle's weapons can be brought to bear. See Richard E. Simpkin, Mechanized Infantry (Oxford and New York: Brassey's, 1980), pp. 46-48.

[65] Headquarters, Department of the Army, Organization of the United States Army (Washington, D.C.: USGPO, December 1988), p. 8.

[66] Wray, op. cit., pp. 16-18.

of the extreme difficulty of spotting covered, concealed infantry from a buttoned-up tank.[67]

On the attack, infantry benefits from its ability to advance over rough or forested ground, its surefootedness in darkness and bad weather, and its ability to move silently.[68] Attacking infantry units can often infiltrate a defensive position at night using concealed approach routes closed to armored vehicles which would in any case be too loud to move far without betraying their location to the enemy.[69] Even in broad daylight, a dismounted infantry advance over a forested approach provides a substantial measure of effective concealment. While the defender may see parts of the advancing formation, it will be difficult to see enough at any one time to formulate an accurate picture of the formation's boundaries and center of mass. Without such information, however, the effectiveness of defensive artillery fire is reduced substantially.[70]

While infantry is thus difficult to see, infantrymen are excellent observers. This is particularly important on the attack, and especially for attacks on dug-in infantry defenses. Dismounted infantry have excellent peripheral vision, no minimum observation distance, numerous observers, and an ability to inspect suspicious terrain directly. Infantry's capability to root out concealed infantry defenses therefore far exceeds that of armored units of comparable size. In cooperation with armor, dismounted infantry offers synergistic benefits: the infantry provides the eyes and pinpoint small arms fire to protect the armor against defending infantry; the armor provides firepower to deal with strong points and opposing armored vehicles too tough for organic infantry weapons, and simultaneously provides supporting machine gun fire with which to help suppress identified opposing infantry.[71]

---

[67] For historical examples, see, e.g., G.D. Sheffield, "Blitzkrieg and Attrition: Land Operations in Europe, 1914-45" in Colin McInnes and G.D. Sheffield, eds., Warfare in the Twentieth Century: Theory and Practice (London: Unwin Hyman, 1988), pp. 51-79, esp. pp. 68, 71-4; English, op. cit., pp. 110, 112-13; Wray, op. cit., pp. 29-30, 100-104. For a more personalized perspective, see the attitude of tank crews toward opposing infantry as described in Ken Tout's memoir of life in a World War II Sherman tank crew, Tank (London: Robert Hale, 1985), esp. pp. 83, 117-118.

[68] On the rough terrain mobility and stealth advantages of infantry over armor, see Headquarters Department of the Army, FM 7-10, The Infantry Rifle Company (Washington, D.C.: USGPO, 8 January 1982), pp. 1-1 to 1-2, 3-2 to 3-3, 4-1; FM 7-30, Infantry, Airborne, and Air Assault Brigade Operations (Washington, D.C.: USGPO, 24 April 1981), pp. 1-3 to 1-4, 2-12.

[69] For historical examples of infantry infiltration attacks, see English, op. cit., pp. 101-2, 172-3, 159-62.

[70] See, for example, Wray, op. cit., pp. 166-7.

[71] See Bidwell, op. cit., pp. 149-50, 170-1; English, op. cit., pp. 200, 202, 110, 112-13, 142; Griffith, op. cit., Chapter 5, "1915-1945: The Alleged Triumph of Armour over Infantry," esp. pp. 97-98; Strachan,

Infantry's strength as observers is also of value on the defense. This is particularly so for defense against dismounted infiltration attacks at night. Tanks are notoriously poor counter-infiltration weapons.[72] Their smaller numbers, restricted field of view and muffled hearing are substantial penalties in opposing stealthy attack in darkness or bad weather. The greater manpower available in an infantry unit of a given size can be distributed to cover more potential infiltration routes, while the greater sensitivity of infantry as sensors offers a better chance of detecting stealthy movement on a given route. Again, however, the combination of infantry and armor offers synergistic advantages—the firepower of armor in a defensive role can be a powerful asset in dealing with infiltrators identified by the defending infantry.

Aircraft also pose a different set of problems with respect to visibility, again largely because of the great difference in speed and exposure time between aircraft and other weapon classes. As weapon platforms (as opposed to their role in reconnaissance, for which they are often specially equipped), aircraft suffer from their relatively brief opportunity to search an area prior to selecting a target. As such, they have particular difficulty acquiring concealed defenders and tend to do better against exposed targets in the open. Fixed-wing aircraft are easily seen—and easily heard. Concealment is thus less important to them than is cover and brevity of exposure. Helicopters, on the other hand, can be very difficult to spot at long range when hovering in the vicinity of a broken tree line. They have the additional advantage of rapid displacement following remasking, thus forcing searchers to spread their efforts over a wide area rather than focusing in on a few meters of horizon. While their noisiness makes most helicopters inappropriate for missions that require true stealth (such as nighttime infiltration), they are thus not without a significant capacity for concealment.[73]

5. Advanced Conventional Munitions

How do ACM differ from traditional weapon classes in these respects? In effect, long range ACM (such as the developmental U.S. ATACMS, or Army TACtical Missile

---

op. cit., pp. 183-6; Ogorkiewicz, op. cit., p. 128; Anthony Farrar-Hockley, <u>Infantry Tactics</u> (London: Almark, 1976), pp. 29-33, 62-68.

[72] See, for example, Wray, op. cit., pp. 26-27, 40.

[73] On helicopter operations, see Headquarters, Department of the Army, <u>FM 17-50, Attack Helicopter Operations</u> (Washington, D.C.: USGPO, 1 July 1977). For Soviet practice, see Isby, op. cit., pp. 432-443; for non-US NATO, see Isby and Kamps, op. cit., pp. 225-6; c.f. pp. 357-8.

System) offer the precision anti-armor firepower, range, and speed of concentration advantages of aircraft, but without the aircraft's constraints with respect to the need for cover in order to survive antiaircraft fire. Long range ACM are thus less vulnerable to loss en route to the target than are aircraft. Short range ACM (such as the MLRS/TGW, or Multiple Launch Rocket System with Terminally Guided Warheads) are somewhat less useful for counterconcentration, in that they still require ground transportation to the point of attack. Like aircraft, however, all ACM are less effective against concealed or stationary targets than they are against moving vehicles in the open. Thus, they are well-suited to engage an advancing assault wave, an administrative march column, or defensive forces while withdrawing or while moving to the point of attack from rearward assembly areas. They are ill-suited to engage rearward forces while in hide positions, or forward defenders in prepared positions.[74]

6. Impact on Force Employment

Tanks, then, offer great hardness to enemy fire, accurate and fairly heavy firepower, and high maximum speeds when operating on favorable terrain—but, they are immobilized by rough terrain, subjec to poor crew visibility, and seen easily by opponents. Dismounted infantry, on the other hand, is difficult to see (especially from moving armored vehicles) and is excellent for observation, it can be hardened by the use of cover, and it retains its mobility on rough terrain. But it, too, has drawbacks: slow maximum speed, limited firepower at long range, and vulnerability to large-scale suppressive fire or to extended artillery bombardment. Artillery offers heavy long range firepower, removal from direct fire, and the capability to evade indirect counterfire by displacement, but it is inaccurate and requires remote target acquisition to locate targets. Aircraft provide a capability to concentrate rapidly over long distances, to reach targets deep in the enemy rear, and to provide accurate, directly observed firepower against targets beyond the reach of friendly ground vehicles. However, they are limited by exposure time and flight speeds to targets that can be acquired very quickly (typically exposed or moving vehicles). ACM offers capability much like that of traditional fixed-wing aircraft, but at

---

[74] Stationary vehicles are substantially easier to conceal against acquisition by top-attack ACM submunition sensors; for an overview of countermeasures to such munitions, see Stephen Biddle, How to Think About Conventional Nuclear Substitution: The Problem of Structural Uncertainty (Alexandria, VA: Institute for Defense Analyses, 1986) IDA P-1884, esp. pp. 10-15.

potentially higher levels of firepower. What does all of this mean for force employment and, thus, for combat outcomes as a whole?

To answer this question, we must look at the differing effects of these weapon classes on the relationship between attacker casualties and velocity. Artillery, for example, affects casualties in direct proportion to the time available for delivering fire. Given its inaccuracy, artillery is essentially a means for covering a specified area with a certain density of shells. For a given firing rate and a given number of firing tubes, the only way to increase the area under fire (or the density of fire in a given area) is to increase the duration of the barrage.[75] Thus, for the attacker, the longer the preparatory barrage, the more extensive the effects on the defender and thus the fewer the direct fire casualties to the assault forces—but the longer the time required from initiation to completion of the attack (and thus the lower the attack velocity). For the defender, the quicker the attack, the less time available for defensive artillery fire and thus the fewer attacker casualties from defensive artillery. Moreover, quick attacks rely mostly on armored vehicles, against which defensive artillery has little effect. Slower attacks typically produce more dismounted targets and provide more time to deliver artillery against those targets. Thus slower attack velocity increases artillery effectiveness for the defender as well as the attacker.

Infantry effectiveness is also strongly a function of attack velocity. For the attacker, for example, infantry can be extremely useful, but only for a slow velocity assault. For maximum impact, infantry must dismount. At best this reduces velocity to that of a walking foot soldier. For infantry to survive dismounted, however, requires further preparations which add to the time requirements for a successful infantry assault. Especially careful reconnaissance is required to identify approach routes which provide the necessary cover and concealment, since exposure has such grave consequences for dismounted infantry. The routes themselves are likely to be circuitous and time-consuming to traverse. It may be necessary to await darkness to provide concealment, or to delay long enough for a smoke screen to be planned, delivered, and fully formed. It is also essential to ensure cooperation with other weapon classes—especially artillery—to

---

[75] Alternatively, if the attacker can more tightly specify the defender's actual locations, then artillery fire can be directed into a smaller area—thus providing higher effectiveness per hour of barrage time. The only way to improve target location accuracy, however, is to spend time in reconnaissance. The conclusion thus remains the same: for a constant number of tubes, artillery effectiveness is directly proportional to the time allotted to its use. On this relationship, see the historical perspective in Bidwell, Modern Warfare, op. cit., pp. 53-9.

provide suppressive fire. As noted in our discussion of defensive depth, rapid advance tends over time to break down the cohesion of units and reduce the coordination of component arms. To use infantry effectively in the assault, time must therefore be allowed to bring the proper supporting weapons into close cooperation with the infantry itself and to maintain that cooperation as the assault advances into depth.[76]

On the defense, infantry effectiveness is again strongly a function of attacker velocity. If an attacker simply charges ahead at maximum velocity—and thus with little preparation and no dismounted support, an entrenched infantry defense can extract a very heavy toll.[77] At the opposite extreme, if an attacker is willing to pound an infantry defense with an extended artillery barrage, that defense can eventually be annihilated at almost no casualties to the attacker—but only at the cost of a very slow-moving attack.[78]

Armor effectiveness, by contrast, is less dependent on attack velocity. For the attacker, casualties can certainly be reduced by taking the time to deploy off-road or to provide smoke obscuration, but the sensitivity of armor losses to such preparations is much less than that of infantry. Moreover, it is unlikely that any amount of scouting, or any delay for circuitous transit will permit armor an unexposed approach to the target. While a more concealed route will still be superior to a less concealed one, the magnitude of the attainable difference will be smaller for armor—as are the consequences of success or failure in finding concealment.

For the defender, armor effectiveness is likewise less sensitive than infantry or artillery to the attacker's choice of velocity. Defensive armor is little affected by the length of the attacker's artillery preparation. It is also less sensitive than defensive infantry to the attacker's choice of mounted or dismounted attack. Defending *infantry* can remain concealed even as the attacker moves through the position unless the attacker dismounts; defending *tanks* will be seen by the attacker once the range closes, whether the attacker dismounts or not.

---

[76] On the potential speed penalties of close combined arms cooperation, see, for example, Macksey, op. cit., p. 245; also FM 71-1, op. cit., p. 3-27.

[77] As, for example, the Israelis discovered in the Sinai in 1973. See Herzog, op. cit., e.g., pp. 182-96. See also Bidwell and Graham, op. cit., p. 288.

[78] This was typically the case, for example, during the Allied "artillery offensives" of 1916 and early 1917. These multi-week-long preparatory barrages essentially wiped out those German positions which were subject to attack. The extraordinarily slow pace, however, permitted German reserves to arrive and occupy positions outside Allied artillery observation range. As the French described such tactics: "l'artillerie conquiert, l'infanterie occupiert" (the artillery conquers, the infantry occupies): John Keegan, The Face of Battle (New York: Random House, 1977), p. 215.

Aircraft, like armor, are relatively insensitive to attacker velocity. Aircraft effectiveness can be improved if the attacker slows sufficiently to permit more elaborate air-ground liaison and if artillery are able to suppress enemy air defenses (SEAD), but neither of these are especially time-demanding functions. Since the aircraft themselves can afford only brief exposure over the target area, there is little opportunity for any interaction that would require large amounts of time. Aircraft are most effective against exposed, moving targets, but the relative rate of those targets' movement is less important.[79]

7. Graphical Analysis

Figures C-17 and C-18 posit the effects of differing attacker and defender weapon mixes (or different *combined arms* balances) on the attacker's casualty-velocity trade-off frontier. In Figure C-17, three alternative weapon mixes are given for the attacker: a balanced (i.e., equal fractional composition) case CVT_{balA}, an armor-heavy mix CVT_{armA}, and an infantry-heavy mix CVT_{infA}, with a balanced defender weapon mix assumed throughout.[80] The balanced attacker case CVT_{balA} is essentially that of the nominal trade-off frontier as given in Figure C-1. The armor-heavy mix CVT_{armA} reduces attacker casualties for high attack velocities since it contains fewer thin-skinned infantry fighting vehicles and ATCs and more tanks, which are better suited to high speed mounted attack. Casualties fall only slightly as velocity decreases, however, since armor effectiveness is relatively insensitive to velocity. The infantry-heavy attacker mix CVT_{infA}, on the other hand, increases attacker casualties at high velocity because it contains a larger fraction of IFVs and ATCs, which suffer more heavily than tanks in a mounted assault. CVT_{infA} has a steeper slope than CVT_{armA}, however, since infantry is better able to reduce its casualties through increased preparation and execution time. Note that for the attacker, artillery and infantry have roughly similar effects on casualties

---

[79] Note, however, that this relative independence of aircraft effectiveness and attacker velocity may be much more characteristic of fixed-wing aircraft than rotary-wing aircraft; helicopters may be substantially susceptible to variations in opposing tactics (such as the increased use of dismounted infantry, artillery, or heavier overwatch by armored vehicles).

[80] As noted above, aircraft and ACM performance is effectively independent of attacker velocity and thus is not explicitly depicted here.

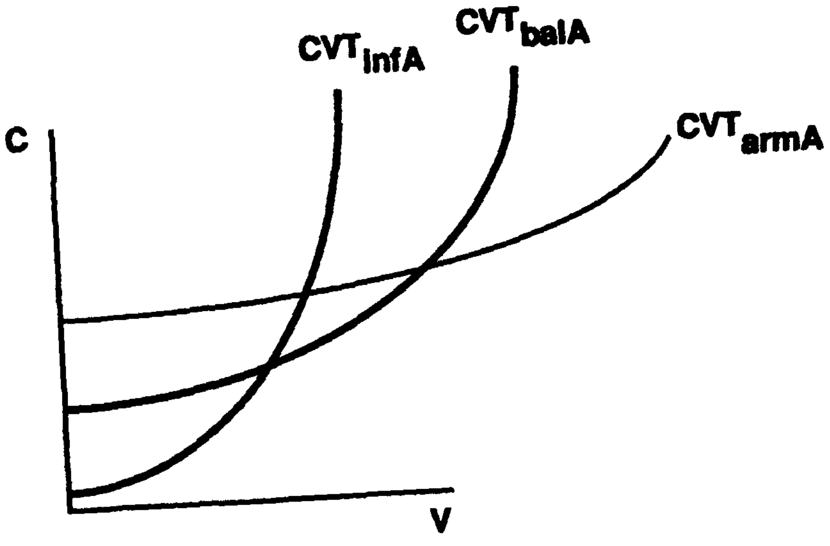

Figure C-17. Effect of Different Attacker Weapon Class Mixes on Casualty-Velocity Trade-off

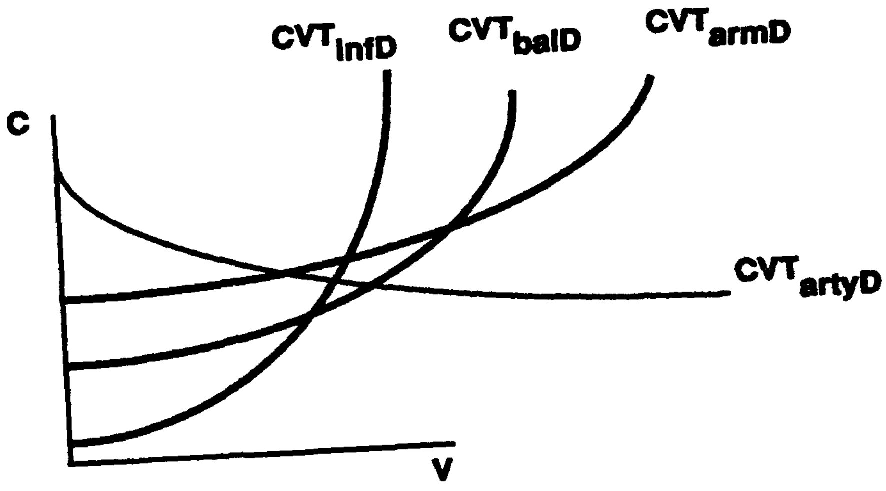

Figure C-18. Effect of Different Defender Weapon Class Mixes on Casualty-Velocity Trade-off as a function of velocity; the effectiveness of each is highly sensitive to velocity, and each tends to reduce casualties when velocity is reduced. CVT_{infA} thus represents the effect of either an infantry-heavy or an artillery-heavy weapon mix for an attacker.[81]

Figure C-18 gives the effects of four different weapon mixes for the defender: a balanced case CVT_{balD}, an armor-heavy mix CVT_{armD}, an infantry-heavy mix CVT_{infD}, and an artillery-heavy mix CVT_{artyD}, with a balanced attacker weapon mix assumed throughout. The balanced defender case CVT_{balD} is identical to that of the balanced attacker case CVT_{balA} in Figure C-17. As with the attacker in Figure C-17, the infantry-heavy mix CVT_{infD} produces higher attacker casualties at high velocity than the balanced case CVT_{balD}: dug-in infantry performs well against a high-speed mounted attack, and CVT_{infD} provides a larger number of such infantry for a given total force size than CVT_{balD}. Infantry is vulnerable to methodical attack, however. Thus the infantry-heavy CVT_{infD} performs less well than the balanced case against lower velocity attacks, and CVT_{infD} thus produces fewer attacker casualties than CVT_{balD} for low values of V. Conversely, armor-heavy defenses are relatively unaffected by more extensive attacker preparation, but neither does their effectiveness increase as fast as infantry-heavy defenses when the attacker neglects to prepare his attack. Thus CVT_{armD} will tend to produce higher attacker casualties at low velocity than the more infantry-heavy CVT_{infD} or CVT_{balD}, but it does not penalize the attacker as heavily for high assault speeds as do the more infantry-heavy cases.

While the effects of infantry- and armor-heavy weapon mixes are thus broadly similar for attackers and defenders, the effects of artillery are opposite. For both sides,

---

[81] Note that Figures 17 and 18 both assume tripartite "combined arms" mixes consisting of armor, infantry and artillery such that, for the purposes of the figures, more of any one means less of the other two. In the equations and the associated code, we will find it convenient to separate artillery (a form of fire suport) from infantry and armor (the maneuver forces). As a consequence, the attrition expression (equation 19) treats artillery additively; that is, it does not assume that more artillery necessarily reduces the armor or infantry available to either side in the tactical engagement at the point of attack. Were artillery treated additively here, the implication of the military logic described above would be an "artillery heavy" casualty-velocity tradeoff frontier with a higher low-velocity value than the balanced curve for artillery heavy defenders vs non-artillery heavy attackers; a lower low-velocity value than the balanced curve for artillery heavy attackers vs non-artillery heavy defenders; and each curve would tend to converge with the "balanced" frontier at high velocity (where artillery is less effective for either attackers or defenders). The functional form of equation 19 is intended to reflect this relationship.

Note also that for both Figures 17 and 18, the specific cross-over points between curves will vary with the quality of the weapons within the class, and the nature of the local terrain. While the cross-overs depicted are consistent with the experimental results described in appendix D, it is thus the shape and particularly the slope of the curves depicted that is of general applicability, rather than the relative height of any given curve.

more barrage time means more casualties from artillery fire for the other side. Since both Figures C-17 and C-18 are denominated in units of *attacker* casualties C, the result is a decrease in C as velocity decreases when the attacker is the side with the artillery-heavy mix (effectively, curve CVT_{infA} in Figure C-17), but a greater impact on casualties when velocity decreases when the *defender* is the side with the artillery-heavy mix, CVT_{artyD} in Figure C-18.[82]

These changes in the casualty-velocity trade-off frontier imply corresponding changes in the attacker's optimal velocity. These changes are illustrated in Figure C-19. Relative to a balanced attacker weapon mix CVT_{bal}, an armor-heavy weapon mix CVT_{arm} produces an increase in the attacker's optimal velocity from V_1 to V_2, an increase in the attacker's optimal casualty level from C_1 to C_2, and, in this case, a decrease in net territorial gain as a result of moving from IG_1 to IG_2. By contrast, an infantry-heavy mix CVT_{inf} produces a decrease in optimal velocity from V_1 to V_3, a decrease in the attacker's optimal casualty level from C_1 to C_3, and, in this case, a further decrease in net territorial gain as a result of moving from IG_1 to IG_3.

Note, however, that these changes in net territorial gain as a result of variations in weapon mix are substantially smaller than would be the case if we were to assume constant attacker velocity. If the attacker fails to modify his behavior, a reduction in attacker tank strength (such as that associated with the transition from CVT_{bal} to CVT_{inf}) would produce an increase in casualties for a constant velocity V_1 from C_1 to C_4, and in the process cut net territorial gains from IG_1 to IG_4. This outcome results in both higher casualties (C_4 vs C_3) and smaller total advances (IG_4 vs IG_3) than would be the case were he to adapt to the altered circumstances.

---

[82] For the purposes of clarity, this effect has been exaggerated in Figure 18; while this sharp an effect might obtain for an infantry-heavy attacker, it is unlikely for a balanced attacker as is assumed here. Moreover, it is not necessarily the case that the slope will be negative as depicted here; we contend here only that it may be smaller than that of the balanced case.

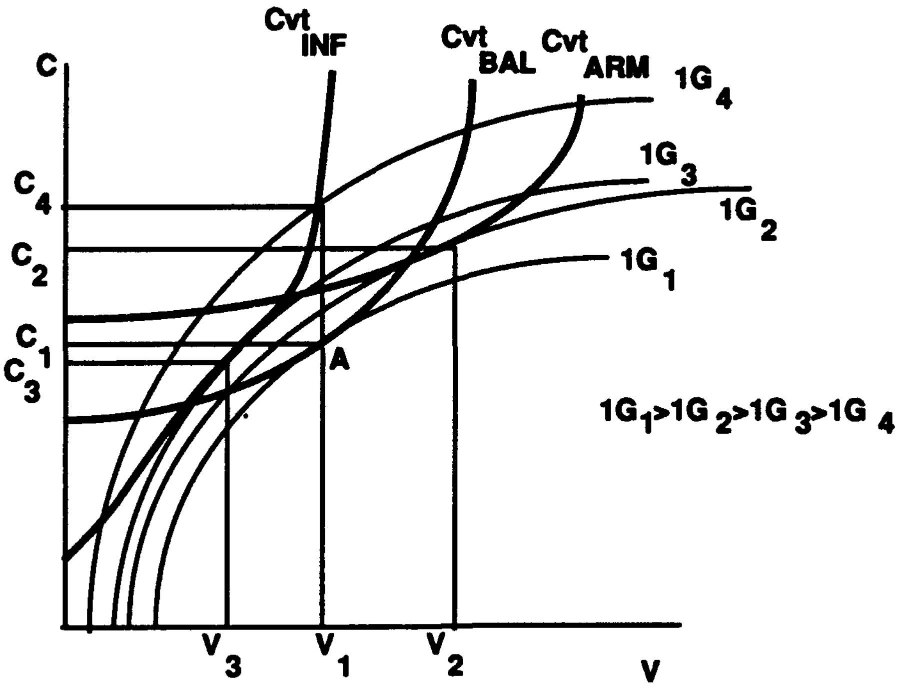

Figure C-19. Optimal Velocity Choice as a Function of Weapon Class Mix

In effect, *the ability to modify force employment behavior offers combatants a substantial opportunity for mitigating the effects of changes in physical circumstances*—in this case a reduction in tank inventories. To the extent that we fail to take proper account of this effect in our analyses of the impact of changes in force-to-space ratio, we thus risk a substantial overestimate of the potential impact of the changes in question.

H. TERRAIN

The final category of independent variable raised in the literature concerns the military geography of the theater of war. As we have noted above, this broader category incorporates two sub-issues: the impact of variations in natural terrain, and the impact of variations in "man-made" terrain in the form of barriers and obstacles. Both are clearly important. For our purposes, however, the natural terrain of the European theater is effectively a constant. While topography does change over time, the pace of such change is slow and relatively insensitive to considerations of national security. Barriers and obstacles, on the other hand, are much more amenable to policy intervention—and have at least as significant an influence on military outcomes. Given this, we will concentrate primarily on the question of man-made terrain in the form of barriers and their effects.

First, however, it is important to note one key function of natural terrain for the dynamics described earlier: terrain establishes an upper bound on the size of a single assault wave for a given frontage. If too large an assault force is crammed into too small a space, it increases its vulnerability to defensive artillery, and loses its ability to take evasive action under fire, to choose the least exposed path between its jump-off point and its objective, to maintain efficient formations that maximize its own firepower, or to change direction quickly to meet unexpected threats. As a result, all terrain has a "carrying capacity." Adding forces beyond the carrying capacity of the terrain produces less and less additional combat power for each unit added—and may even *reduce* total combat power in extreme cases. As a rule, the more open the terrain, the higher the carrying capacity for mounted attack. The rougher the terrain, the lower the carrying capacity.[83]

Central Europe, while no Switzerland, is nevertheless not ideal terrain for large-scale mounted warfare. The Federal Republic is substantially closer country than, for example, the steppes of western Russia on which the Red Army learned its craft.[84] Moreover, the continuing urbanization of Germany is gradually reducing the amount of open space in the border zone, further constricting the available maneuver area.[85] In conjunction with the threat of nuclear attack and conventional artillery fire, these considerations of terrain tend to confine Central European attackers to local force

---

[83] For recent discussions of the role of terrain constraints on local force concentrations, see Mearsheimer, Conventional Deterrence, op. cit., pp. 181-183; and John J. Mearsheimer, "Numbers, Strategy and the European Balance" International Security, Vol.12, No.4, Spring 1988, pp. 174-185. For a discussion of a USAREUR study suggesting that terrain constraints on force concentration would significantly hamper Soviet offensive operations in Central Europe, see Charles D. Odorizzi and Benjamin F. Schemmer, "An Exclusive AFJ Interview with General Glenn K. Otis," Armed Forces Journal International, January 1987, pp. 44-47. For a more general treatment of the role of terrain in modern ground force operations, see, for example, Richard Simpkin, Race to the Swift, op. cit., pp. 57-78.

[84] For a general description see Mearsheimer, Conventional Deterrence, op. cit., pp. 176-181. The North German plain is often assumed to offer the best tank country in the Central Region, but even it is hardly ideal. See General James H. Polk (ret.), "The North German Plain Attack Scenario: Threat or Illusion?" Strategic Review, Summer 1980, pp. 60-66. For a discussion of the role of Russian topography in the formation of current Soviet doctrine for mobile armored warfare, see S. Labuc and C.N. Donnelly, "Modeling the Red Force: Simulating Soviet Responses in Battle" in Reiner K. Huber, (ed.), Systems Analysis and Modeling in Defense (New York and London: Plenum, 1984), pp. 829-843.

[85] See Paul J. Bracken, "Urban Sprawl and NATO Defense," Survival, November/December 1976, pp. 254-260; and Bracken, "Models of West European Sprawl as an Active Defense Variable" in Reiner K. Huber, Lynn F. Jones, and Egil Reine, (eds.), Military Strategy and Tactics (New York and London: Plenum, 1975), pp. 219-230.

densities no greater than about 10 to 20 armored vehicles per kilometer for a single assault wave.[86]

With respect to man-made terrain, military barriers and obstacles can include anything from trees felled across a forest trail, to blown bridges, minefields, antitank ditches and barbed wire, all the way to reinforced concrete bunkers and pillboxes.[87] Barriers and fortifications have been a major element of land warfare for centuries, and constitute a traditional *economy of force* technique—i.e., one which enables small forces to hold wide frontages so as to free manpower for other uses. They are thus a natural option for defenders attempting to cope with low force-to-space ratios.

How, then, are we to assess their effects? We must begin by recognizing some key characteristics of barriers. First, barriers delay an attacker, but rarely do barriers in and of themselves bring an attacker to an outright halt. Any barrier can be cleared given sufficient time and effort, and most barriers can be cleared (albeit more slowly) by maneuver units without engineering support. Moreover, many barriers, such as minefields, can be overcome *without* delay if an attacker is willing to pay the price in terms of additional casualties. The purpose of a barrier is thus not to prevent movement per se, but to slow that movement, to channel it in directions that lead attackers onto prepared defensive positions, or to force attackers to accept heavier casualties if they chose not to delay or be detoured.

Second, barrier effectiveness is closely related to the strength of the forces defending the barriers. The barrier may, for example, delay an attacker at long range where the defender will enjoy a better loss-exchange ratio, but the actual casualty impact on the attacker will depend on the number of defenders present to fire, and on the number of attackers available to return that fire. A pillbox can only kill attackers if manned and armed, and any pillbox can be overcome given sufficient numerical odds against it.

---

[86] For representative Soviet densities, see Headquarters, Department of the Army, FM 100-2-1, The Soviet Army: Operations and Tactics (Washington, D.C.: USGPO, July 1984), pp. 5-11, 5-12. For representative US densities, see Headquarters, Department of the Army, FM 71-1, The Tank and Mechanized Infantry Company Team (Washington, D.C.: USGPO, November, 1988), p. 3-11.

[87] For a detailed treatment of obstacle types and construction techniques, see Headquarters, Department of the Army, FM 5-102, Countermobility (Washington, D.C.: USGPO, March 1985). For obstacle clearance techniques, see Headquarters, Department of the Army, FM 5-101, Mobility (Washington, D.C.: USGPO, March 1985). For survivability enhancement "barriers" for defenders, see Headquarters, Department of the Army, FM 5-103, Survivability (Washington, D.C.: USGPO, June 1985).

Third, it is difficult to measure the magnitude, or scope, of a given defender's barrier system in absolute terms. Few defenses are wholly devoid of barriers, given the broad definition of the term as it occurs in the literature. An infantryman who positions a claymore mine in front of his foxhole constitutes a "barrier" of a sort, and most defending units will create ad-hoc barriers and survivability improvements as a matter of routine upon occupying a position.

To assess their effects on combat outcomes, however, it will be necessary to develop some measure of the quantity or extent of a given barrier system; for graphical purposes, equivalent man-hours of engineering effort can be used as a rough first approximation.[88] While any soldier can create an obstacle, specialized assets are capable of producing more extensive barriers in less time with fewer people. By the same token, large forces of unspecialized manpower (e.g., infantry units) can nevertheless create a substantial obstacle system given sufficient time. A scale for measurement by which any combatant can be expressed as some multiple of the barrier-creation capability of a specialized engineer permits us to put such alternatives on a common yardstick.

Thus, in evaluating barrier effects, we must recognize that attackers exercise an important element of choice in contending with any given barrier, and that we cannot assess the effectiveness of a barrier in isolation from the strength of the defensive maneuver units manning it. Taken together, these two observations enable us to represent the effects of defensive barrier employment in terms of its effects on the attacker's trade-off between velocity and casualties.

1. Graphical Analysis

This effect is described graphically in Figure C-20. CVT_1 and CVT_2 represent posited casualty-velocity trade-off frontiers for defenses of constant size and weapon mix, but differing levels of barrier preparation—with CVT_2 representing the greater engineer-equivalent man-hours of effort. As shown in Figure C-20, for any given velocity, greater defensive barrier effort produces greater attacker casualties, but this effect is not uniform across all values of V. In particular, for low attack velocities, a given difference in barrier effort produces the smallest increase in casualties; the higher the velocity, the greater the

---

[88] The equations below and the associated computer code, on the other hand, will use the nominal effect of an assumed barrier system on attacker:defender loss exchange ratios at the point of attack as a more direct approximation of the net effect of a given barrier system.

difference in attacker casualties for a given level of barrier construction effort. In effect, barriers present attackers with a choice between maintaining a high velocity at a substantial cost, or slowing down to clear the barrier—thereby limiting losses but delaying the attack accordingly.

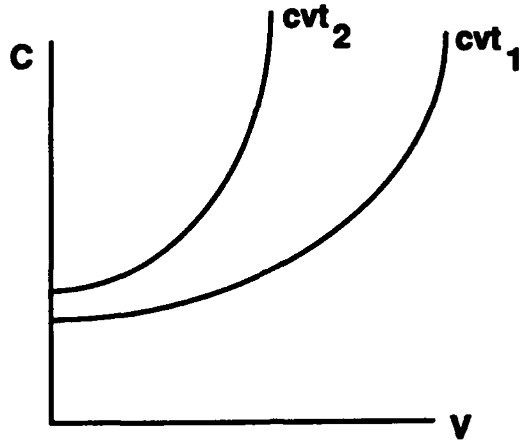

Figure C-20. The Impact of Barrier Defenses on the Casualty-Velocity Trade-off Frontier

Within this framework, the effect of barriers can thus be combined with the effects of other changes as described above. In particular, reductions in the number of defenders available at the point of attack for a constant level of barrier effort would shift CVT_1 and CVT_2 to the right (as well as potentially reducing the level of barrier effort itself by limiting the labor available to the defender for barrier construction, and hence reducing the extent of the resulting defensive works to a level below that represented by CVT_2).

For any given set of circumstances, the implied change in the attacker's optimal velocity choice—and the consequent change in net territorial gain—is as depicted in Figure C-21. For the increase in engineer-equivalent man-hours represented by CVT_2 relative to CVT_1, the optimal attacker velocity falls from V_1 to V_2, yet casualties nevertheless increase from C_1 to C_2, resulting in a decrease in net territorial gain from IG_1 to IG_2. If the attacker chooses instead to maintain his velocity and simply accept the resulting increase in casualties, the outcome is given by point B in Figure C-21. Casualties increase much more substantially to C_3, with a consequent further decrease in net territorial gain to IG_3. In effect, barriers therefore both slow the optimal attacker's advance, and increase the defensive loss-exchange ratio—but do so proportionately to the force-to-force ratio and the weapon mix at the point of attack.

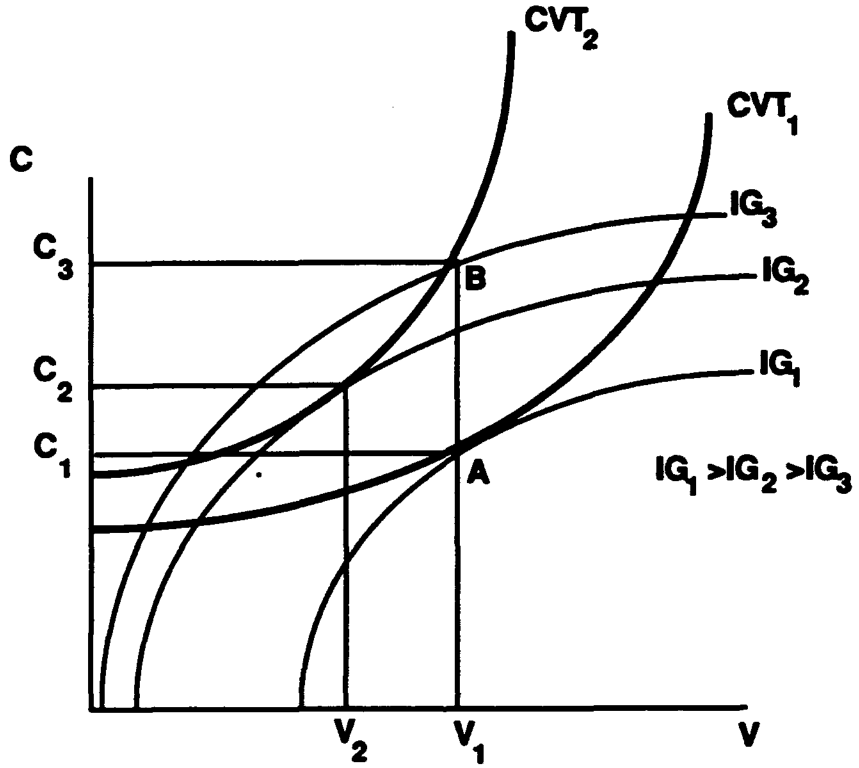

Figure C-21. Optimal Velocity Choice as a Function of Barrier Defense

I. EQUATIONS

Given the analysis presented above, we must now develop a set of equations to embody these trade-offs and enable us to compute an explicit value for net territorial gain as a function of our specified independent variables. In most instances, the functional forms required to represent the trade-offs we have described will be too complex to permit direct, closed form solution for the tangency points and intersection points identified in the graphical analyses above. Our strategy in developing a quantitative expression of those relationships will therefore be to develop an equation to predict net territorial gain given any particular set of force employment choices; ensure that the predicted ground gains respond in the manner described above to variations in those choices; then use numerical approximation techniques to estimate the optimal values for the endogenous choice variables given specified values for exogenous variables (such as technology or barriers).

For our purposes, the nature of this optimum is game theoretic. That is, we can describe a theater offensive as a two-person, zero sum game. The players are the two combatants, Blue and Red. Player strategies are the force employment choices given above. A Blue strategy thus consists of a unique quadruple of values for Blue's fraction of forces deployed forward, predeployed depth, fraction of forward forces withdrawn, and fraction of reserves used for counterattack.[89] A Red strategy consists of a unique assault velocity. The payoff for a given combination of a Red and a Blue strategy is the net territorial gain that Red would achieve in an offensive fought with the given force employment choices by Red and Blue. Red tries to maximize this payoff; Blue tries to minimize it. As a point of departure, we will assume that Blue chooses a strategy which is then observed by Red, enabling Red to choose its own strategy with prior knowledge of Blue's choice. Red's optimal strategy is thus simply the velocity which maximizes net territorial gain for the given Blue strategy. Blue's optimal strategy, given that Blue must choose without prior knowledge of Red's choice, is the quadruple for which the maximum net territorial gain as a function of Red's velocity choice is lowest (i.e., the minimax strategy).[90]

Given this game theoretic context, our task here is thus to develop a set of equations to compute a payoff (i.e., a net territorial gain) for a given set of Red and Blue force employment choices and exogenous independent variables. (For more detail on the algorithm used to identify the optimal choices given this functional relationship between choices and payoffs, see appendix E). We will compute this net territorial gain payoff in five steps. First, we will define some additional terms. We will then develop an expression for the attacker's attrition, after which we will describe the attacker's rate of advance,

---

[89] See specific definitions below.

[90] Note that the use of a sequential move game is conservative with respect to Blue's ability to hold (which is our primary interest here). In a real confrontation, the defender would have some (albeit imperfect) prior knowledge as to the attacker's likely assault velocity, while the attacker would have only imperfect knowledge of the defender's strategic choices prior to launching his attack. By assuming a sequential move game in which the defender knows only the functional relationship between Blue strategy, Red strategy and outcomes, but the attacker observes the defender's actual choices prior to moving, we can therefore be confident that any Blue strategy which holds Red short of breakthrough would also do so in a simultaneous move game in which neither side could observe the other's behavior in advance. The same does not hold true for strategies that produce Red breakthrough, however—that is, some Blue strategies that produce breakthroughs under the sequential game structure assumed here would not produce breakthroughs under a simultaneous game assumption.

the density of flank defenses the attacker requires to thwart defensive counterattack, and finally, we will combine these to obtain an explicit expression for net territorial gain.[91]

1. Preliminary Definitions

To begin with, let B be the total Blue maneuver (i.e., infantry and armor) strength available for combat in the theater at the initiation of hostilities, measured in AFVEs. Let B_{ART} be the total Blue artillery available for combat, measured in tubes. Let us further define B_{FWD} to be the number of Blue AFVEs allocated to forward positions, and B_{RSV} to be the number of Blue AFVEs allocated to theater reserve. If ϕ_{FWD} is the fraction of his total AFVEs Blue chooses to allocate forward, then $B_{FWD} = \phi_{FWD}B$. Since Blue has only a total of B AFVEs, $B_{RSV} = B - B_{FWD}$.

As for Red, let us define R as the total Red maneuver strength available for combat in the theater at the initiation of hostilities, measured in AFVEs. Let R_{ART} denote the total Red artillery available for combat, measured in tubes. Let R_{OFV} be the number of AFVEs allocated to the sector of main attack by Red. Note, however, that R_{OFV} will often be greater than the maximum number of AFVEs Red can simultaneously present in an attack, since the terrain limits the density of forces an attacker can mass on a given frontage to a finite value ρ_{MAX}. We will therefore define R_{ECH} as the number of AFVEs Red presents simultaneously within direct fire range of the defender in a single assault wave (or "echelon").[92]

If we define λ_{THR} as the length of the theater frontier to be defended (measured in kilometers), and λ_{ATK} as the combined length of Red main attack frontage in the theater, then we obtain a simple abstraction of the military geography of a theater of war, as

---

[91] Attrition to the defender will be treated implicitly via the number of engagements fought during the attacker's advance, and the number of defenders withdrawn from each of these engagements—plus defender losses due to deep interdiction by the attacker.

[92] In fact, terrain constraints on concentration are unlikely to be as absolute as suggested here; far more likely is a diminishing marginal return relationship in which additional forces beyond a given concentration yield smaller (but non-zero) benefits the larger their number. As a simplifying assumption, however, we will assume here that there is a finite limit to concentration ρ_{MAX}. It should also be noted that while R_{ECH} cannot exceed ρ_{MAX}, it need not equal ρ_{MAX}. Casualties to an engaged echelon, for example, will tend to decrease R_{ECH} to levels below ρ_{MAX} as an operation proceeds; replacement of one echelon by another will tend to restore R_{ECH} to levels approximating ρ_{MAX} at regular intervals, but since this exchange is time consuming it is executed far less frequently than casualties are suffered.

depicted in Figure C-22[93]. Ultimately, λ_{ATK} would be chosen optimally by the Red attacker;[94] as a point of departure, we will assume a λ_{ATK} determined as a simple function of the size of the Red theater force:

$$\lambda_{ATK} = k_0 + k_\lambda R \qquad (1)$$

where k_0 and k_λ are constants.[95]

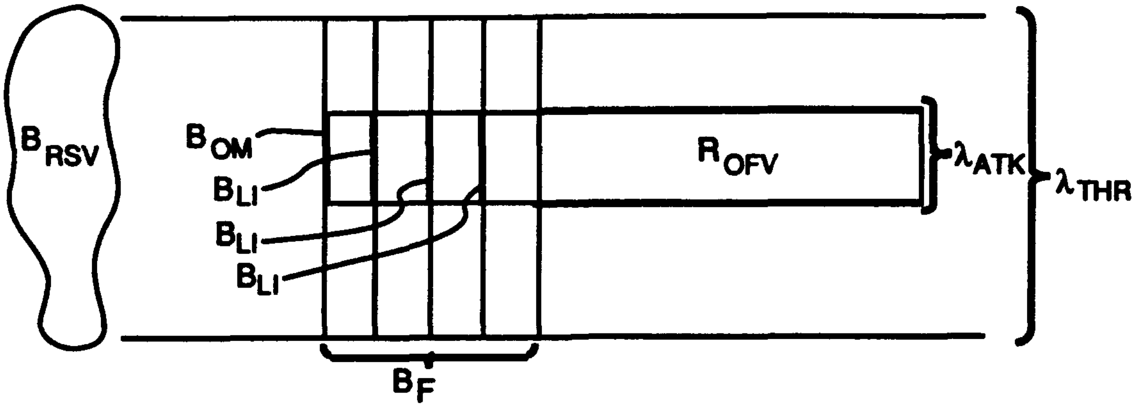

Figure C-22. Theater Geometry

This geometry introduces two additional terms, B_{LI} and B_{OM}. To define these, let us assume, as an initial simplifying assumption, that Blue deploys his forward forces in several discrete, identical lines (e.g., four, as illustrated in Figure C-22). This enables us to describe Red's advance through the defended zone as a series of discrete engagements in which Red assaults successive Blue lines. In this context, B_{LI} denotes the number of Blue AFVEs deployed on a given line within the Red attack frontage λ_{ATK}, and thus

---

[93] While this abstraction could be made more detailed in further work (for example, by differentiating the troop strengths of Blue defensive lines, by distinguishing different numbers of lines in different parts of the theater, or so on), the simple version given here is sufficient for our purposes, and will be utilized as an analytic point of departure.

[94] Moreover, in actual combat, Red would ordinarily attack in several sectors at once, rather than a single, combined frontage as shown here. For analytic purposes, however, we will describe the theater outcome in terms of combined attack frontage rather than in terms of geographically localized subsets.

[95] In particular, the form given here is intended to provide the simplest possible formulation that would facilitate downward scaling from a known frontage value on the basis of observed (Soviet) practice with respect to offensive frontage as a function of force size; for a more detailed discussion, including values (and sources) for constants, see appendices F and D.

constitutes the size of the Blue force opposing each assault. Since the lines are identical, B_{LI} is the same for each line at the outset of hostilities. This implies:

$$B_{LI} = \frac{\phi_{FWD}\, B\, \lambda_{ATK}}{n_{LI}\, \lambda_{THR}} \tag{2}$$

where n_{LI} is the number of pre-deployed Blue defensive lines.

Blue, however, will move reserve AFVEs to the attack sector once he discovers its location. As a further simplifying assumption, let us assume that Blue allocates those arriving reserves which are to be used for reinforcement (as opposed to counterattack) to the last line which Red will be able to reach. That is, Blue neither wastes his reinforcements by deploying them so far from the international border that Red is halted prior to reaching them, nor does Blue spread his reinforcements out in penny packets among forward lines likely to be overrun by Red. Rather, Blue masses his arriving reinforcements for a single, decisive battle on the line that he expects will—with the benefit of reinforcement—be the point at which he has the best chance of halting the attacker. Let us define the strength of this line as B_{OM} (measured in AFVEs). Thus, while B_{LI} is the same for each line and is constant over time, B_{OM} is initially the same as B_{LI}, but increases over time as Blue reinforcements arrive. In particular, if we assume that these reserves are initially distributed more or less uniformly across the theater:

$$\psi_{BOMT}(t) = \begin{cases} 0 \quad \text{if } t < t_{BPREP} + t_{BST} \\ \max\left[0, \dfrac{V_{RSV} B_{RSV}(1-\phi_{CA})}{\lambda_{THR}} - \delta_4 (1-\phi_{CA})\right] \\ \quad \text{if } \dfrac{\lambda_{THR}}{V_{RSV}} + t_{BPREP} + t_{BST} > t > t_{BPREP} + t_{BST} \\ 0 \quad \text{if } t > \dfrac{\lambda_{THR}}{V_{RSV}} + t_{BST} + t_{BPREP} \end{cases} \tag{3}$$

where $\psi_{BOMT}(t)$ is the rate at which Blue forces arrive on the decisive line, t_{BPREP} is the time required by arriving defenders to prepare their positions before undergoing attack, t_{BST} is the time at which Blue reserves begin to move toward the point of attack, V_{RSV} is the speed of Blue's reserves (in kilometers per hour); $(\lambda_{THR}/V_{RSV} + t_{BST} + t_{BPREP})$ approximates the time at which the last of the blue reserves arrives; δ_4 denotes losses to Red long range tacair and ACM (in AFVEs per hour); and ϕ_{CA} denotes the fraction of

Blue reserves designated for use in counterattack (thus $1-\phi_{CA}$ provides the fraction of reserves to be used in reinforcing B_{OM}).[96] It follows that:

$$B_{OM}(t)=\begin{cases} B_{LI}+\max\left[0,(t-t_{BPREP}-t_{BST})\,\psi_{BOMT}(t)\right] \\ \quad \text{if } t-t_{BPREP}-t_{BSTART}<\dfrac{\lambda_{THR}}{V_{RSV}} \\ B_{LI}+B_{RSV}(1-\phi_{CA})-\delta_4\,(t-t_{BPREP}-t_{BST}) \\ \quad \text{if } t-t_{BPREP}-t_{BST}\geq\dfrac{\lambda_{THR}}{V_{RSV}} \end{cases} \tag{4}$$

We will denote the number of Blue artillery tubes available for the support of a given defensive line as B_{ARTLI}; similarly, R_{ARTECH} gives the number of Red artillery tubes available to support a given assault echelon. Note that because of its range, artillery organic to several lines of defenders or to several assault echelons of attackers can be used to support any given line or any given echelon, thus:

$$B_{ARTLI}=\frac{\phi_{FWD}B_{ART}\lambda_{ATK}}{\lambda_{THR}} \tag{5}$$

$$R_{ARTECH}=\lambda_{ATK}\,\rho_{MAX}\,k_{AE}\,\frac{R_{ART}}{R} \tag{6}$$

where k_{AE} is a constant representing the number of assault echelons of artillery the attacker masses forward for support of a single echelon.

The depth of the Blue defense, D, is a function of two forms of depth—predeployed and rolling, or withdrawal-induced. To relate these two, note that predeployed depth is determined by the number of Blue defensive lines, n_{LI}, and the depth of a single line, D_{LI}; that the number of AFVEs assigned to each Blue line is the same, B_{LI}, and that the fraction of AFVEs withdrawn from battle on any given line is given by w (like n_{LI},

---

[96] Ideally, in a more complicated model we would be more sensitive to the particular location of Blue reserve assembly areas relative to the international border, the position of Blue's prepared defenses and the location of the attacker's chosen point of attack; for the simple treatment here we have instead employed, on the one hand, a defender best-case assumption as to reserve assembly area setback (i.e., that setback distance = 0), but on the other hand, a defender worst-case assumption as to the location of the point of attack (i.e., that it is very near a flank). While this is not unreasonable as a first order approximation, refinement could profitably be pursued in further work.

chosen by the defender). If w_{SURV} denotes the fraction of AFVEs on any given line that survive attack by Red tacair and short range ACM during withdrawal and thus redeploy successfully, then when any given line is taken by Red, a force of (w_{SURV} B_{LI}) is made available to extend the depth of the Blue defended zone, where:

$$w_{SURV} = \frac{w\left(B_{LI} - \mu_{RS}\right)}{B_{LI}} \tag{7}$$

and where:

$$\mu_{RS} = \left(\delta_{1R}\hat{R}_{ARTECH} + \delta_{2R}\right)k_{ACMC} \tag{8}$$

with δ_{1R} denoting kills of Blue AFVEs by Red short range ACM (in units of Blue AFVEs killed per assault per Red artillery tube); δ_{2R} denoting kills by Red close air support aircraft (in units of Blue AFVEs killed per assault per two kilometers); and k_{ACMC} giving the ratio of ACM and CAS effectiveness against moving targets to ACM and CAS effectiveness against stationary targets.[97] $\hat{R}_{ARTECH}$ denotes a scaled quantity to correspond to a nominal two kilometer attack frontage for the purposes of casualty estimation.[98]

Let us assume that Blue has prepared secondary positions of comparable defensive potential to those of the initially manned forward lines, and that Blue will fight from these secondary positions in the same manner as he does from the initial lines—and in particular, that he continues to withdraw a fraction w of the strength on any given line prior to its capture. Let us further assume that Blue cannot prepare an infinite number of potential positions, and thus limits preparations to a depth such that Blue can continue to mount a defense of size B_{LI} on the secondary lines as well. It follows that the ultimate number of prepared lines in the Blue defended zone will be:

---

[97] ACM will ordinarily be less effective against stationary, concealed targets than against exposed, moving ones. Similarly, CAS is less effective when targets are not exposed. k_{ACMC} provides in effect a measure of the difference in effectiveness against concealed and exposed targets.

[98] Scaling is necessary to exploit fully the JANUS runs by which attrition constants were fit, inasmuch as the JANUS experiments were fought at the battalion level on a two kilometer frontage. For scaling factors and a more detailed treatment, see the discussion accompanying equation 19, below.

$$n_{LITOT} = n_{LI} + w_{SURV} n_{LI} + w^{2}_{SURV} n_{LI} + \dots + w^{x}_{SURV} n_{LI} \qquad (9)$$

where:

$$W^{x}_{SURV}\, n_{LI} \geq 1 \qquad (10)$$

$$\text{and} \quad W^{x+1}_{SURV}\, n_{LI} < 1$$

We can rewrite (9) as the sum of a finite series:

$$n_{LITOT} = n_{LI} \sum_{i=0}^{x} w^{i}_{SURV} \qquad (11)$$

from which it follows that:

$$n_{LITOT} = n_{LI} \left(\frac{1 - w^{x+1}_{SURV}}{1 - w_{SURV}} \right) \qquad (12)$$

Given that (10) implies:

$$W^{x}_{SURV}\, n_{LI} + \in\, = 1 \qquad (13)$$

and since our definition of Blue's defense preparation criterion implies that defenders surviving the xth withdrawal will not be able to occupy prepared positions (and hence will be relatively combat ineffective), we will therefore assume that for combat purposes, the defenders surviving the xth withdrawal, ∈, are of negligible value; hence we will assume that for combat purpose, ∈ = 0, and thus, by (13):

$$w^{x}_{SURV} = \frac{1}{n_{LI}} \qquad (14)$$

which enables us to rewrite (12) as:[99]

---

[99] Note that equation 10 implies:

$$x = \mathrm{INT}\left[\frac{-\ln(n_{LI})}{\ln(w_{SURV})} \right]$$

and thus the more exact but somewhat less computationally efficient:

$$n_{LITOT} = \frac{n_{LI} - w_{SURV}}{1 - w_{SURV}} \qquad (15)$$

Since the total depth of the defended zone in kilometers is the product of the total number of lines n_{LITOT}, and the depth of each line D_{LI}, we can thus write:

$$D = D_{LI}\left(\frac{n_{LI} - w_{SURV}}{1 - w_{SURV}}\right) \qquad (16)$$

2. Attrition

We will now develop an expression to quantify the casualty-velocity tradeoff relationship depicted graphically in Figures C-1 through C-12 and C-17 through C-21. In particular, predicted casualties must reflect the effects of attacker velocity, defender depth and withdrawal, weapon class, terrain preparation, and local force-to-force ratios in the manner described above. We will then use that expression to describe Red attrition in terms of the number of echelons of Red forces that are lost in taking any given Blue line, a form which we will find convenient for computing net territorial gain, below.

To this end, we must define some additional terms. We will consider attacker casualties in terms of the losses in maneuver unit AFVEs required in order to take a single defended line, for a given frontage λ_{ATK}. Let us define this quantity as C, a real number between zero and infinity. Given this definition, we may consider a given defended line "taken" for any assault in which Red's available assault forces at the point of attack exceed C. Note, however, that as C is defined in absolute, rather than percentage terms, it will vary with the scope of the engagement, and thus must be expressed as a function of λ_{ATK}.

$$n_{LITOT} = n_{LI}\left\{\frac{1 - w_{SURV}^{\left(1 + \mathrm{INT}\left[\frac{-\ln(n_{LI})}{\ln(w_{SURV})}\right]\right)}}{1 - w_{SURV}}\right\}$$

In preference to this form, the code in appendix E uses the approximation given in equation (15), resulting in a difference of between zero and 1.0 in the value of n_{LITOT}, and thus, by equation (16), a difference of between zero and D_{LI} in the depth of the Blue defense.

Attacker assault velocity V will be defined as per the discussion above: the time between initiation of attack preparation and arrival of the first successful assault wave on the objective, divided by the distance covered by that assault wave (V is thus a real number between zero and infinity, in kilometers per hour). Note that the attacker's *choice* of V is thus not necessarily the same as his *realized* rate of advance ψ_{ROA}. V represents the pace of the assault itself while that assault is actually underway, but a given engagement may consist of a combination of assault and lull time, in which attackers re-group after an unsuccessful attempt, or in which the commander and his staff plan commitment of additional echelons to the assault. Engagements in which many waves, or echelons, of attackers are required to defeat a stubborn defense may thus involve a sizeable interval in which no attackers are visibly pushing forward. The larger the number of echelons required to take a given defensive line, the longer the total lull time involved in the engagement. The attacker's realized rate of advance, ψ_{ROA} (in effect, the overall rate at which defensive lines fall to the attacker) is thus a function both of (a) the rate at which assault waves close with the enemy, and (b) the number of such waves required, and thus the total duration of the lull time. V as given here refers only to the first of these; the rate at which defensive lines fall to the attacker, ψ_{ROA}, will be discussed in greater detail under "Attacker Rate of Advance" below.

Our description of attacker casualties must also include the effects of defensive depth on attacker losses. To capture this interaction, we must reflect both the mass effect and the entropic effect of depth, and we must account for the reduction in attacker casualties associated with "rolling depth" via defender withdrawal. The mass effect is already incorporated in the definition of B_{LI} as given in equation 1 (i.e., B_{LI} is given as a decreasing function of n_{LI}). To capture the entropic effect, let us define a real number γ, between one and infinity, as a scalar multiple by which to represent the increase in attacker casualties in a given assault as a result of entropy induced by the lead echelon's advance through defended depth prior to the assault in question, relative to an attack conducted with perfect coherence under otherwise identical circumstances. In particular:

$$\gamma = 1 + k_8 D_{LI} n_{ASLT} \tag{17}$$

where D_{LI} is the depth of a single defended line (measured in kilometers from the forwardmost occupied position of one line to the forwardmost occupied position of the line behind it); n_{ASLT} is the number of successive attacks the given assault wave has completed prior to the attack in question (and thus the number of defended lines it has traversed); and k_8 is a constant.[100] A γ value of 1.5, for example, would signify that an assault conducted after an advance through ($D_{LI}n_{ASLT}$) kilometers of defended territory would suffer 50 percent higher casualties than one conducted with perfect coherence. A γ value of 1.0, on the other hand, would correspond to the commitment of a fresh assault echelon which had not been subject to entropic effects and thus would suffer no additional, or excess casualties as a consequence of lost coherence.

To account for the reduction in attacker casualties as a result of defensive withdrawal, let us define a quantity α between zero and one, as a scalar multiple by which to represent the decrease in attacker casualties in a given assault as a result of early termination of defensive fire upon withdrawal, relative to a fight to the finish under otherwise identical circumstances. In particular:

$$\alpha = 1 - w^{k_5} \tag{18}$$

where w is the fraction of the defending force withdrawn, and k_5 is a constant. An α value of 0.5 would signify that an assault against a defense which withdrew a fraction w of its AFVEs would suffer only 50 percent of the losses it would have taken if the defender had fought to the finish. (Equation 18 thus corresponds to the curve depicted in Figure C-10 for $k_5>1$).

With respect to the effects of terrain preparation, it will be necessary here to be somewhat circumspect. Whereas it is possible to test functional forms and fit values for constants with respect to phenomena such as withdrawal or entropy on the basis of JANUS investigation, the limitations of the LLNL version of JANUS available at IDA make a detailed exposition of the impact of particular levels of barrier construction effort problematic. While the general impact of increased barrier deployment will be taken to correspond to that depicted in Figures C-20 and C-21, we will thus not attempt to compute a specific effect as a function of a specific labor contribution. Pending a more detailed study of this effect, we will in the meantime approximate its impact by defining a parameter ß, a real number between one and infinity, to be a scalar multiple by which to denote the increased slope of the attacker's casualty-velocity tradeoff frontier as a result

---

[100] Representing the increase in attacker casualties associated with entropic effects per kilometer of defensive depth penetrated by a single assault echelon. Since k_8, D_{LI} and n_{ASLT} are all positive, nonzero real numbers, γ therefore cannot be less than 1.0. Values for key constants are fit statistically from the JANUS results described in appendix D. For values and procedures, see appendix D below.

of the availability to the defender of additional barrier preparation labor not organic to the defending maneuver units themselves.

As for the effects of differing combined arms balances, let us define a quantity ϕ_{INF}, which is the sum of the fraction of the attacker's total AFVEs which are infantry, and the fraction of the defender's total AFVEs which are infantry. Recall that the effects on the casualty-velocity trade-off of increasing infantry content were broadly similar, whether the increased infantry belonged to the attacker or the defender (see Figures C-17 and C-18). While one could imagine differences in detail sufficient to justify separate treatment, statistical tests on the results of the JANUS runs conducted to evaluate these equations failed to support such a formulation (see appendix D). In the absence of further experimental work, ϕ_{INF} is thus presented here as a single, summary value. Note also that since maneuver force strength is expressed in terms of notional infantry and armor components, it is therefore redundant to represent armor fractions explicitly; they are implicit in ϕ_{INF}. Note that artillery and tacair/ACM strength, however, are *not* implied by ϕ_{INF} and thus require explicit representation, as given in equations (5) and (6) above.

Finally, we will represent the local force-to-force ratio at the point of attack in terms of AFVEs actually in contact with the enemy at any given time—that is, R_{ECH} and B_{LI} as defined above. The effect of attacker echelonment, for example, is thus to provide forces for a series of firefights, each of whose local force-to-force ratio is determined solely by the strength of the forces within direct fire range of one another at the time of the firefight itself. Similarly, artillery and other fire support will be represented in terms of systems immediately available for participation in the given firefight.

Given the above definitions, we may now write:[101]

---

[101] Note that this equation can produce unrealistically high casualty estimates for extreme values of some independent variables—in particular, for very low values of ϕ_{INF}, the combined attacker and defender infantry fractions, and $\hat{R}_{ARTBCH}$, the number of Red artillery tubes per two kilometers of attack frontage. The particular form of this equation emerged from the empirical work described in appendix D, where the range of values represented in the experimental data extended from 0.1 to 0.9 for the fraction of the attacker's AFVEs that were *infantry* and from 0.1 to 0.9 for the fraction of the defender's AFVEs that were infantry; hence, the data represents values of ϕ_{INF} between 0.2 and 1.8. Similarly, the experimental data included Red artillery tube totals ranging from six to 52 tubes per two kilometers. Caution must therefore be exercised in interpreting results for infantry fractions or Red artillery inventories very far outside those bounds. Similarly, caution should be exercised for values of the other independent variables which greatly exceed the bounds of the experimental data from which the equation was fit, which included velocities (V) ranging from one to eight kilometers per hour, local

$$C = \frac{\lambda_{ATK}}{2}\left[\alpha\gamma\left(\frac{\hat{B}_{LI}-\mu_{RS}}{\hat{R}_{BCH}-\mu_{BS}}\right)\left(\beta\, k_3\, \phi_{INF}\, V + \frac{k_4}{\phi_{INF}} + \frac{k_1 \hat{B}_{ARTLI}}{(1+V)} + \frac{k_6}{\hat{R}_{ARTECH}(1+V)}\right) + \mu_{BS}\right] \tag{19}$$

where:

$$\mu_{BS} = \delta_{1B}\, \hat{B}_{ARTLI} + \delta_{2B} \tag{20}$$

$$\mu_{RS} = \left(\delta_{1R}\, \hat{R}_{ARTECH} + \delta_{2R}\right) k_{ACMC} \tag{21}$$

and where k_1, k_3, k_4, and k_6 are constants; δ_{1B} denotes the contribution of Blue short range ACM (in units of Red AFVEs killed per assault per Blue artillery tube).[102] δ_{2B} denotes the contribution of Blue close air support aircraft (in units of Red AFVEs killed per assault per two kilometers of front). δ_{1R} gives the kills of Blue AFVEs by Red short range ACM (in units of Blue AFVEs killed per assault per Red artillery tube); δ_{2R} gives the kills by Red close air support aircraft (in units of Blue AFVEs killed per assault per two kilometers); k_{ACMC} gives the ratio of ACM and CAS effectiveness against moving targets to ACM and CAS effectiveness against stationary targets.[103] As noted above, C is given as a function of attacker velocity, α, γ, β, weapon class availability and the local force-to-force ratio, $\hat{B}_{LI}/\hat{R}_{BCH}$ (where $\hat{B}_{LI}$ and $\hat{R}_{BCH}$ are reduced by the effect of preliminary engagement by ACM and CAS). C is also scaled to account for the scope of the engagement by reference to λ_{ATK}; inasmuch as the JANUS experiments to be

---

force to force ratios $\hat{B}_{LI}/\hat{R}_{BCH}$ ranging from 7.5:1 to 1:1.5, and Blue artillery inventories ranging from zero to 104 tubes per two kilometers. For a more detailed discussion, see the treatment under "Limitations" below.

[102] Since short range ACM is organic to ground combat formations, its availability will consequently be a function of the availability of the combat units with which it is associated. This is represented here by expressing ACM contributions in terms of of traditional artillery availability.

[103] As noted above, ACM will ordinarily be less effective against stationary, concealed targets than against exposed, moving ones. Similarly, CAS is less effective when targets are not exposed. k_{ACMC} provides in effect a measure of the difference in effectiveness against concealed and exposed targets.

described in appendix D were conducted at the battalion level on a two kilometer attack frontage, (and since constants were estimated on the basis of these results), the expression is scaled against a baseline of a two kilometer front. Similarly, $\hat{B}_{LI}$, $\hat{R}_{ECH}$, $\hat{B}_{ARTLI}$ and $\hat{R}_{ARTECH}$ denote scaled quantities to correspond to a nominal two kilometer attack frontage for the purposes of casualty estimation.[104]

Note that since γ in equation 19 is given as a function of n_{ASLT}, C is thus dependent on the number of assaults carried out by the lead echelon prior to the reference engagement. As a more general measure, we will therefore derive from (19) an expression for Red attrition in terms of the average number of echelons required to take a line, and the related quantity N_{ASLT}, the total number of assaults carried out by any given echelon. To obtain N_{ASLT}, we will substitute for γ using (17), decompose (19) into separate functions of n_{ASLT} and R_{OFV}, integrate and solve for N_{ASLT}. Observing that $C = dR_{OFV}/dn_{ASLT}$; that for a given echelon's first engagement $\gamma = 1$, and that the number of engagements a given echelon will survive is independent of λ_{ATK};[105] treating the direct contribution of Blue short range ACM and tacair implicitly (via the limits of integration in 26, since μ_{BS} does not change as a function of n_{ASLT}); and substituting for γ, we obtain:

$$\frac{dR_{OFV}}{dn_{ASLT}} = -\left[1 + Q_2 (n_{ASLT} - 1)\right] \left[\alpha \frac{(\hat{B}_{LI} - \mu_{RS})}{\hat{R}_{ECH}} Q_1\right] \qquad (22)$$

---

[104] The JANUS runs by which the constants were fit consisted of a series of engagements fought on a two-kilometer assault frontage (see appendix D for a more complete description). The constants estimated from those runs therefore represent quantities per two kilometers of front. Since red's theater attack frontage as a whole will ordinarily be greater than two kilometers, the casualty equation is written in such a way as to compute casualties per two kilometers of front; this figure is then scaled up to the actual theater attack frontage (by multiplying the bracketed quantity by one-half the actual frontage in kilometers). To do this, it is necessary to express red and blue forces at the point of attack in terms of AFVEs per two kilometers of front, i.e.: $\hat{B}_{LI} = B_{LI} * 2/\lambda_{ATK}$; $\hat{B}_{ARTLI} = B_{ARTLI} * 2/\lambda_{ATK}$ etc.

[105] Although marginal casualties—and thus C in equation 19—are not. In effect, as λ_{ATK} increases, both the number of Red AFVEs in a single echelon (of fixed density ρ_{MAX}) and the absolute number of Red casualties suffered per line taken, increase in direct proportion. Hence, they cancel for the purposes of determining the number of lines a given echelon can take, and n_{ASLT} is thus independent of λ_{ATK} even though C is not.

where:

$$Q_1 = \beta k_3 \phi_{INF} V + \frac{k_4}{\phi_{INF}} + \frac{k_1 \hat{B}_{ARTLI}}{(1+V)} + \frac{k_6}{\hat{R}_{ARTECH}(1+V)} \tag{23}$$

$$Q_2 = k_8 D_{LI} \tag{24}$$

which implies:

$$\frac{dR_{OFV} \hat{R}_{BCH}}{\alpha\, Q_1 (\hat{B}_{LI} - \mu_{RS})} = -(1 + Q_2 n_{ASLT} - Q_2)\, dn_{ASLT} \tag{25}$$

and:[106]

$$\int_{2R_{BPT}}^{(2\rho_{MAX} - \mu_{BS})} \frac{dR_{OFV} \hat{R}_{ECH}}{\alpha Q_1 (\hat{B}_{LI} - \mu_{RS})} = - \int_{N_{ASLT}}^{0} (1 + Q_2 n_{ASLT} - Q_2)\, dn_{ASLT} \tag{26}$$

where R_{BPT} is the cut-off (or breakpoint) strength per kilometer for a given echelon at which Red will replace it with another (R_{BPT} is thus a real number between 0 and ρ_{MAX}), and *($2\rho_{MAX}-\mu_{BS}$) is the initial strength* (per reference two kilometers) of a Red echelon as it is committed to direct fire action against Blue. The quantities ($2\rho_{MAX}-\mu_{BS}$) and $2R_{BPT}$ thus define the initial and final strengths of the committed Red echelon. N_{ASLT}, in these terms, gives the total number of assaults the given echelon will be able to complete before reaching its end strength. Integrating, we obtain:

$$\frac{(2\rho_{MAX} - \mu_{BS})^2 - (2R_{BPT})^2}{2\, \alpha\, Q_1\, (\hat{B}_{LI} - \mu_{RS})} = (1 - Q_2)\, N_{ASLT} + \frac{Q_2}{2} N_{ASLT}^2 \tag{27}$$

which is a readily solvable quadratic in N_{ASLT}; we will use the inverse of the largest root of this equation as the expression for Red attrition in terms of the number of echelons of Red forces that are lost in taking any given Blue line (n_{ECH}):

---

[106] *Given that* $dR_{OFV}/dn_{ASLT} = d\hat{R}_{BCH}/dn_{ASLT}$, which is necessarily so, since Red's casualties as a result of an additional assault are suffered by the engaged echelon, $\hat{R}_{ECH}$.

$$n_{ECH} = \frac{1}{N_{ASLT}} \qquad (28)$$

Finally, let us address the related question of the attacker:defender loss exchange ratio, h, for the special case of the attacker's assault on the final Blue line. In particular, for this decisive, final battle we can make a number of simplifying assumptions. First, we can assume a fight to the finish—Blue has no further pre-deployed positions behind the final line to which he could withdraw. Second, we can assume rough numerical parity for forces in direct contact, given the substantial degree of reinforcement characteristic of this case (Red's total surviving force in the attack sector may still substantially exceed B_{OM}, but the force that Red can present simultaneously within direct fire range of the Blue defenses is unlikely to greatly outnumber Blue's local defenders). Third, it follows that this local numerical parity occurs at a level approximating ρ_{MAX}.[107] Finally, we can assume minimum entropy for the Red assault waves—to the extent that this is truly the decisive battle, Red has a substantial incentive to husband his resources carefully, and thus to replace spent assault echelons rather than wasting scarce troops.

Given these assumptions, given that the strength of the nominal defense (and thus the number of Blue casualties suffered for the loss of C attackers) is $(\rho_{MAX}\lambda_{ATK})$,[108] and observing that for these circumstances Blue's defensive density (and thus Blue's proportional artillery availability) is equivalent to Red's assault density, we can write:

---

[107] It is also possible that B_{OM} could exceed ρ_{MAX} if Blue's arrival rate were high enough, and the duration of the fighting long enough. Arrivals in excess of ρ_{MAX} would be forced to deploy to the rear of the final line as initially defined, but there need be no visible gap between the line and the dispositions of the later arrivals; in effect, if $B_{OM} > \rho_{MAX}$ then Red will confront a final line of density ρ_{MAX} but extended continuously in a belt (rather than a "line") with greater tactical depth. While it is possible that Red could continue to advance beyond the nominal final line by fighting through part of this belt, it would be extremely difficult to hold the additional ground taken in this way (since the "breathing space" normally afforded Red by Blue's inter-line separation distance is zero here; thus consolidation of taken positions would be extremely difficult). Whether Red expends valuable forces at a high rate to advance temporarily, partway into a continuous belt that cannot be taken and then withdraw to more defensible positions to the rear, or whether Red conserves strength and halts at the edge of such a belt, the net territorial gain, G, will be the same.

[108] If B_{OM} exceeds ρ_{MAX}, absolute casualties (for both sides) would be higher if Red chose to assault the final "belt", but since neither side can cause more than ρ_{MAX} AFVEs to be simultaneously engaged per kilometer of front, the *ratio* of Red to Blue casualties will stay the same regardless of the absolute size of either force overall. Thus h, the loss exchange ratio, is as defined in equation 29 whether B_{OM} exceeds ρ_{MAX} or not.

$$h = \frac{1}{2\,\rho_{MAX}} \left[\begin{array}{l} \beta k_3 \phi_{INF} V + \dfrac{k_4}{\phi_{INF}} + \dfrac{k_1 \hat{R}_{ARTECH}}{(1+V)} + \\ \dfrac{k_6}{\hat{R}_{ARTECH}(1+V)} + \delta_{1B} \hat{R}_{ARTECH} + \delta_{2B} \end{array} \right] \qquad (29)$$

3. Rate of Attacker Advance

The attacker's realized rate of advance, ψ_{ROA}, can be described in terms of the depth of a single Blue line (and thus the distance advanced by Red per line taken) and the overall time required to take a Blue line.

As for the latter, the overall time required to take a line is the sum of the total combat time required to place an assault wave successfully onto the objective, t_{CBT}, and the time required to commit a sufficient number of assault waves to defeat the defenders on the objective, t_{CMT}. As for the former, the definition of velocity given above implies:

$$t_{CBT} = \frac{D\lambda}{V} \qquad (30)$$

As for the latter, for any single echelon, the time required to carry out a commitment to action is the sum of two components: a more or less constant preparation time, t_{OPREP},[109] and a variable movement time for the conduct of the approach march between the echelon's assembly area in the rear and the jump-off line for the assault, t_{MV}. This movement time is a function of the number of echelons required to take the position at hand. If two echelons are needed, the second need travel only the set-back distance of its own assembly area to reach the jump-off line. If three are needed, however, the third echelon must traverse not only the set-back distance of the second, but also the separation

---

[109] During which staff planning for the movement to contact, and especially, for the execution of the assault itself is completed. While some advance planning can be done prior to an actual commitment order, it is not always possible to anticipate correctly the point of commitment or the indicated axis of advance. Moreover, the circumstances of the battle at the point of attack will inevitably change over time, requiring modification of orders to subordinate commanders within the echelon to be committed, and possibly the modification of coordination and support arrangements with higher command authorities. Thus, it is unlikely that any commitment of a previously unengaged formation can be commenced without a preliminary delay for command and staff planning.

distance between the second wave's assembly area and the third's.[110] Thus, for a given number of echelons n_{ECH} required to take a single defensive line, the average travel time required for an echelon to reach the jump-off line can be written as:

$$t_M = \frac{n_{ECH}}{2} \frac{D_{ASY}}{V_{RSV}} \tag{31}$$

where D_{ASY} denotes the depth (or setback separation) of a single echelon's assembly area. Given that echelons which fail to take the defended line must still advance a substantial distance under fire before it is recognized that a further echelon will be required, we thus obtain:

$$\psi_{ROA} = \frac{D_{LI}}{t_{CBT} + n_{ECH}\left(t_{OPREP} + t_M\right)} \tag{32}$$

where t_{OPREP} is the preparation time required prior to committing a given echelon; and by substitution:

$$\psi_{ROA} = \frac{D_{LI}}{\frac{D_{LI}}{V} + n_{ECH}\left(t_{OPREP} + \frac{n_{ECH} D_{ASY}}{2\, V_{RSV}}\right)} \tag{33}$$

4. Red Flank Density

To develop an expression for ρ_{FLK}, Red's flank density, we will consider the issues of flank defense and counterattack as special cases of the general defense and attack problem. As we noted in our earlier discussion of counterattack, Red's objective with respect to flank defense is to deploy only enough force to prevent a Blue

---

[110] Of course, if Red's *station-keeping* were perfect, every time an echelon advanced into action, all other echelons behind it would instantaneously advance a distance equal to the inter-assembly-area setback—with the result that movement time for the approach march would be constant with respect to the number of echelons required to take a single line. Alternatively, if no such station-keeping movement took place at all, the final Red echelon committed prior to culmination would require a preliminary approach march equivalent to the entire distance G plus the cumulative setbacks of all prior echelons. Although neither bound is plausible, it is difficult to say where actual performance will fall between these bounds. For our purposes, we will therefore assume that station-keeping moves are executed only when a *line* is taken, but that such moves are executed perfectly when they are attempted. Thus, while engaged against a given Blue line, no station-keeping occurs and the final echelon called for against that line must conduct an approach march of a distance equal to the cumulative setbacks of the other echelons committed before it against that particular Blue line.

counterattack from breaking through. Any more than this minimum represents an unnecessary diversion of potential assault forces from the essential task of penetrating Blue's defense; any less than this is to risk catastrophic failure should Blue break through and cut off Red's spearhead forces. How then can we specify this minimum, given the military logic of attack and defense described above? Our approach will be to develop a general expression for net territorial gain as a function of the strength of the defenses occupying the attacker's chosen frontage, set this net territorial gain equal to the distance Red can afford to allow the Blue counterattack to advance without cutting off the Red spearhead, then solve for the minimum flank defensive strength consistent with this Blue advance distance.

To do this, we must introduce additional notation to denote the reversed identity of attackers and defenders under a condition of counterattack. In particular, B_{CA} will represent the total AFVEs available to Blue for the counterattack (and is thus the counterpart to R_{OFV} in the theater attack case). B_{CAECH} will signify the number of Blue AFVEs committed to a single assault echelon, while B_{ARTECH} will denote the number of Blue artillery tubes available to support that echelon. Blue's counterattack frontage will be given by λ_{CA}. R_{LI} will represent the Red AFVEs defending a single line; C_{CA}, the number of counterattacker casualties suffered in taking a *single* Red defensive line; $R_{OM}(t_{CA})$, the number defending the reinforced line at time t_{CA} (where t_{CA} denotes time in hours since the initiation of the counterattack—note that unless Blue's counterattack and Red's theater attack begin at the same moment, $t_{CA} \neq t$); and n_{LICA} the number of lines with which Red defends a given flank. Correspondingly, $\psi_{ROM}(t_{CA})$ denotes the rate at which Red flank reinforcements arrive, and ψ_{BCAT} the rate at which Blue loses counterattack strength over time during the counterattack. V_{CA} will be Blue's counterattack velocity, and ψ_{ROACA}, the overall rate at which Blue advances through the Red flank defense.

Let us begin the development by specifying $B_{CAST}(t)$, the initial size of the Blue counterattack force against which Red must defend. Note that this quantity changes over time as Blue builds up a larger force of arriving reserves in positions suitable for counterattack. The timing of the counterattack thus affects the size of the counterattack force. To define $B_{CAST}(t)$, let us therefore begin by addressing the issue of counterattack timing.

In a sense, the ideal time for Blue's counterattack is the culminating point of Red's offensive. As Clausewitz argued, this is Red's moment of greatest vulnerability; moreover, the later the jump-off time for the counteroffensive, the larger the counterattack force Blue will have assembled at the point of attack. Yet if Blue delays too long, he runs the risk that Red will ignore his counterattack threat, concentrate a larger force at the spearhead point, and attempt to break through before Blue's counterattack can be completed. For the threat of counterattack to be serious enough to compel a real diversion of Red assault forces, it must thus be launched early enough to break through Red's flank defenses before Red breaks through Blue's theater defense. It follows that Blue must begin his counterattack early enough that:

$$t_{BRKB} \le t_{BRKR} \tag{34}$$

where t_{BRKB} denotes the projected time at which Blue's counterattack would break through Red's flank defense, and t_{BRKR} denotes the projected time at which Red's theater offensive would break through Blue's theater defense. As a rough approximation:

$$t_{BRKR} = {}^{D}\!/\!_{\Psi_{ROA}} \tag{35}$$

$$t_{BRKB} = t_{JO} + \frac{D_{CA}}{K_{CROA}} \tag{36}$$

where k_{CROA} is a constant denoting the blue theater commander's preliminary estimate of what his rate of advance will be once the counterattack is launched, and where t_{JO} is the ("jump off") time at which the counterattack is launched, implying:

$$t_{JO} = \frac{D}{\Psi_{ROA}} - \frac{D_{CA}}{K_{CROA}} \tag{37}$$

To determine the size of the counterattack force available as of t_{JO}, note that (like B_{OM}), the size of the counterattack force, $B_{CAST}(t)$, is bounded. As an upper limit, Blue can amass no more counterattackers than he has reserves—no matter how much time he has available before jump-off—and a fraction of these reserves, $(1-\phi_{CA})$ are to be used in reinforcement rather than counterattack. At the lower limit, no reserves are available for any purpose until t_{BST}, the time at which Blue begins to move those reserves to the point of attack. Thus:

$$B_{CAST}(t) = \begin{cases} 0 \text{ if } t_{JO} \le t_{BST} & (38) \\ \max\left[0, \left(t_{JO} - t_{BST}\right)\left(\frac{V_{RSV} B_{RSV} \phi_{CA}}{2\lambda_{THR}} - \frac{\delta_4 \phi_{CA}}{2}\right)\right] & \\ \text{if } \frac{\lambda_{THR}}{V_{RSV}} + t_{BST} > t_{JO} > t_{BST} & \\ \max\left[0, \frac{B_{RSV} \phi_{CA}}{2} - \left(\frac{\lambda_{THR}}{V_{RSV}} - t_{BST}\right)\frac{\delta_4 \phi_{CA}}{2}\right] & \\ \text{if } t_{JO} \ge \frac{\lambda_{THR}}{V_{RSV}} + t_{BST} & \end{cases}$$

To continue our development, note that the special circumstances of counterattack enable us to make a number of simplifying assumptions relative to the treatment of the theater attack problem. As noted above, for example, the width of the penetration corridor limits the depth available for flank defense; in fact, not even this distance is really available, inasmuch as a clear channel must be maintained through the center of this corridor to permit continued resupply and reinforcement of the Red spearhead. The width of this clear channel, λ_{LOC}, will ordinarily be proportional to λ_{ATK}, in that the wider the attack frontage, the larger the number of attackers occupying that frontage (for a constant ρ_{MAX}), the larger the number of roads required for resupply, and thus the wider the corridor required to provide the necessary number of non-intersecting routes.[111] Given this, we can define the total depth available to Red's flank defenders, D_{CA}, as:

$$D_{CA} = \frac{\lambda_{ATK} - \lambda_{LOC}}{2} \qquad (39)$$

where λ_{ATK} is the Red theater attack frontage (in kilometers), and thus the width of the Red penetration corridor, while λ_{LOC} denotes the the width (in kilometers) of the minimum clear channel through that corridor required for the continued supply and reinforcement of the Red spearhead. Since Blue may counterattack both flanks simultaneously, the depth available for defense against a single pincer is thus half the total. The constrained width of this line of communications through the penetration

---

[111] For a map analysis suggesting an average ratio of about 1:3 between λ_{LOC} and λ_{ATK}, see S. Biddle, "Non-Intersecting Routes per Kilometer, FRG," unpublished manuscript, IDA.

corridor also suggests that for Red, the rate of reserve arrival ($\psi_{ROM}(t_{CA})$) can be approximated by a constant, rather than computed as a function of the number of reserves in the theater as for Blue (i.e., since road availability through a narrow corridor is likely to be the binding constraint on Red reserve arrivals, it is unlikely that large increases in reserve numbers would produce corresponding increases in reserve arrivals for Red).

This restricted depth implies a number of consequences. First, it suggests that the number of Red defensive lines is essentially fixed by the width of the penetration corridor, i.e.:

$$n_{LICA} = 1 + \frac{D_{CA}}{D_{LI}} \tag{40}$$

where D_{LI}, the depth of a single defensive line, is assumed to be the same for Blue and Red.

These depth constraints suggest that Red will typically be unable to afford an elastic defense on the flanks of his penetration corridor, but rather must stand and fight to the finish. This in turn implies that for the special case of counterattack, $\alpha = 1$; in effect, we will *assume that* Red *does not* withdraw when there is so little depth to withdraw into. As a consequence, we would expect a modest number of high-intensity local engagements rather than extended, rolling delays in depth. Thus, we can further assume $\gamma = 1$, since these conditions are least likely to produce the extended advances by single echelons through multiple engagements that produce high γ values. Relative to Blue's theater defenders, Red's flank defenders will have significantly less time to prepare defensive barriers and obstacles; we will therefore assume $\beta = 1$ for the purposes of counterattack. Finally, we will not explicitly consider counter-counterattacks; that is, the flanks of Blue's counterattack penetration are likely to be short enough that Red attacks against Blue's counterattack force would be dealt with by the Blue spearhead itself, rather than by dedicated, passive flank defenders.[112] As such, there need be no diversion of Blue strength into flank defense per se.

---

[112] Moreover, the length of Blue's flanks would in any case be short enough that the marginal value of allocating arriving reserves to counter-counterattack would presumably be low for Red relative to the marginal value of allocating those forces to reinforcement of the Red omega line (see the analysis of counterattack vs reinforcement above). Thus, even if we considered counter-counterattack explicitly as a Red option, it would rarely be an optimal choice.

Given this notation and these simplifying assumptions, we will now develop a number of key expressions which will enable us to describe net territorial gain for the counterattack as a function of R_{LI}. Rather than solving directly for R_{LI}, however, we will find it algebraically more compact to begin this process by relating R_{LI} and C_{CA}, the casualties suffered by the counterattacker in taking a single Red defensive line; we will then express the counterattacker's net territorial gain first as a function of C_{CA}, solve this expression for C_{CA}, and only then substitute for R_{LI}, and the corresponding density of Red flank defenders ρ_{FLK}.

To do this, let us first observe that, given our simplifying assumptions, we may rewrite (19) for the special case of counterattack as:

$$C_{CA} = \frac{\lambda_{CA}}{2} \frac{(\hat{R}_{LI} - \mu_{BS2})}{(2\rho_{MAX} - \mu_{RS2})} Q_{10} \tag{41}$$

where:

$$Q_{10} = k_3 \phi_{INF} V_{CA} + \frac{k_4}{\phi_{INF}} + \frac{k_1 \hat{R}_{ARTLI}}{(1+V_{CA})} + \frac{k_6}{\hat{B}_{ARTECH}(1+V_{CA})} + \mu_{RS2} \tag{42}$$

$$\mu_{BS2} = \delta_{1B} \hat{B}_{ARTECH} k_{ACMC} + \delta_{2B} k_{ACMC} \tag{43}$$

$$\mu_{RS2} = \delta_{1R} \hat{R}_{ARTLI} + \delta_{2R} \tag{44}$$

$$\hat{R}_{LI} = {}^{2R_{LI}}\!/\lambda_{CA} \tag{45}$$

and where $\hat{R}_{LI}$ denotes the scaled equivalent of $\hat{R}_{LI}$ corresponding to the two kilometer benchmark frontage used in the JANUS runs from which the constants were estimated, $2\rho_{MAX}$ gives the initial strength of a single Blue counterattack echelon (for the same two kilometer benchmark frontage), V_{CA} denotes the Blue assault velocity, $\hat{R}_{ARTLI}$ gives the scaled number of artillery tubes available to support a single Red defensive line, and

$\hat{B}_{ARTECH}$ gives the number of tubes supporting a single assault echelon.[113] Given our assumption that for counterattack, $\gamma = 1$, and given that the strength of a single Blue counterattack echelon is $(\rho_{MAX}\lambda_{CA})$, it follows that:

$$n_{ECHCA} = \frac{C_{CA}}{\rho_{MAX}\lambda_{CA}} \tag{46}$$

and:

$$\psi_{BCAGCA} = -\frac{C_{CA}}{D_{LI}} \tag{47}$$

where n_{ECHCA} is the number of Blue counterattack echelons required to take a single Red defensive line, and ψ_{BCAGCA} is the counterattacker's loss rate per kilometer of realized advance.[114] This in turn enables us to define the counterattacker's realized rate of advance in terms of C_{λ} (and thus, implicitly, in terms of R_{LI}). In particular, by analogy to (33):

$$\psi_{ROACA} = \frac{D_{LI}}{\dfrac{D_{LI}}{V_{CA}} + \dfrac{C_{CA}}{\rho_{MAX}\lambda_{ATK}}\left(t_{OPREP} + \dfrac{C_{CA}D_{ASY}}{2V_{RSV}\rho_{MAX}\lambda_{ATK}}\right)} \tag{48}$$

where t_{OPREP} is assumed to be the same for Blue counterattackers and Red theater attackers.[115]

Given this, we can define the Blue counterattacker's attrition rate in AFVEs per hour of counterattack, ψ_{BCAT}, as:

$$\psi_{BCAT} = -\frac{C_{CA}\,\psi_{ROACA}}{D_{LI}} \tag{49}$$

and thus the strength of the Blue counterattack force at any given time t_{CA}, subsequent to the initiation of the counterattack as:

---

[113] Thus, $\hat{R}_{ARTLI} = 2R_{ARTLI}/\lambda_{CA}$, and $\hat{B}_{ARTECH} = 2B_{ARTECH}/\lambda_{CA}$.

[114] Assuming that Blue counterattack echelons are replaced only when annihilated; i.e., that $B_{BPT} = 0$.

[115] Note that in our simplifying assumptions above we have argued that n_{BCHCA} will not ordinarily fall far below 1; it can, however, be variably greater than 1 depending upon circumstances.

$$B_{CA}\left(t_{CA}\right) = \mathrm{Max}\left[0, B_{CAST}(t) + t_{CA}\,\psi_{BCAT}\right] \qquad (50)$$

Note also that we can now use (41), by analogy from (29), to define h_{CA}, the loss exchange ratio (now Blue casualties/Red casualties) for the decisive engagement on Red's reinforced line, as:

$$h_{CA} = \frac{Q_{10}}{2\rho_{MAX}} \qquad (51)$$

The final key expression in R_{LI} is the strength of the final Red defensive line as a function of t_{CA}. Assuming that Red will move unengaged reserves to reinforce this final line once the Blue point of counterattack is known,[116] then we may write, by analogy to (4):

$$R_{OM}\left(t_{CA}\right) = \begin{cases} R_{LI} \text{ if } t_{CA} \leq t_{RST} + t_{RPREP} \\ R_{LI} + \left(t_{CA} - t_{RST} - t_{RPREP}\right)\,\psi_{ROM}(t_{CA}) \\ \qquad \text{if } t_{CA} > t_{RST} + t_{RPREP} \end{cases} \qquad (52)$$

We will now use these expressions to describe the net territorial gain attained by Blue's counterattack, G_{CA}, as a function of C_{CA}. In particular, let us define:

$$G_{CA}\left(t_{CA}\right) = t_{CA}\,\psi_{ROACA} \qquad (53)$$

We earlier described the dynamics of theater-level attack and defense in terms of a "culminating point" at which a weakening attacker is no longer able to advance against a reinforced local defense at the attacker's chosen point of attack; let us denote the time at which this culminating point is reached for the Blue counterattack as t^*_{CA}. In particular, we may define t^*_{CA} as the earliest time at which:[117]

---

[116] Note, however, that unlike (4), we will not specify an explicit upper bound here. In effect, we will use an explicit bound on the size of the Red reserve force itself (which will be defined to be only as large as necessary to prevent a Blue breakthrough) to prevent Red from moving more reserves to the final line than he has forces available in the theater to be moved.

[117] Equations (50), (51) and (52) together imply that where t^*_{CA} does not exist, Red's allocation of forces to flank defense is inadequate to halt the Blue counterattack prior to its breaking through. Inasmuch as we will define Red's offensive force availability as the excess of total forces over the requirement for adequate flank defense of the ground taken by Red in the offensive, and inasmuch as

$$h_{CA}\,R_{OM}\left(t^*_{CA}\right) = B_{CA}\left(t^*_{CA}\right) \tag{54}$$

which implies, given (50) and (52):[118]

$$t^*_{CA} = \frac{B_{CAST}(t) - h_{CA}\,R_{LI} + h_{CA}\left(t_{RST} + t_{RPREP}\right)\psi_{ROM}\left(t_{CA}\right)}{h_{CA}\,\psi_{ROM}\left(t_{CA}\right) - \psi_{BCAT}} \tag{55}$$

and thus, the net territorial gain for the Blue counterattack:

$$G_{CA}\left(t^*_{CA}\right) = \frac{\psi_{ROACA}\begin{bmatrix} B_{CAST}(t) - h_{CA}\,R_{LI} + \\ h_{CA}\left(t_{RST} + t_{PREP}\right)\psi_{ROM}\left(t_{CA}\right)\end{bmatrix}}{h_{CA}\,\psi_{ROM}\left(t_{CA}\right) - \psi_{BCAT}} \tag{56}$$

To rewrite this expression in C_{CA} and solve, note that :

$$\psi_{BCAT}/\psi_{ROACA} = \psi_{BCAGCA} \tag{57}$$

and that (41) implies that:

$$R_{LI} = C_{CA}\,\frac{(2\rho_{MAX} - \mu_{RS2})}{Q_{10}} + \frac{\lambda_{CA}\mu_{BS2}}{2} \tag{58}$$

Given this, rearranging terms, and substituting, using (56), (58), (47) and (57), and given that Red's optimal flank density obtains when $G_{CA}\,(t^*_{CA}) = D_{CA}$, we obtain:

$$\bar{a}\,C^2_{CA} + \bar{b}\,C_{CA} + \bar{c} = 0 \tag{59}$$

where:

$B_{CA}(t_{CA})$ as defined in equation (50) is always a positive number, any condition under which t^*_{CA} does not exist for some value of R_{LI} is therefore a condition under which Red net territorial gain will be zero -- which is an outcome consistent with the logic of permitting Red to take only as much ground as he can hold against Blue counterattack.

[118] And assuming for the sake of compactness that the resulting t^*_{CA} lies between the bounds specified in (52) above; the alternative development for t^*_{CA} outside these bounds, while tedious, is elementary.

$$\bar{a} = \frac{D_{ASY}\, D_{CA}\, h_{CA}\, \psi_{ROM}\left(t_{CA}\right)}{2\, \rho_{MAX}^{2}\, \lambda_{CA}^{2}\, V_{RSV}\, D_{LI}} \tag{60}$$

$$\bar{b} = \frac{t_{OPREP}\, D_{CA}\, h_{CA}\, \psi_{ROM}\left(t_{CA}\right)}{\rho_{MAX}\, \lambda_{CA}\, D_{LI}} + \frac{D_{CA}}{D_{LI}} + \frac{h_{CA}\left(2\, \rho_{MAX} - \mu_{RS2}\right)}{Q_{10}} \tag{61}$$

$$\bar{c} = \frac{D_{CA}\, h_{CA}\, \psi_{ROM}\left(t_{CA}\right)}{V_{CA}} + \frac{h_{CA}\, \lambda_{CA}\, \mu_{BS2}}{2} - B_{CAST}(t) - h_{CA}\, \psi_{ROM}\,(t_{CA})\,(t_{RST} + t_{RPREP}) \tag{62}$$

a readily solvable quadratic in C_{CA}, which enables us to solve for R_{LI} by substitution using (58). Given that, we obtain the Red flank density:

$$\rho_{FLK} = \frac{R_{LI}\, n_{LICA}}{\lambda_{CA}} \tag{63}$$

5. Red Net Territorial Gain

It remains to interrelate the terms developed above to provide an expression for Red net territorial gain. In particular, we seek here to describe the conditions under which Red will reach a culminating point, to distinguish these from conditions that produce breakthrough, and for circumstances where culmination obtains, to specify Red's net territorial gain at that culminating point.

First, however, we must define a few final quantities relating to Red force availability for offensive use. In addition to the flank defense "overhead cost" described in the previous section, Red must also deploy forces along the international border away from Red's intended point of attack, and Red must also withhold unengaged reserves at least sufficient to meet the demand for reinforcement of in-place flank defenses in the event of Blue counterattack. Forces allocated to such duties cannot be used in the Red offensive spearhead. As for the latter, the number of Red AFVEs required for reinforcement of flank defenses is given by:

$$R_{RSV} = \begin{cases} 0 \text{ if } t^{*}_{CA} < t_{RST} + t_{RPREP} \\ 2\, \psi_{ROM}\,(t_{CA})\,(t^{*}_{CA} - t_{RST} - t_{RPREP}) \text{ otherwise} \end{cases} \tag{64}$$

Red's required force density (in AFVEs per kilometer) for defense of the border away from the intended point of attack, ρ_{MIN}, is determined by two related requirements: the need to pin Blue forward forces away from the main attack sector to prevent their easy disengagement and displacement to the point of attack, and the need to defend against a possible cross-border counterstroke by uncommitted Blue reserves. Unlike Red's flank defense requirements, we will not attempt to model these quantities in detail here; rather, we will simply assume that Red must set aside a certain, constant number of AFVEs to pin each opposing forward AFVE, a certain, constant number of AFVEs to defend the international border against attack by each Blue counterattacker, and that these requirements are additive in nature. Thus:

$$\rho_{MIN} = \frac{k_{PIN} B_{LI} n_{LI}}{\lambda_{ATK}} + \frac{k_{DEF} B_{RSV} \phi_{CA}}{\lambda_{THR}} \qquad (65)$$

where k_{PIN} and k_{DEF} are constants.[119]

Given these definitions, we may now specify the size of the offensive forces with which Red opens the operation, R_{OFVST}:

$$R_{OFVST} = R - \rho_{MIN}(\lambda_{THR} - \lambda_{ATK}) - R_{RSV} \qquad (66)$$

This initial quantity of available offensive forces diminishes over time, however, as a combined result of casualties in taking Blue defensive lines, losses to Blue tacair and ACM, and as a result of ongoing diversions of otherwise-offensive forces into flank defense, thus:

$$\psi_{ROFVT} = -\psi_{ROA}\left[\frac{n_{ECH}\lambda_{ATK}\left(\rho_{MAX} - R_{BPT}\right)}{D_{LI}} + \frac{\delta_3}{\psi_{ROA}} + 2\rho_{FLK}\right] \qquad (67)$$

where ψ_{ROFVT} is the net rate at which Red's offensive spearhead changes size (in AFVEs per hour), ψ_{ROA} is the Red rate of advance (in kilometers per hour), and δ_3 denotes Red

---

[119] These expressions are given here in terms of two constants; a more satisfactory approach would be to compute Red's off-axis requirements by the same logic used to determine combat outcomes generally, assuming putative attackers in the form of Blue counterstroke forces, and other circumstances as a function of Red's need to prevent a breakthrough off-axis while prosecuting his own offensive at the chosen point of attack. This approach is, of course, utilized to compute Red's requirements for defense against Blue counterattacks directed against the Red penetration itself; to extend this logic to Red's defense of the international border away from the penetration corridor would increase computer run-times substantially, and has not been implemented here.

AFVE losses to Blue long range tacair and ACM (per hour).[120] Thus, Red's available offensive strength at any given time, t, is:

$$R_{OFV}(t) = \max\left[0, R_{OFVST} + t\,\psi_{ROFVT}\right] \qquad (68)$$

Given these definitions, we may now describe Red's net territorial gain at any given time t as:

$$G(t) = t\,\psi_{ROA} \qquad (69)$$

As we have discussed above, this ground gain may produce either breakthrough or culmination short of breakthrough; let us begin by considering the case in which culmination obtains. In particular, let us define a time t* at which Red reaches its culminating point; that is, the earliest time at which:[121]

$$h\,B_{OM}(t^*) = R_{OFV}(t^*) \qquad (70)$$

Substituting for $B_{OM}(t^*)$ and $R_{OFV}(t^*)$ using (4) and (68), and rearranging terms, we obtain:[122]

$$t^* = \frac{\left[R_{OFVST} - h\,B_{LI} + h\,\psi_{BOMT}(t)\left(t_{BPREP} + t_{BST}\right)\right]}{h\,\psi_{BOMT}(t) - \psi_{ROFVT}} \qquad (71)$$

and thus, Red net territorial gain for conditions under which the Red offensive reaches culmination prior to breakthrough is given by:

---

[120] Red AFVE losses to Blue CAS are included in the determination of n_{ECH}, the number of Red echelons required to take a Blue defensive line; see (19) above.

[121] Where no such time exists—that is, where the strength of Red's offensive spearhead (a declining function of time) never comes into equilibrium with the strength of Blue's defenses on the omega line (an increasing function of time)—or where this equilibrium is reached only after Red has advanced beyond the depth of Blue's prepared defense, the result is by definition breakthrough, rather than culmination (breakthrough being the consequence of the absence of a culminating point; see the discussion under "theater dynamics" above). The case of Red breakthrough is discussed in greater detail below.

[122] For $t^* < \frac{\lambda_{THR}}{V_{RSV}} - t_{BPREP} + t_{BST}$; development for the alternative case follows readily from (4) by substitution.

$$G(t^*) = \frac{\psi_{ROA}\left[R_{OFVST} - h\,B_{LI} + h\,\psi_{BOMT}(t)\left(t_{BPREP} + t_{BST}\right)\right]}{h\,\psi_{BOMT}(t) - \psi_{ROFVT}} \qquad (72)$$

As for the alternative outcome—in which the defender fails to impose culmination and a breakthrough occurs—the conditions defining this case can be described using the culminating point expression given in equation 72. Recall that a breakthrough, as described earlier, is in effect an attacker penetration that exceeds the depth of the defended zone prior to the defender bringing sufficient reserves to the point of attack to halt the attacker's advance. This outcome will obtain if and only if:

$$G(t^*) \geq D \qquad (73)$$

where G is as given in equation 72, and D is defined as the total depth of the defended zone in kilometers, as given in equation 16. As a point of departure for analysis, we will assume here that if this condition holds and Red thus breaks through, Blue is effectively defeated; in effect, we will assume an arbitrarily high net territorial gain for any offensive that breaks through the Blue defense.[123]

6. Limitations

Equation 73 completes the development of the relationship between net territorial gain and the given independent variables. These equations have important limitations, however, and must be used with discrimination. As we noted in the introduction to this appendix, for example, the class of phenomena addressed by these equations is limited to high intensity land warfare between sophisticated opponents. As a result, these equations cannot be used to determine counter-infiltration or administrative/logistical requirements that might limit a combatant's ability to reduce forces to very low levels in a theater or war. Likewise, while these equations may be of some utility with respect to high intensity conflicts outside the European theater (e.g., the Mideast or southwest Asia), they

---

[123] This is of course not necessarily the case: attackers could be too near exhaustion at the point of breakthrough to exploit their victory, arriving defender reserves could succeed in re-establishing a defensive perimeter in front of the attacker's exploitation force, or arriving reserves could counterattack an unconstrained attacker with sufficient success to contain the exploitation. None of these are impossible, though some may be substantially unlikely. It seems appropriate, however, to begin with a simple assumption that attacker breakthrough defeats the defender in modern warfare. This assumption could be relaxed to consider some or all of the alternatives noted here, but this has not been attempted in this appendix.

are not appropriate to consideration of low intensity conflict, and their applicability to combat between low-sophistication opponents is unclear.

The equations above employ a number of simplifying assumptions to streamline the analysis—while many of these could be relaxed in further work, they must be taken into account for appropriate use of the existing formulation. These include the inability of forward defensive forces to displace laterally for counterconcentration; the summary treatment of Red's off-axis force requirement to defend against Blue cross-border invasion; the restrictions on Red's range of choice with respect to attack frontages and the employment of forces in flank defense; and the absence of a requirement for residual Red forces to exploit fully an accomplished breakthrough or a detailed treatment of operations subsequent to such a breakthrough.

There were also a number of potentially significant issues which, due to time and resource constraints, received limited attention here. The effects of variations in natural terrain or weapon quality (as opposed to weapon class); the highly aggregate treatment of tacair and ACM; constraints (other than en route air or ACM attack) on the ability of defenders to withdraw successfully under fire; and the absence of an explicit consideration of weapon classes beyond armor, infantry and artillery (e.g., attack helicopters or air defense systems), or of logistical or command systems per se, are all issues which would benefit from more detailed consideration than was possible here. The extent of formal proof for deduced properties is quite limited in the discussion above; a more thoroughgoing mathematical argument would be a valuable addition. The effects of intangibles such as morale or training are not considered here, nor are the potential roles of organizational or social variables. We also do not consider here the possibility of variations in the fundamental war aims of the two sides. That is, it is assumed throughout that the underlying objective of the theater invader is to seize and assert political control over the territory of the invaded (and/or to annihilate the opposing armed forces as an essential means to this end). Yet it is possible that at very low force levels—where seizure and control of territory (or annihilation of opposing forces) can become very difficult—that an aggressor would instead choose what Archer Jones has termed a "raiding strategy," where opposing forces are avoided and where military power is used for coercive purposes, to threaten with destruction economic and political assets which are vulnerable, but which an aggressor could not hold against counterattack.[124] An unconventional aim of this sort, while not without historical precedent, and while potentially worthy of further analysis, nevertheless lies beyond the scope of the theory developed here. Similarly, protracted guerilla-style infiltration or terror campaigns have not been considered here.

Finally, it should be noted that less attention was devoted to the equations' behavior in some regions of the potential independent variable space than in others. In particular, given the focus of the study on the implications of low force levels, the consequences of very high force levels received limited consideration. While we might expect, for example, diminishing marginal return effects with respect to various phenomena addressed in the equations (e.g., reserve arrival rates as a function of theater force size; or planning and command response times as a function of the size of committed forces), these effects were not given explicit attention. Where the implications of large force size were clearly crucial—as in the case of the attacker's local AFVE concentration—phenomena that are probably best represented by diminishing marginal return relationships are instead approximated by imposed ceilings.

Similarly, the nature of the functional forms which emerged from the JANUS testing display particular sensitivity for certain values of certain independent variables. In particular, for very low levels of attacker artillery R_{ARTECH}, equation 19 will tend to predict unreliably high attacker casualties. To a lesser degree, very low values for ϕ_{INF}, the combined attacker and defender infantry fractions, will tend to inflate casualties. An exponential form for either variable would eliminate this over-sensitivity, but the tests conducted did not provide sufficient variance in these parameters to estimate an exponential form with sufficient confidence. We have thus retained the form given in equation 19 as a point of departure (for a more detailed discussion see appendix D); caution should be exercised, however, in interpreting results where independent variable values diverge significantly from the range of values considered in the experimental data from which this form was fit.[125] Finally, the particular form chosen for the denominator

---

[124] Archer Jones, The Art of War in the Western World (Chicago: University of Illinois Press, 1987), pp. 666-7.

[125] Where, as noted above, that range extended from from 0.1 to 0.9 for the fraction of the attacker's AFVEs that were infantry and from 0.1 to 0.9 for the fraction of the defender's AFVEs that were infantry; hence, the data represents values of ϕ_{INF} between 0.2 and 1.8. For Red artillery, the experimental data included inventories ranging from six to 52 tubes per two kilometers. Ranges for other variables included velocities (V) ranging from one to eight kilometers per hour, local force to

in the red artillery term of the casualty expression (equation 19) can create difficulties at very high theater force-to-force ratios (or very high red artillery inventories). In particular, for theater force-to-force ratios in excess of about 2.0, the (1+V) formulation creates a casualty-reduction incentive for red to choose very low velocities that can be so strong as to overwhelm the disadvantages of allowing blue to counterconcentrate fully in the meantime—and strong enough to cause a disproportionate increase in net territorial gain for force-to-force ratios at which this effect is possible. If instead, constants are re-estimated using the same experimental data but according to a functional form in which (1+V) is replaced with (.1+V), the resulting casualty equation is nearly as strong a fit to the experimental data, but the incentive for very low red velocity choices is greatly reduced and as a consequence, net territorial gain for force-to-force ratios above 2.0 is greatly reduced. Further experimental work—and in particular, the accumulation of additional data for very low assault velocities—will be required to resolve this instability;[126] in the meantime, however, caution is warranted in interpreting results for force-to-force ratios above 2.0.

This completes our formal description of net territorial gain G, as a function of force employment (i.e., the attacker's choice of V, and the defender's choices of ϕ_{FWD}, ϕ_{CA}, n_{LI} and w), weapon mix (B_{ART}, R_{ART}, ϕ_{INF}, δ_1, δ_2, δ_3, δ_4,), terrain (ß), and of course the force-to-space and force-to-force ratios (B, R, and λ_{THR}). For convenience, variable definitions have been collected in Table C-1. Computations associated with these equations have been automated in the form of a FORTRAN code described in appendix E; the validity of these equations in experimental testing is addressed in appendix D.

---

force ratios $\hat{B}_{LI}/\hat{R}_{ECH}$ ranging from 7.5:1 to 1:1.5, and Blue artillery inventories ranging from zero to 104 tubes per two kilometers.

[126] The lower the assault velocity, the greater the computer time required to complete a single JANUS experiment. As a result, resource constraints have made it impossible to complete a sufficiently large number of very low velocity runs to distinguish conclusively between the fits obtained for the respective functional forms. Additional experimental work is ongoing, but incomplete at the present time.

Table C-1. Variable Definitions[127]

B	Blue maneuver forces in theater (AFVEs), B: $\mathbf{R} \rightarrow [0,\infty]$ (#)
B_{ART}	Blue artillery in theater (tubes), B_{ART}: $\mathbf{R} \rightarrow [0,\infty)$ (#)
B_{ARTECH}	Blue artillery supporting a single counterattack echelon (tubes), B_{ARTECH}: $\mathbf{R} \rightarrow (0,\infty)$
$\hat{B}_{ARTECH}$	Blue artillery supporting a single counterattack echelon (tubes, scaled to two-kilometer benchmark frontage), $\hat{B}_{ARTECH}$: $\mathbf{R} \rightarrow (0,\infty)$
B_{ARTLI}	Blue artillery supporting a single defensive line (tubes), B_{ARTLI}: $\mathbf{R} \rightarrow [0,\infty)$
$\hat{B}_{ARTLI}$	Blue artillery supporting a single defensive line (tubes, scaled to two-kilometer benchmark frontage), $\hat{B}_{ARTLI}$: $\mathbf{R} \rightarrow [0,\infty)$
$B_{CA}(t_{CA})$	Blue maneuver forces assigned to counterattack surviving at time t_{ca} (AFVEs), $B_{CA}(t_{CA})$: $\mathbf{R} \rightarrow [0,\infty)$
$B_{CAST}(t)$	Blue maneuver forces available for counterattack at time counterattack begins (AFVEs), $B_{CAST}(t)$: $\mathbf{R} \rightarrow [0,\infty)$
B_{CAECH}	Blue single assault echelon maneuver force initial strength (AFVEs), B_{CAECH}: $\mathbf{R} \rightarrow [0,\infty)$
$\hat{B}_{CAECH}$	Blue single assault echelon maneuver force initial strength (AFVEs, scaled to two-kilometer benchmark frontage), $\hat{B}_{CAECH}$: $\mathbf{R} \rightarrow [0,\infty)$
B_{FWD}	Blue maneuver forces allocated to forward positions (AFVEs), B_{FWD}: $\mathbf{R} \rightarrow [0,\infty)$
B_{LI}	Blue maneuver forces defending a single defensive line, (AFVEs), B_{LI}: $\mathbf{R} \rightarrow (0,\infty)$
$\hat{B}_{LI}$	Blue maneuver forces defending a single defensive line (AFVEs, scaled to two-kilometer benchmark frontage), $\hat{B}_{LI}$: $\mathbf{R} \rightarrow (0,\infty)$

---

[127] Exogenous independent variables are denoted by (#); endogenous independent variables by (##); constants by ($); the dependent variable is G. All other variables given are endogenous instrumental variables.

Table C-1 (continued)

B_{RSV}	Blue maneuver forces allocated to theater reserve (AFVEs), B_{RSV}: $\mathbf{R}\rightarrow[0,\infty)$
$B_{OM}(t)$	Blue maneuver forces defending final defensive line at time t (AFVEs), $B_{OM}(t)$: $\mathbf{R}\rightarrow[0,\infty)$
C	Red casualties required to take a single blue defensive line (AFVEs), C: $\mathbf{R}\rightarrow(0,\infty)$
C_{CA}	Blue casualties required to take a single red flank defensive line (AFVEs), C_{CA}: $\mathbf{R}\rightarrow(0,\infty)$
D	Overall depth of the blue defended zone (kilometers), D: $\mathbf{R}\rightarrow(0,\infty)$
D_{ASY}	Depth of an assembly area (kilometers), D_{ASY}: $\mathbf{R}\rightarrow(0,\infty)$ (\$)
D_{CA}	Depth of red's flank defense against blue counterattack (kilometers), D_{CA}: $\mathbf{R}\rightarrow(0,\infty)$
D_{LI}	Depth of a single defensive line (kilometers), D_{LI}: $\mathbf{R}\rightarrow(0,\infty)$ (\$)
G(t)	Red net territorial gain at time t (kilometers), G: $\mathbf{R}\rightarrow[0,\infty)$
$G_{CA}(t_{CA})$	Ground gained by blue counterattack at time t_{CA} (kilometers), $G_{CA}(t_{CA})$: $\mathbf{R}\rightarrow[0,\infty)$
h	Loss exchange ratio in combat on final blue defensive line (dimensionless: [red AFVEs lost/blue AFVEs lost]), h: $\mathbf{R}\rightarrow[0,\infty)$
h_{CA}	Loss exchange ratio in counteroffensive combat on final red defensive line (dimensionless: [blue AFVEs lost/red AFVEs lost]), h_{CA}: $\mathbf{R}\rightarrow[0,\infty)$
k_0	Constant, minimum attack frontage (kilometers) (\$)
k_1	Constant, related to C (dimensionless) (\$)
k_{PIN}	Constant, red AFVEs required to pin one forward blue AFVE away from the point of attack (dimensionless) (\$)
k_{DEF}	Constant, red AFVEs required to defend against one reserve blue AFVE away from the point of attack (dimensionless) (\$)
k_3	Constant, related to C (dimensionless) (\$)

Table C-1 (continued)

k_4	Constant, related to C (dimensionless) (\$)
k_5	Constant, related to α (dimensionless) (\$)
k_6	Constant, related to C (dimensionless) (\$)
k_8	Constant, (fractional increase in attacker casualties due to entropy effect of depth per kilometer advanced through defended territory) (\$)
k_{ACMC}	Constant, ratio of ACM and CAS lethality vs stationary targets to ACM and CAS lethality vs moving targets (dimensionless) (\$)
k_{AE}	Constant, number of offensive echelons whose organic artillery is available to support a single assault by the lead echelon (dimensionless) (\$)
k_{λ}	Constant, increase in attack frontage per offensive AFVE (kilometers/AFVE) (\$)
n_{ASLT}	Number of successive assaults completed successfully by a single offensive echelon prior to reference assault (dimensionless), n_{ASLT}: $\mathbf{R} \to (0,\infty)$
N_{ASLT}	Total number of successive assaults completed successfully by a single offensive echelon (dimensionless), N_{ASLT}: $\mathbf{R} \to (0,\infty)$
n_{ECH}	Number of offensive echelons required to take a single defensive line (dimensionless), n_{ECH}: $\mathbf{R} \to (0,\infty)$
n_{ECHCA}	Constant, estimated number of blue echelons required to take a single red defensive line; related to t_{JO} determination (dimensionless), n_{ECHCA}: $\mathbf{R} \to (0,\infty)$ (\$)
n_{LI}	Number of predeployed blue defensive lines (dimensionless), n_{LI}: $\mathbf{R} \to [1,\infty)$ (##)
n_{LICA}	Number of predeployed red flank defense lines (dimensionless), n_{LICA}: $\mathbf{R} \to [1,\infty)$
n_{LITOT}	Total blue defensive lines, predeployed and subsequently occupied (dimensionless), n_{LITOT}: $\mathbf{R} \to [1,\infty)$
Q_1	Instrumental quantity, related to red casualties per average assault

Table C-1 (continued)

Q_2	Instrumental quantity, related to red casualties per average assault
Q_{10}	Instrumental quantity, related to C_{CA}
R	Red maneuver forces in theater (AFVEs), R: $\mathbf{R}\to(0,\infty)$ (#)
R_{ART}	Red artillery in theater (tubes), R_{ART}: $\mathbf{R}\to[0, \infty)$ (#)
R_{ARTECH}	Red artillery supporting a single assault echelon (AFVEs), R_{ARTECH}: $\mathbf{R}\to[0,\infty)$
$\hat{R}_{ARTECH}$	Red artillery supporting a single assault echelon (tubes, scaled to two-kilometer benchmark frontage), $\hat{R}_{ARTECH}$: $\mathbf{R}\to(0,\infty)$
R_{ARTLI}	Red artillery supporting a single defensive line (tubes), R_{ARTLI}: $\mathbf{R}\to[0,\infty)$
$\hat{R}_{ARTLI}$	Red artillery supporting a single defensive line (tubes, scaled to two-kilometer benchmark frontage), $\hat{R}_{ARTLI}$: $\mathbf{R}\to[0,\infty)$
R_{BPT}	Red residual maneuver strength at which a single assault echelon will break off an attack (AFVEs), R_{BPT}: $\mathbf{R}\to[0, \rho_{MAX}]$ (#)
R_{ECH}	Red single assault echelon maneuver force initial strength (AFVEs), R_{ECH}: $\mathbf{R}\to[0,\infty)$
$\hat{R}_{ECH}$	Red single assault echelon maneuver force strength (AFVEs, scaled to two-kilometer benchmark frontage), $\hat{R}_{ECH}$: $\mathbf{R}\to(0,\infty)$
R_{LI}	Red maneuver forces defending a single defensive line (AFVEs), R_{LI}: $\mathbf{R}\to[0,\infty)$
$\hat{R}_{LI}$	Red maneuver forces defending a single defensive line (AFVEs, scaled to two-kilometer benchmark frontage), $\hat{R}_{LI}$: $\mathbf{R}\to[0,\infty)$
R_{RSV}	Red contingency reserve for augmenting flank defenses (AFVEs), R_{RSV}: $\mathbf{R}\to[0,\infty)$
$R_{OFV}(t)$	Red maneuver forces assigned to offensive use surviving at time t (AFVEs), $R_{OFV}(t)$: $\mathbf{R}\to[0,\infty)$

Table C-1 (continued)

R_{OFVST}	Red maneuver forces available for offensive use at time theater offensive begins (AFVEs), $R_{OFVST}: \mathbf{R} \rightarrow (0, \infty)$
$R_{OM}(t)$	Red maneuver forces defending final defensive line at time t (AFVEs), $R_{OM}(t): \mathbf{R} \rightarrow [0, \infty)$
t	Time (hours, measured from initiation of theater offensive), $t: \mathbf{R} \rightarrow [0, \infty)$
t^*	Time red culminating point is reached (hours, measured from initiation of theater offensive), $t^*: \mathbf{R} \rightarrow (0, \infty)$
t^*_{CA}	Time blue counterattack reaches culminating point (hours, measured from blue counterattack jump-off), $t^*_{CA}: \mathbf{R} \rightarrow (0, \infty)$
t_{BPREP}	Time required to prepare blue reinforcement positions for combat (hours, measured from blue reinforcement arrival), $t_{BPREP}: \mathbf{R} \rightarrow [0, \infty)$ (#)
t_{BRKB}	Estimated time that blue will break through red flank rear defense line if not halted (hours, measured from blue counterattack jump-off), $t_{BRKB}: \mathbf{R} \rightarrow [0, \infty)$
t_{BRKR}	Estimated time that red will break through blue theater rear defense line if not halted (hours, measured from initiation of theater offensive), $t_{BRKR}: \mathbf{R} \rightarrow [0, \infty)$
t_{BST}	Time that blue begins to move reserves toward point of attack (hours, measured from initiation of theater offensive), $t_{BST}: \mathbf{R} \rightarrow (-\infty, \infty)$ (#)
t_{CA}	Time (hours, measured from blue counterattack jump-off), $t_{CA}: \mathbf{R} \rightarrow [0, \infty)$
t_{JO}	Time blue counterattack jumps off (hours, measured from initiation of theater offensive), $t_{JO}: \mathbf{R} \rightarrow [0, \infty)$
t_{MV}	Time required for uncommitted assault echelon to complete approach march (hours, measured from completion of preparation), $t_{MV}: \mathbf{R} \rightarrow [0, \infty)$
t_{OPREP}	Time required to prepare uncommitted assault echelon to begin approach march (hours, measured from termination of preceeding echelon's assault), $t_{OPREP}: \mathbf{R} \rightarrow (-\infty, \infty)$ (#)
t_{RPREP}	Time required to prepare red flank reinforcement positions for combat (hours, measured from red reinforcement arrival), $t_{RPREP}: \mathbf{R} \rightarrow [0, \infty)$ (#)

Table C-1 (continued)

t_{RST}	Time that red begins to move reserves toward point of counterattack (hours, measured from initiation of blue counterattack), t_{RST}: $\mathbf{R}\rightarrow(-\infty,\infty)$ (#)
V	Assault velocity (kilometers per hour), V: $\mathbf{R}\rightarrow(0,\infty)$ (##)
V_{CA}	Counterattack assault velocity (kilometers per hour), V_{CA}: $\mathbf{R}\rightarrow(0,\infty)$ (#)
V_{RSV}	Reserve road march velocity (kilometers per hour), V_{RSV}: $\mathbf{R}\rightarrow(0,\infty)$ (#)
w	Fraction of maneuver forces defending a given line to be withdrawn (dimensionless), w: $\mathbf{R}\rightarrow[0,1]$ (##)
w_{SURV}	Fraction of maneuver forces defending a given line that survive withdrawal (dimensionless) w_{SURV}: $\mathbf{R}\rightarrow[0,1)$
$\bar{a}$	Instrumental quantity, related to C_{CA}
$\bar{b}$	Instrumental quantity, related to C_{CA}
$\bar{c}$	Instrumental quantity, related to C_{CA}
α	Scalar multiple representing the decrease in attacker casualties in a given assault as a result of early termination of defensive fire upon withdrawal, relative to a fight to the finish under otherwise identical circumstances (dimensionless), α: $\mathbf{R}\rightarrow(0, 1]$
β	Scalar multiple representing the increased slope of the attacker's casualty-velocity tradeoff frontier as a result of the availability to the defender of additional barrier preparation labor not organic to the defending maneuver units themselves (dimensionless), β: $\mathbf{R}\rightarrow[1, \infty)$
γ	Scalar multiple representing the increase in attacker casualties in a given assault as a result of entropy induced by the lead echelon's advance through defended depth prior to the assault in question, relative to an attack conducted with perfect coherence under otherwise identical circumstances (dimensionless), γ: $\mathbf{R}\rightarrow[1,\infty)$
δ_{1B}	Blue short range ACM contribution (red AFVE kills per blue artillery tube per assault per two kilometers), δ_{1B}: $R\rightarrow\left[0, \left(\hat{\mathbf{R}}_{ECH}-\delta_{2B}\right)\right]$ (#)

Table C-1 (continued)

δ_{2B}	Blue CAS contribution (red AFVE kills per assault per two kilometers), δ_{2B}: $R \rightarrow \left[0, \left(\hat{R}_{ECH} - \delta_{1B}\right)\right]$ (#)
δ_{1R}	Red short range ACM contribution (blue AFVE kills per red artillery tube per assault per two kilometers), δ_{2R}: $R \rightarrow \left[0, \left(\hat{B}_{LI} - \delta_{2R}\right)\right]$ (#)
δ_{2R}	Red CAS contribution (blue AFVE kills per assault per two kilometers), δ_{2R}: $R \rightarrow \left[0, \left(\hat{B}_{LI} - \delta_{1R}\right)\right]$ (#)
δ_3	Blue BAI-long range ACM contribution (red AFVE kills per hour), δ_3: $R \rightarrow [0, R_{OFVST}]$ (#)
δ_4	Red BAI-long range ACM contribution (blue AFVE kills per hour), δ_4: $R \rightarrow [0, B_{RSV}]$ (#)
ϕ_{CA}	Fraction of blue reserve AFVEs allocated to counterattack (dimensionless), ϕ_{CA}: $R \rightarrow [0,1]$ (##)
ϕ_{FWD}	Fraction of blue AFVEs deployed forward (dimensionless), ϕ_{FWD}: $R \rightarrow [0,1]$ (##)
ϕ_{INF}	Sum of fraction of red and fraction of blue maneuver AFVEs that are infantry (dimensionless), ϕ_{INF}: $R \rightarrow [0,2]$ (#)
μ_{BS}	Contribution of blue CAS and short range ACM on defense (red AFVE kills per red assault per two kilometers), μ_{BS}: $R \rightarrow \left[0, \hat{R}_{ECH}\right]$
μ_{BS2}	Contribution of blue CAS and short range ACM on offense (red AFVE kills per blue assault per two kilometers), μ_{BS2}: $R \rightarrow \left[0, \hat{R}_{LI}\right]$
μ_{RS}	Contribution of red CAS and short range ACM on offense (blue AFVE kills per red assault per two kilometers), μ_{RS}: $R \rightarrow \left[0, \hat{B}_{LI}\right]$

Table C-1 (continued)

μ_{RS2}	Contribution of red CAS and short range ACM on defense (blue AFVE kills per blue assault per two kilometers), μ_{RS2}: $R \rightarrow \left[0, \hat{B}_{ECH} \right]$
λ_{ATK}	Length of red theater attack frontage (kilometers), λ_{ATK}: $R \rightarrow \left[0, \lambda_{THR} \right)$(##)
λ_{CA}	Length of blue counterattack frontage (kilometers), λ_{CA}: $R \rightarrow (0,\infty)$ (##)
λ_{LOC}	Length of red frontage required as clear channel to resupply and reinforce assault elements (kilometers), λ_{LOC}: $R \rightarrow (0,\lambda)$
λ_{THR}	Length of theater (kilometers), λ_{THR}: $R \rightarrow (0,\infty)$ (#)
ρ_{FLK}	Density of red flank defense (AFVEs per kilometer), ρ_{FLK}: $R \rightarrow [0, \infty)$
ρ_{MAX}	Maximum maneuver force density for single assault echelon at point of attack (AFVEs per kilometer), ρ_{MAX}: $R \rightarrow (0,\infty)$ (#)
ρ_{MIN}	Minimum maneuver force density required by red away from point of attack (AFVEs per kilometer), ρ_{MIN}: $R \rightarrow [0,\infty)$
ψ_{ROA}	Rate of Red theater advance (km per hour), ψ_{ROA}: $R \rightarrow (0,\infty)$
ψ_{ROFVT}	Rate of change in available Red attack forces (AFVEs per hour), ψ_{ROFVT}: $R \rightarrow (-\infty, 0]$
ψ_{ROFVG}	Rate of change in available Red attack forces (AFVEs per kilometer of penetration), ψ_{ROFVG}: $R \rightarrow (-\infty, 0]$
$\psi_{BOMT}(t)$	Rate of arrival of Blue reserves on final line (AFVEs per hour), $\psi_{BOMT}(t)$: $R \rightarrow (0,\infty)$
ψ_{ROACA}	Rate of Blue counterattack advance (km per hour), ψ_{ROACA}: $R \rightarrow (0,\infty)$
ψ_{BCAT}	Rate of change in Blue counterattack force due to losses (AFVEs per hour), ψ_{BCAT}: $R \rightarrow (-\infty,0]$
$\psi_{BCAST}(t)$	Rate of buildup of Blue reserves for counterattack at time t (AFVEs per hour), $\psi_{BCAST}(t)$: $R \rightarrow (0,\infty)$
ψ_{BCAGCA}	Rate of change in Blue counterattack force due to losses (AFVEs per kilometer of penetration), ψ_{BCAGCA}: $R \rightarrow (-\infty, 0]$

Table C-1 (continued)

$\Psi_{ROM}(t_{CA})$	Rate of arrival of Red reserves on final line at time t_{CA} (AFVEs per hour), $\Psi_{ROM}(t_{CA})$: $R \rightarrow [0,\infty)$

Appendix D

TESTS FOR VALIDITY

Stephen D. Biddle and David G. Gray

A. INTRODUCTION

Appendix C describes a formal theory of force to space ratios, and motivates that theory in terms of inherent, or a priori plausibility. A priori plausibility alone, however, is a weak basis for confidence in a theory or its policy implications. To merit greater confidence, a theory must survive attempted falsification through systematic comparison with real world outcomes. This appendix describes the testing methodology, experimental design and procedures, and results obtained in the falsification attempts undertaken in this study. The equations described in Appendix C are those that survived this process of test; potential alternative formulations that failed under testing are described. Finally, constant parameter values for the equations described in Appendix C are estimated from experimental results.

B. TEST METHODOLOGY

Three broad alternatives were considered as test methodologies. *Ex post facto* techniques use past events as evidence; "large n" or *statistical ex post* methods control for extraneous effects by exploiting partial correlation within a large database of many past events, while "small n" or *comparative ex post* methods obtain control by careful selection of a small number of events for in-depth study. *Ex ante* techniques create new events for evidence by conducting experiments in which control is obtained by deliberate manipulation of the circumstances of the observed event.[1]

For our purposes, none of these methods is wholly sufficient. Statistical ex post techniques require a suitable data base. Existing data on historical combat results, however, are of uneven quality and lack coverage of variables crucial to the phenomenon under study here.[2]

---

[1] For comparisons of these approaches, see Arend Lijphart, "Comparative Politics and the Comparative Method," American Political Science Review, September 1971, pp.682-693; Harry Eckstein, "Case Study and Theory in Political Science" in Fred I. Greenstein and Nelson W. Polsby, (eds.), Strategies of Inquiry, Vol.7 of The Handbook of Political Science (Menlo Park, CA: Addison-Wesley, 1975), pp.79-137; Richard A. Brody and Charles N. Brownstein, "Experimentation and Simulation," in Greenstein and Polsby, op. cit., pp.211-264; and Alexander L. George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul G. Lauren, (ed.), Diplomacy: New Approaches in History, Theory, and Policy (New York: Free Press, 1979), pp.43-68.

[2] Probably the best currently available database on historical combat results is that compiled by the Historical Evaluation and Research Organization (HERO), and described in Robert L. Helmbold and Aqeel A. Khan, Combat History Analysis Study Effort (CHASE): Progress Report, (Bethesda, MD: U.S. Army Concepts Analysis Agency, August, 1986), CAA-TP-86-2. The quality of these data is

Comparative ex post techniques offer the potential for superior data quality and more appropriate variable coverage. Comparative methods, however, are labor intensive, and thus are most practical where the number of cases can be kept very small. This is a problematic requirement for a multivariate theory such as that proposed in Appendix C. Under the right circumstances, a small number of cases can conclusively disprove even a multivariate theory, but rarely can a multivariate theory be lent substantial credibility by a small number of cases alone.[3]

Ex ante techniques offer superior flexibility and span of control and enable a multivariate theory to be more fully examined. For our purposes, however, a true ex ante experiment would require actual combat and thus is clearly impractical. As a consequence, experiments must be conducted via simulation, rather than in the real world itself; thus ex ante techniques involve some loss of verisimilitude relative to ex post facto observation of real combat results.

Any given technique poses a particular set of tradeoffs. Ideally, a combination would offer the greatest potential confidence in the validity of the outcome; time and resource constraints, however, make simultaneous pursuit of multiple methods impossible

---

disputed; see T.N. Dupuy, et. al., Analysis of Factors that have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements (Bethesda, MD: U.S. Army Concepts Analysis Agency, September 1984), CAA-SR-84-6, Vol.1, Main Report, for reviews of data samples by referees from the the U.S. Army Military History Institute; the U.S. Army Center of Military History; the Department of Military History, U.S. Military Academy; and the U.S. Army Combat Studies Institute. A similar data collection effort conducted by the RMC/Vertex Corporation and sponsored by the SHAPE Technical Center produced statistical results consistent with the presence of significant data errors. For a description of the data collection effort and statistical results, see Rex Goad, "Predictive Equations for Opposed Movement and Casualty Rates for Land Forces," and James K. Cockrell, "Prediction of Advances in Combat," in Reiner K. Huber, Lynn F. Jones, and Egil Reine, (eds.), Military Strategy and Tactics: Computer Modeling of Land War Problems (New York and London: Plenum Press, 1975), pp.267-285, and 153-165, respectively. In neither case, moreover, are the data structured in such a way as to facilitate an examination of hypotheses focusing on force employment variables ("depth," and "reserves," for example, are addressed in the HERO database by a summary assessment that for the battle in question, the variable was either "an advantage decisively affecting the outcome," "a disadvantage decisively affecting the outcome," or neither. See T.N. Dupuy, et. al., op. cit., Vol.2, pp.1-24; neither effect is addressed by the Vertex data). While data appropriate for a large n statistical approach to this inquiry eventually may become available, current sources are problematic for our purposes.

[3] See Eckstein, op. cit., pp.79-137; George, op. cit., pp.43-68. An exception concerns instances where a small number of alternative theories constitute an exhaustive set of plausible explanations for the phenomenon in question. Here, a well-selected "critical case" approach can at times conclusively disprove one or more of the alternatives, thus lending substantial weight to the other in the absence of any other plausible explanation. Alternatively, a "least likely" critical case can sometimes offer a

for this study.[4] Pending such a broader guage approach, we have relied here on ex ante computer simulation for initial testing.

Simulation has proven to be a valuable "in vitro" experimental tool in a variety of disciplines; it is widely exploited, for example, in aeronautical engineering, astrophysics, and a growing range of other applications in which true experiments are too risky or too expensive for routine use.[5] Simulation can involve the use of analogous physical systems, such as a wind tunnel; mathematical representations of physical systems, as in computer models of fluid dynamics; or combinations of the two, as in man-machine computer-assisted conflict games.[6] Although simulation is only a partial substitute for ex post facto observation, and is best used in conjunction with at least selective ex post examination, it can be an invaluable tool under the proper circumstances. In particular, computer simulation is most appropriate for applications in which fundamental properties are well understood, but in which the propositions under test involve interactions among component elements too complex for direct deduction from those fundamental properties. To the degree that these preconditions obtain, simulation enables theories pertaining to aggregate behaviors to be tested subject to the validity of the disaggregate representation of the simulated entities themselves.[7]

The theory in Appendix C, for example, contends in part that increased attacker velocity increases the minimum casualties the attacker will suffer in an assault on a defended position. To test this proposition, we need a disaggregate simulation that correctly represents the fundamental properties of movement, observation, target acquisition, hit and kill for particular weapons, and that enables the user to control assault velocity (and the related variables of barrage duration for artillery, or dismounting of infantry). We then can test propositions about the aggregate behavior of many such

---

powerful corroboration for a true theory. For our purposes, however, neither an exhaustive set of plausible alternatives nor a suitable "least likely" case have yet been identified.

[4] Further evaluation of comparative ex post techniques is highly desirable as a complement to this initial testing, as is being pursued as a follow-on effort to this study.

[5] For an informal survey of this growing field, see "The World in a Grain of Silicon," The Economist, June 10, 1989, pp.79-82.

[6] See Brody and Brownstein, op. cit., pp.211-264.

[7] Ibid. That degree may be incomplete—as is the case, for example, in the current state of the art in combat simulation—in which case simulation testing will be useful but necessarily inconclusive as the sole basis for the validity of the theory under test.

weapons as velocity changes, subject to the validity of the simulation of individual shooter-target duels and individual movement or observation phenomena.

The particular computer simulation chosen for this purpose was the Lawrence Livermore National Laboratory's Janus model.

> Janus is an interactive, brigade level, two sided game created to explore relationships of combat and tactical processes using a stand-alone, event sequenced, stochastic, computer simulation....Janus is an event-driven simulation that models fighting systems as entities (tank, helicopter, howitzer, etc). Entity characteristics include descriptions of the weapons carried, weapon capabilities, movement speeds and how they are attenuated by terrain effects, accountability of ammunition and fuel, crew performance, sensor data describing how the battlefield is observed, as well as supply/resupply data.[8]

Janus is highly disaggregate and extremely detailed with respect to physical interactions between weapon systems. Individual weapons move, acquire targets, hide, shoot and die. Target acquisition, determination of hit and determination of kill (given hit) are computed stochastically for each interaction. Movement and intervisibility are determined using digitized Defense Mapping Agency representations of actual Central European terrain. Surface features such as rivers, roads, towns and vegetation, and terrain elevation features such as hills, plains and ravines are all incorporated explicitly. Janus displays the progress of the battle across this terrain with high resolution, real-time graphics.

Whereas the physics of weapon interactions are handled by the model, force employment is determined directly by the user. Weapon deployments, movement orders and firing authorization are all controlled through a graphical input-output system that permits the experimenter a high degree of control over variables such as depth, formation, direction of advance, withdrawal, suppressive fire, speed of closure, and, of course, force levels or weapon types. Combat outcomes can be observed directly on the screen as the engagement unfolds and systems advance, retreat and are lost to enemy fire.

While Janus has not been validated to the same degree of confidence as have, for example, aeronautical engineering models, it has nevertheless survived substantial empirical testing—especially through systematic comparison of simulation output and the results of U.S. Army field exercises on the instrumented test range of the Army's National

---

[8] Janus Users Manual, Version 4.0, (Livermore, Ca.: Lawrence Livermore National Laboratories, January 4, 1988), Introduction, p. 1.

Training Center at Fort Irwin, California.[9] The U.S. Army, the U.S. Marine Corps, and the Canadian National Defense Force have each evaluated the simulation and certified it as an appropriate representation of small unit combat for purposes of analysis, doctrinal development, and for the training of field grade officers.[10] While it is unlikely that any simulation of combat can ever attain the degree of validity associated with purely physical engineering models, the Janus simulation has been validated to an unusually high degree. It thus substantially meets the dual demands of control and disaggregate validity, and is well-suited for the conduct of controlled experiments to test the theoretical propositions advanced in Appendix C.[11]

C. EXPERIMENTAL DESIGN

Although Janus provides the control and disaggregate validity required for sound ex ante experimentation, its scale restricts the potential scope of testing. Janus is a brigade level simulation; it cannot directly accomodate a 40-division theater force.

---

[9] See, for example, L. Ingber, H. Fujio, and M.S. Wehner, "Mathematical Comparison of Combat Computer Models to Exercise Data," forthcoming in Mathematical and Computer Modelling; L. Ingber and D.D. Sworder, "Statistical Mechanics of Combat with Human Factors," forthcoming in Mathematical and Computer Modelling; and L. Ingber, "Mathematical Comparison of Janus(T)," in S. E. Johnson and A. H. Lewis, (eds.), The Science of Command and Control: Part II- Coping with Complexity, (Washington, D.C.: AFCEA International Press, 1989), pp. 165-176.

[10] Current Janus users include the U.S. Army Training and Doctrine Command (TRADOC), the U.S. Army Command and General Staff College (Fort Leavenworth), U.S. Army Southern Command (Panama), U.S. Army I Corps (Fort Lewis), the Berlin Brigade, the U.S. Marines Corps (Quantico Marine Base), the Canadian National Defense Headquarters, the Institute for Defense Analyses, and Lawrence Livermore National Laboratories.

[11] Note, however, that no matter how valid or useful a depiction of small unit combat, JANUS is not a predictive theory for the outcomes of theater level warfare. This is partially due to its scale—JANUS is a brigade level simulation, and cannot accomodate a theater-size terrain file or force structure. More importantly, however, JANUS per se—as are all small scale combat simulations—is insufficiently specified to address the kinds of theoretical questions of interest here. To conduct a JANUS run, the user must provide detailed unit deployments, movement orders, battle formations, and fire plans. The computed combat results are then specific to the particular circumstances chosen for that run. Since an infinite number of possible configurations could be established, the task of choosing appropriate circumstances is clearly crucial, yet without some further theoretical guidance it is not necessarily clear how to establish sets of circumstances appropriate for addressing, e.g., the effects of theater level force to space ratios. JANUS is much like a wind tunnel in this regard; a wind tunnel is not a theory of lift, or even a model of aircraft performance that enables one to deduce directly the optimal configuration of a wing. Rather, it is a very detailed device for obtaining knowledge about the behavior of a complex system by trial and error. If the hypotheses to be tried are well-chosen, it can be an extremely valuable tool, but it is not in itself a predictive causal theory for the phenomena of interest to us here.

Simulations capable of handling larger forces, however, lack the detail and flexibility required for effective experimentation.[12]

To exploit Janus' unique capabilities, a "critical node" strategy was employed. Janus alone cannot directly address the entire chain of inference embodied in the theory, but not all steps in that chain are equally important for the validity of the outcome. Some, such as the theater dynamic structure that interconnects assault, reserve counterconcentration and counterattack, are matters of substantial consensus among writers and analysts.[13] Others, such as the description of overall advance rate given assault velocity (see Appendix C), are necessary components, but are not of fundamental significance for the theory as a whole. A limited number of key propositions, however, are both novel and crucial to the entire chain of inference. In particular, the tradeoff between casualties and velocity serves as the analytic foundation of the theory. This relationship drives advance rates, losses, and counterattack effectiveness, and thus plays the central role in determining optimal force employment choices and the resulting net territorial gain outcome. While the theory as a whole could still be invalid even if this relationship is true, the theory would be fundamentally unsound if this relationship were false. Moreover, this relationship is a novel contribution, and thus is not subject to circumstantial support by any existing analytic consensus. The casualty-velocity tradeoff is a "critical node" for the validity of the larger chain of inference; by focusing testing on this issue, we can obtain a substantial degree of insight into the validity of the theory as a whole pending complementary test by ex post facto techniques.

The Janus experiments are designed to disprove this proposed relationship between casualties and velocity if it is in fact false, and to provide experimentally fit values for constant parameters if the relationship is not disproven. With respect to the former, the experiment must test each of six contentions essential to the description of the casualty-velocity relationship in Appendix C:

---

[12] For surveys of relevant models, see for example James G. Taylor, "Attrition Modeling" in Reiner K. Huber, et. al., (eds.), Operational Research Games for Defense (Munich: R. Oldenbourg, 1979), pp.139-89; Alan F. Karr, "Lanchester Attrition Processes and Theater-Level Combat Models" in Martin Shubik, (ed.), The Mathematics of Conflict (New York: Elsevier, 1983), pp.89-126; Garry D. Brewer and Martin Shubik, The War Game: A Critique of Military Problem Solving (Cambridge, MA: Harvard University Press, 1979); U. Candan, L.S. Dewald, and L.R. Speight, Present NATO Practice in Land Wargaming, (The Hague: SHAPE Technical Center, 1987), Professional Paper STC-PP-252

[13] Which does not, of course, necessarily make them true. It does, however, provide some relative measure of confidence pending formal test—and thus it does suggest that the presumed theater dynamics are appropriately regarded as a secondary target for validation efforts.

(1) that minimum attacker casualties increase as velocity increases;[14]

(2) that the slope of the resulting casualty-velocity curve increases as the fraction of infantry or artillery in the attacker force mix increases (and, thus, that the slope decreases as the fraction of armor increases);

(3) that the slope increases as the fraction of infantry in the defender force mix increases, and that the slope decreases as the defender fraction of armor and artillery increases;

(4) that minimum casualties increase as the local attacker:defender force-to-force ratio decreases, for all attack velocities;

(5) that minimum casualties increase as the depth of the defense (and thus the average distance travelled by a given assault echelon prior to launching a given assault) increases; and

(6) that minimum casualties decrease as the fraction of defenders withdrawn increases.

To test these contentions, a modified complete factorial design was employed. Two hundred and seventy-seven separate Janus engagements were fought under 55 unique scenarios comprising three different force-to-force ratios, three different weapon mixes for attackers, three different weapon mixes for defenders, at four different assault velocities. Not all permutations were militarily feasible, however, and certain combinations of characteristics led to disproportionately lengthy Janus execution times. Where a slow-running scenario was unlikely to add significant variance to the resulting attacker casualties or where the scenario was infeasible, a full program of experimental engagements was not run, but a sufficiently broad sampling was provided as to enable gross effects to be perceived.[15] A total of five randomly selected terrain samples were drawn from among the set provided by LLNL; preliminary assessments, however, failed to yield

---

[14] As discussed in Appendix C, it is not required that the casualty-velocity relationship be monotonically positive for all velocities and all weapon mixes. In instances wherein a slowly advancing, infantry-heavy attacker faces an artillery-heavy defender, casualties could increase as velocity decreases. However, the theory does contend that, for all velocities above some value, attacker casualties increase as *velocity increases.*

[15] For example, it was discovered that at low force-to-force ratios, extreme velocities cannot be attained by armor-heavy attackers against infantry-heavy defenders. Such scenarios were infeasible. Alternatively, at very low velocity, low force-to-force ratio scenarios in which both attacker and defender are armor-heavy produce results very similar to higher velocity assaults (since the attacker lacks the supporting arms to exploit the longer preparation and the defender is largely invulnerable to such support were it available); yet run times for these scenarios are prohibitively long. To achieve a complete factorial design including a full complement of such scenarios thus would have been impossible, with the result that fewer velocity increments could be run for such combinations given available resources.

statistically significant variance in casualty results across these particular samples, and the bulk of the experimental engagements were consequently fought on a single, randomly selected terrain file. Defensive depth was assessed separately for depths beyond one engagement per assault echelon; six engagements were fought in which assault waves advanced through one level of depth prior to the recorded engagement, and five engagements were fought in which assault waves advanced through two separate levels of depth prior to the recorded engagement. These engagements were characterized by a single force-to-force ratio and a single (balanced) weapon mix for attackers and defenders. Defensive withdrawal was assessed separately by correlating attacker and defender casualties over time during a given engagement; it was assumed that withdrawal at an arbitrarily selected time would successfully terminate the engagement with final attacker casualties being those incurred as of the reference time.

Casualty data from these engagements were then fit to several candidate functional forms consistent with the curves developed in Appendix C. Falsification criteria consisted of coefficient values outside the bounds implied by those curves for the functional form that best fit the experimental data (for detailed criteria, see below). In the event that the experimental data failed to falsify, those coefficient values provided fitted constants for use in the VFM model that implements the equations developed in Appendix C.

D. EXPERIMENTAL PROCEDURE

For each scenario (a unique combination of force levels, weapon mixes, velocity and terrain), a variety of assault configurations were examined subject to the proviso that each configuration met the required velocity and represented a plausible use of forces in the context of known military doctrines. The lowest casualty configuration was accepted as the efficient assault for that velocity. Defenders were deployed along standard doctrinal lines.[16] Defensive deployment was held constant across scenarios with a given defensive force composition and terrain sample. Although Janus can be run as an interactive game, all experimental runs were conducted as closed simulations; i.e., movement and engagement orders, dismount points and artillery preparations were determined as

---

[16] See, for example, Headquarters, Department [illegible]he Army, FM 71-2, The Tank and Mechanized Infantry Battalion Task Force (Washington, D.C.: USGPO, 30 June 1977); Headquarters, Department of the Army, FM 71-1, The Tank and Mechanized Infantry Company Team (Washington, D.C.: USGPO, 22 November 1988).

scenario conditions and not altered during the course of experimental runs for the given scenario.[17]

Velocity was defined as the distance to be covered by the assault (measured in kilometers from the jump-off point of the initial assault wave to the objective line), divided by the time required to cover the given distance and defeat the defenders on the objective.[18] "Defeat" was defined as the destruction of 60 percent or more of the defending AFVE score. Elapsed time was measured from initiation of preparatory artillery fire to the arrival on the objective line of the first assault wave for which the defender defeat criterion had been met. Given that Janus is a stochastic simulation, velocity by this definition can vary for individual runs within a scenario. Four broad classes of *attempted* velocities were considered, however: a slow case, in which available infantry were dismounted following 60 minutes of preparatory artillery by the accompanying artillery complement and four minutes of smoke preparation; a moderate case, in which infantry were dismounted following 12 minutes of preparatory artillery and four minutes of smoke; a fast case, in which a mounted assault followed four minutes of smoke and a two volley suppressive artillery barrage; and a very fast case, representing a hasty attack in which a mounted assault proceeded directly from the march with only the support of suppressive artillery that could be brought to bear during the advance itself.[19]

Attacker casualties were assessed as AFVEs lost prior to engagement termination, where termination was determined by either defensive withdrawal or the defeat of defenders on the position.

For scenarios involving depths beyond one engagement per assault echelon, depth was measured as the distance between the initial jump-off point and the ultimate objective line, in kilometers. The purpose of these runs was to test the entropic effects of depth on attack coherence as described in Appendix C, where this effect is measured in terms of fractional increase in casualties for assaults conducted after a preliminary combat advance to the given depth, relative to an assault conducted at an identical force-to-force ratio at zero depth (i.e., with no preliminary contested advance prior to jump-off for the reference

---

[17] As would be the case for non-interactive simulations such as CARMONETTE or BASIS. Interactive operation would dramatically increase the set-up and execution times for the given runs, in addition to doubling the personnel required to conduct the runs themselves.

[18] Attacker jump-off points and objective lines were held constant across scenarios conducted on a given terrain sample.

[19] As noted above, not all of these proved feasible for all scenarios.

assault). To obtain the proper control (i.e., to obtain two engagements differing only in the opposed distance covered prior to the engagement of interest and not differing, for example, in strength as of final jump-off), assault echelons advancing into depth were opposed by non-firing defenders prior to the assault's reaching its final jump-off line. In this way, equal jump-off strengths could be ensured between deep and shallow attacks, yet the entropic effects on the attacker's coherence of encountering defenders in prepared positions could be obtained, at least in part. In any case, these effects are only partially addressable through Janus, which can assess the loss of orderly formation through evasive maneuver and advance over broken terrain, but cannot assess commander overload or fatigue. All of these are contributors to the entropic effect described in Appendix C, however; thus the coefficient estimates resulting from Janus testing must be considered underestimates of this effect.

E. RESULTS AND STATISTICAL ANALYSIS

Data for scenarios involving depths beyond one engagement per assault echelon and data for defensive withdrawal were compiled separately and analyzed using Minitab on a WIN TurboAT personal computer. All other data were compiled and analyzed using SAS on a VAX 8600.

The SAS procedure used for this study was PROC NLIN, a technique used by SAS to fit nonlinear regression models by least squares.[20] PROC NLIN uses the Gauss-Newton iterative process to estimate parameter values.[21] With this method, initial values

---

[20] Although the final functional form is itself linear, the majority of functional forms that were explored by this study were non-linear. For this reason, it was necessary to use PROC NLIN.

[21] The description given in the text, and the following, more mathematical explanation, are drawn from SAS User's Guide: Statistics, Version 5 Edition, (Cary, N.C.: SAS Institute Inc.,1985), pp. 584 - 586.

For a general non-linear function Y = F (B0, B1, B2, ... Bk, X0, X2, ...Xj), in which B0...Bk are the k parameters and X0...Xj are the j independent variables, initial values for the parameters B0...Bk are estimated from the data. Call these initial estimates B0'...Bk'. The function Y is then approximated by the following Taylor series expansion of Y using B0'...Bk':

Y = F(B0'...Bk',X0...Xk) + D0(B0-B0') + D1(B1-B1') + ... + Dk(Bk-Bk')

where Dk, the partial derivative of Y with respect to the parameter Bk, is evaluated for B0 = B0', B1 = b1', ... Bk = Bk'. This approximation of Y is a linear function of the k variables (B0-B1'), (B1-B1'), ... (Bk- Bk'). Using the principle of least-squares, values d0...dk can be estimated for (B0-B1'),...(Bk-Bk'). These values, d0...dk, represent corrections to the initial estimates for the parameters B0...Bk. SAS then calculates a second approximation to B0...Bk, namely B0"=B0'+d0, B1"=B1'+d1,..., Bk"=Bk'+dk. A new Taylor series expansion of Y is generated using B0"...Bk" in place of B0'...Bk'. This expansion of Y is used to estimate corrections to B0"...Bk". These corrections, d1'...dk', are used to calculate B0'''...Bk'''. This iterating process is repeated until the error sum of squares for the ith iteration meets the criterion given in the following footnote.

for the constants are estimated from the data. A linear approximation of the nonlinear function is then generated using these initial parameter estimates and the partial derivatives of the nonlinear function with respect to these parameters. The principle of least squares is used to find, from the linear approximation, corrections to the initial estimates for the constants. These corrections are used to find new estimates for the parameter values, which in turn are used to generate a second linear approximation to the nonlinear function. This second linear approximation is used to generate new corrections to the parameter estimates. The process is repeated until the error sum of squares is minimized.[22]

1. Statistical Fits

For the functional form,

casualties=
[(B/R)^Krb]*[(K3*inf*vel)+(k4/inf)+((k1*Ba)/(vel+1))+(k6/(Ra*(vel+1)))], [23]

where

inf = infa + infd,
infa = fraction of attacker AFVEs that are infantry,
infd = fraction of defender AFVEs that are infantry,
B = blue maneuver force strength (AFVEs),
R = red maneuver force strength (AFVEs),
vel = assault velocity (kilometers per hour),
Ba = blue artillery strength (tubes), and
Ra = red artillery strength (tubes),

the treatment in Appendix C implies the following falsification criteria: [24]

---

[22] For PROC NLIN, the iterating process is terminated if, for the ith iteration, the following condition has been met:

(SSEi-1 - SSEi) / (SSEi + 10^-6) < 10 ^ -8.

[23] Note that tacair and ACM effects are not explicitly included here; the treatment in Appendix C denotes these effects in terms of AFVEs removed from direct fire combat by these weapons, where the determination of the number of AFVEs so removed is exogenously determined. As such, it was not experimentally assessed here.

[24] The value of k4 has no bearing on the falsification criteria. While statistically significant, it plays no part in the six essential contentions of the theory (given earlier in this appendix).

k3 < = 0 (implying that casualties do not increase with velocity, and that the slope of the casualty-velocity curve does not increase with increasing infantry fractions); [25]

k1 < = 0 (implying that the slope of the casualty-velocity curve does not decrease as the defender's artillery fraction increases); [26]

k6 < = 0 (implying that the slope of the casualty-velocity curve does not increase as the attacker's artillery fraction increases); [27] and

Krb < = 0 (implying that casualties do not increase as the attacker:defender force-to-force ratio decreases). [28]

SAS produced the following statistics regarding the constants k3, k1, k6, and k4:

---

[25] These criteria derive from contentions 1, 2, (as pertains to the fraction of attacker infantry) and 3 (as pertains to the fraction of defender infantry). Contentions 2 and 3 imply that if the partial derivative of dC/d(vel) with respect to the fraction of infantry, $d^2C/d(vel)d(infa)$ and $d^2C/d(vel)d(infd)$ respectively, were not each monotonically positive, the result would tend to disconfirm. Equivalently, if $d^2C/d(vel)d(inf)$ were not monotonically positive, this result would also tend to disconfirm. The form of the partial derivative is

$d^2C/d(vel)d(inf)$ = k3 * (B/R)^Krb.

As [(B/R)^Krb] is always positive, disconfirmation requires that k3 be negative.

Contention 1 implies that if the partial derivative of C with respect to vel, dC/d(vel), were not monotonically positive (at least for velocities above some value), then the result would tend to disconfirm. The form of this partial derivative is

dC/d(vel) = [(B/R)^Krb] *
{ (k3*inf) - ((k1*Ba)/((vel+1)^2)) - (k6/(Ra*((vel+1)^2)) }.

As [(B/R)**Krb] is always positive, above some velocity (defined by k3, k1, k6, inf, Ba, and Ra), this partial derivative is monotonically positive (indicating that higher velocity will give higher casualties) if (and only if) k3 is positive; thus, disconfirmation requires that k3 be less than, or equal to, zero.

[26] This criterion derives from contention 3 (as pertains to artillery), which implies that if the partial derivative of dC/d(vel) with respect to Ba were positive, the result would tend to disconfirm. The form of this partial derivative is

$d^2C/d(vel)d(Ba)$ = [(B/R)^Krb] * [- (k1/((vel+1)^2))].

If k1 is less than or equal to zero, the partial derivative will be positive, and thus imply disconfirmation.

[27] This criterion derives from contention 2 (as pertains to artillery), which implies that if the partial derivative of dC/d(vel) with respect to Ra were negative, the result would tend to disconfirm. The form of this partial derivative is

$d^2C/d(vel)d(Ra)$ = [(B/R)^Krb] * { k6/((Ra*(vel +1))^2) }.

If k6 is less than or equal to zero, the partial derivative will be negative, and thus imply disconfirmation.

[28] This criterion derives from contention 4, which implies that if the partial derivative of C with respect to B were negative, the result would tend to disconfirm. The form of this partial derivative is

dC/d(B) = [Krb * {(B/R)^(Krb - 1)}] * [positive factors].

If Krb is less than or equal to zero, the partial derivative will be negative, and thus imply disconfirmation.

parameter	estimate	std.err	t-ratio [29]
k3	14.81	0.37	40.03
k1	0.54	0.42	1.29
k6	727.97	100.51	7.24
k4	7.05	1.12	6.29

sum of squares (corrected total): 183387.2
error sum of squares: 49185.9
degrees of freedom: 277
adjusted R^2: [30] 0.73

In the mathematical expression that produced the best fit to the data, Krb was not estimated. In effect, therefore, Krb was constrained to be equal to one. All mathematical expressions in which Krb was not so constrained produced estimates of Krb in the range 0.9 to 1.1.

All fitted parameter estimates thus fall outside the falsification range identified above, with confidence in excess of the .01 level for k3, k4, and k6; and with confidence in excess of the .20 level for k1. Thus, the observed experimental data tend to corroborate the hypothesized relationship.

Representative comparisons of observed data and fitted casualty curves are provided in Figures D-1, D-2[31], and D-3. Each figure illustrates one of the key relationships postulated by the theory. Figure D-1 provides experimental results and predicted casualties for a single weapon mix (i.e., the "balanced" attacker and defender case) as a function of velocity and local force to force ratio. Figure D-2 provides experimental results and predicted casualties for a single local force to force ratio (i.e., the "medium" force to force ratio) and attacker weapon mix (i.e., the "balanced" attacker case) as a function of velocity and the defender's weapon mix. Figure D-3 provides experimental

---

[29] This is the approximate T-ratio, calculated as estimate/std.error.

[30] The meaning of the R^2 value is somewhat ambiguous for functional forms that have no constant term. The given R^2 is therefore illustrative only. The R^2 was calculated as 1 - (error sum of squares/ total sum of squares). The sum of squares has been corrected to account for the lack of a constant, and is identical to the true total sum of squares for the data.

[31] With respect to the impact of an increase in the defender's artillery fraction, Figure D-2 does not clearly resemble the curve postulated in Appendix C, in which casualties actually decreased as velocity increased. As noted in the text, that curve represented an extreme case. We contend only that the slope of the casualty-velocity curve (as compared to the actual number of casualties suffered) decreases as the defender's artillery fraction increases.

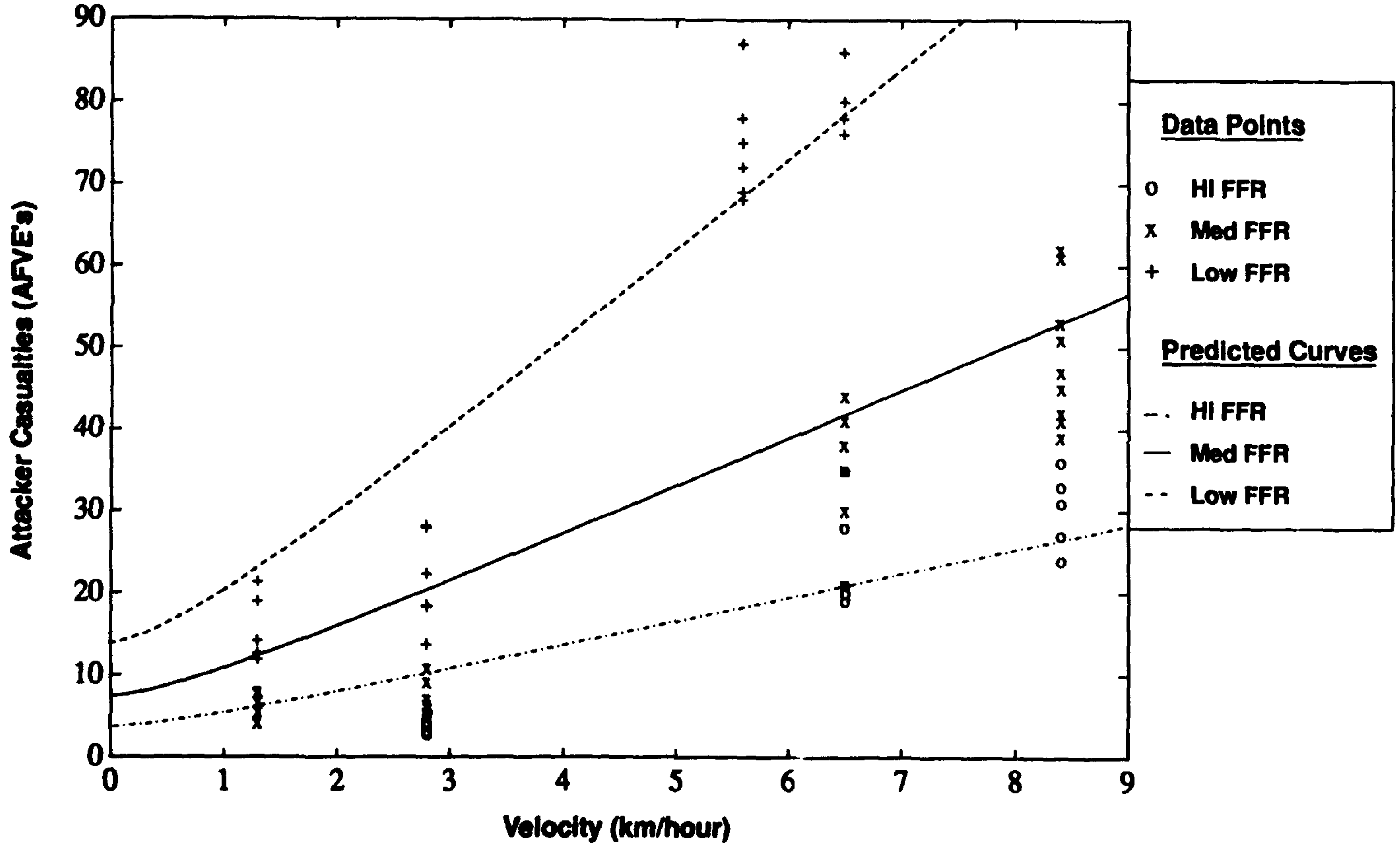

Figure D-1. Experimental Results: Effect of FFR (Force to Force Ratio, Red:Blue)

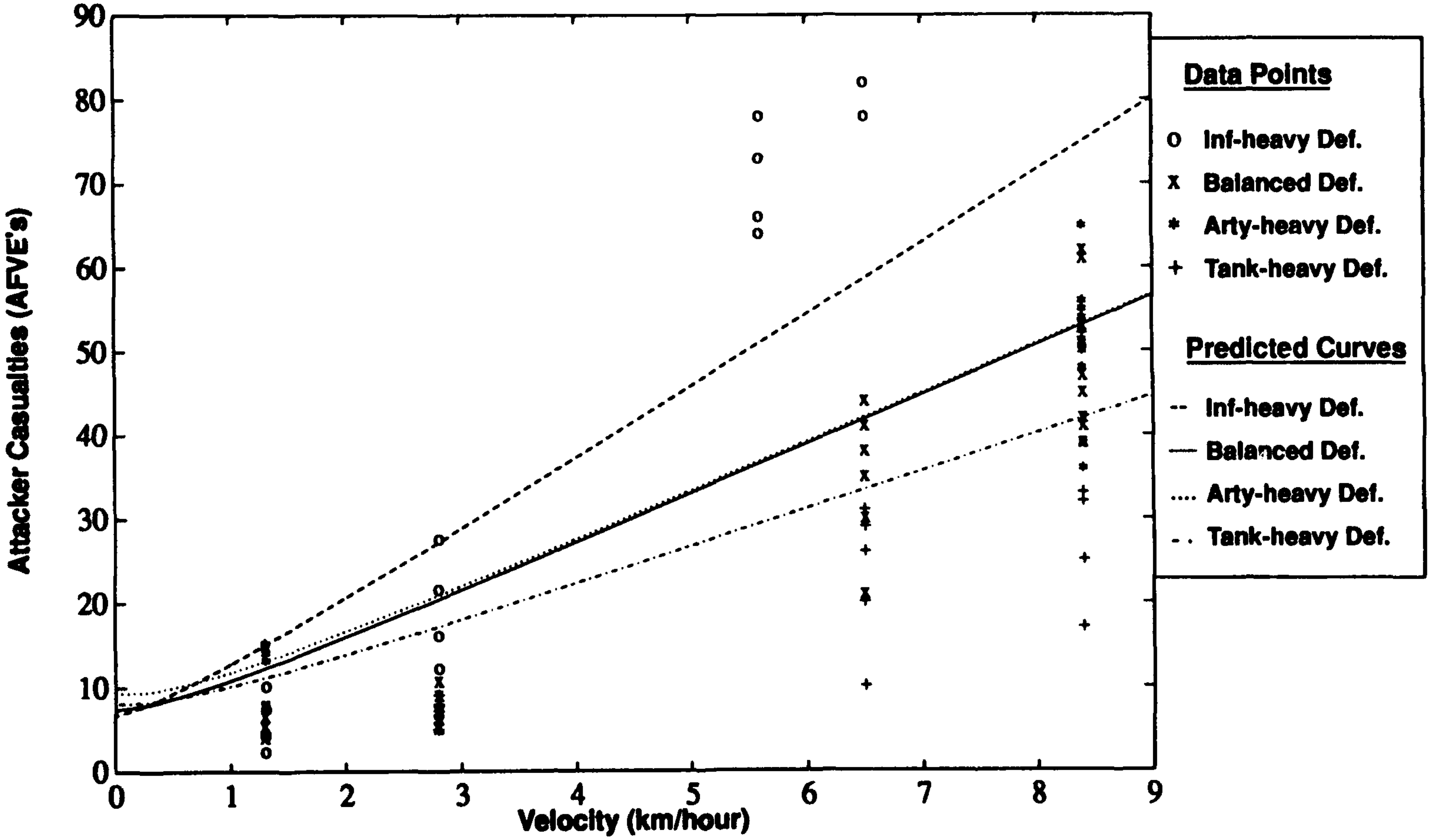

Figure D-2. Experimental Results: Effect of Defender Weapon Mix

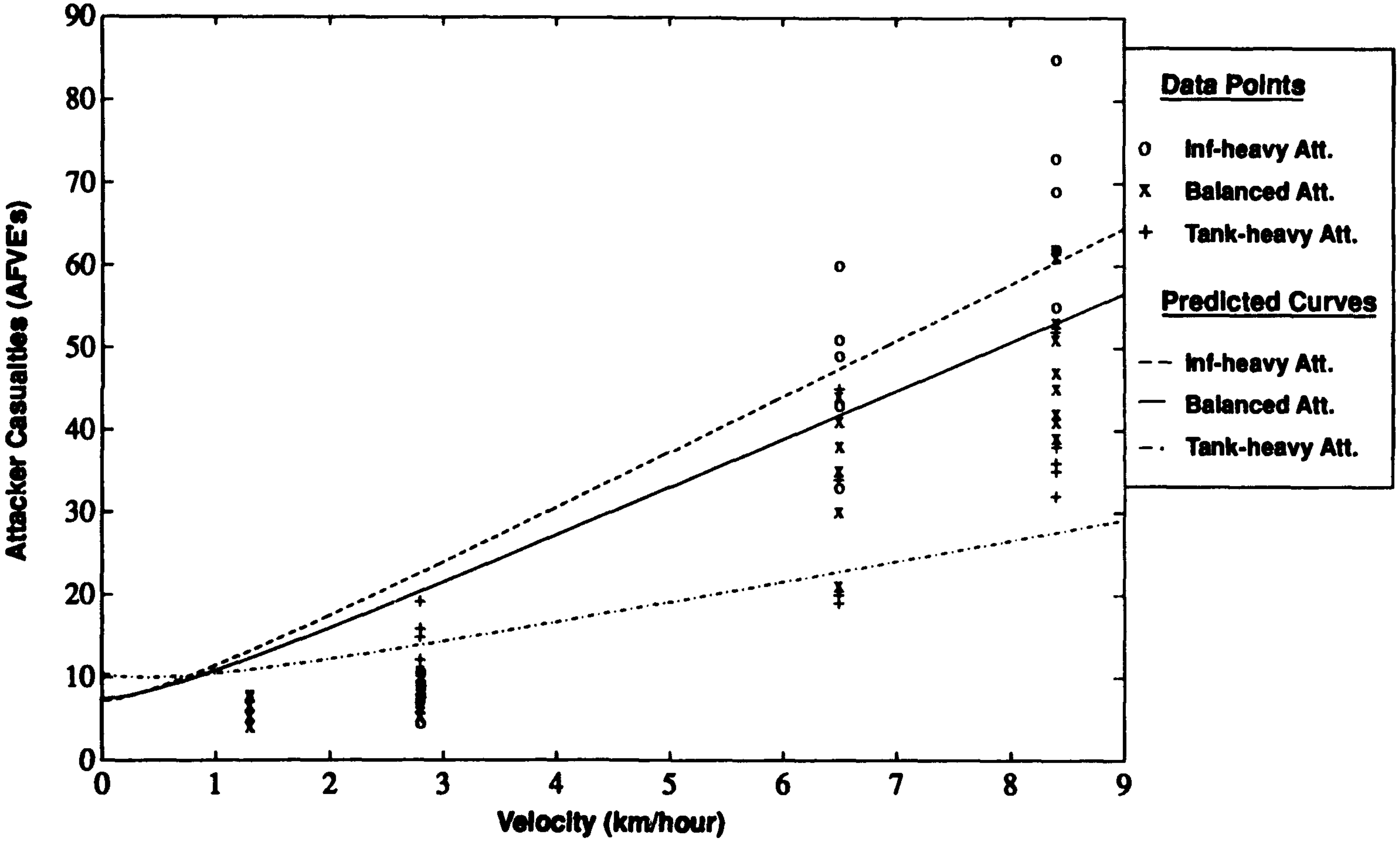

Figure D-3. Experimental Results: Effect of Attacker Weapon Mix results and predicted casualties for a single local force to force ratio (i.e., the "medium" force to force ratio) and defender weapon mix (i.e., the "balanced" defender case) as a function of velocity and the attacker's weapon mix. Not all possible permutations of these variables have been illustrated here—the figures are intended to be representative of the general quality of fit, rather than to provide an exhaustive summary of the data.

For the functional form

$$\text{gamma} = (1 + (k8 * \text{depth})),$$

where

gamma = fraction of zero-depth casualties suffered in an assault conducted at the given, non-zero depth, and
depth = length of opposed advance prior to the reference engagement (kilometers),

the treatment in Appendix C implies a falsification criterion of

$$k8 <= 0.$$ [32]

This form expresses gamma as a percentage increase in casualties relative to an hypothetical engagement fought under conditions varying only in depth of engagement. However, the Janus experiments measured absolute casualties suffered by an attacker during an assault. It was therefore necessary to re-express gamma in terms that would permit direct estimation of k8 from absolute casualty data:

$$\text{casualties} = Co + (k8'*\text{depth}),$$

where Co is the number of attacker casualties that would be expected at zero depth under otherwise identical conditions. In appendix C, attacker casualties and gamma are related in the following way:

$$\begin{aligned}\text{casualties} &= \text{gamma} * Co;\\ &= (1 + (k8 * \text{depth})) * Co\end{aligned}$$

Thus:

$$\text{casualties} = (1 + (k8 * \text{depth})) * Co = Co + (k8'*\text{depth})$$

which implies:

---

[32] This criterion derives from contention 5, which implies that if the derivative of gamma with respect to depth were negative, the result would tend to disconfirm. As d(gamma)/d(depth) = k8, disconfirmation requires that k8 be less than or equal to zero.

k8 = (k8' / Co).

Correspondingly, the falsification criterion given above for k8 can be restated in terms of k8': to disconfirm, k8' must be less than or equal to zero.[33]

For the functional form

casualties = Co + (k8'*depth),

Minitab generated the following statistics for k8' and Co:

parameter	**estimate**	**std. error**	**T-ratio**
Co	58.61	2.73	21.50
k8'	5.56	1.11	5.02

sum of squares (corrected total): 7272.8
error sum of squares: 3763.1
degrees of freedom: 28
adjusted R^2: 0.46

The fitted estimate for the parameter k8' lies outside the falsification range identified above, with confidence in excess of the .01 level; thus the observed experimental data tend to corroborate the hypothesized relationship.

These parameter estimates imply a value of 0.09 for k8.

Observed data and the fitted gamma curve are compared in Figure D-4.[34]

---

[33] This follows directly from the definition of k8 in terms of Co and k8'. If k8 is to be less than or equal to zero, then k8' must be less than or equal to zero, as Co must always be greater than or equal to zero (it is not possible to suffer negative casualties).

[34] Figure D-4 provides an exhaustive, rather than a representative, depiction of the results obtained in the withdrawal experimentation (note, however, that overstrikes are present, but difficult to discern visually).

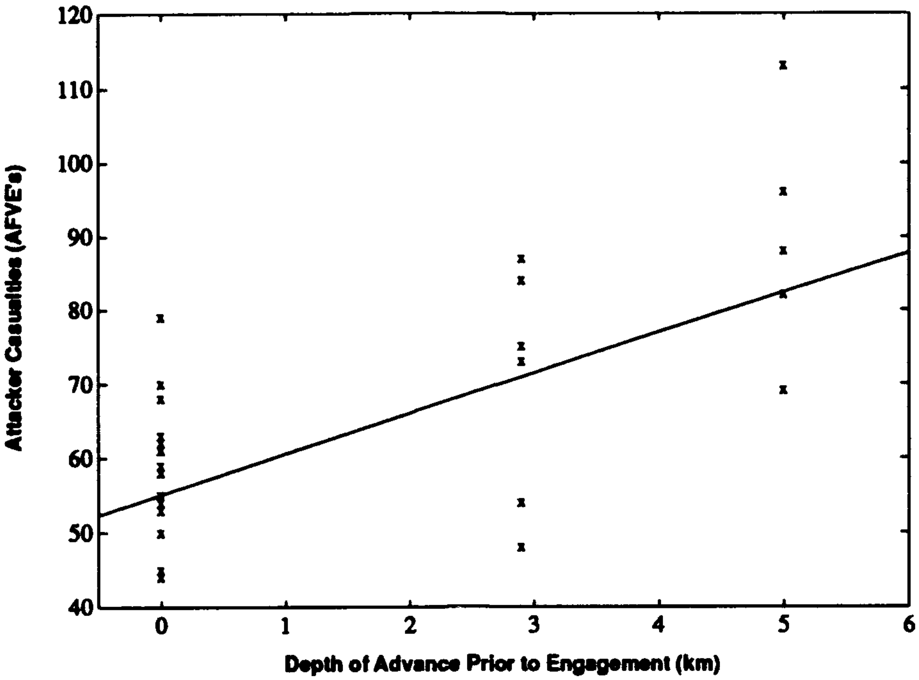

Figure D-4. Experimental Results: Entropic Effect of Depth

For the functional form

alpha = 1 - [kW * (w^k10)],

where

alpha = fraction of fight-to-the-finish attacker casualties suffered when the defender withdraws a fraction of his strength, w; and

w = fraction of defending AFVE strength withdrawn,

the treatment in Appendix C implies the falsification criterion: [35]

1) k10 < = 0; or
2) kW < = 0.

---

[35] This criterion derives from contention 6, which implies that a positive derivative of C with respect to w would tend to disconfirm. As dC/dw = (-1*k10*kW*(w^(k10-1))), if either k10 or kW is negative (but not both), then dC/dw would be greater than zero and the results would tend to disconfirm (if both k10 and kW were negative, then dC/dw would be negative, as postulated in appendix C, although the resulting form would produce the problematic behavior of rising attacker casualties as increasing fractions of defenders are withdrawn, reaching a maximum value when all defenders withdraw prior to the initiation of combat).

Minitab generated the following statistics:

parameter	**estimate**	**std. error**	**T-ratio**
k10	2.51	0.08	31.86

sum of squares (corrected total): 154.58
error sum of squares: 62.71
degrees of freedom: 218
adjusted R^2: 0.59

The mathematical expression that provided the best fit did not estimate kW. In effect, kW was thereby constrained to equal one. Mathematical expressions that did not so constrain kW produced estimates for kW that were close to one, but the inclusion of kW lowered the adjusted R^2.

The fitted parameter estimate falls outside the falsification range identified above, with confidence in excess of the 0.01 level. Thus, the observed experimental data tend to corroborate the hypothesized relationship.

Observed data and the fitted alpha curve are compared in Figure D-5.[36]

---

[36] Figure D-5 provides an exhaustive, rather than a representative, depiction of the results obtained in the gamma experimentation.

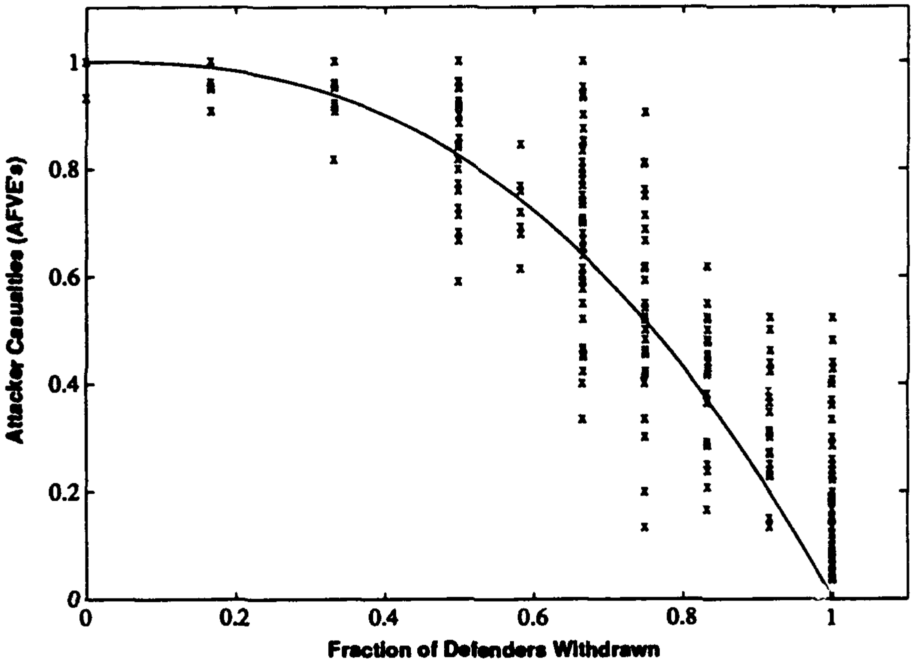

Figure D-5. Experimental Results: Effect of Defender Withdrawal

2. *Alternative Functional Forms*

Many other equations were tested against the data. An abridged list of alternative functions, accompanied by the reason that each was rejected, follows:

1) A nonlinear form for the (inf*vel) relationship: Three formulations provided statistically significant parameter estimates (at a confidence level in excess of .05): (vel^inf), (inf*e^vel), and ((inf^k)*(vel^k)). The first two expressions resulted in substantially lower adjusted R^2. The third expression lowed the adjusted R^2 somewhat, and produced parameter estimates very close to 1.0 (e.g., 1.11 and 1.04 respectively).

2) An isolated velocity term ([(B/R)*K*vel], for example): No tested formulations yielded significant parameter estimates (at the .05 level).

3) Alternative forms for the infantry term: Three expressions were attempted: (K*inf), (K*inf^k), (K/(inf+1)). Only the first resulted in a slightly higher adjusted R^2, but only for values of K such that zero velocity would imply negative attacker casualties.

4) Alternative forms for the Ba term: Four expressions were attempted: (K*Ba), (K*Ba*e^vel), (K*Ba*vel^k), (K*Ba^k). Each of these expressions reduced the adjusted R^2.

5) Alternative forms for the Ra term: The following expressions were attempted for both Ra and (Ra^k): (K*Ra), (K*Ra*e^vel), (K*Ra*vel^k), (K/((Ra(vel+1))+1)). Each of these expressions reduced the adjusted R^2.

Appendix E

VARIABLE FORCE EMPLOYMENT (VFM) MODEL DOCUMENTATION

D. Sean Barnett
Dennis DeRiggi

A. INTRODUCTION

Appendix C describes a methodology for computing Red's net territorial gain as a function of given values for the force to space and force to force ratios, weapon technology, military geography and force employment. While all computations associated with the theory are elementary, they are numerous and tedious. Moreover, to solve for optimal red and blue force employment choices, and thus to compute optimal red ground gain, it is necessary to solve a two-person zero sum game with a potentially very large strategy space.

To facilitate the calculations involved in this process, a small computer program was developed. The resulting VFM (Variable Force eMployment) model calculates minimax Red ground gain by generating Blue strategy vectors, computing ground gained as a function of Red velocity for each vector, and selecting the vector that results in the smallest maximum Red gain. The resulting solution is an approximation of the true optimum; the user controls the accuracy of the approximation (and the running time of the program) by specifying the number of vectors within the set of allowable Blue strategies to be generated.

The VFM model was written in Microsoft FORTRAN (5.0) and developed on a COMPAQ 386/25 personal computer. It was compiled using the FL/C command and linked with the LLIBFOR7.LIB FORTRAN library.

The model returns an approximation to both the optimal net territorial gain, and the strategy vector that produced it, as a function of the force to space ratio. It thus enables the user to vary any of the parameters embodied in the theory and generate tables of consequent optimal ground gain and the associated optimal blue and red employment choices.

1. Approximation Algorithm

The set of feasible Blue strategies lies within a four dimensional rectangular polytope. The size of the polytope depends on the admissible values for the four components of the Blue strategy vector. These components are the fraction of Blue forces committed forward, the fraction of Blue reserves allocated to counterattack missions, the fraction of Blue forward forces withdrawn from any given defensive line, and the number of Blue defensive lines. In Appendix C, these variables are denoted by ϕ_{FWD}, ϕ_{CA}, w, and n_{LI}, respectively.

The approach by which VFM approximates Red's minimax ground gain is as follows. The code calculates ground gain as a function of Red velocity for all possible combinations of the values of the components of the Blue strategy vector that have been selected by the user. The optimum Blue strategy is the combination of vector component values that results in the smallest maximum ground gain.

The user selects the values of the components of the Blue strategy vector through the VFM input values that control the iterative process by which VFM finds the approximate optimum Blue strategy. The first iteration covers the entire allowable Blue option space and produces a first approximation of the optimum strategy. The second and subsequent iterations cover areas around the immediately preceding approximations to produce closer and closer approximations, until the last iteration produces the closest. For the first iteration, the user sets the allowable range of each component of the Blue strategy vector and the number of points between the endpoints of each range at which VFM will calculate ground gain. The user also sets the number of iterations that VFM will perform and the number of points between the endpoints of the vector component ranges of the second and subsequent iterations at which VFM will calculate ground gain. In all iterations, the values of the vector components include the endpoints of the allowable ranges and the values of the interior points. The interior points are evenly spaced and are separated by a distance equal to the size of the range divided by the number of interior points. In each iteration, VFM calculates an approximation of the minimax ground gain for all possible combinations of values of the vector components.

In the second and each subsequent iteration VFM defines new ranges and new component values and calculates minimax ground gain. In the second iteration the new range of each vector component is defined by the value of the same component of the vector that produced the first approximation plus and minus the distance between the interior points of that component in the first iteration. The number of evenly spaced interior points is set by the user; the value of the interior points is determined as for the first iteration. The process of the second iteration is repeated for all subsequent iterations, using the values of the vector components that produced the immediately preceding approximation and the user-selected number of interior points to define the new vector component ranges and values.

B. USER'S MANUAL

1. Data Preparation

The VFM input data set consists of 41 records in list-directed format. Each record begins with a character string, enclosed in single quotes, that identifies the variable or variables contained in that record. (A sample data set follows this section.) Data records are grouped into two categories: range value and single value. Range value records contain the minimum, maximum, and increment of the particular variable in question. For example, the fraction of forces that are forward deployed would be specified in a range value record. The minimum allowable fraction, the maximum allowable fraction, and the size of the steps the user requires would appear on the record. In contrast, single value records contain the one value assigned to the relevant variable during program execution. The ratio of BLUE artillery tubes to BLUE maneuver forces would be specified in a single value record.

All variables in a range record, with the exception of the record name, are FORTRAN type REAL. Most single value records contain REAL variables. The sole exception is NGRID, which is an integer.

Minimum and maximum values for each component of the Blue strategy vector are specified in one data record. The first entry in the record is the name of the variable. The minimum and maximum allowable values are the second and third entries, respectively. The number of points between the minimum and maximum values is the fourth entry.

Red velocity and Blue force size are assigned values in increasing order from a given minimum to a given maximum. The step size is the specified increment. Incremental variables are specified in the data set as a record with four entries. The first entry is the name of the variable; the second and third entries are the minimum and maximum allowable values; the fourth entry is the step size.

Single valued variables are specified by a record with two entries. The first entry is the variable name; the second is the single value to be assumed throughout the program.

2. Variables

The following is a list and short description of each of the VFM input files. Records (and variables) are presented in the same order as they appear in the file. The

FORTRAN type (real or integer) of each variable is specified after the name of the record in which it appears. Range and single valued records are identified as such.

FCODE	(real range) an interval of admissible values for the fraction of Blue forces deployed forward. Specified by lower and upper bounds, and number of interior points.
WCODE	(real range) an interval of admissible values for the fraction of Blue forces that are withdrawn whenever Red overruns a Blue defensive line. Specified by lower and upper bounds and number of interior points.
NL0	(real range) an interval of admissible values for the number of defensive lines along which Blue is initially deployed. Specified by lower and upper bounds and number of interior points.
FCA	(real range) an interval of admissible values for the fraction of Blue forces that will be used for counterattack missions. Specified by lower and upper bounds and number of interior points.
VEL	(real range) the interval and step size of possible Red velocities. Specified by lower and upper bounds and step size.
REP	(real range) the set of Blue force sizes for which minimax computations are to be performed. Specified by a minimum, maximum, and step size.
FORCE	(real single) the theater ratio of Red to Blue maneuver forces.
BETA	(real single) value of BETA (see Appendix C).
FBI	(real) fraction of Blue maneuver forces that are infantry.
FRI	(real)fraction of Red maneuver forces that are infantry.
IHAT1B	(real) Blue short range ACM contribution (Red AFVEs killed per blue artillery tube per assault).
IHAT2B	(real) Blue CAS contribution (Red AFVEs killed per assault).
IHAT1R	(real) Red short range ACM contribution (Blue AFVEs killed per red artillery tube per assault).
IHAT2R	(real) Red CAS contribution (Blue AFVEs killed per assault).
IHAT3	(real) Blue BAI/Long Range ACM contribution (Red AFVEs killed per hour).
IHAT4	(real) Red BAI/Long Range ACM contribution (Blue AFVEs killed per hour).
HATFAC	(real) variable indicating whether or not IHAT variables will be scaled according to force size (0.0 implies no scaling, 1.0 implies scaling).
TDSTB	(real) delay time for Blue reserve movement (hours).
TDSTR	(real) delay time for Red reserve movement (hours).
BRAT	(real) ratio of blue artillery (in tubes) to blue maneuver forces (in AFVEs).
RRAT	(real) ratio of red artillery (in tubes) to red maneuver forces (in AFVEs).
LAMA	(real) default width of Red assault frontage: used if zero value input for LAMFAC (kilometers).
LAMCA	(real) default width of Blue counterattack frontage: used if zero value input for LAMFAC (kilometers).
DROMDT	(real) rate at which Red forces arrive at Red 'omega', or final, line (AFVEs per hour).
DENSMAX	(real) maximum allowable Red force density in main attack (AFVEs per kilometer).

DNSMXCA	(real) maximum allowable Blue force density in counterattack (AFVEs per kilometer).
LLOC	(real) default width of Red resupply corridor: used if zero value input for LAMFAC (kilometers).
NECA	(real) number of Blue echelons required to overrun a Red defensive line during a Blue counterattack operation.
VCA	(real) velocity of Blue counterattack (kilometers per hour).
TDPREP	(real) defensive preparation time (Red and Blue, hours).
TOPREP	(real) offensive preparation time (Red and Blue, hours).
DASY	(real) distance between assembly areas for successive echelons (kilometers).
DL	(real) spacing between defensive lines (same value for Blue and Red) (kilometers).
KPIN	(real) parameter related to 'DENSMIN' (see Appendix C).
KDEF	(real)parameter related to 'DENSMIN' (see Appendix C).
KA	ratio of ACM lethality vs stationary targets to ACM lethality vs moving targets
RMIN	(real) residual force level at which a given Red echelon breaks off assault (AFVEs per kilometer).
VR	(real) speed of reserve forces (kilometers per hour).
LAMFAC	(real) scaling coefficient for LAMA, the width of the Red assault.
FINEDIV	(real) number of interior points in iterations after the first.
NGRID	(integer) number of iterations.

3. Sample Input File

The following is a sample file containing all variables and variable names as they would appear in an actual input file. Typically, a user would modify such a file to fit specific needs and interests. All records begin with the name of the variable in question. Single valued records have one number (real or integer) per record. Records for Blue strategy vector components have the minimum and maximum values and the number of points between them at which VFM will calculate ground gain in the first iteration. Records for incremental range variables list the minimum, maximum and increment.

```
'FCODE'      0.01      0.99     10.0
'WCODE'      0.01      0.99     10.0
'NLO'        1.00     10.10     10.0
'FCA'        0.01      0.99     10.0
'VEL'        0.10     10.10      0.25
'REP'      2500.0   50000.0   2500.0
'FINEDIV'    4.0
'NGRID'      4.0
'DUMP'       1.0
'FORCE'      1.00
```

```
'BETA'          1.5
'FBI'           0.6
'FRI'           0.6
'THAT1B'        0.0
'THAT2B'        8.0
'THAT1R'        0.0
'THAT2R'        0.0
'THAT3R'       60.0
'THAT4R'       25.0
'HATFAC'        1.0
'TDSTB'         4.0
'TDSTR'         6.0
'BRAT'          0.30
'RRAT'          0.30
'LAMA'          0.0
'LAMCA'         0.0
'DROMDT'       65.0
'DENSMAX'      15.0
'DNSMXCA'      15.0
'LLOC'         12.0
'NECA'          2.0
'VCA'           5.0
'LAMTH'       850.0
'TDPREP'        6.0
'TOPREP'        2.0
'DASY'         22.5
'DL'            5.0
'KPIN'          0.5
'KDEF'          0.5
'Ka'            0.5
'RMIN'          6.0
'VR'           10.0
'LAMFAC'    0.00045
'FINEDIV'       4.0
'NGRID'           3
```

4. Output

Output from VFM is directed to both the monitor and a user-specified file. No graphical functions are performed. The VFM model writes one standard record for each Blue force size. This record consists of the Blue force size, Red ground gain, minimax Blue strategy and optimal Red velocity. Recall that Blue strategy is a vector whose components are the Blue forward fraction, the Blue counterattack fraction, the initial number of lines along which Blue deploys forces and the Blue withdrawal rate. If the variable 'DUMP' is set equal to 1.0, VFM produces an expanded record of the intermediate quantities used to calculate Red ground gain and the optimum Blue strategy.

5. Running the Model

There are two methods of running VFM: interactively or in batch mode. To run interactively, the user simply types the word "vfm" (without quotes), then strikes the "enter" key. VFM will respond by querying the user for the name of an input file. The user must respond by typing the name (in single quotes) of an existing file. Next, VFM will query the user for the name of an output file. The user must respond with a *new* file name (again, in single quotes). If a file already exists with the name supplied for the output file, the program will abort. (File names, including suffix, should not exceed twelve characters).

The following is a typical sequence of user prompts and responses. Prompts are upper case and responses are lower.

```
vfm
ENTER NAME OF INPUT FILE
'inp1.dat'
ENTER NAME OF OUTPUT FILE
'out1.dat'
```

If all file names are acceptable to VFM, the program will run to completion and create an output file with the specified name. No further prompts will be sent to the user.

To run in batch mode, the user submits a "BAT" file (i.e., batch file) that essentially contains all the information that would be entered directly in interactive mode. Typically, the batch file contains a single line with the name of the program to be executed (in this case "vfm") and the name of a data file containing the responses to program prompts.

For example, a sample batch file might contain the single line

```
vfm<pipe.dat
```

where "pipe.dat" is the file containing the names of the VFM input and output files. Specifically, "pipe.dat" might look like this:

```
'inp1.dat'
'out1.dat'
```

Again, if all file names are acceptable to VFM, the program will run to completion and create an output file with the specified name.

6. Code Modification

The VFM code can be modified easily to meet user needs. All source code is writen in Microsoft FORTRAN 5.0.

C. PROGRAM LISTING

```
C       ---------------------------------------------------
C       THIS FORTRAN PROGRAM IS AN IMPLEMENTATION OF THE IDA
C       VARIABLE FORCE EMPLOYMENT MODEL. IN MOST CASES, THE
C       VARIABLE NAMES USED IN THIS PROGRAM ARE CLOSE OR
C       IDENTICAL TO THE NAMES USED IN APPENDIX C. THE FEW
C       EXCEPTIONS ARE THESE:
C
C         CODED 'F' CORRESPONDS TO 'PHIF' IN THE REPORT.
C         CODED 'FCA' CORRESPONDS TO 'PHICA'.
C         CODED 'BRKTH' CORRESPONDS TO 'D'.
C
C
C     ------------------------------------------------------

      IMPLICIT REAL (A-Z)

      INTEGER JREP, MREP, HIT, ENUF
      INTEGER MIMXROW, ROW, TOOMNY, HALF, CYCLES
      INTEGER FINEG, FINEI, NGRID
      INTEGER * 4  NSEED, HOUR, MINUTE, SECOND, HNDRD
      DIMENSION F1(9), F2(9), FCA1(9), FCA2(9), NL01(9),
     1     NL02(9), W1(9), W2(9), FINC(9), FCAINC(9),
     1     NL0INC(9), WINC(9)
      LOGICAL PFLAG
      LOGICAL QFLG
C     ------------------------------------------------------
      COMMON /BLACK/
     1 FIRSTF,   LASTF,   FIRSTFCA,LASTFCA,
     2 FIRSTNL0, LASTNL0, FIRSTW,  LASTW,
     3 FIRSTV,   LASTV,   INCRV,
     4 FIRSTREP, LASTREP, INCREP

       COMMON /BLUE/
     5 FORCE, BETA,   BONN,   FBI,   FRI,
     6 DLTA1B,DLTA2B, DLTA1R, DLTA2R,DLTA3,DLTA4, HATFAC,
```

```
     7 TDSTB, TDSTR,  BRAT,   RRAT,  LAMA, LAMCA, DROMDT,
     8 LLOC,  DENSMAX,DNSMXCA,NECA,  VCA, LAMTH

      COMMON /YELLOW/
     1 TDPREP, TOPREP, DASY,DL,KPIN,KDEF, KA,RMIN, VR, LAMFAC

      COMMON /GREEN/
     1 DCA, HCA,HTERM, NEQ11,NEQ12,NEQ13,QCA,Q10, UBS2,URS2

      COMMON /PURPLE/
     1 FAHB, FAHR, FAMB, FAMR, FP, HLOMAX, HAVB, HELONS

      COMMON /RED/
     1 FDIV, FCADIV, NLODIV, WDIV, FINEDIV, NGRID, DUMP
C    -----------------------------------------------------------

      DATA K1, K3,K4, K5 /0.544, 14.8, 7.047, 2.51/
      DATA K6,    K8     /727.97, 0.086/

      DATA  ZERO /0.0/
      DATA KAE /  6.0/
C    ------------------------------------------------------------
      CALL INPUT (NEAR, ENUF, TOOMNY)
C    ------------------------------------------------------------
      HATFAC= 1.0 - HATFAC
      FIN    = FBI + FRI
      Q2     = K8 * DL
      NEQ11 = BETA*K3*FIN
      NEQ12 = K4/FIN
      RMAX  = 2.0*DENSMAX
C    ------------------------------------------------------------
      MREP = (LASTREP-FIRSTREP)/INCREP + 1
      REP = FIRSTREP

C    ------------------------------------------------------------
      DO 700 JREP = 1, MREP
      WRITE (*,1004) 'G','F','FCA','NLO','W','V'

      MIMXG = 5.0E6
C    ------------------------------------------------------------
      B = REP
      R = FORCE * B

      IF(HATFAC .LT. 0.999) THEN
         HATFAC =   B/60000.0
      ELSE
```

```
         HATFAC = 1.0
      END IF
      IF(LAMFAC.GT.0) THEN
         LAMA = 20.0 + LAMFAC*R
         LLOC = LAMA/3.0
      ELSE
      END IF

      NS     = LAMTH/LAMA
      RE     = DENSMAX * LAMA
      DCA  = 0.5*(LAMA-LLOC)
      BCACON= DL/VCA + NECA*(TOPREP + 0.5*NECA*DASY/VR)
      BCACON= DCA*(BCACON)/DL
C    -------------------------------------------------------------
C    LOOPS FOR BLUE FORCE EMPLOYMENT OPTIMIZATION:
C     F
C      FCA
C       NL0
C        W
C
C       DETERMINE SIZES OF VARIABLE INCREMENTS AND
C       INITIALIZE VARIABLES:
C
      FINEG=1
      FINC(1)=(LASTF-FIRSTF)/FDIV
      FCAINC(1)=(LASTFCA-FIRSTFCA)/FCADIV
      NL0INC(1)=(LASTNL0-FIRSTNL0)/NL0DIV
      WINC(1)=(LASTW-FIRSTW)/WDIV
      F1(1)=FIRSTF
      F2(1)=LASTF
      FCA1(1)=FIRSTFCA
      FCA2(1)=LASTFCA
      NL01(1)=FIRSTNL0
      NL02(1)=LASTNL0
      W1(1)=FIRSTW
      W2(1)=LASTW
C
C     RESET VARIABLES AND INCREMENTS FOR FINE GRIDS
  900 IF(FINEG.GT.1)THEN
        FINEI=FINEG-1
        F1(FINEG)=MIMXF-FINC(FINEI)
         IF(F1(FINEG).LE.FIRSTF)F1(FINEG)=FIRSTF
        F2(FINEG)=MIMXF+FINC(FINEI)
         IF(F2(FINEG).GE.LASTF)F2(FINEG)=LASTF
        FCA1(FINEG)=MIMXFCA-FCAINC(FINEI)
         IF(FCA1(FINEG).LE.FIRSTFCA)FCA1(FINEG)=FIRSTFCA
```

```
        FCA2(FINEG)=MIMXFCA+FCAINC(FINEI)
         IF(FCA2(FINEG).GE.LASTFCA)FCA2(FINEG)=LASTFCA
        NL01(FINEG)=MIMXNL0-NL0INC(FINEI)
         IF(NL01(FINEG).LE.FIRSTNL0)NL01(FINEG)=FIRSTNL0
        NL02(FINEG)=MIMXNL0+NL0INC(FINEI)
         IF(NL02(FINEG).GE.LASTNL0)NL02(FINEG)=LASTNL0
        W1(FINEG)=MIMXW-WINC(FINEI)
         IF(W1(FINEG).LE.FIRSTW)W1(FINEG)=FIRSTW
        W2(FINEG)=MIMXW+WINC(FINEI)
         IF(W2(FINEG).GE.LASTW)W2(FINEG)=LASTW
        FINC(FINEG)=(F2(FINEG)-F1(FINEG))/FINEDIV
        FCAINC(FINEG)=(FCA2(FINEG)-FCA1(FINEG))/FINEDIV
        NL0INC(FINEG)=(NL02(FINEG)-NL01(FINEG))/FINEDIV
        WINC(FINEG)=(W2(FINEG)-W1(FINEG))/FINEDIV
      ENDIF
C       RUN LOOPS

C       F LOOP
      F=F1(FINEG)
  910 CONTINUE

C       FCA LOOP
      FCA=FCA1(FINEG)
  920 CONTINUE

C       NL0 LOOP
      NL0=NL01(FINEG)
  930 CONTINUE

C       W LOOP
      W=W1(FINEG)
  940 CONTINUE
      BL    = B*F/(NL0*NS)
      BR    = B*(1-F)
      BRBAR= BR*(1-FCA)
      BCAP = FCA*(0.5*BR - DLTA4*HATFAC*LAMTH/VR)
      DBCADT = FCA*(VR*BR/LAMTH - DLTA4*HATFAC)
      BCAP   = AMAX1(BCAP, 0.0)
      DBCADT= AMAX1(DBCADT, 0.0)

      RAE = RE * RRAT * KAE
      BAL = BL * BRAT * NL0

      BHATAL = 2.0 * BAL/LAMA
      BHATL  = 2.0 * BL/LAMA
      RHATAE = 2.0 * RAE/LAMA
```

```
      NEQ13 = K1 * BHATAL + K6 / RHATAE
      HTERM = K1 * RHATAE + K6/RHATAE
      UB    = DLTA1B * RHATAE + DLTA2B*HATFAC
      UBS   = DLTA1B * BHATAL + DLTA2B*HATFAC
      URS   = DLTA1R*RHATAE*KA+ DLTA2R
C     -------------------------------------------------------
      RMIN2 = 4.0* RMIN*RMIN
      RMAX2 = (RMAX- UBS)*(RMAX-UBS)

      DBDT =    (1.0-FCA)*(VR*BR/LAMTH - DLTA4*HATFAC)
      DBDT =     AMAX1(DBDT, 0.0)
      DBTRNS=    DBDT * (TDPREP + TDSTB)
      DENSMIN = (KPIN*B*F + KDEF*BR*FCA)/LAMTH
      RSTERM =   R - DENSMIN*(LAMTH - LAMA)
      RSTERM =   AMAX1(RSTERM, 0.0)
C
      ALPHA =    1.0 - W**K5
      WSURV  =   W * MAX(1.0 - URS/BL,  0.0)
      BRKTH =    DL*(WSURV - NL0)/(WSURV - 1.0)
      IF(RMAX2 .GT. RMIN2 + 1.0E-2) THEN

          QRAT  = ALPHA*AMAX1(0.0,BHATL - URS)/(RMAX2-RMIN2)
          NECCO = QRAT *(1.0-Q2)

          GMAX = 0.0
          ROW  = ROW + 1
          GLAST = -1.0
          V     = AMAX1(FIRSTV,-1.0+SQRT( NEQ13/NEQ11))
C         "THIS IS THE V LOOP"

  100                 CONTINUE

          Q1    = (NEQ11 * V + NEQ12 + NEQ13/(1+V))
          NE    = Q1*(NECCO+SQRT(NECCO*NECCO + Q2*QRAT/Q1))
          CAV   = NE*LAMA*(DENSMAX-RMIN)

          H     =  NEQ11*V + NEQ12 + HTERM/(1+V) + UB
          H     =  0.5* H/DENSMAX

          DTDG  = 1.0/V + NE*(TOPREP + 0.5*NE*DASY/VR)/DL
          DGDT  = 1.0/DTDG

          BCA    = 0.5*DBCADT*(BRKTH*DTDG-BCACON)
          BCA    = AMAX1(BCA,0.0)
          BCA    = AMIN1(BCA, BCAP)
```

```
          IF(LAMFAC.GT.0)LAMCA = 10.0 + LAMFAC*BCA
C     -------------------------------------------------------
      ZBAR = 1.0/VCA
      YBAR = TOPREP/(DL*DNSMXCA*LAMCA)
      XBAR = 0.5*DASY/(VR*DL*DNSMXCA*DNSMXCA*LAMCA*LAMCA)

          CALL CASUAL    (ZERO,  DROMDT, K1, K6,
     3                    TDSTR, TDPREP,
     1                    XBAR,  YBAR, ZBAR,
     2                    LITA,  LITB, LITC)
          RLFAC   = AMAX1((2.0*DNSMXCA -URS2)/Q10 , 0.0)
C     -------------------------------------------------------

          CTERM   = LITC - BCA
          CASCA   = QUAD (LITA, LITB, CTERM, QFLG)
          CASCA   = AMAX1(CASCA, 0.0)

C         IF QFLG = .FALSE., THEN QUAD = -1.

          DTDGCA = XBAR* CASCA*CASCA + YBAR * CASCA + ZBAR
          TCA = DCA*DTDGCA
          REDRES = DROMDT*AMAX1(TCA -TDSTR - TDPREP, 0.0)
          REDRES = AMIN1(REDRES, RSTERM)
          RS = RSTERM - REDRES
C         THIS IS RS(0)

          IF(  TCA .LT. TDSTR + TDPREP) THEN
                  CASCA = BCA - 0.5*HCA*LAMCA*UBS2
                  CASCA = DL*CASCA/(DCA + DL*HCA*RLFAC)
                  CASCA = AMAX1(CASCA, 0.0)

          ELSE
          END IF

          DENSFLK =   FLANK(CASCA, LAMCA, RLFAC, RL)
          DRDG    = -(CAV/DL+ 2.0*DENSFLK + DLTA3*DTDG)
          DRDT    =   DRDG * DGDT

          TOMEGA  =   TSTAR (BL, BRBAR, DBDT, DBTRNS, DRDT,
     1                       H,  RS,    TDPREP, TDSTB)

          GRND = DGDT * AMAX1(TOMEGA,0.0)
C          CHECK FOR BREAKTHROUGH; INCREASE W IF NEEDED
          IF (GRND .GT. BRKTH) GO TO 951
          IF(GRND  .GE. GLAST)THEN
```

```
            IF (GRND .GT. GMAX) THEN
               MAXIV= V
               GMAX = GRND
               MIMXNE=NE
               MIMXDF=DENSFLK
               MIMXBC=BCA
               MIMXT=TOMEGA
               MIMXDT=DTDG
               MIMXD=BRKTH
               MXDIFF=DBDT*(TOMEGA-TDPREP-TDSTB)
               MXBAR=BRBAR
               MIMXDB=DBDT
               MIMXLA=LAMA
               MIMXTC=TCA
               MIMXC=CAV
               MIMXRS=RS
            END IF
          GLAST1=GLAST
          GLAST = GRND
          END IF
          IF(.NOT.((GRND.LT.GLAST1).AND.(V.GT.2.0)))THEN
            IF(V .LT. 1.0)THEN
              DELV=INCRV
            ELSE
              DELV=INCRV*V
            ENDIF
            V = V + DELV
            IF(V .LT. LASTV)GOTO 100
          ENDIF
C   AT THIS POINT, RED HAS CALCULATED THE OPTIMAL V FOR
C   MAXIMUM GROUND GAIN FOR THE CURRENT SET OF BLUE
C   PARAMETERS (F,NL0, ETC).
            IF (GMAX .LT. MIMXG) THEN
               MIMXG = GMAX
               MIMXV = MAXIV
               MIMXROW = ROW
               MIMXFCA = FCA
               MIMXW   = W
               MIMXNL0 = NL0
               MIMXF   = F
       WRITE (*,1007)MIMXG, F,FCA,NL0,W, MIMXV
               MNE=MIMXNE
               MDF=MIMXDF
               MBC=MIMXBC
               MT=MIMXT
               MDT=MIMXDT
```

```
                    MD=MIMXD
                    MDIFF=MXDIFF
                    MBAR=MXBAR
                    MDB=MIMXDB
                    MLA=MIMXLA
                    MTCA=MIMXTC
                    MC=MIMXC
                    MRS=MIMXRS
                  END IF

             ELSE
             WRITE (*,*) 'RMAX^2 < RMIN^2'
             WRITE (1,*) 'RMAX^2 < RMIN^2'
             END IF
C  END OF W LOOP
  951 W=W+WINC(FINEG)
      IF(W.GT.W2(FINEG))GO TO 941
      GO TO 940
C  END OF NL0 LOOP
  941 NL0=NL0+NL0INC(FINEG)
      IF(NL0.GT.NL02(FINEG))GO TO 931
      GO TO 930
C  END OF FCA LOOP
  931 FCA=FCA+FCAINC(FINEG)
      IF(FCA.GT.FCA2(FINEG))GO TO 921
      GO TO 920
C  END OF F LOOP
  921 F=F+FINC(FINEG)
      IF(F.GT.F2(FINEG))GO TO 911
      GO TO 910
  911 IF(FINEG.LT.NGRID)THEN
        FINEG=FINEG+1
        GO TO 900
      ELSE
C         FORCE OPTIMIZATION LOOPS ARE FINISHED
C           WRITE MINIMAX VALUES TO THE OUTPUT FILE
              WRITE (*, 1001) ' GROUND ', MIMXG
              WRITE (1, 1003) REP , MIMXG, MIMXF,
     1        MIMXNL0,  MIMXV,  MIMXW,  MIMXFCA
        IF(DUMP.EQ.1.0)THEN
          WRITE(1,*)'NE',MNE
          WRITE(1,*)'RHO FL',MDF
          WRITE(1,*)'BCA',MBC
          WRITE(1,*)'T*',MT
          WRITE(1,*)'DT/DG',MDT
          WRITE(1,*)'D',MD
```

```
            WRITE(1,*)'DB/DT*T',MDIFF
            WRITE(1,*)'BR*(1-FCA)',MBAR
            WRITE(1,*)'DBDT',MDB
            WRITE(1,*)'LAMA',MLA
            WRITE(1,*)'TCA',MTCA
            WRITE(1,*)'CAV',MC
            WRITE(1,*)'RS(0)',MRS
          ENDIF
        ENDIF
C  END OF REP LOOP
          REP = REP + INCREP
  700 CONTINUE

 1000  FORMAT (A9, 1X, 3F10.3)
 1001  FORMAT (  1X,A9,F10.0)
 1002  FORMAT (2(1X,A9,E10.3))
 1003  FORMAT (  1X, F8.0,',', F12.1, 5(',',F9.5))
 1004  FORMAT (7(6X,A4)/)
 1006  FORMAT (7I5)
 1007  FORMAT (7F10.3)
      STOP
      END

C    -------------------------------------------------------
      SUBROUTINE CASUAL (DBCADT, DROX, K1, K6, TRS, TRP,
     1                   XBAR,   YBAR, ZBAR,
     2                   BIGA,   BIGB, BIGC)

C     NOTE: USER MUST SUBTRACT 0.5*DBRKTH*DBCADT*DTDG FROM
C     THE VARIABLE BIGC BEFORE CALLING QUAD.

      REAL DBCADT,  DROX, K1, K6, TRS, TRP
      REAL XBAR, YBAR, ZBAR
      REAL BIGA, BIGB, BIGC
      REAL
     1 TDPREP, TOPREP, DASY,DL,KPIN,KDEF, KA,RMIN, VR, LAMFAC

       REAL
     5 FORCE, BETA, BONN, FBI, FRI,
     6 DLTA1B, DLTA2B,DLTA1R,DLTA2R, DLTA3R, DLTA4R, HATFAC,
     7 TDSTB, TDSTR, BRAT, RRAT, LAMA, LAMCA, DROMDT,
     8 LLOC, DENSMAX,DNSMXCA, NECA, VCA

      REAL EHAT,   HTERM,  TERM,   NEQ11, NEQ12, NEQ13
      REAL HCA,    KBA,    Q10,    QCA
      REAL BHATAE, RHATAL, UBS2,   URS2
```

```
C    -------------------------------------------------------
       COMMON /BLUE/
     5 FORCE, BETA, BONN,FBI, FRI,
     6 DLTA1B, DLTA2B, DLTA1R, DLTA2R,DLTA3R, DLTA4R,HATFAC,
     7 TDSTB, TDSTR, BRAT, RRAT, LAMA, LAMCA, DROMDT,
     8 LLOC, DENSMAX,DNSMXCA, NECA, VCA, LAMTH

       COMMON /YELLOW/
     1 TDPREP, TOPREP, DASY,DL,KPIN,KDEF, KA,RMIN, VR, LAMFAC

      COMMON /GREEN/
     1 EHAT, HCA,HTERM, NEQ11,NEQ12,NEQ13,QCA,Q10, UBS2,URS2
C    -------------------------------------------------------
      DATA RHATAL, KBA /50.0, 6.0/
C    -------------------------------------------------------
      HCA = NEQ11*VCA + NEQ12 + HTERM/(1+VCA)
      HCA = HCA + DLTA1R * BHATAE + DLTA2R
      HCA = 0.5*HCA/DNSMXCA

      BHATAE = 2.0*DNSMXCA *BRAT *KBA

      Q10 = K1*RHATAL + K6/BHATAE
      Q10 = Q10/(1.0 + VCA) + NEQ11 * VCA/BETA + NEQ12

      UBS2 = DLTA1B * BHATAE * KA  + DLTA2B*HATFAC
      URS2 = DLTA1R * RHATAL + DLTA2R

      QCA  = HCA*DROX*(TRS + TRP)
      TERM = HCA*DROX + 0.5*DBCADT
      BIGA = EHAT*XBAR*TERM
      BIGB = EHAT*YBAR*TERM + EHAT/DL
      BIGB = BIGB + HCA*AMAX1(2.0*DNSMXCA-URS2, 0.0)/Q10
      BIGC = EHAT*ZBAR*TERM + 0.5*HCA*LAMCA*UBS2
      BIGC = BIGC - QCA
      BIGC = BIGC - 0.5* DLTA1R*RHATAL*LAMCA*EHAT/DL
      BIGC = BIGC - 0.5* DLTA2R*LAMCA*EHAT/DL
      RETURN
      END
C    -------------------------------------------------------
      REAL FUNCTION QUAD (A,B,C, FLAG)
      REAL A,B,C, DISC, EPSILON
      LOGICAL FLAG
      DATA EPSILON /1.0E-6/

      DISC = B*B - 4.0 * A * C
      IF (DISC .LT. 0.0) THEN
```

```
      FLAG = .FALSE.
      ELSE

          IF (ABS(A). GT. EPSILON) THEN
          QUAD = 0.5*(-B + SQRT(DISC) )/A
          FLAG = .TRUE.
          ELSE
               IF(ABS (B) .GT. EPSILON)THEN
               QUAD = -C/B
               FLAG = .TRUE.
               ELSE
                  FLAG = .FALSE.
               END IF

          END IF

      END IF
      IF(.NOT.FLAG) QUAD = -1.0
      RETURN
      END

C     -------------------------------------------------------
       REAL FUNCTION FLANK (CASCA, LAMCA, RLFAC, RL)

       REAL
     1 TDPREP, TOPREP, DASY, DL, KPIN, KDEF, RMIN, VR, ILAM
       REAL
     1 DCA, HCA,HTERM, NEQ11,NEQ12,NEQ13,QCA,Q10, UBS2,URS2
       REAL CASCA, LAMCA, RLFAC, RL

       COMMON /YELLOW/
     1 TDPREP, TOPREP, DASY, DL, KPIN, KDEF, RMIN, VR, ILAM
       COMMON /GREEN/
     1 DCA, HCA,HTERM, NEQ11,NEQ12,NEQ13,QCA,Q10, UBS2,URS2

       RL = AMAX1(CASCA, 0.0)*RLFAC
       RL = RL + 0.5*LAMCA*UBS2
       RL = AMAX1(RL, 0.0)

       FLANK  = RL*(1.0+DCA/DL)/LAMCA
       RETURN
       END
C     ------------------------------------------------------
       REAL FUNCTION TSTAR (BL, BRBAR, DBDT, DBTRNS, DRDT,
     1                          H, RS, TDPREP, TDSTB)
       REAL BL, BRBAR, DBDT, DBTRNS, DRDT,H,RS,TDPREP, TDSTB
```

```
          REAL T1, T2, TOMEGA
          LOGICAL TFLG

                T1 =  DBTRNS - BL
             T1 =  (RS + H * T1)/(H*DBDT - DRDT)

             IF(DBDT*T1- DBTRNS .LT. BRBAR) THEN
                      IF(T1.GT. TDPREP + TDSTB)THEN
                      TOMEGA = T1
                      ELSE
                      TOMEGA = (H*BL - RS)/DRDT
                      END IF
                      TFLG = .FALSE.
             ELSE
                      T2 = H* (BL + BRBAR) - RS
                      T2 = T2/DRDT
                      TOMEGA = T2
                      TFLG = .TRUE.
             END IF
             TSTAR = TOMEGA
             RETURN
             END


         SUBROUTINE INPUT (NEAR, ENUF, TOOMNY)
         IMPLICIT REAL (A-Z)
         CHARACTER * 9 FINAME, FONAME, CODE

         INTEGER  ENUF, TOOMNY
         INTEGER NGRID
         REAL LAMFAC
C       -------------------------------------------------------
          COMMON /BLACK/
        1 FIRSTF, LASTF,FIRSTFCA,LASTFCA,
        2 FIRSTNL0, LASTNL0, FIRSTW, LASTW,
        3 FIRSTV,   LASTV, INCRV,
        4 FIRSTREP, LASTREP, INCREP

          COMMON /BLUE/
        5 FORCE, BETA, BONN,FBI, FRI,
        6 DLTA1B, DLTA2B, DLTA1R, DLTA2R, DLTA3, DLTA4, HATFAC,
        7 TDSTB, TDSTR, BRAT, RRAT, LAMA, LAMCA, DROMDT,
        8 LLOC, DENSMAX,DNSMXCA,  NECA, VCA, LAMTH

          COMMON /YELLOW/
        1 TDPREP, TOPREP, DASY,DL,KPIN,KDEF, KA,RMIN, VR, LAMFAC
```

```
      COMMON /RED/
     1 FDIV, FCADIV, NLODIV, WDIV, FINEDIV, NGRID, DUMP
C     ---------------------------------------------------------

      WRITE (*,*) ' TYPE NAME OF INPUT FILE '
      READ  (*,*)   FINAME
      WRITE (*,*) ' TYPE NAME OF OUTPUT FILE '
      READ  (*,*)   FONAME
      OPEN (UNIT=31,FILE=FINAME, STATUS='OLD')
      OPEN (UNIT=1, FILE=FONAME, STATUS='NEW')
C     ---------------------------------------------------------
      READ (31, *) CODE, FIRSTF, LASTF, FDIV
      READ (31, *) CODE, FIRSTW, LASTW, WDIV
      READ (31, *) CODE, FIRSTNLO, LASTNLO, NLODIV
      READ (31, *) CODE, FIRSTFCA, LASTFCA, FCADIV
      READ (31, *) CODE, FIRSTV, LASTV, INCRV
      READ (31, *) CODE, FIRSTREP, LASTREP, INCREP
      READ (31, *) CODE, FINEDIV
      READ (31, *) CODE, NGRID
      READ (31, *) CODE, DUMP

      READ (31, *) CODE, FORCE
      READ (31, *) CODE, BETA
      READ (31, *) CODE, FBI
      READ (31, *) CODE, FRI

      READ (31, *) CODE, DLTA1B
      READ (31, *) CODE, DLTA2B
      READ (31, *) CODE, DLTA1R
      READ (31, *) CODE, DLTA2R
      READ (31, *) CODE, DLTA3
      READ (31, *) CODE, DLTA4

      READ (31, *) CODE, HATFAC
      READ (31, *) CODE, TDSTB
      READ (31, *) CODE, TDSTR

      READ (31, *) CODE, BRAT
      READ (31, *) CODE, RRAT

      READ (31, *) CODE, LAMA
      READ (31, *) CODE, LAMCA
      READ (31, *) CODE, DROMDT

      READ (31, *) CODE, DENSMAX
```

```
      READ (31, *) CODE, DNSMXCA
      READ (31, *) CODE, LLOC

      READ (31, *) CODE, NECA
      READ (31, *) CODE, VCA
      READ (31, *) CODE, LAMTH
      READ (31, *) CODE, TDPREP
      READ (31, *) CODE, TOPREP
      READ (31, *) CODE, DASY
      READ (31, *) CODE, DL

      READ (31, *) CODE, KPIN
      READ (31, *) CODE, KDEF
      READ (31, *) CODE, KA
      READ (31, *) CODE, RMIN
      READ (31, *) CODE, VR
      READ (31, *) CODE, LAMFAC

      WRITE (1,1004)' REP','GMIN','OPTF',' NLO','
VEL','OPTW', ' FCA'
      RETURN
 1004 FORMAT (7(6X,A4))
      END
```

Appendix F

BASE CASE DATA

Stephen D. Biddle

A. INTRODUCTION

This appendix lists and documents the VFM data file used to produce the base case results illustrated in the main body of the paper (for a listing of the VFM source code, see Appendix E). Inasmuch as the base case is intended to correspond roughly to the ground weapon mix and tacair balance that would obtain in Central Europe in the aftermath of a CFE agreement, the data given here are based where possible on Central European weapons effectiveness values and orders of battle, as these would be modified by the draft CFE treaty. Documentation is provided here for those data treated as a constant for the discussion in the paper. Endogenous force employment variables are given here as ranges and thus require no documentation, while values for exogenous independent variables treated in the main text are motivated in the text itself.

Values for the accuracy-of-approximation parameters are somewhat arbitrary; larger numbers of force employment increments, and higher values for FINEDIV or NGRID will always yield a closer approximation of the true optimum G (and the true optima for force employment choices) at the cost of increased run times. The values given here were intended to provide a very close approximation at the cost of very long run times.[1] A more typical run with 10 increment steps per blue force employment choice and a velocity step size of 0.25 would produce an execution time of about 45 minutes when run on a COMPAQ 386/25 with a standard math co-processor and would yield an approximation within about five to ten percent of the red net territorial gain reported here.

---

[1] The run described here, if executed on a single COMPAQ 386/25 with a standard math co-processor, would require almost 150 hours to complete. The results reported here were consequently obtained by dividing each such run among many PCs (e.g., by assigning the first PC REP values from 2500 to 5000; the second from 7500 to 10000, etc.). The result is an extremely close approximation of optimal net territorial gain, and the smooth curves shown in figures I-1 through I-2 and I-6 through I-8. Optimal force employment approximation accuracy is somewhat more sensitive than net territorial gain to increment size; the run described above produces smooth net territorial gains as a function of force to space ratios, but is not accurate enough to produce smooth curves for force employment optima as a function of force to space ratios. Consequently, an even closer approximation was used to produce figures I-3 through I-5, in which force employment increments exceeded 200 per .01-.99 force employment variable range. Similarly, these runs were produced by dividing each run among many PCs, and in this case, by conducting multiple runs for each REP, reducing in each successive run the range of variation for each force employment choice around the previously estimated optimum while holding the number of increments constant. The resulting net territorial gains, on the other hand, were for all REPs within two percent of those produced by the input file given above.

B. DATA FILE LISTING

```
'FCODE'          0.01       0.99      25.0
'WCODE'          0.01       0.99      25.0
'NL0'            1.0       10.1       25.0
'FCA'            0.01       0.99      25.0
'VEL'            0.10      10.1        0.05
'REP'         2500.0    50001.0     2500.0
'FINEDIV'        4.0
'NGRID'          4.0
'DUMP'           0.0
'FORCE'          1.00
'BETA'           1.5
'FBI' [2]        0.6
'FRI' [3]        0.6
'IHAT1B'         0.0
'IHAT2B' [4]     8.0
'IHAT1R'         0.0
'IHAT2R' [5]     0.0
```

---

[2] Assuming a post-CFE ratio of 30,000 armored troop carriers (with associated infantry) to 20,000 tanks for each side in the Central Region. See Alan Riding, "Arms Pact to Codify Europe's New Power Balance," The New York Times, 18 November 1990, pp.1ff.

[3] Ibid.

[4] Figure given is for West German, British, Belgian, Dutch, French and U.S. aircraft committed to NATO and either present in the theater in peacetime or immediately available upon mobilization, and excludes interceptor and recce aircraft. David G. Gray, IDA Unclassified Conventional Forces Database: Atlantic to the Urals, 1990, (Alexandria, VA: Institute for Defense Analyses, October, 1990) IDA D-708, pp.34-36 provides data for aircraft numbers; aircraft numbers were translated into AFVE kills per two kilometer front per engagement (as required for IHAT2B), on the basis of the following assumptions: that attack frontages are as computed by VFM, and that all NATO aircraft are used against the main attack sector; that NATO aircraft fly three sorties per day, suffer five percent losses and kill 0.5 red AFVEs per sortie (see Joshua M. Epstein, The 1988 Defense Budget (Wasington, D.C.: Brookings, 1987), p.44); that NATO ground attack aircraft are allocated evenly between close air support and battlefield air interdiction; that an average ground engagement lasts about five hours; and that the performance of the aircraft surviving after two days of combat constitutes a reasonable average value for the blue air contribution across the duration of a nominal theater offensive operation (as posited in VFM; see Appendix C).

[5] Only the SU-25 was assumed to be used for close air support; for SU-25 inventory, see Gray, op. cit., pp.37-41. Half of all Soviet SU-25s were assumed to be committed to the NATO Central region; it was further assumed that these would fly two sorties per day, suffer five percent attrition and kill .25 blue AFVEs per sortie (see Epstein, The 1988 Defense Budget op. cit., p.44); that attack frontages are as computed by VFM, and that all SU-25's are used in the main attack sector; that an average ground engagement lasts about five hours; and that the performance of the aircraft surviving after two days of combat constitutes a reasonable average value for the blue air contribution across the duration of a nominal theater offensive operation (as posited in VFM; see Appendix C). The result was a negligible contribution of 0.16 AFVE kills per two kilometers per engagement; this was rounded to an input value of zero.

'IHAT3R' [6]	60.0
'IHAT4R' [7]	25.0
'HATFAC' [8]	1.0
'TDSTB' [9]	4.0
'TDSTR' [10]	6.0
'BRAT' [11]	0.3
'RRAT' [12]	0.3
'LAMA' [13]	0.0
'LAMCA' [14]	0.0
'DROMDT' [15]	65.0

---

[6] Data sources and assumptions as per note 4 above; note that whereas IHAT2B is defined in terms of AFVE kills per two kilometers per engagement, IHAT3 is defined in terms of AFVE kills per hour (theaterwide). IHAT2B and IHAT3 are thus different values, although the underlying assumptions with respect to aircraft numbers and performance are the same here.

[7] See Gray, op. cit., pp.37-41. Figure given is for Soviet, East German, Polish, and Czech aircraft presumed committed against the NATO Central region and either present in the theater in peacetime or immediately available upon mobilization, and excludes interceptor, recce, and dedicated close air support aircraft (i.e., the SU-25). Performance assumptions are as per note 5 above; note that whereas IHAT2R is defined in terms of AFVE kills per two kilometers per engagement, IHAT4 is defined in terms of AFVE kills per hour (theaterwide).

[8] It is assumed in the base case that any ground force reductions are accompanied by corresponding reductions in tacair (there is no ACM in the base case force structures).

[9] See Statement of General Fred K. Mahaffey, Director, Requirements Office of the Deputy Chief of Staff for Operations and Plans in Department of Defense Authorization for Appropriations for Fiscal Year 1981, Hearings Before the Committee on Armed Services, United States Senate, Ninety Sixth Congress Second Session, Part 5 (Washington, D.C.: U.S. Government Printing Office, 1980), p.3030. Gen. Mahaffey estimates three hours of command and control time required from the moment a decision is reached by higher command to begin counterconcentration to the time a brigade or larger reserve formation could be given movement orders; in addition, we assume here that one hour is required for the theater commander to process the necessary data and make that decision. Thus, if the stimulus for action is the initiation of the Soviet attack, it follows that reserve units would receive movement orders four hours after the attack begins.

[10] It is assumed here that in the midst of an ongoing theater offensive, WTO command, control and decision time would be slightly longer for response to NATO counterattack than would be NATO's response to the initial WTO theater attack.

[11] Gray, op. cit., pp. 5-11, assuming theater parity.

[12] Ibid., pp.16-31.

[13] Constant frontage option was not used; frontages were instead computed via non-zero LAMFAC.

[14] Constant frontage option was not used; frontages were instead computed via non-zero LAMFAC.

[15] Assuming Soviet-style march column density, as per Headquarters, Department of the Army, FM 100-2-1, The Soviet Army: Operations and Tactics (Washington, D.C.: USGPO, 16 July 1984), p. 5-5; figure given assumes 32 AFVEs per nominal Soviet battalion, two routes available for moving reserves to the point of counterattack, and a reserve velocity of 10 kilometers per hour (see below). For a sensitivity analysis, see Appendix G.

'DENSMAX' [16]	15.0
'DNSMXCA' [17]	15.0
'LLOC' [18]	0.0
'NECA' [19]	2.0
'VCA' [20]	5.0
'LAMTH'	850.0
'TDPREP' [21]	6.0
'TOPREP' [22]	2.0
'DASY' [23]	22.5
'DL' [24]	5.0
'KPIN' [25]	0.5
'KDEF' [26]	0.5
'Ka' [27]	0.5

---

[16] Ibid., pp.5-11 and 5-12, gives a range of 12.5 to 17.5 AFVEs per kilometer for the Soviet Army; Headquarters, Department of the Army, FM 71-1, The Tank and Mechanized Infantry Company Team (Washington, D.C.: USGPO, November 1988), p.3-11, implies a range of 11 to 17.5 AFVEs per kilometer for the U.S. Army; the figure given is intended as a rough mean within these bounds.

[17] See note 16 above.

[18] Constant LOC width option was not used; LOC width was instead computed via non-zero LAMFAC (i.e., as one-third the computed attack frontage).

[19] Estimate; based on observation of VFM-computed optimal ne for theater attacker (for sensitivity, see Appendix G).

[20] Estimate; based on observation of VFM-computed optimal V for theater attacker (for sensitivity, see Appendix G).

[21] See Mahaffey, op. cit., p.3030, which estimates three hours to prepare the reinforcing formation for movement (after receipt of orders), and three hours to prepare defensive positions after arrival (for sensitivity, see Appendix G).

[22] FM 100-2-1, op. cit., p.5-14 estimates one to three hours' reaction time for a Soviet regimental assault echelon. The figure given is intended as a rough mean within these bounds; NATO single-assault wave response time is assumed to be roughly similar (for sensitivity, see Appendix G).

[23] Ibid., p.5-18, estimates an inter-echelon separation distance of 15-30 kilometers. The figure given is intended as a rough mean within these bounds; NATO separation distance is assumed to be roughly similar (for sensitivity, see Appendix G).

[24] Ibid., p.6-7, estimates a WTO first echelon battalion defensive position of two kilometers' depth, separated from the second echelon position by two kilometers, and with greater separation distances between larger formations. The figure given is intended as a rough overall theater average; NATO spatial distribution is assumed to be roughly similar (for sensitivity, see Appendix G).

[25] Estimated; see Appendix C, note accompanying equation 65 (for sensitivity, see Appendix G).

[26] Estimated; see Appendix C, note accompanying equation 65 (for sensitivity, see Appendix G).

[27] Estimated on the basis of Stephen D. Biddle, How to Think About Conventional Nuclear Substitution (Alexandria, VA: Institute for Defense Analyses, May 1986), IDA P-1884, Volume I, pp.10-15 (for sensitivity, see Appendix G).

'RMIN'[28]	6.0
'VR'[29]	10.0
'LAMFAC'[30]	0.00045

---

[28] Figure given corresponds to 40 percent of a DENSMAX value of 15 (see note 16 above). For estimates of percentage losses at which assault forces become ineffective, see Headquarters, Department of the Army, Soviet Army Operations (Washington, D.C.: USGPO, April 1978), IAG-13-U-78, p.5-6; also FM 100-2-1, op. cit., p.8-1 (for sensitivity, see Appendix G).

[29] See Marshall Hoyler, "Notes on Reserve Movement," Institute for Defense Analyses, unpublished manuscript (for sensitivity, see Appendix G).

[30] Corresponds to a total main effort attack frontage for a putative three-front, 88,000 AFVE offensive of 60 kilometers, given k0 = 20 kilometers (see Appendix C, equation 1). For frontages, see Michael Sadykiewicz, Soviet-Warsaw Pact Western Theater of Military Operations: Organization and Missions (Santa Monica, CA: RAND, August 1987) RAND N-2596-AF, pp.39, 78; also John Hines, "The Operational Calculations for Equal Security Under Arms Control," Conference Paper presented at the "International Symposium on Conventional Stability in Europe: Prerequisites and Analysis Requirements," 10-13 October, German Armed Forces University, Munich. For alternative estimates, see for example FM 100-2-1, op. cit., pp.4-2 to 4-6, and 5-18 to 5-20. For a sensitivity analysis, see Appendix G.

Appendix G

SENSITIVITY ANALYSES

Stephen D. Biddle

A. INTRODUCTION

Not all values required as input for the VFM model can be known with certainty. The values given in Appendix F represent best estimates, given data available in the literature, but in some cases the uncertainties associated with those best estimates can be substantial. To what degree do the conclusions developed in the main body of this paper depend on the particular best-estimate data values given in Appendix F? This appendix addresses this question by providing the results of a series of sensitivity analyses for those input variables not addressed explicitly in the main text.

B. DISCUSSION

For each input variable considered, a set of VFM runs were conducted for values of that variable representing the plausible range of uncertainty associated with that variable. To show the effects of variation most clearly, a higher theaterwide force to force ratio of 1.75:1 was employed as a base case;[1] sensitivity runs were conducted as univariate excursions from this base. Given the number of variables involved, interaction effects resulting from simultaneous changes in several variables were not explicitly considered. In effect, these sensitivities thus represent partial derivatives of net territorial gain (G) with respect to the variables considered.

The results of these runs are illustrated in Figures 1 through 12. In each figure, net territorial gain is plotted as a function of the force to space ratio and of the value of the parameter under study. The result is a surface showing how the relationship between force to space ratios and combat outcomes responds to changes in the value of the given parameter. Two such surfaces are depicted for each figure: a dark surface representing the base case control (for which the value of the parameter is held constant at its base case value), and a lighter surface showing the response of the combat outcome (G) as the parameter changes, other variable values held constant. The degree of divergence between the surfaces thus indicates the degree of sensitivity of the model to the parameter in question; the more nearly coincident the surfaces, the more nearly insensitive is the

---

[1] For a theaterwide force to force ratio of 1:1, many sensitivity runs produce degenerate net territorial gains of zero or near zero. A higher theaterwide force to force ratio provides a larger range of non-zero territorial gains, and is thus better suited to demonstrate the underlying sensitivities of the model.

model to the parameter. Surfaces were plotted using the IDA Response Surface Methodology from the Advanced Technology Combat Simulation Project.[2]

Overall, the analyses conducted show mostly modest sensitivity to variations in parameter values. In Figure 2, for example, a Red reserve arrival rate ($\psi_{ROM}(t_{CA})$) of about half its base case value of 65 AFVEs per hour produces net territorial gains (G) only about 15 percent below base at low force to space ratios, with results being less divergent elsewhere. Similarly, in Figure 3, a reduction in the mean depth of an assembly area (D_{ASY}) to a value less than half that of the base case produces a divergence in outcomes of about 15 percent at low force to space ratios, with less divergence elsewhere. More broadly, Figures 2, 3, 4, 7 and 9 show maximum divergences of less than 20 percent of the base case value for any point on the surface. Figures 1, 8, 10 and 11 show maximum divergences of between 20 and 50 percent. Figure 5 demonstrates substantial sensitivity to Red assault termination criteria above about 8 to 10 surviving AFVEs per kilometer, for force to space ratios above about 30,000 to 40,000 Blue AFVEs per 850 kilometers, but substantially smaller divergences elsewhere.[3]

Figures 6 and 12, however, represent exceptions. In Figure 6, reductions in reserve velocity below about 8 kilometers per hour can result in net territorial gains more than twice those of the base case for a wide range of force to space ratios. Increases in reserve velocity produce a smaller, but still non-trivial response relative to the base case. For all values of V_{RSV} examined, however, the relationship between the force to space ratio and net territorial gain remains continuous, relatively smooth, and relatively shallow for force densities above the G-maximizing level. The underlying phenomenon is thus

---

[2] For a detailed description, see Peter S. Brooks, et. al., The IDA Advanced Technology Combat Simulation Project, IDA P-2329, (Alexandria, VA: Institute for Defense Analyses, forthcoming).

[3] In effect, Red cannot gain ground without suffering casualties. If an assault is terminated soon enough (i.e., at a high enough value of R_{MIN}--and thus a low enough level of casualties, $\rho_{MAX} - R_{MIN}$), then Red gains little ground. In the limit, if $R_{MIN} = \rho_{MAX}$, then Red would be forced to terminate the engagement before being exposed to fire and could gain no ground at all. As Figure 5 shows, however, as long as this termination criterion is not set very close to Red's initial force size (i.e., ρ_{MAX}; equal to 15.0 for the base case), then G is relatively insensitive to its precise value. Although R_{MIN} is treated as a constant in this paper, it is properly regarded as a force employment option for the attacker; the decision as to when to call off an assault is doctrinally (or morally) rather than physically determined. Thus, if R_{MIN} values above a certain level posed insurmountable problems for Red, Red in theory could choose to press the assault harder. This potential force employment option has not been implemented in this version of VFM. Nevertheless, the value for R_{MIN} used in the base case puts Red in a relatively insensitive region of the response surface, as illustrated in Figure 5.

essentially the same, regardless of V_{RSV}—what changes is the absolute distance Red can advance at any given Blue force density.

In Figure 12, net territorial gain is shown to be highly sensitive to changes in K_{ACMC} (the ratio of ACM effectiveness against stationary targets to ACM effectiveness against moving targets) in the neighborhood of 0.20 to 0.30. To examine K_{ACMC}, however, it was necessary to introduce a further change relative to the base case. Since the base included no advanced conventional munitions, these sensitivities were run with both Red and Blue assumed to deploy short range ACM systems capable of killing 0.5 AFVEs per assault per traditional artillery tube available at the point of attack. Thus the excursions and the base are not comparable cases here in the sense that they are in Figures 1 through 11; the dark surface in Figure 12 is intended to provide a point of reference rather than to serve as a base for sensitivity assessment as elsewhere in this appendix.

Nevertheless, it is clear that variation in K_{ACMC} can have a major impact on net territorial gain. For a force to space ratio of about 25,000 Blue AFVEs per 850 kilometers, an increase in K_{ACMC} from 0.0 to 0.33 increases G by more than a factor of two. For lower force to space ratios, the result was to transform a defense equal or superior to that of the no-ACM base case into an offensive breakthrough (the data were truncated at a value of G = 200 to simplify presentation). As a result, we must conclude that recommendations with respect to the utility of short range ACM as a hedge against the effects of lower NATO force levels are substantially sensitive to assumptions as to ACM's relative effectiveness against moving and stationary targets.

The VFM model thus appears to be only moderately sensitive to variations in the parameters considered here, exceptions with respect to the role of target posture for ACM effectiveness and the velocity of reserve formations notwithstanding. In no case were the basic conclusions of the study found to be dependent on particular values of these parameters. The predicted net territorial gain is subject to change as a result of changes in uncertain inputs, but the relationship between net territorial gain and the force to space ratio remains fundamentally the same. Thus, while specific outcomes may vary, the central conclusions of the study are substantially robust with respect to uncertainty in input data values.

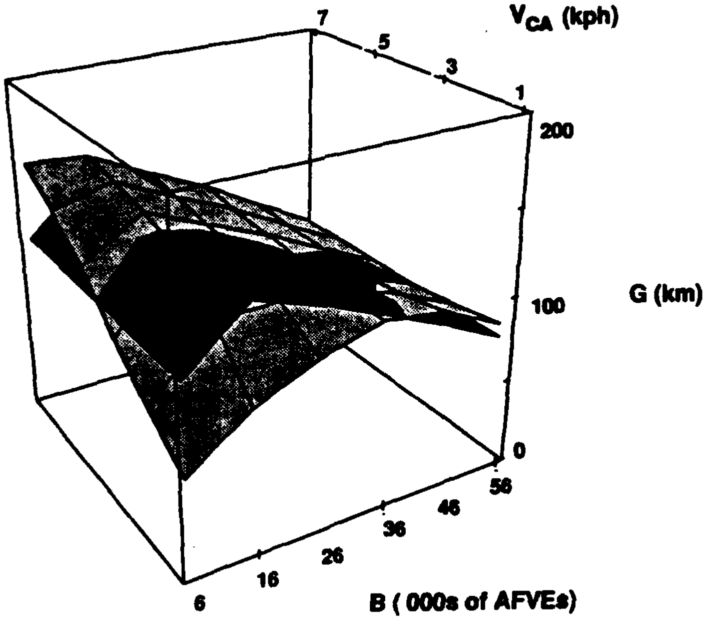

Figure G-1. Sensitivity of G to V_{CA}

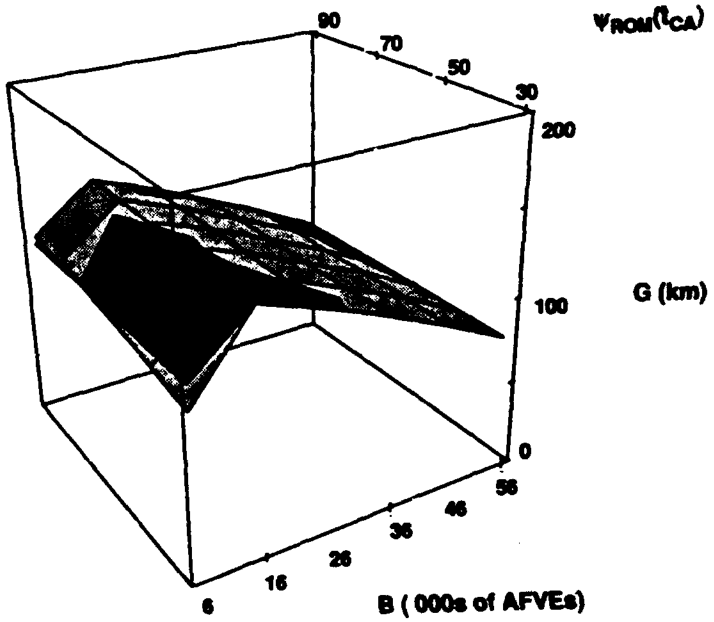

Figure G-2. Sensitivity of G to $\psi_{ROM}(t_{CA})$

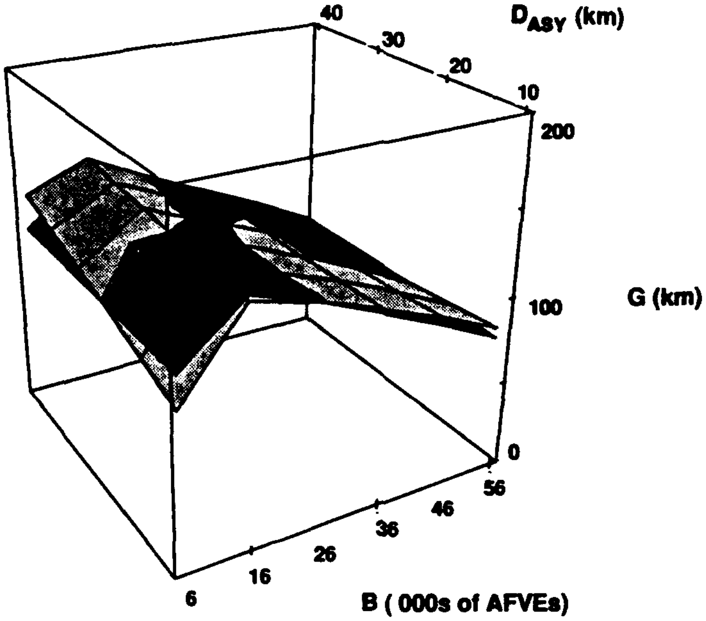

Figure G-3. Sensitivity of G to D_{ASY}

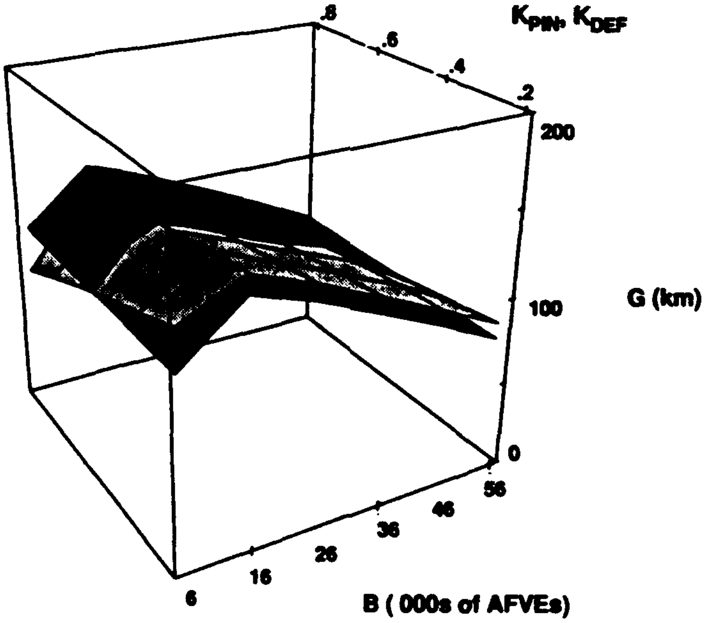

Figure G-4. Sensitivity of G to K_{PIN}, K_{DEF}

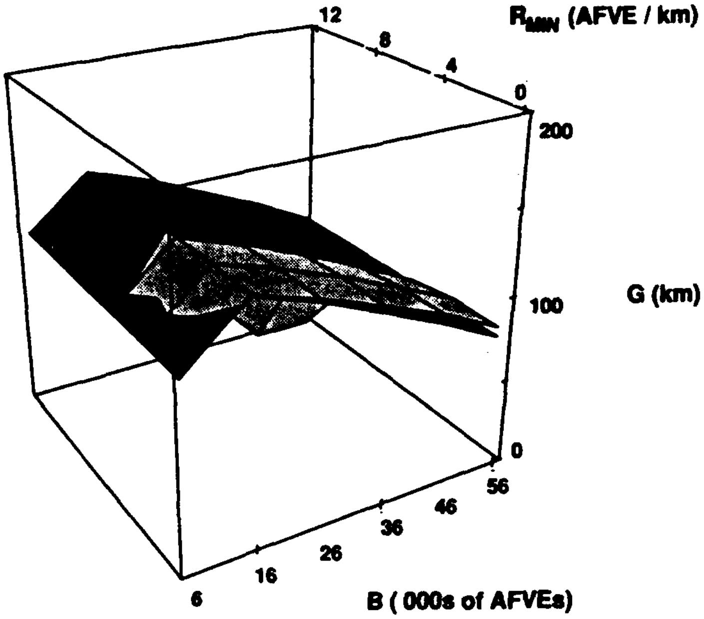

Figure G-5. Sensitivity of G to R_{MIN}

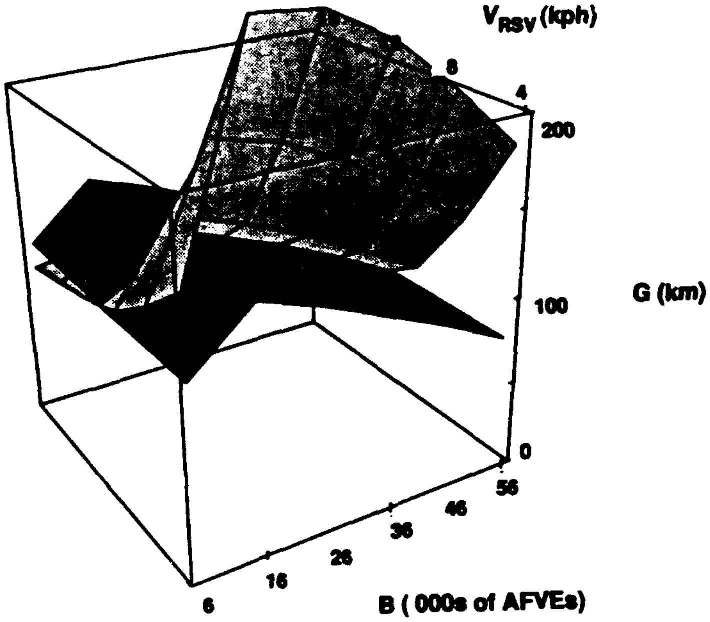

Figure G-6. Sensitivity of G to V_{RSV}

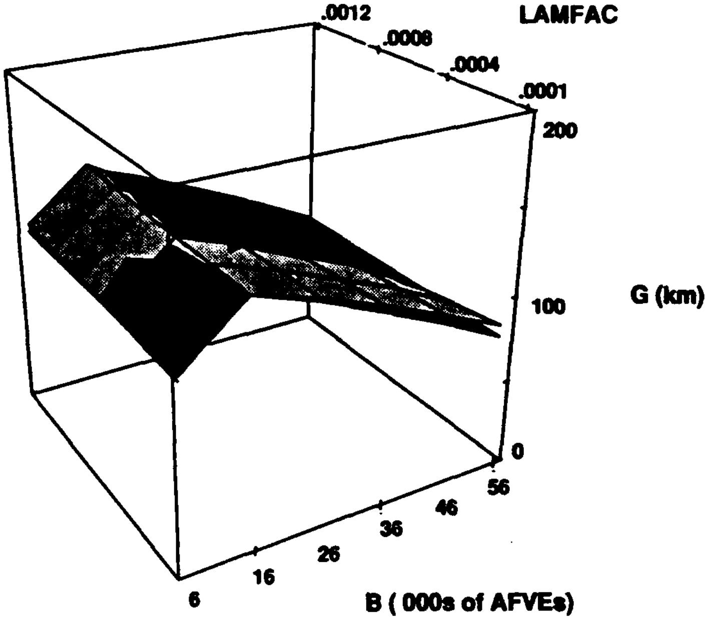

Figure G-7. Sensitivity of G to LAMFAC

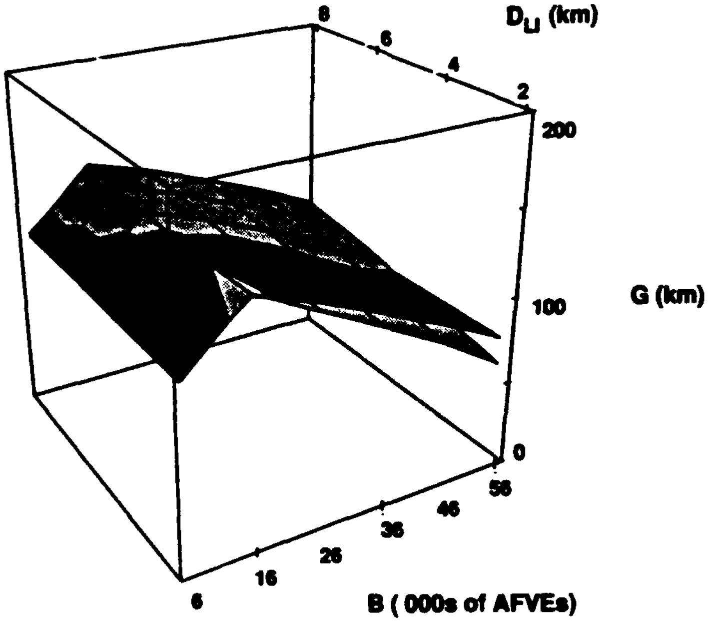

Figure G-8. Sensitivity of G to D_{LI}

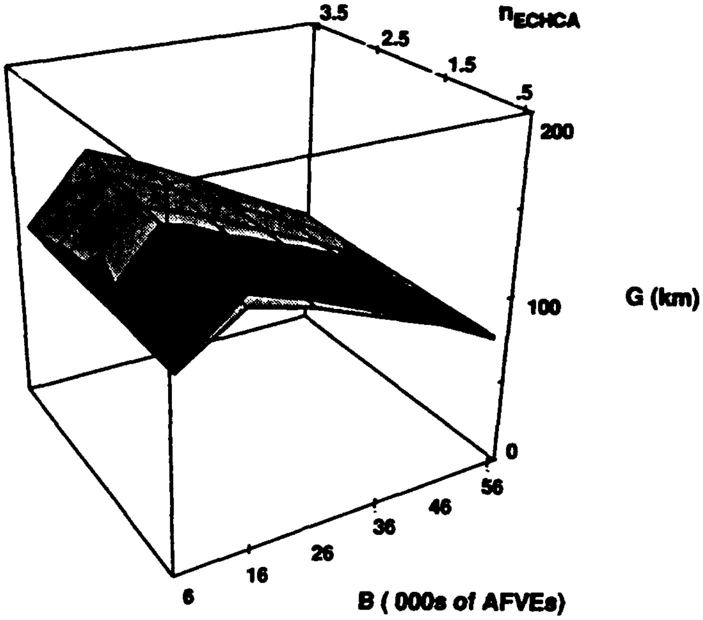

Figure G-9. Sensitivity of G to n_{ECHCA}

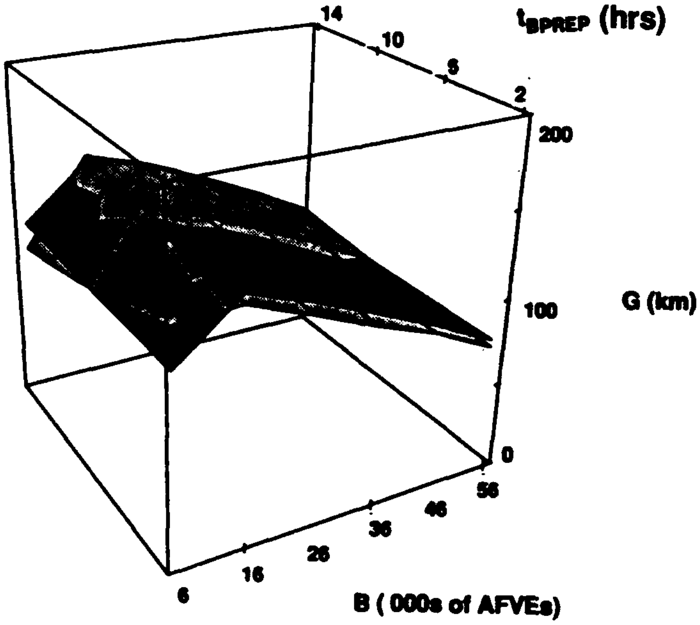

Figure G-10. Sensitivity of G to t_{BPREP}

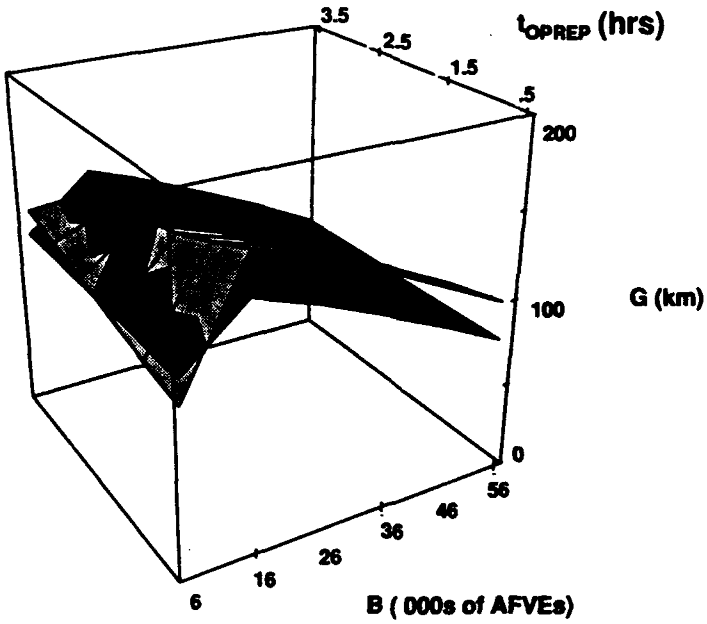

Figure G-11. Sensitivity of G to t_{OPREP}

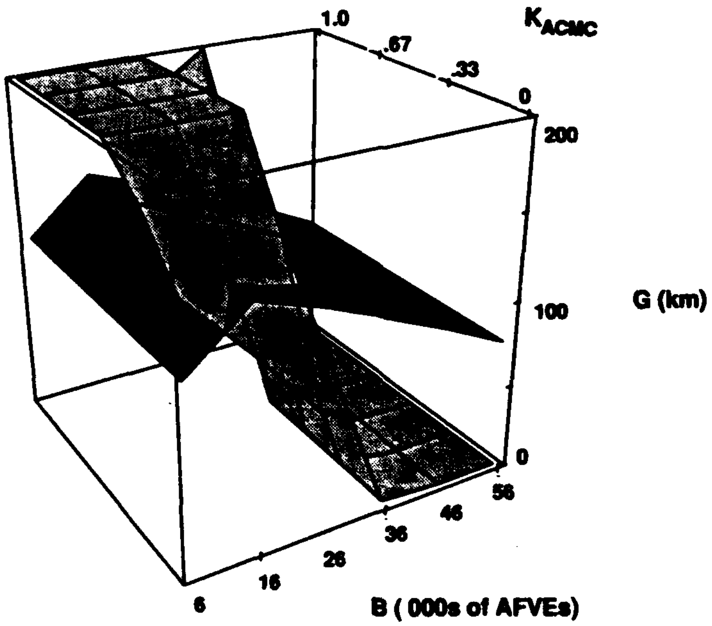

Figure G-12. Sensitivity of G to K_{ACMC}

Appendix H

BIBLIOGRAPHY

BIBLIOGRAPHY

Ananev, I.M. Tankovie Armiy v Nastupleniy (Moscow: Voenizdat, 1988)

Andrushkevich, Col I. "Combat Against Tanks in Modern Operations," Voennaya Mysl', No. 4, 1969

Babakov, A.A., and V.V. Larionov. Evolutsia voennogo iskusstva: etapy, tendentsy printsipy, (Moscow: Voenizdat, 1987)

Bailey, J.B.A. Field Artillery and Firepower, (Oxford: The Military Press, 1989)

Balabolkin, Engineer-Lt Col. R. "The Future of the Infantry of the Main Capitalist Countries." Voennaya Mysl' No. 12, 1973

Balck, Wilhelm. Tactics, translated by Walter Kruger (Ft. Leavenworth KS: U.S. Army Cavalry Association, 1915 translation of the fourth edition of 1908)

Beatty, Jack. "The Exorbitant Anachronism," The Atlantic Monthly, June 1989, pp.40-52

Bellamy, Chris. Red God of War: Soviet Artillery and Rocket Forces, (New York: Brassey's, 1986)

Bennett, Bruce W. et.al., Main Theater Warfare Modeling in the RAND Strategy Assessment System (3.0), (Santa Monica, CA: The Rand Corporation, 1988)

Bentson, Major William. The Problem of Width - Division Tactics in the Defense of an Extended Front, (Fort Leavenworth, KS: U.S. Army School of Advanced Military Studies, 1987)

von Bernhardi, Freidrich. On War of Today, (New York: Dodd, Mead & Company, 1914)

Betts, Richard K. Surprise Attack, (Washington, D.C.: Brookings, 1982)

______. "Conventional Deterrence: Predictive Uncertainty and Policy Confidence," World Politics, January 1985, pp.153-79

Biddle, Stephen D. "The European Conventional Balance: A Reinterpretation of the Debate" Survival, March/April 1988, pp.99-121

______. How to Think About Conventional-Nuclear Substitution: The Problem of Structural Uncertainty, (Alexandria, VA: Institute for Defense Analyses, 1986) IDA P-1884

Bidwell, Shelford, and Dominick Graham. Firepower: British Army Weapons and Theories of War, 1904-1945, (London: Allen and Unwin, 1985)

Bidwell, Shelford. Modern Warfare: A Survey of Men, Weapons, and Theories, (London: Allen Lane, 1973)

Biryukov, Major-General of Artillery G. "Nadezhnoe Ognevoe Porazhenie Protivnika--Osnova Vysokikh Tempov Nastuplenia," Voenniy Vestnik, No. 5, 1977

Biryukov, Major General G. and Colonel G. Melnikov. Antitank Warfare, (Moscow: Progress Publishers, 1972)

Bond, Brian. Liddell Hart: A Study of his Military Thought, (New Brunswick, NJ: Rutgers University Press, 1977)

Bracken, Paul J. "Models of West European Sprawl as an Active Defense Variable," in Reiner K. Huber, Lynn F. Jones, and Egil Reine (eds.), Military Strategy and Tactics, (New York and London: Plenum, 1975), pp.219-230

______. "Urban Sprawl and NATO Defense," Survival, November/December 1976, pp.254-260

Brately, P., B. Fox, and L. Schrage. A Guide to Simulation, (New York: Springer-Verlag, 1983)

Brewer, Garry D., and Martin Shubik. The War Game: A Critique of Military Problem Solving, (Cambridge, MA: Harvard University Press, 1979)

Brody, Richard A., and Charles N. Brownstein. "Experimentation and Simulation," in Fred I. Greenstein and Nelson W. Polsby (eds.), Strategies of Inquiry, Vol.7 of "The Handbook of Political Science," (Menlo Park, CA: Addison-Wesley, 1975), pp.211-264

Brooks, Peter S., et. al. The IDA Advanced Technology Combat Simulation Project, (Alexandria, VA: Institute for Defense Analyses, forthcoming), IDA P-2329

von Caemmerer. The Development of Strategical Science During the 19th Century, (London: Hugh Rees, Ltd., 1905)

Candan, U., L.S. Dewald, and L.R. Speight. Present NATO Practice in Land Wargaming, (The Hague: SHAPE Technical Center, 1987), Professional Paper STC-PP-252

Cimral, Colonel Ted A. "Moving the Heavy Corps," Military Review, Vol.68, No.7, July 1988, pp.28-34

Chernyaev, Major-General V. "Razvitie Taktiki Oboronitel'nogo Boya," Voenno-Istoricheskiy Zhurnal, No. 6, 1976

von Clausewitz, Carl. On War, translated and edited by Michael Howard and Peter Paret (Princeton, NJ: Princeton University Press, 1976)

Cockrell, James K. "Prediction of Advances in Combat," in Reiner K. Huber, Lynn F. Jones, and Egil Reine, (eds.), Military Strategy and Tactics: Computer Modeling of Land War Problems, (New York and London: Plenum Press, 1975), pp.153-165

Cole, Hugh M. The Ardennes: Battle of the Bulge (Washington, D.C.: Department of the Army, Office of the Chief of Military History, 1965)

Colin, Jean. The Transformations of War, translated by L.H.R. Pope-Hennessy (London: Hugh Rees, Ltd., 1912)

Danto, Arthur, and Sidney Morgenbesser, "Laws and Theories," in Arthur Danto and Sidney Morgenbesser, eds., Philosophy of Science (New York: Meridian, 1960), pp.177-181

Davis, Paul K. et. al., Variables Affecting the Central Region Stability: The "Operational Minimum" and Other Issues at Low Force Levels, (Santa Monica, CA: The Rand Corporation, 1989)

Dean, Jonathan. "Alternative Defense: Answer to NATO's Central Front Problems?" International Affairs Winter 1987/88 (Vol.64, No.1), pp.61-82

Defining Stability in the European Theater, Hearings Before the Defense Policy Panel of the Committee on Armed Services, House of Representatives, One Hundredth Congress, Second Session, HASC No. 100-104, 1989

Deitchman, Seymour J. Military Power and the Advance of Technology, (Boulder, CO: Westview, 1983)

Department of Defense Authorization for Appropriations for Fiscal Year 1981, Hearings Before the Committee on Armed Services, United States Senate, Ninety Sixth Congress, Second Session, Part 5 (Washington, D.C.: USGPO, 1980)

Department of the Army Historical Study No. 20-233, German Defense Tactics Against Russian Break-Throughs, (Washington, D.C.: US Army Center of Military History, 1984 reprint of 1951 orig.)

Department of the Army Historical Study No.20-234, Operations of Encircled Forces, (Washington, D.C.: USGPO, 1952)

Department of the Army, United States Army Training and Doctrine Command Systems Analysis Activity. Effects of Barriers in a Combat Environment, (White Sands, NM: TRASANA, 1978), as referenced in US Army Corps of Engineers, Engineer Studies Center, Survivability -- The Effort and the Payoff, R-81-8, June 1981, pp.21-23

DePuy, General William E. "Technology and Tactics in Defense of Europe," Army, April 1979, pp.15-23

Dick, C.J. "Catching NATO Unawares: Soviet Army Surprise and Deception Techniques," International Defense Review, No. 1, 1986

______. "Soviet Operational Art, Part I: The Fruits of Experience," International Defense Review, No. 7, 1988

______. "Soviet Operational Concepts, Part I," Military Review, September 1985

Diesing, Paul. Patterns of Discovery in the Social Sciences, (Chicago and New York: Aldine-Atherton, 1971)

Donnelly, C.N. "Future Soviet Military Policy, Part 2: Where and How," International Defense Review, No. 2, 1989

______. "Soviet Tactics for Overcoming NATO Anti-Tank Defenses," International Defense Review, No. 7, 1979

______. "Tactical Problems Facing the Soviet Army," International Defense Review, No. 9, 1978

______. "The Soviet Operational Maneuver Group: A New Challenge for NATO," Military Review, March 1983

______. "The Wind of Change in Sovet Artillery," International Defense Review, No. 6, 1982

Doughty, Robert A. The Evolution of U.S. Army Tactical Doctrine, 1946-76, (Ft. Leavenworth, KS: U.S. Army Combat Studies Institute), Leavenworth Paper No.1

Dupuy, Col. T.N. (ret'd.), et. al. Analysis of Factors that Have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements, (Bethesda, MD: U.S. Army Concepts Analysis Agency, September 1984), CAA-SR-84-6

______. Understanding War: History and Theory of Combat (New York: Paragon House, 1987)

Eckstein, Harry. "Case Study and Theory in Political Science," in Fred I. Greenstein and Nelson W. Polsby (eds.), Strategies of Inquiry, Vol.7 of "The Handbook of Political Science" (Menlo Park, CA: Addison-Wesley, 1975), pp.79-137

Egorov, Colonel A. "Bundeswehr Armored Division on the Defense," Zarubezhnoe Voennoe Obozrenie, No. 1, 1987

English, John A. On Infantry, (New York: Praeger, 1984)

Epstein, Joshua M. The Calculus of Conventional War: Dynamic Analysis Without Lanchester Theory, (Washington, D.C.: Brookings, 1985)

______. "The 3:1 Rule, Adaptive Modeling, and the Future of Security Studies," International Security, Vol. 13, No. 4 (Spring, 1989)

______. The 1988 Defense Budget, (Wasington, D.C.: Brookings, 1987)

Erickson, John. The Road to Berlin, (Boulder, CO: Westview, 1983), pp.87-136

Facer, Roger L. L. Conventional Forces and the NATO Strategy of Flexible Response, (Santa Monica, CA: The Rand Corporation, 1985)

Farrar-Hockley, Anthony. Infantry Tactics, (London: Almark, 1976)

Fitzgerald, Mary C. "Marshal Ogarkov on the Modern Theater Operation," Naval War College Review, Autumn 1986

Flanagan, Stephen J., and Andrew Hamilton. "Arms Control and Stability in Europe: Reductions are not Enough," Survival, September/October 1988, p.448-463

Fuller, J.F.C. Lectures on F.S.R. III, (London: Sifton Praed and Co., Ltd., 1932)

______. "What is an Aggressive Weapon?" English Review, June 1932, pp.601-5

Galvin, General John R. "Some Thoughts on Conventional Arms Control," Survival, April 1989, pp. 99-107

Gaponov, Col. A. "Correlation of Forces and Rate of Advance," *Voennaya Mysl'*, No. 10, 1971

Gareev, Colonel General Makhmut Akhmetovich. *M.V. Frunze, Military Theorist,* (London: Pergamon-Brassey's, 1988)

George, Alexander L. "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul G. Lauren (ed.), *Diplomacy: New Approaches in History, Theory, and Policy*, (New York: Free Press, 1979), pp.43-68

Glantz, Col. David M. "Operational Art and Tactics," *Military Review*, December 1988

______. *Soviet Operational Art and Tactics in an Era of Reform*, (Fort Leavenworth, KS: Soviet Army Studies Office, April 1989)

Goad, Rex. "Predictive Equations for Opposed Movement and Casualty Rates for Land Forces," in Reiner K. Huber, Lynn F. Jones, and Egil Reine (eds.), *Military Strategy and Tactics: Computer Modeling of Land War Problems*, (New York and London: Plenum Press, 1975), pp.267-285

Goldhamer, Herbert. *The Soviet Soldier: Soviet Military Management at the Troop Level*, (New York: Crane, Russak, 1975)

Goodpaster, General Andrew J. *Gorbachev and the Future of East-West Security: A Response for the Mid-Term*, (Washington, D.C.: The Atlantic Council of the United States, 1989)

Gray, David G. *IDA Unclassified Conventional Forces Database*, (Alexandria, VA: Institute for Defense Analyses, 1989) IDA D-708

Griffith, Paddy. *Forward into Battle,* (London: Anthony Bird, 1981)

Hansen, James H. "Countering NATO's New Weapons: Soviet Concepts for War in Europe," *International Defense Review*, No. 11, 1984

Headquarters Department of the Army. *FM 5-101, Mobility*, (Washington, D.C.: USGPO, March 1985)

______. *FM 5-102, Countermobility*, (Washington, D.C.: USGPO, March 1985)

______. *FM 5-103, Survivability*, (Washington, D.C.: USGPO, 10 June 1985)

______. *FM 6-20, Fire Support in Combined Arms Operations*, (Washington, D.C.: USGPO, 1977)

______. *FM 6-20-1, Field Artillery Cannon Battalion*, (Washington, D.C.: USGPO, 27 December 1983)

______. *FM 7-10, The Infantry Rifle Company*, (Washington, D.C.: USGPO, 8 January 1982)

______. *FM 7-20, The Infantry Battalion*, (Washington, D.C.: USGPO, 28 December 1984)

______. FM 7-30, Infantry, Airborne, and Air Assault Brigade Operations, (Washington, D.C.: USGPO, 24 April 1981)

______. FM 17-50, Attack Helicopter Operations, (Washington, D.C.: USGPO, 1 July 1977)

______. FM 71-1, The Tank and Mechanized Infantry Company Team, (Washington, D.C.: USGPO, 22 November 1988)

______. FM 71-2, The Tank and Mechanized Infantry Battalion Task Force, (Washington, D.C.: USGPO, 30 June 1977)

______. FM 100-2-1, The Soviet Army: Operations and Tactics, (Washington, D.C.: USGPO, July 1984)

______. FM 100-5, Operations, (Washington, D.C.: USGPO, 1982)

______. FM 100-5, Operations, (Washington, D.C.: USGPO, 1986 edition)

______. Organization of the United States Army, (Washington, D.C.: USGPO, December 1988)

______. Soviet Army Operations, (Washington, D.C.: USGPO, April 1978), IAG-13-U-78

Helmbold, Robert L., and Aqeel A. Khan. Combat History Analysis Study Effort (CHASE): Progress Report, (Bethesda, MD: U.S. Army Concepts Analysis Agency, August, 1986), CAA-TP-86-2

Herbert, Paul H. Toward the Best Available Thought: The Writing of Field Manual 100-5, "Operations" by the United States Army, 1973-1976, PhD. dissertation, The Ohio State University, 1985

Herzog, Chaim. The War of Atonement, (Boston: Little Brown and Company, 1975)

Hines, John. "The Operational Calculations for Equal Security Under Arms Control," Conference Paper presented at the "International Symposium on Conventional Stability in Europe: Prerequisites and Analysis Requirements," 10-13 October 1989, German Armed Forces University, Munich

House, Captain Jonathan M. Toward Combined Arms Warfare: A Survey of Twentieth Century Tactics, Doctrine, and Organization, (Ft. Leavenworth KS: U.S. Army Combat Studies Institute, 1984)

Howard, Michael. "Men Against Fire: Expectations of War in 1914," in Steven Miller (ed.), Military Strategy and the Origins of the First World War, (Princeton, New Jersey: Princeton University Press, 1976)

Ingber, Lester, and D.D. Sworder. "Statistical Mechanics of Combat with Human Factors," forthcoming in Mathematical and Computer Modeling

Ingber, Lester, H. Fujio, and M.S. Wehner. "Mathematical Comparison of Combat Computer Models to Exercise Data," forthcoming in Mathematical and Computer Modeling

Ingber, Lester. "Mathematical Comparison of Janus(T)," in S.E. Johnson and A.H. Lewis (eds.), The Science of Command and Control: Part II- Coping with Complexity, (Washington, D.C.: AFCEA International Press, 1989), pp. 165-176

International Institute for Strategic Studies. The Military Balance, 1989-1990, (London: Brassey's, 1989)

Ionin, Col. G. "Sovremennaya Oborona," Voenniy Vestnik, No. 4, 1981

______. "Foundations of Modern Defensive Battle," Voenniy Vestnik, No. 3, 1988

Ionin, Col G., and Col J. Kushch-Zharko. "Defense in the Past and Present," Voennaya Mysl', No. 7, 1971

Isby, David C. and Charles Kamps, Jr. Armies of NATO's Central Front, (New York and London: Jane's, 1985)

Isby, David C. Weapons and Tactics of the Soviet Army, (New York: Jane's, 1988)

Jomini, Antoine Henri. A Summary of the Art of War, translated and edited by J.D. Hittle (Harrisburg, PA: The Military Service Publishing Co., 1947)

Jones, Archer. "The New FM 100-5: A View From the Ivory Tower," Military Review, Vol.58, No.2, February 1978, pp.27-36

______. The Art of War in the Western World, (Chicago: University of Illinois Press, 1987)

Kaczmarek, Col J. "Concerning the Density of Artillery," Voennaya Mysl', No. 12, 1971, reprinted from Mysl Wojskowa, (Warsaw) No. 4, 1971

Karber, Phillip A. "In Defense of Forward Defense," Armed Forces Journal International, May 1984

______. "The Soviet Anti-Tank Debate," Survival, 1976

Kardashevskiy, Lieutenant General of Artillery Yu. "Tvorcheski Planirovat' Ognevoe Porazhenie Tseley," Voennyy Vestnik, July 1978

Karr, Alan F. "Lanchester Attrition Processes and Theater-Level Combat Models," in Martin Shubik (ed.), The Mathematics of Conflict, (New York: Elsevier, 1983), pp.89-126

Keegan, John. The Face of Battle, (New York: Random House, 1977)

Kireev, Colonel N. "Primenenie Tankovykh Podrazdeleniy i Chastey pri Proryve Oborony Protivnika," Voenno-Istoricheskiy Zhurnal, No.2, 1982

Konoplya, Col. P."Preodolenie Oborony, Nasyshchennoy protivotankovymi sredstvami," Voenniy Vestnik, No.6, 1980

Kusch-Zharko, Col K. "Principles of the Art of Warfare in Defense," Voennaya Mysl', No. 9, 1973

______. "Armeyskaya Oboronitel'naya Operatsia," in Marshal of the Soviet Union A. A. Grechko, (ed.), Sovetskaya Voennaya Entsiklopedia, Vol. 1 (Moscow, 1976)

______. "Oborona," in Marshal of the Soviet Union N.V. Ogarkov (ed.), Sovetskaya Voennaya Entsiklopedia, Vol. 5 (Moscow, 1978), pp. 661-2

Labuc, S., and C.N. Donnelly. "Modeling the Red Force: Simulating Soviet Responses in Battle," in Reiner K. Huber (ed.), Systems Analysis and Modeling in Defense, (New York and London: Plenum, 1984)

Lanchester, Frederick William. "Mathematics in Warfare," reprinted in James R. Newman, The World of Mathematics, (New York: Simon and Schuster, 1956), Vol.4, pp.2139-2157

Lee R.G., et. al. Guided Weapons, (Oxford: Brassey's, 1983)

von Leeb, Ritter Wilhelm. Defense, edited and translated by Stefan T. Possony and Daniel Vilfroy (Harrisburg, PA: The Military Service Publishing Company, 1943 edition of 1938 original)

Liddell Hart, Basil H. "The Ratio of Troops to Space," Military Review, Vol.XL, April 1960

______. Deterrent or Defense: A Fresh Look at the West's Military Position, (New York: Praeger, 1960)

______. The Defense of Britain, (London: Faber and Faber, Ltd., 1939)

______. The Liddell Hart Memoirs, Vol.II, (New York: G.P. Putnam's Sons, 1965)

Lijphart, Arend. "Comparative Politics and the Comparative Method," American Political Science Review, September 1971, pp.682-693

Lupfer, Captain Timothy. The Dynamics of Doctrine: Changes in German Tactical Doctrine During the First World War, (Ft. Leavenworth KS: U.S. Army Combat Studies Institute, 1981), Leavenworth Paper No.4

Macksey, Kenneth. Tank Warfare: A History of Tanks in Battle, (New York: Stein and Day, 1972)

Makarychev, Lt. Gen. of Artillery M. "Artillery in Overcoming an Anti-Tank Defense in an Offensive," Voennaya Mysl', No. 1, 1972

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

Maryshev, Major General (res) A.P. "Breaching Enemy Defenses," Voenniy Vestnik, No. 3, 1988

McInnes, Colin, and G.D. Sheffield (eds.). Warfare in the Twentieth Century: Theory and Practice, (London: Unwin Hyman, 1988)

Mearsheimer, John J. "Numbers, Strategy, and the European Balance," International Security, Spring 1988 (Volume 12, No. 4)

______. "Why the Soviets Can't Win Quickly in Central Europe," International Security, Vol.7, No.1 (Summer 1982)

______. Liddell Hart and the Weight of History, (Ithaca and London: Cornell University Press, 1988)

Messenger, Charles. The Art of Blitzkrieg, (London: Ian Allen, Ltd., 1976)

Moore, James W. "The Estimation of Optimum Force Size and Force Reduction Potential in Conventional Arms Reduction Negotiations," Arms Control, September 1988 (Volume 9, No. 2)

Nicholson, Walter. Micro Economic Theory, (Hinsdale, IL: Holt, Rinehart and Winston/ Dryden Press, 1978)

Nikitin, N. "Motopekhotnye Batal'ony Bundesvera v boyu," Voenniy Vestnik, No. 9, 1988

Odorizzi, Charles D., and Benjamin F. Schemmer. "An Exclusive AFJ Interview with General Glenn K. Otis," Armed Forces Journal International, January 1987, pp.44-47

Ogarkov, N.V. (ed.). Voennyy Entsiklopedicheskiy Slovar', (Moscow: Voenizdat, 1983)

Ogorkiewicz, Richard. Armoured Forces, (New York, Arco Publishing Company, 1970)

Pastukhov, Colonel A.A. "Development of Fire Plan in Defensive Combat," Voenno-Istoricheskiy Zhurnal, No. 2, 1986

Pengelley, R. B. "HELBAT Strikes Back," International Defense Review, May 1981, Vol.14, No.5, pp.555-578

Petersen Phillip A., and John G. Hines. "The Conventional Offensive in Soviet Theater Strategy," Orbis, Fall 1983

Polk General James H. (ret.). "The North German Plain Attack Scenario: Threat or Illusion?" Strategic Review, Summer 1980, pp.60-66

Posen, Barry R. "Measuring the European Conventional Balance," in Steven Miller, (ed.), Conventional Forces and American Defense Policy, (Princeton, N.J.: Princeton University Press)

Postlethwait, Lt. Colonel E.M. "Corps Defense on a Broad Front," Military Review, July, 1949

Radzievskiy, A.I. Proryv (Moscow: Voenizdat, 1979)

Reid, Brian Holden. "J.F.C. Fuller's Theory of Mechanized Warfare," Journal of Strategic Studies, December 1978

______. J.F.C. Fuller: Military Thinker, (New York: St. Martin's Press, 1987)

Reznichenko, Lt. Gen. V. "Characteristic Features and Methods of Conducting an Offensive," Voennaya Mysl', No. 1, 1972

______. "Sovetskie Vooruzhennye Sily v poslevoennyy period," Kommunist Vooruzhennykh Sil, No. 1, January 1988

______. Tactics, 1984, Translation Bureau, Secretary of State Department, Ottowa, Canada, (Washington: USGPO, 1987)

Ritter, Gerhard. The Schlieffen Plan: Critique of a Myth, (Westport, CT: Greenwood Press, 1958)

Romjue, John L. From Active Defense to AirLand Battle: The Development of Army Doctrine, 1973-1982, (Ft. Monroe, VA: Historical Office, U.S. Army Training and Doctrine Command, 1984)

Sadykiewicz, Michael. Soviet-Warsaw Pact Western Theater of Military Operations: Organization and Missions, (Santa Monica, CA: The Rand Corporation, August 1987), RAND N-2596-AF

Sapozhnikov, Lieutenant General of Artillery A. M. "Deystviya Diviziona pri Prorive Sil'noy Protivotankovoy Oborony," Voenniy Vestnik, No.8, 1980

Savushkin, Colonel R.A., and Colonel N.M. Ramanichev, "Development of Combined-Arms Tactics Between Civil and Great Patriotic Wars," Voenno-Istoricheskiy Zhurnal, No. 11, 1985

Shirer, William L. The Collapse of the Third Republic: An Inquiry into the Fall of France in 1940 (New York: Simon and Schuster, 1969)

Shkarubskiy, Col. P. "Artillery in Modern Combat Operations of the Ground Forces," Voennaya Mysl', No. 6, 1968

Simpkin, Richard E. Mechanized Infantry, (Oxford and New York: Brassey's, 1980)

______. Antitank, (Oxford and New York: Brassey's, 1982)

______. Race to the Swift: Thoughts on Twenty-First Century Warfare, (Oxford and New York: Brassey's, 1985

______. Red Armor, (Oxford and New York: Brassey's 1984)

______. Tank Warfare, (New York: Crane Russack, 1979)

Skorodumov, Major-General Ivan "Attack Momentum," Soviet Military Review, No. 8, 1987

Stepanenko, Lieutenant Colonel Georgy "In the Interests of Stable Defence," Soviet Military Review, No. 3, 1989

Snyder, Jack. "Limiting Offensive Conventional Forces," International Security, Vol. 12, No. 4 (Spring, 1988)

Stoeckli, Fritz. "Soviet Operational Planning: Superiority Ratios and Casualties in Soviet Front and Army Operations," RUSI Journal, Spring 1989

Strachan, Hew. European Armies and the Conduct of War, (London: Allen and Unwin, 1983)

Sullivan, Leonard Jr. Security and Stability in Conventional Forces: Differing Perceptions of the Balance, (Washington, D.C.: The Atlantic Council of the United States, 1988)

Sun Tzu. The Art of War, trans. Samuel B. Griffith (London: Oxford University Press, 1963)

Taylor, James G. "Attrition Modeling," in Reiner K. Huber, et. al. (eds.), Operational Research Games for Defense, (Munich: R. Oldenbourg, 1979), pp.139-89

______. Lanchester Models of Warfare, 2 Vols., (Arlington VA: Operations Research Society of America, 1983)

Thompson, James A., and Nanette C. Gantz. Conventional Arms Control Revisited: Objectives in the New Phase, (Santa Monica, CA: The Rand Corporation, 1987), Rand Note N-2697-AF

Tout, Ken. Tank, (London: Robert Hale, 1985)

Tuchman, Barbara. The Guns of August, (New York: Macmillan Publishing Co., 1962)

United States General Accounting Office. NATO-Warsaw Pact: Assessment of the Conventional Balance, (Washington, D.C.: USGPO, 1988), Main Report and Supplement, GAO/NSIAD-89-23 and 23A

Van Evera, Stephen. "The Cult of the Offensive and the Origins of the First World War," in Steven Miller (ed.), Military Strategy and the Origins of the First World War, (Princeton, New Jersey: Princeton University Press, 1976)

Veselov, Lt. Col L., and Lt. Col. V. Selyavin. "The Question of the Correlation of Forces and the Rate of Advance," Voennaya Mysl', No. 5, 1972

Waltz, Kenneth N. Theory of International Politics (New York: Random House, 1979)

Wass de Czege, Col. Huba, and Lt. Col. L.D. Holder. "The New FM 100-5," Military Review, July 1982, pp.53-70

Weeks, John. Men Against Tanks, (New York: Mason/Charter, 1975)

Wittman, Klaus. Adelphi Paper 239: Challenges of Conventional Arms Control, (London: Brassey's, 1989)

Wray, Major Timothy A. Standing Fast: German Defensive Doctrine on the Russian Front During World War II, (Ft. Leavenworth KS: U.S. Army Combat Studies Institute, 1986)

Wynne, G.C. If Germany Attacks: The Battle in Depth in the West, (London: Faber and Faber, Ltd., 1940; reprinted by Greenwood Press, 1976)

Ziemke, Earl F. Stalingrad to Berlin: The German Defeat in the East, (Washington, D.C.: U.S. Army Center of Military History, 1987)