DTIC
SELECTE
JAN 0 3 1996
S D
F

RAND

Aggregation, Disaggregation, and the 3:1 Rule in Ground Combat

Paul K. Davis

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

19951228 047

Project AIR FORCE, Arroyo Center, National Defense Research Institute

DTIC QUALITY INSPECTED 1

The research described in this report was sponsored by RAND's federally funded research and development centers for national security studies: Project AIR FORCE, sponsored by the United States Air Force; the Arroyo Center, sponsored by the United States Army; and the National Defense Research Institute (NDRI), sponsored by the Office of the Secretary of Defense, the Joint Staff, and the defense agencies.

Library of Congress Cataloging in Publication Data

Davis, Paul K., 1943–
Aggregation, disaggregation, and the 3:1 rule in ground combat / Paul K. Davis.
p. cm
"Prepared for the United States Air Force, the United States Army, the Office of the Secretary of Defense."
"MR-638-AF/A/OSD."
Includes bibliographical references.
ISBN 0-8330-2321-7 (alk. paper)
1. Combat—Mathematical models. I. United States. Dept. of the Air Force. II. United States. Dept. of the Army. III. United States. Dept. of Defense. Office of the Secretary of Defense. IV. Title.
U21.2.D268 1995
355.4´8—dc20 95-39769
CIP

© Copyright 1995 RAND

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

RAND is a nonprofit institution that helps improve public policy through research and analysis. RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.

Published 1995 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact Distribution Services: Telephone: (310) 451-7002; Fax: (310) 451-6915; Internet: order@rand.org

RAND

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Aggregation, Disaggregation, and the 3:1 Rule in Ground Combat

Paul K. Davis

Prepared for the
United States Air Force
United States Army
Office of the Secretary of Defense

Project AIR FORCE, Arroyo Center, National Defense Research Institute

Approved for public release; distribution unlimited

PREFACE

This report illustrates certain generic issues involved in aggregating and disaggregating models of combat. It was written for RAND's Defense Planning and Technology Department using research support funds provided by RAND's federally funded research and development centers for national security studies: Project AIR FORCE, sponsored by the United States Air Force; the Arroyo Center, sponsored by the United States Army; and the National Defense Research Institute, sponsored by the Office of the Secretary of Defense, the Joint Staff, and the defense agencies. Comments are welcome and should be addressed to the author at RAND, 1700 Main Street, Santa Monica, CA 90407, or by electronic mail to pdavis@rand.org.

CONTENTS

FIGURES

TABLES

SUMMARY

Validity of Aggregation. In this report I illustrate some generic subtleties of aggregation and disaggregation in combat models by starting with an analytically tractable model at one level of detail and then attempting to derive an aggregate model. In particular, suppose that a Lanchester square law is valid for ground combat in each of a number of individual sectors. What equations then describe events at a higher, more aggregate, level? What factors determine whether a closed aggregate-level model exists (i.e., a reasonably accurate model dependent only on aggregate-level variables and with any coefficients being independent of time)? The answer is that what matters is "outside" the "detailed model" altogether, notably (1) higher-level strategy, (2) command and control, and (3) the relative durations of several time scales for battle and maneuver. These factors have major effects on whether a valid aggregate-level model exists and, if it does, what values its coefficients should have.

The 3:1 Rule. A bonus of this analysis is a clarification of when the famous 3:1 rule applies. If it applies at the sector level, then it may or may not apply at a more aggregate level. Indeed, in a theater with multiple corps sectors (e.g., the old Central Region of Europe), the theater-level break-even ratio will typically be more like 1.5:1 than 3:1. By contrast, it is possible for the same 3:1 rule to apply at several lower levels (e.g., corps, division, brigade, and even battalion). In mobile combat in which there is no particular defense advantage, the theater-level break-even force ratio may be about 0.8 or 0.9.

Maneuver, Tempo Control, and Reequilibration. One of the major factors determining outcomes at lower levels is the relative ability of the sides to control where and when to have decisive engagements. If a side can readily break off battle, collect forces, and reengage, then temporary concentrations by the opponent will be less significant. Conversely, the side can itself choose to have decisive engagements under favorable circumstances. These considerations are very important quantitatively and help explain why operational commanders have long tended to focus more on maneuver than, say, on the advantages of static defenses so often praised by analysts. They also illustrate again the disadvantages of military strategies that constrain the defender to fight in particular places and times (e.g., the old NATO forward-defense strategy). Such defenses are quite feasible, but they require more forces.

Aggregation and Disaggregation in Distributed Simulation. Although the specific analysis presented here is narrow, it suggests a broader conclusion of particular interest for distributed simulation, including distributed interactive simulation (DIS). A recurring question is whether it is legitimate and desirable to disaggregate and reaggregate processes during the course of a given simulation run (e.g., decomposing a battle into a higher resolution view, noting the results, reaggregating to a higher level, and later disaggregating again). Extrapolating from the example worked out in detail here, it seems that such temporary disaggregation would be most defensible if the real-world forces are able to reequilibrate at the aggregate level in between the events the simulation describes at a disaggregated level. A simple example might be a division fighting a battle, maneuvering to a different position, taking up new positions, and fighting again. Typically, this sequence would include reequilibration such as combining partly degraded units, balancing across subunits, and assigning new functions. Thus, it might be legitimate to use a disaggregated description for the battles and an aggregate description for the maneuver. By contrast, if the same unit underwent two periods of intensive battle separated by only a few hours (division level) or a day (corps level), the unit's initial state at the time of the second battle might be much the same as at the end of the first battle, in which case aggregating and disaggregating would sacrifice important information.

Families of Models. The simplified analysis of this report also motivates a number of other generic principles. In particular, and contrary to current trends in developing families of models, it is desirable to work top-down rather than bottom-up (or, more accurately, to work *both* top-down and bottom-up rather than only the latter). The reason is fundamental: The allegedly detailed models are only selectively detailed. In particular, they typically do not contain the information most critical in developing valid aggregate expressions. By contrast, top- or aggregate-level issues such as strategy often set context and critical boundary conditions for events at the detailed level.

Validation. This has implications for testing as well. Efforts to validate aggregate-level models should focus on the treatment of strategy, command-control, constraints, time scales, and uncertainties rather than on efforts to calibrate aggregate results against those predicted by a detailed model in which these factors are not even well represented. Interestingly, this is why comparing with historically based and insightful commercial board games of combat can sometimes be more useful to validation than comparisons with high resolution models. Detailed models, however, can be very useful in selectively calibrating specific parameters of higher-level models.

Further, they are essential in understanding cause-effect relationships and determining which aggregate-level variables are important. Although I do not discuss such matters in this report, detailed models can also help generate statistical distributions that should be used to inform the calibration of deterministic or stochastic aggregate models. Ultimately, then, developing sound families of models requires giving proper respect to both higher- and lower-level perspectives of the same problem.

1. INTRODUCTION

This report illustrates a number of basic principles about aggregation and disaggregation in combat modeling by working through the mathematics and phenomenology of a concrete example in which simplified ground combat takes place in a number of sectors and subsectors within a theater.[1] Even with the simplifications, the model is rich enough to demonstrate the importance to combat modeling, in this era of distributed simulation and model families, of approaching aggregation and disaggregation with care. There is need for a strong dose of theory and mathematics rather than the usual dash to programming. It is also unwise to rely solely upon intuition, because aggregation and disaggregation relationships are often much more complicated than original intuition would have it (Davis and Hillestad, 1993; Horrigan, 1992).

I assume in most of what follows that combatail is dictated by the Lanchester square law. I do this merely for simplicity and analytical tractability, despite appreciating well the limitations of the description.[2] I then discuss whether an aggregate law, Lanchester or otherwise, applies at the next level up (i.e., a level with more aggregation and less detail). The answer is that "it depends." Discussing these issues for the simple case suggests broader principles for building model families, principles involving the treatment in models of strategies, command and control, and time scales.

A bonus of the discussion is an explanation of how the famous (or infamous) 3:1 rule does and does not apply at different levels of combat. This is particularly interesting in modern times, because American army forces are likely to be engaged in counteroffensive operations. Will they really need 3:1 force ratios to succeed?

---

[1]I would like to thank Lou Moore, James Dewar, and Robert Howe for reviews of the manuscript in its draft form. I also thank those who responded to an earlier version presented on the World Wide Web in ELECSIM 1995, the Electronic Conference on Scalability in Training Simulation sponsored by the Society for Computer Simulation.

[2]The appendix provides background on Lanchester equations and the related issue of how one "scores" forces.

2. THE MICROSCOPIC MODEL: THE LANCHESTER SQUARE LAW ON COMBAT SECTORS

DEFINITIONS AND STRUCTURE

Let us assume that ground combat occurs in N essentially independent sectors. The attacker and defender in sector i are characterized by strengths A_i and D_i (see also the appendix). Both sides have reserves, which constitute fractions f_a and f_d of their total capabilities A and D. Figure 1a depicts this schematically for a notional theater with rough terrain and largely isolated lines of communications (LOCs). Although the

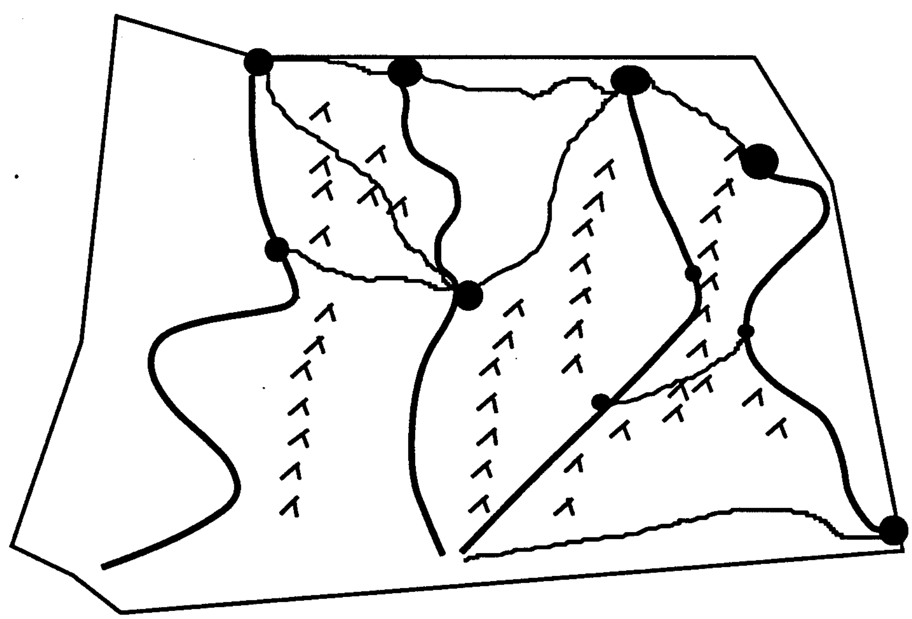

Figure 1a—A Theater with a Network of Independent LOCs

major LOCs are connected by some minor roads, flanking operations can be prevented with small forces so that the major battles are conducted on the principal LOCs, the "sectors." Figure 1b shows a once-standard piston-model representation of a theater with contiguous sectors. In such a case (e.g., the old NATO Central Region), the simple treatment requires the assumption that both sides fall back as necessary to avoid overexposing flanks. The battles in the various sectors are then assumed to be independent frontal attacks characterized by close combat between opposing ground forces. Thus, the battles described involve classic close-combat attrition warfare

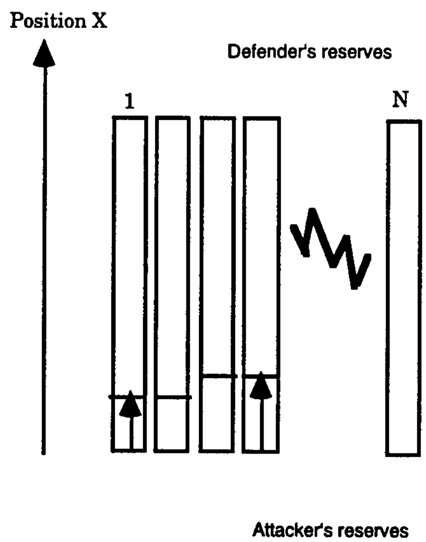

Figure 1b—A Piston-Model Representation of a Theater

without the complications of air forces and long-range fires. And, because we use scalar strengths, we ignore unit structure and the interactions of various weapons types.

In what follows we consider a number of special cases of increasing complexity. The initial examples imagine very simple static battles in which the forces in given sectors "slug it out." Later examples consider concentration and intra-battle reinforcement and maneuver. Although we could readily add a movement model and show forward lines of troops (FLOTs) advancing over time, it would add nothing to the present discussion.

THE LANCHESTER SQUARE LAW

General Version of the Lanchester Square Law

Let us now assume that the "detailed" model of combat is given by the Lanchester square law for a given sector of combat. On each sector i of intense combat, attrition of attacker and defender forces is given by:

$$\frac{dA_i}{dt} = -K_d D_i \qquad \frac{dD_i}{dt} = -K_a A_i \tag{1}$$

where the coefficients K_{ai} and K_{di} are sector-dependent constants reflecting the lethality of shooters and the vulnerability of targets. The usual rationale for this formulation is that the rate of kills should be proportional to the number of shooters and, for aimed fire, should not depend on the number of targets.

It is often more revealing to work with the *fractional* loss rates defined below:

$$ALR_i \equiv \frac{dA_i / dt}{A_i} \quad ; \quad DLR_i \equiv \frac{dD_i / dt}{D_i} \tag{2}$$

A particularly useful variable is the ratio of fractional loss rates *RLR* in the i^{th} sector. If F_i is force ratio, A_i/D_i, then

$$RLR_i \equiv \frac{ALR_i}{DLR_i} = \frac{K_{di}}{K_{ai}} \frac{1}{F_i^2} \tag{3}$$

The ratio of loss rates is a measure of who is winning the slugfest. At a value of 1, both sides' forces are shrinking at the same rate and one can say that the battle is at the break-even point. The attacker, of course, seeks an *RLR* much less than 1. This expression illustrates why (1) is called the Lanchester *square law*: *RLR* varies as the inverse *square* of the force ratio.

For convenience, Table 1 summarizes in one place the principal variables used not only in this section, but in what follows as well.

Special Case: The 3:1 Rule

The famous 3:1 rule in ground combat is represented in Lanchester equations by requiring that the ratio of loss rates be 1 at a force ratio of 3:1 (see also the appendix). That is, a force ratio of 3 implies a break-even situation. For the square law this means that the ratio of kill coefficients K_d/K_a must be 9, in which case

$$RLR_i = \frac{9}{F_i^2} \qquad \text{(square law and 3:1 rule)} \tag{4}$$

The basis of the 3:1 law (see the appendix) is the notion that the defender has a substantial, factor-of-three, advantage if he has prepared positions and good defensive terrain, which reduce his vulnerability and increase the vulnerability of the attacker (e.g., by channelization) (Dupuy, 1987). The square law is assuredly not a statement of

Table 1

Variables Used in Mathematical Analysis

Symbol	Definition
A, D	Attacker and defender strengths
A_i, D_i	Attacker and defender strengths on sector i
ALR	Attacker loss rate, $(dA/dt)/A$
DLR	Defender loss rate, $(dD/dt)/D$
F	Overall force ratio, A/D
F^*	Break-even force ratio: force ratio at which RLR=1
f_a, f_d	Fractions of attacker and defender forces in reserve
F_i	Force ratio on sector i
F_{main}, F_{other}	Force ratio on main sectors and other sectors
K_a, K_d	Attrition coefficients: rates at which each side kills opponent
$\overline{A}_{main}$, $\overline{D}_{main}$	Average force levels on each main sector
m	Reduced-intensity factor for non-main sectors
N	Number of sectors
N_{main}	Number of main sectors
p_d, p_a	Attacker and defender concentration factors
RLR	Ratio of loss rates, ALR/DLR
T_1	Time required to commit reserves
T_2	Time for defender to counterconcentrate by redeploying

general truth. For example, it does not apply to meeting engagements or mobile warfare more generally. Later, we shall consider cases where the defender has no such advantage.

3. DERIVING AGGREGATE MODELS FOR CASES WITHOUT REINFORCEMENT OR REDEPLOYMENT OF FORCES

DERIVING AGGREGATE EXPRESSIONS

General Expressions

The challenge in what follows is to derive closed expressions for the aggregate, theater-level, variables A and D, that is, expressions in which dA/dt and dD/dt are functions of A and D, but not microscopic variables such as the sector-level force strengths or the reserve fractions. These expressions will depend on various statistical averages over microscopic phenomena.

In general we can write

$$\frac{dA}{dt} = \sum_{i=1}^{N} \frac{dA_i}{dt} \equiv -Q_d(A,D,t)$$
$$\frac{dD}{dt} = \sum_{i=1}^{N} \frac{dD_i}{dt} \equiv -Q_a(A,D,t) \qquad (5)$$

which *define* the functions Q_d and Q_a in terms of the sum over sector-level attritions. They may be functions of time. That is, we cannot *assume* them to be constant as desired for an aggregate-level formulation. The aggregate-level ratio of loss rates *RLR* can now be expressed as follows:

$$RLR \equiv \frac{ALR}{DLR} = \frac{(dA/dt)/A}{(dD/dt)/D} = \frac{(dA/dt)}{(dD/dt)}\frac{1}{F}$$

$$(6)$$

$$RLR = \frac{Q_d(A,D,t)}{Q_a(A,D,t)}\frac{1}{F}$$

Special Case: Lanchester Square Law and Uniform Distribution of Forces

In developing aggregate expressions it is usually helpful to begin with idealized situations because the resulting expressions may form a baseline on which to build. That is, the aggregated equations of interest may be similar to those of the idealized situation, with the differences appearing as one or two understandable and well-localized correction factors. Knowing the desired form makes derivations much easier.

With this in mind, let us first consider the case in which the two sides spread their on-line forces evenly across the front (no concentrating or massing), where they engage in combat described by the Lanchester square law. Let us further assume that the circumstances on the various sectors are the same so that the coefficients K_{di} and K_{ai} are in fact sector independent. We then have from (1) and (5) that

$$\frac{dA}{dt}=\sum_{i=1}^{N}\frac{dA_i}{dt}=-\sum_{i=1}^{N}K_dD_i=-NK_dD_1 \qquad \text{square law, no massing} \qquad (7)$$

where D_1 is the defender force level on the first (or any) sector. But ND_1 is then defender's on-line (or in-sector) force level $(1-f_d)/D$. Thus,

$$\frac{dA}{dt}\equiv -Q_a=-(1-f_d)K_dD \qquad \text{square law, no massing} \qquad (8)$$

By virtue of symmetry we can write an equivalent expression for Q_a and use (6) to express the ratio of loss rates. We obtain

$$\begin{aligned} &Q_d=\{(1-f_d)K_d\}D \qquad Q_a=\{(1-f_a)K_a\}A \\ &RLR=\frac{Q_d}{Q_a}\frac{1}{F}=\left\{\frac{(1-f_d)K_d}{(1-f_a)K_a}\right\}\frac{1}{F^2} \end{aligned} \qquad \text{square law, no massing} \qquad (9)$$

In this particular case the functional form of the aggregate model happens to be a Lanchester square law also—i.e., it is identical in form to the detailed model. Note, however, that the coefficients of the model (i.e., the parameters that multiply the variables A, D, or F) depend not only on the coefficients of the detailed model (K_d and K_a), but also on some strategic features of the microscopic problem: the reserve fractions f_a and f_d. If one side has a larger fraction of its forces on line than the other, it has an aggregate advantage, because results depend on the number of shooters.

In the special case in which the 3:1 rule applies, we merely replace K_d/K_a by 9.

A MORE TYPICAL CASE WITH CONCENTRATION EFFECTS

Deriving Basic Expressions

There have been historical instances of uniform attacks across a front, but attackers usually concentrate forces and effort on some sectors while fighting a holding battle on others. Indeed, concentration of force is a central feature of maneuver warfare.

Assume now that the attacker concentrates on N_{main} of the N sectors, conducting low-intensity feints on the other sectors to tie down defenders without suffering excessive attrition himself. Assume also that the coefficients K_{di} and K_{ai} each have two values, one for all main sectors and one for all of the other sectors. That is,

$$
\begin{aligned}
&K_{di} = K_d \text{ on main sectors and } K_{di} = m_d K_d \text{ on other sectors} \\
&K_{ai} = K_a \text{ on main sectors and } K_{ai} = m_a K_a \text{ on other sectors}
\end{aligned}
\tag{10}
$$

As before, let us now see about *deriving* aggregate equations. Since there are two kinds of sectors, main and other, we obtain the equations below. The first equation is straightforward except that we have to renumber the sectors so that the main sectors are 1, 2,...N_{main}, regardless of whether they are contiguous.

$$
\begin{aligned}
&\frac{dA}{dt} = -Q_d = \sum_{i=1}^{N_{main}} \frac{dA_i}{dt} + \sum_{i=N_{main}+1}^{N} \frac{dA_i}{dt} \\
&Q_d = \left\{ N_{main}\overline{D}_{main} + m_d(N - N_{main})\overline{D}_{other} \right\} K_d
\end{aligned}
\tag{11}
$$

Comparing with (9) we would like to restructure this so that there is a factor multiplying $(1-f_d)K_dD$. That would allow us to express Q_d as Q_d for the uniform case, but for a correction factor. It also seems as though the term involving the main sectors should logically be proportional to N_{main}/N, so we may want to make that explicit as well. With this in mind, we can manipulate (11) as follows.

$$
Q_d = \left\{ \frac{\overline{D}_{main}}{(1-f_d)\dfrac{D}{N}} \frac{N_{main}}{N} + \frac{m_d(N - N_{main})\overline{D}_{other}}{(1-f_d)D} \right\} (1-f_d)K_dD \tag{12}
$$

Upon inspection we see that the first term is the fraction of the defender's forward forces that are on main sectors. It is written as a product of two factors, the ratio of the actual average strength per main sector to the amount one would expect if the defender's forward forces were spread evenly, and the fraction of sectors that are main sectors. We can express the first as a concentration factor p_d. From physical considerations we realize that the second term must then be expressible in terms of the remaining forward defense forces. We then have

$$Q_d = \left\{ p_d \frac{N_{main}}{N} + m_d\left(1 - p_d \frac{N_{main}}{N}\right) \right\} (1 - f_d) K_d D \tag{13}$$

where

$$p_d = \frac{N\overline{D}_{main}}{(1 - f_d)D} = \frac{N\,D(on\,main)\,/\,N_{main}}{D(forward)} = \frac{D(on\,main)}{D(forward)} \frac{N}{N_{main}} \tag{14}$$

As is clear from the rightmost expression, the concentration factor p_d can vary from 1 to N_{main}/N.

It follows by symmetry that

$$\begin{aligned} p_a &= \frac{N\overline{A}_{main}}{(1 - f_a)A} \\ Q_a &= \left\{ p_a \frac{N_{main}}{N} + m_a\left(1 - p_a \frac{N_{main}}{N}\right) \right\} (1 - f_a) K_a A \end{aligned} \tag{15}$$

It follows that we have again derived valid aggregate equations, even though the distribution of forces is not uniform.[3]

We can also solve for the ratio of loss rates *RLR*, obtaining

$$\begin{aligned} RLR &= \frac{Q_d}{Q_a} \frac{1}{F} \\ RLR &= \left\{ \frac{\frac{p_d N_{main}}{N} + m_d\left[1 - \frac{p_d N_{main}}{N}\right]}{\frac{p_a N_{main}}{N} + m_a\left[1 - \frac{p_a N_{main}}{N}\right]} \right\} \frac{(1 - f_d)K_d}{(1 - f_a)K_a} \frac{1}{F^2} \end{aligned} \tag{16}$$

Once again, the aggregate equations happen to have the same form as the detailed equations (Lanchester square). The coefficients are defined in terms of the coefficients of the detailed model and some gross features of the microscopic problem—

[3]It is common to hear the claim that an aggregate model of a process is only valid if events at the microsopic level are uniform. That is quite wrong, as this example illustrates. However, a sound aggregation must retain information about microscopic *configuration*. For dramatic examples of configuration effects in aggregation, see Horrigan (1992).

in this case, the reserve fractions, the reduced-intensity factors m_d and m_a, the concentration factors p_d and p_a, and the fraction of sectors on which concentration occurs, N_{main}/N.[4]

Although it is plausible that we could estimate reserve fractions from doctrine, how are we going to estimate the concentration factors and the fraction of sectors on which concentration occurs? We have an aggregate equation, but we do not know how to evaluate or even estimate the coefficients. Furthermore, it is not apparent that there will be good "representative values" of the coefficients, because there may be a great deal of variation across battles. Different attacking generals may use different strategies; some defenders may be more clever about anticipating the points of attack; and so on.

From the mathematics alone, then, we can justify an aggregate formulation, but in the absence of more information—and, in particular, a deeper understanding of the military phenomenology—we have no basis for believing that the coefficients are reliably predictable.

A Reasonable Approximation: Low-Intensity Feints on Non-Main Sectors

For the purposes of this report, and to a reasonable approximation in any case, let us assume that m_a<<1 and m_d<<1 (or, at least, that the factors containing them are small). We then have

$$Q_d \approx (1-f_d)p_d K_d D \quad Q_a \approx (1-f_a)p_a K_a$$
$$RLR \approx \frac{(1-f_d)}{(1-f_a)}\frac{p_d}{p_a}\frac{K_d}{K_a}\frac{1}{F^2} \tag{17}$$

To eliminate some of the variables (or to substitute variables that are more useful) we need to use additional information. Another expression for *RLR* in this case follows from physical considerations. If the only attrition is in main sectors, then the ratio of loss rates for the theater is the exchange ratio $(dA_{main}/dt)/(dD_{main}/dt)$ in the main sectors divided by F (not F_{main}). But $(dA_{main}/dt)/(dD_{main}/dt)$ is $(K_d/K_a)/F_{main}$. It therefore follows that

[4]In some formulations it is more appropriate to focus on the fraction of the militarily useful geographical front on which there are main thrusts. The main-thrust sectors may have differing widths, averaging w. The relevant fraction, then, is $g=wN_{main}/L$ rather than N_{main}/N. The formulations are equivalent if the sectors have equal width.

$$RLR \approx \frac{K_d}{K_a} \frac{1}{F_{main}} \frac{1}{F} \tag{18}$$

Comparing (17) and (18) we see a simple expression for F_{main}:

$$F_{main} = \frac{p_a(1-f_a)}{p_d(1-f_d)} F \tag{19}$$

Solving for the Break-Even Force Ratio F^*

One of the basic questions to be asked at the aggregate level is what force ratio constitutes break-even. Or, to put it differently, what is the force ratio F^* at which RLR is 1? From (18) we can relate F^* to the main-sector force ratio F_{main} at the break-even point by setting F=F^* and RLR=1. We obtain

$$F^* \approx \frac{K_d}{K_a} \frac{1}{F_{main}(F^*)} \tag{20}$$

Thus, F^* is inversely proportional to $F_{main}(F^*)$. We can use (18) and (19) to find another expression for F^*.

$$\begin{aligned} 1 &= \frac{K_d}{K_a} \frac{p_d(1-f_d)}{p_a(1-f_a)F^{*2}} \\ F^* &= \sqrt{\frac{K_d}{K_a} \frac{p_d(1-f_d)}{p_a(1-f_a)}} \end{aligned} \tag{21}$$

For many purposes it is more convenient to have an expression that depends not on p_a, but on other variables that are arguably more natural for the attacker as he sets strategy, notably F_{other} and N_{main}/N.

Accordingly, consider that the attacker's force strength is the strength on main sectors plus the force on other sectors plus the force in reserves. However, the force strengths on the main sectors and other sectors can be expressed as main-sector and other-sector force ratios times the defender strengths. With this and some algebra we can derive an expression for F_{main}:

$$A = F_{main}(1-f_d)D\left(\frac{p_d N_{main}}{N}\right) + F_{other}\left[(1-f_d)D\left(1-\frac{p_d N_{main}}{N}\right)\right] + f_a A$$

$$(1-f_a)A = F_{main}(1-f_d)D\left(\frac{p_d N_{main}}{N}\right) + F_{other}\left[(1-f_d)D\left(1-\frac{p_d N_{main}}{N}\right)\right]$$

$$\frac{(1-f_a)F}{(1-f_d)} = F_{main}\left(\frac{p_d N_{main}}{N}\right) + F_{other}\left(1-\frac{p_d N_{main}}{N}\right)$$

$$F_{main} = \frac{\frac{(1-f_a)F}{(1-f_d)} - F_{other}\left(1-\frac{p_d N_{main}}{N}\right)}{\frac{p_d N_{main}}{N}} \qquad (22)$$

If we require that F=F^* at RLR=1, and use (18), we have a quadratic equation for F^* that can be solved analytically.

$$F^* = \frac{F_{other}\left(1-p_d\frac{N_{main}}{N}\right) + \sqrt{\left[F_{other}\left(1-p_d\frac{N_{main}}{N}\right)\right]^2 + 4\frac{K_d}{K_a}p_d\frac{N_{main}}{N}\frac{(1-f_a)}{(1-f_d)}}}{2\frac{(1-f_a)}{(1-f_d)}} \qquad (23)$$

Only the positive root is physically meaningful, since F^* must be positive and the negative root is negative (the square root is always larger than the first term).[5]

From (20), (21), and (23) we then have three expressions for F^* if m<<1.

$$F^* = \frac{K_d}{K_a} / F_{main}(F=F^*)$$

$$F^* = \sqrt{\frac{K_d}{K_a}\frac{(1-f_d)p_d}{(1-f_a)p_a}} \qquad (24)$$

$$F^* = \frac{F_{other}\left(1-p_d\frac{N_{main}}{N}\right) + \sqrt{\left[F_{other}\left(1-p_d\frac{N_{main}}{N}\right)\right]^2 + 4\frac{K_d}{K_a}p_d\frac{N_{main}}{N}\frac{(1-f_a)}{(1-f_d)}}}{2\frac{(1-f_a)}{(1-f_d)}}$$

The first expression makes it clear that a low break-even point implies, at that force ratio, a large main-sector force ratio, as one would expect intuitively. Figure 2

[5]For large values of K_d/K_a the second term within the square root normally dominates the calculation. Thus, we see that F^* goes roughly as the square root of N_{main}/N.

shows the results graphically for three different assumptions about K_d/K_a. The 3:1, 2:1 and 1:1 rules correspond to K_d/K_a=9, 4, and 1, respectively. Note that if the 3:1 rule applies, then when the attacker has the break-even force ratio, he will have a main-sector force ratio of about 6.

The last of the expressions in (24) is the most useful because it poses the issue in terms of strategy variables. Thus, we can use military reasoning rather than approach the problem as one of pure mathematics.

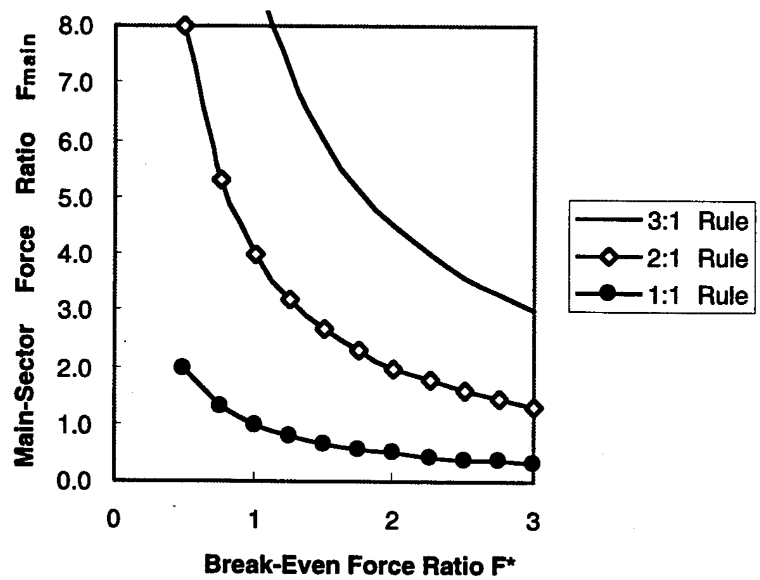

Figure 2—Main-Sector Force Ratio at the Break-Even Force Ratio

The way to view the factors in the third expression is perhaps as follows. An attacker must decide how to concentrate his force. To do this he probably *estimates* p_d (it will be 1 if his concentration is a surprise and the defender has not preferentially defended the main sectors) and f_d (which might be about 1/4 to 1/3 if the defender has a forward defense). He may establish a minimum value of F_{other}, taking the view that any lower value would endanger his operation by making counterattacks too feasible.[6] He

[6]Some of the principal reasons for maintaining a reserve force are "outside the model." At any level of combat, a side with no reserves is exceedingly vulnerable to a random penetration of his line. By constraining f_a and f_d to be non-zero, perhaps on the order of 1/3, we are compensating realistically for inadequacies of the deterministic Lanchester equations. A better model might have stochastic attrition coefficients that are functions of the reserve fraction.

may specify a minimum value of f_a, below which he would be endangering the operation by having too few reserves. Finally, he can derive the value of N_{main}/N that will achieve the break-even force ratio (Davis, 1990). He may choose, of course, to concentrate further to win decisively on the main sectors. However, if N_{main}/N is too small to be strategically significant—i.e., if a breakthrough on so small a portion of the front would leave too much of the defender's army unscathed and too much of his territory unconquered, then he could reduce further the values of F_{other} and f_a, and reconsider his estimates of p_d and f_d, which might initially have been conservative. There is no clear-cut optimizing algorithm, because there is no general utility function to optimize.[7]

Illustrative Results Assuming the 3:1 Rule

Canonical View. Figure 3 shows a relatively canonical view of the problem assuming the 3:1 rule. It assumes the defender has not anticipated the attack and that the defender and attacker have 1/3 and 1/4 in reserve, respectively. We see that the break-even force ratio depends on the fraction of the sectors on which the attacker concentrates (and on the force ratio maintained on other sectors) in a straightforward way.

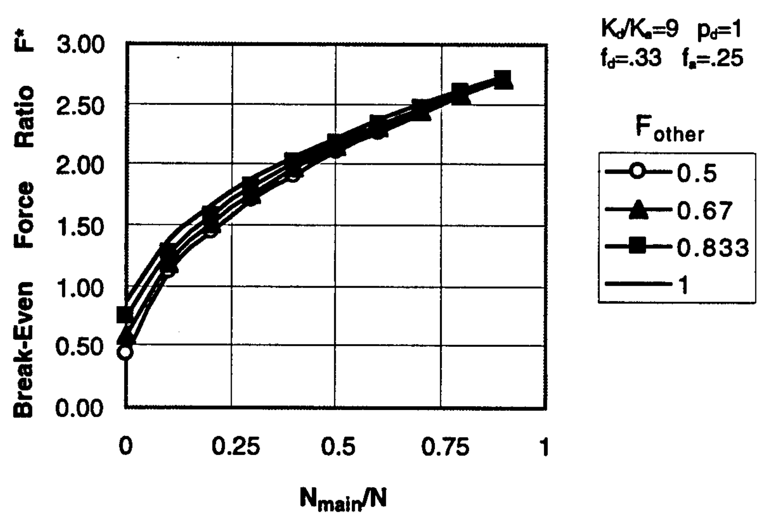

Figure 3—Break-Even Ratio Versus Force Ratio and N_{main}/N for Canonical Case

[7]The Soviet army long used "correlation of force" calculations comparable to those discussed here to make operational decisions about concentration. See, e.g., Hines (1990) or RDA (1990). Presumably, the Russian army is continuing the tradition.

We see immediately that

- Even if the Lanchester square law and 3:1 rule are exact at the sector level, there is no unique analog at the aggregate level. The coefficients (e.g., F^*) depend on issues of strategy, particularly N_{main}/N.
- On the other hand, the attacker will probably need an overall force ratio of *at least* 1.5—if one can argue that a successful attack will need to cover at least, say, 15% of the frontage.

Figure 3 is only one of many possible charts that could be drawn varying different combinations of the parameters. Table 2 shows a range of cases taking nominal, defender-conservative, and two attacker-conservative perspectives. The nominal and defender-conservative cases take the view that the main attack might be on as little as 15%-20% of the front. Further, the attacker might tolerate a 2:1 ratio against him in other sectors (F_{other}=0.5), especially if he could be confident that the defender was not particularly mobile and aggressive. The first attacker-conservative case considers a somewhat larger main frontage, larger reserves, and a less adverse ratio on the other sectors. The last estimate is not unreasonably conservative either; in it the attacker reasons that the defender would surely do *some* anticipatory counterconcentration on the basis of intelligence on massive troop movements. Even a modest counterconcentration (a p_d value of 1.5) changes the break-even point substantially.

Table 2

Representative Bounding Cases Assuming a 3:1 Defender Advantage

p_d	f_d	f_a	N_{main}/N	F_{other}	F^*	Description
1	0.33	0.33	0.20	0.667	1.6	Nominal
1	0.33	0.17	0.15	0.5	1.2	Defense conservative
1	0.33	0.25	0.3	0.67	1.8	Attacker conservative (1)
1.5	0.33	0.25	0.3	0.67	2.1	Attacker conservative (2)

Summary on the 3:1 Rule and Aggregation. In summary, if the 3:1 rule applies at the sector level, which assumes the defender has major advantages from terrain and preparations, then the defender can tolerate only a much smaller aggregate, theater-level, force ratio—something nominally on the order of 1.5. The attacker will seek a larger number, perhaps out of concern that the defender will observe his large-scale maneuvers and do at least some counterconcentration before battle commences. From the attacker's viewpoint, a break-even force ratio of about 2 might seem reasonable. To win decisively, an even larger force ratio might be needed.[8] The requirement would, however, be much lower if the attacker were qualitatively much more capable than the defender (e.g., better morale, training, support forces, air forces, and experience with maneuver). This was the case for the U.S. attack on Iraqi forces in 1991.

This is as far as we can go in the abstract; real-world details matter. For example, in Operation Desert Storm the United States had total information dominance; U.S. generals knew with certainty that the Iraqis had not detected the concentration and mounted a counterconcentration (p_d=1). Nor *could* the Iraqis have done so readily, because of U.S. air supremacy and the lethality of U.S. air forces. Under these circumstances, even a much smaller U.S.-led coalition army could have safely concentrated on a narrow front, broken through, and begun encircling operations to "bag" defender forces.[9]

ALTERNATIVES TO THE 3:1 RULE

Revised Break-Even Curves

Most of the equations derived above are expressed in terms of K_d/K_a, rather than assuming the 3:1 rule. What happens, then, if we do not assume the rule? Suppose, for example, that battle is conducted in relatively open terrain with a great deal of tactical maneuver. There might be some advantage to the defender, but not much. Indeed, the attacker might have the advantage by virtue of having the initiative and associated tactical surprise and momentum. In any case, Figure 4 illustrates the consequences of assuming a 1:1 rule rather than a 3:1 rule. This curve is particularly important for the United States in thinking about maneuver warfare.

To illustrate how this can be used, suppose that the attacker wanted to concentrate on at least 30% of the frontage and to maintain a force ratio of .83:1 on

---

[8]In the mid-to-late 1980s there was rancorous debate about the adequacy of NATO's conventional defense posture in the Central Region. The theater force ratio was variously estimated in the range of 1.5 to 2.2. From the current analysis, one can see why there was a debate. Ultimately, the wartime balance of forces would have depended critically on NATO's response to warning indicators and on whether the Warsaw Pact had improved the readiness of its lowest-quality reserves before beginning mobilization per se (Davis, 1990).

[9]This conclusion is more general than that of the current report, which assumes the approximate validity of Lanchester equations.

"other" sectors. The break-even force ratio would then be about 0.84 (it would be higher if the attacker had the same reserve fraction as the defender).

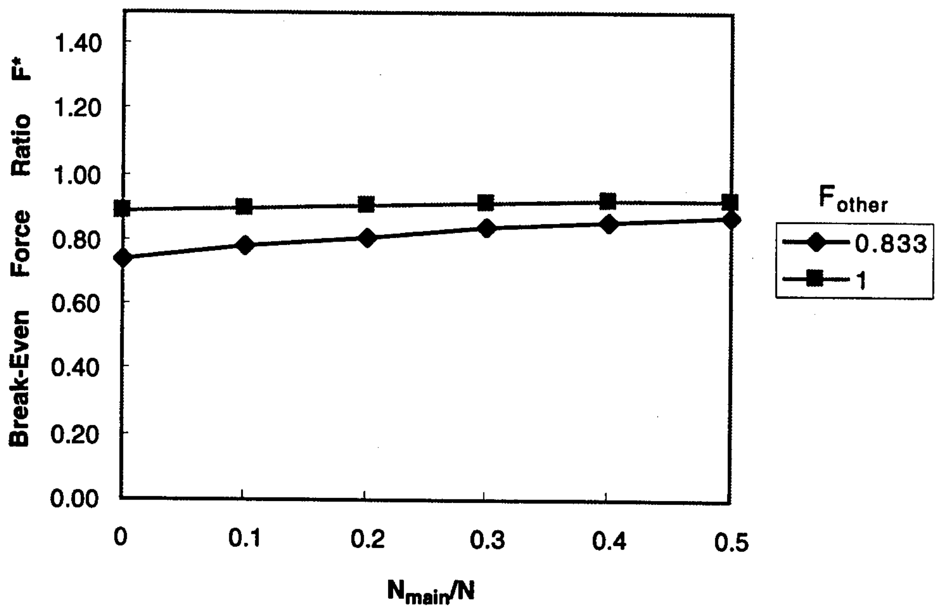

Figure 4—Break-Even Force Ratio Versus Concentration for No Defender Advantage (If No 3:1 Rule Applies)

Minimizing Losses

So far, the discussion has focused on break-even force ratios. But what if the attacker wants to have a decisive victory with minimum losses? From (18) and (19) we see that *RLR* goes roughly as the inverse square of the force ratio, but that assumes a constant value of p_a. By combining (18) and (22) we obtain an expression for *RLR* that depends instead on the strategy variables F_{other} and (N_{main}/N). We obtain

$$RLR = \frac{K_d}{K_a} \frac{p_d \frac{N_{main}}{N}}{\frac{(1-f_a)}{(1-f_d)} F^2 - F_{other} F (1 - p_d \frac{N_{main}}{N})} \tag{25}$$

which, for typical values of the parameters, has the approximate form

$$RLR = \frac{C_1}{F^2 - C_2 F} \approx \frac{C_1}{F^2}\left(1 + \frac{C_2}{F}\right) \tag{26}$$

where C_2 is small. We see, then, that *RLR* actually varies faster than inversely with *F*, for constant attacker strategy as defined by N_{main}/N and F_{other}.

Using (25) for the case shown above, with the break-even point being about 0.83, it turns out that the attacker would need a force ratio of 1.3 for a ratio of loss rates of 1/4, or a force ratio of about 2 for a ratio of loss rates of 1/9. Although estimates of force ratio should reflect qualitative differences in fighting capability, not just equipment counts, this conclusion should nonetheless be sobering for those estimating the capabilities the United States might need in future major regional contingencies.

4. GENERALIZING: EFFECTS ON AGGREGATION OF REINFORCEMENT AND MANEUVER

REINFORCEMENT AND REDEPLOYMENT

A key assumption of the previous cases is that only the forces initially present in the main sectors conduct the battle. But what about reinforcement and redeployment? What if the sides commit their reserves? What if the defender redeploys forces from other sectors? What if the attacker also redeploys forces?[10]

One way to investigate such issues is to develop a simulation model. For current purposes, however, let us instead make some points more qualitatively by looking at the analytics. Repeating (20), we have again that

$$F^* = \frac{F_{other}(1 - p_d \frac{N_{main}}{N}) + \sqrt{[F_{other}(1 - p_d \frac{N_{main}}{N})]^2 + 4\frac{K_d}{K_a} p_d \frac{N_{main}}{N} \frac{(1-f_a)}{(1-f_d)}}}{2\frac{(1-f_a)}{(1-f_d)}} \qquad (27)$$

Now let us account for reinforcement and redeployment as follows. Assume that:

- The defender commits his reserves to the main sectors at an even rate over a period T_1.
- The defender counter concentrates his entire force at an even rate over a period $T_2 (T_2 > T_1)$. That is, over a time T_2 he increases the concentration factor on main sectors at a constant rate until all his forces are on main sectors.
- The attacker follows the defender, maintaining a constant force ratio on non-main sectors. Thus, the attacker also commits his reserves to main sectors over a period T_1 and redeploys additional forces, eventually all of his forces, to the main sectors over a period T_2

[10]These issues are critical in understanding the differences among defensive strategies and, as part of that, appreciating how much more demanding static forward defense strategies are in terms of the force levels needed for success. See, for example, Davis (1990) and Huber (1990). The former paper emphasizes the distinctions between static forward defenses with few reserves and strategies with larger fractions in operational reserve. The latter gives relatively more emphasis to redeployment among sectors, which Huber calls mobile defense. Huber and Helling (1995) summarize extensive recent work in Germany to estimate "stable theater-level force ratios" in a multipolar security environment.

- We express T_2 as a multiple of T_1.

Figure 5 shows the consequences of such assumptions graphically for an aggressive-attacker case. Although real-world changes would be more complex dynamically (e.g., attrition would affect force levels over time, reserves might be initially committed at a higher rate, and the defender might concentrate faster than indicated), this approximate treatment illustrates the basic features. In the example, both sides commit their reserves in time T_1 as shown, and the defender proceeds to counterconcentrate over time (the x axis only goes to 2 T_1, however, so some of the counterconcentration is incomplete).

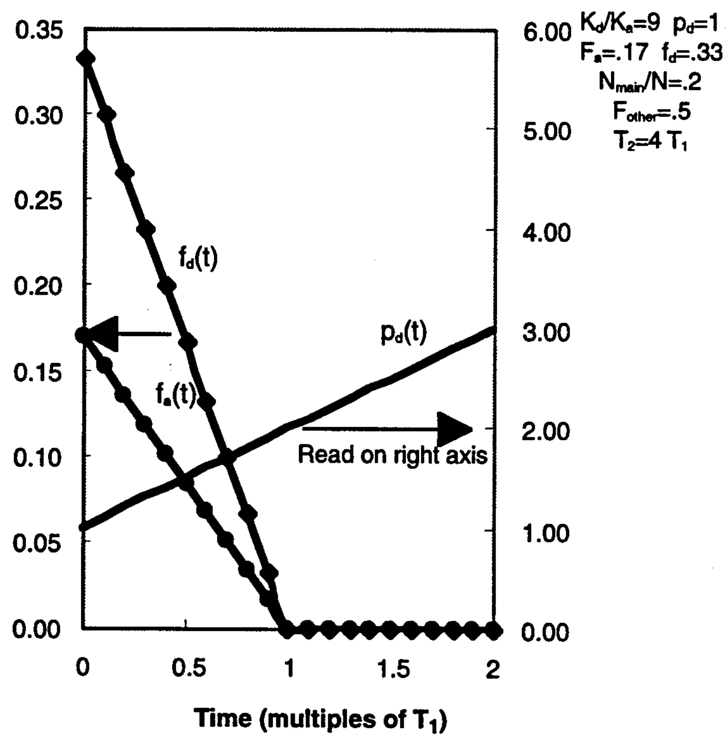

Figure 5—Time Dependence of Reserve and Concentration Factors

EFFECTS OF TIME SCALE

As $p_d(t)$, $f_d(t)$ and $f_a(t)$ change, so also does $F^*(t)$. If we use the time-dependent versions of these variables in (20) we can generate Figure 6.

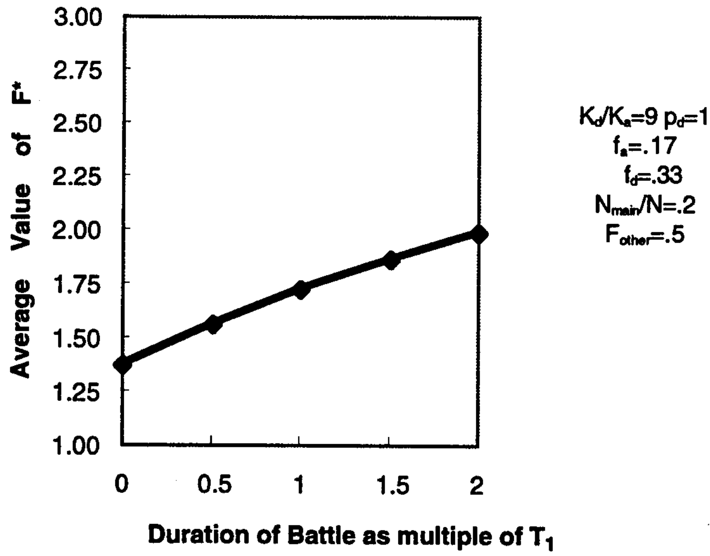

Figure 6—Approximate Average Value of F^* over the Course of Battle as a Function of Battle Duration Relative to Reserve-Commitment Times

The conclusion here is that if the battle is short relative T_1, then the break-even ratio is essentially the same as in the earlier section that ignored intra-battle reinforcement and maneuver (i.e., around 1.5 or so). However, if the battle is not so intense (lower attrition rates and longer duration), then the average value of the break-even ratio rises sharply, reflecting the fact that much of the battle will be fought under circumstances much less congenial to the attacker than intended. Indeed, if the battle lasts long enough, the sides will concentrate all their forces on the main sectors and the value of the original concentration will be greatly reduced—unless, of course, the attacker reconcentrates on a set of new main sectors (or the defender does similarly and goes on the attack).[11]

[11]Note that T_1 should be considered to be a stochastic variable, which would mean that the effective coefficients of any theater-level model would be stochastic. There are many other examples of where stochastic considerations should be made explicit. For example, the probability distribution for sector-level combat is probably distinctly bimodal if the force ratio is 3:1. Thus, the situation at the "break-even force ratio" is not well described as "break-even" in the sense that the sides would stalemate. Instead, it is better described as implying a 50-50 chance of winning or losing. This is discussed, for example, in Huber (1990) and Huber and Helling (1995), based on detailed simulations by Hans Hoffmann and others at the University of the Bundeswehr.

EFFECTS OF AGGREGATION SCALE: PHYSICAL SCALE AND COMMAND STRUCTURE

These observations are reasonable in the abstract, but how do they apply to real-world combat? Roughly speaking, the key point is that forces within isolated sectors with independent generals may be able to call in theater reserves quickly, but they will not be able to draw upon forces from other sectors. That is, in most theater conflicts, T_2 is large compared with T_1, probably much larger even than shown (e.g., T_2 could easily be 10 times greater than T_1 because of factors such as terrain, logistics, disagreements between sector commanders, coalition problems when different nations' forces are on different sectors, or confused intelligence). As a result, a break-even 3:1 ratio at the sector level translates into something more like half that at the theater level.

By contrast, if one were to try to do the same analysis with subsectors, one would conclude that intra-subsector maneuver would probably happen quickly relative to the duration of the sector's battle. Not necessarily, but plausibly. Thus, if the Lanchester equation and 3:1 rule applied at the subsector level (e.g., battalion-level battle), they would probably apply also at the sector level if the higher-level defensive command could reallocate forces within its control on a short time scale compared with the duration of the lower-level battle. An important subtlety here is that the relevant duration of lower-level battle includes the times associated with movement, reconnaissance, engagement and disengagement, and full-out battle. A given full-out battle may be remarkably short in modern warfare (e.g., ten minutes). If the defender is good at maneuver, however, and able to engage and disengage readily, he can drag out the duration of battle to improve his opportunity to "reequilibrate" forces. This ability to control tempo and the point of key battles by maneuver has even more leverage than that of prepared defenses, which helps to explain why field officers have long been much less enamored of static defenses than have analysts.

The conclusion here is that strategy variables (e.g., N_{main}/N) and relative time scales determine the aggregation coefficients. These vary a lot from one level of combat to another.

5. IMPLICATIONS FOR TEMPORARY DISAGGREGATION IN SIMULATIONS, INCLUDING DISTRIBUTED INTERACTIVE SIMULATION

One of the motivations for this report was to illuminate a problem arising in distributed simulation, the problem of connecting models of different resolution meaningfully (Davis, 1995). Crossing levels of resolution is notoriously difficult (see Davis and Hillestad, 1992, 1993), but it is even more difficult to do so frequently in the course of a simulation—sometimes disaggregating and then reaggregating—as when, for example, an aggregate object must do battle with an item-level object, after which the war proceeds.

Is it reasonable to do such disaggregation and reaggregation? Based on the foregoing, a key criterion would appear to be whether the real-world aggregate-level object would "reequilibrate" after one of its components had been in battle. If not, there could be important correlations from one battle to the next and the procedure of disaggregating, aggregating, and disaggregating would be improper. But if the reequilibration is realistic, then the procedure may be reasonable—although other kinds of errors can be introduced if, for example, the disaggregation procedure always assumes the same standard formation and tactics.

6. CONCLUSIONS AND SUMMARY

This report has demonstrated many of the challenges in aggregating and disaggregating descriptions of combat processes by working through an analytically tractable model that assumes the Lanchester square law for ground combat in an individual sector such as that controlled by a corps or, in relatively rough terrain, a division. Even though this assumption is certainly not rigorous, it is nonetheless useful for the purposes here and leads to the following conclusions.

AGGREGATING FROM A SECTOR LEVEL

Given a Lanchester law at the individual sector level, there may or may not be a valid aggregate model at the theater level. If the attacker and defender apply forces uniformly across sectors and maintain constant reserve fractions, an aggregate model exists and is itself Lanchesterian. If the attacker concentrates forces on a fraction of the sectors, conducting mere holding actions on the others, an aggregate model with constant coefficients is still valid so long as there is no change in the reserve fractions or the allocation of forces across sectors. Again the model is Lanchesterian. In this case, however, the key coefficient governing the ratio of loss rates is a complex function of the attacker's strategy and the defender's anticipation of the attacker's strategy (a function of information and decisionmaking). If the break-even force ratio at the sector level is 3:1, then the break-even force ratio at the theater level is about 1.5, 1.2, or 2.1 for canonical, defense-conservative, and attacker-conservative assumptions, respectively.

EFFECTS OF INTRA-BATTLE REINFORCEMENT

If in the course of battle the sides commit their reserves and redeploy forces from other sectors to the main sectors, there is no exact aggregate model with constant coefficients except in extreme cases. What matters are the ratios of several time scales: the duration of battle, the time to reinforce with theater reserves, and the time to redeploy from other sectors. The reinforcement and redeployment times depend not only on physical distances, roads, and movement rates, but also on intelligence, decision times, logistics, and the effects of air power. If, for example, the sector-level battle is intense enough or decisions slow enough, then the intra-battle maneuver and reinforcement will be too late and sector outcomes will depend on the initial sector-level force ratios, thereby favoring the attacker. By contrast, if the defender in a main sector can diagnose events quickly and control the pace of events, perhaps by virtue of multiple prepared lines or giving up space for time, then this will not be so and the value of the initial concentration will be less.

BREAK-EVEN FORCE RATIOS AT DIFFERENT SCALES (DIFFERENT LEVELS OF COMBAT)

While the importance of the relative time scales may seem obvious, these scales are seldom discussed explicitly, even though it has been mysterious to many observers over the years why the 3:1 rule is applied at some levels of combat but not others. As discussed above, if the 3:1 rule is valid at a sector level, the corresponding rule at a theater level may be more like 1.5:1. On the other hand, it is plausible for a 3:1 rule to apply not only at the sector level, but at the subsector level as well. Thus, the same rule might apply to battalion- and division-level battles. The general principle is that if a 3:1 rule applies at a given level, then it will also apply reasonably well at the next-higher level if the higher level's defensive resources can be reallocated in a much shorter time than the duration of lower-level battles (or if the attacker is unable to enforce concentration systematically). If defending forces can break off battles quickly, this increases the effective duration of the low-level battles, thereby allowing more time for "reequilibration." This gives "active" and "mobile" defense concepts advantages over purely static defenses, although static defenses can often exploit fortifications better.

IMPLICATIONS OF MOBILE COMBAT

In mobile warfare the defender has less advantage. In this case, the sector-level break-even force ratio is 1:1 and the break-even force ratio at the aggregate level may be on the order of 0.8—that is, even an outnumbered side can win. The risks of doing so are considerable, however, because holding actions are more difficult. Battles may be more intense and their durations correspondingly shorter. As a result, concentration of force can be decisive—again, unless defending commanders are deft at breaking off battle when outnumbered and maneuvering quickly to reinforce troubled units. Such maneuver issues are especially important today, because the United States is more likely than not to be engaging in mobile warfare rather than a rigid prepared defense of a fixed line.

DISAGGREGATION AND REAGGREGATION WITHIN COMBAT SIMULATION RUNS

Using the insights about aggregation relationships, it is possible to draw conclusions about temporary disaggregation in the course of a simulated battle (e.g., in a distributed simulation). By and large, disaggregating from a theater level in which the independent variables are total attacker and defender force levels is arbitrary and unnatural: It amounts to assuming a particular attack strategy for the entire campaign. Such an assumption cannot then be forgotten as one reaggregates, because in the real world theater-level strategies are highly correlated over time (i.e., if the main attack is through the Ardennes on *D*+1, the Ardennes is probably still a main sector on *D*+2). By contrast, it is not unreasonable to disaggregate temporarily from a sector-level description to a representative subsector-level depiction, and then reaggregate, if the time scales are such that one would expect forces in the sector to "reequilibrate" before the next time period requiring a disaggregated description.

GENERIC PRINCIPLES

The purpose of the analysis is more to illustrate methods of aggregation and disaggregation than to work through the implications of the Lanchester square law. Among the more important principles illustrated are the following:

- Even approximate mathematical analysis can clarify aggregation and disaggregation issues by suggesting functional forms and likely sensitivities.
- However, aggregation typically depends sensitively on assumptions outside the detailed model, notably assumptions about higher-level strategy, command-control, maneuver, and time scales. These cannot generally be determined in advance, making uncertainty analysis necessary at the aggregate level.
- The often dominating role of these higher-level factors is the reason that aggregate models (even board games) can often be quite respectable without being derived in detail from, or calibrated against, detailed models.
- Aggregation may also depend sensitively on other assumptions outside the detailed model, assumptions so implicit as to be largely forgotten. The "detailed" models may, for example, be deterministic because of implicitly assumed tactics such as maintaining reserves that hedge against the consequences of random events. These assumptions must be reflected as constraints when aggregating or using automated methods such as neural nets or mathematical programming to find "optimal" tactics.
- Temporary disaggregation within simulated campaigns may or may not be reasonable, depending on the objectives of the simulation and, importantly,

the time scales involved. By and large, temporary disaggregation is defensible if, in the real world, forces would "reequilibrate" at the aggregate level between periods in which the simulation disaggregates. The reequilibration concept is general, not restricted to ground-force maneuver. The "reequilibration" may involve, e.g., alertness, allocation of fires, redeployment of command and control assets, or maneuver of aircraft and ships.

- Validation of aggregation/disaggregation relationships should focus on the treatment of strategy, command-control, constraints, time scales, and uncertainties. It should not pivot around whether the aggregate model has been fully calibrated against a detailed model, because in many cases such calibration is impossible without mischievous assumptions. On the other hand, experiments with detailed models can often reveal issues and sensitivities that would be missed in even a moderately careful mathematical analysis. Further, they may be a good basis for calibrating *some* parameters of the aggregate model, even though other parameters are outside the model.

A corollary of the last point is that in developing families of models, it may be better to start with more aggregate concepts and develop consistent disaggregated representations and only partial calibrations than to attempt to work exclusively from the bottom up. This may be a radical concept to those wedded to bottom-up approaches. It is contrary to much current discussion, especially by some enthusiasts of distributed interactive simulation who happen to be more acquainted with training and distributed technology than with modeling.

At the same time, work with high-resolution models can be extremely important in clarifying underlying cause-effect relationships, defining the form of aggregate-level models, and calibrating specific parameters within them. As in this report, it is important to work from both directions and to fully appreciate what each level's perspective brings to the problem.

APPENDIX

LANCHESTER EQUATIONS AND SCORING SYSTEMS

GENERAL DISCUSSION

Lanchester equations are differential equations describing the time dependence of attacker and defender strengths *A* and *D* as a function of time, with the function depending only on *A* and *D*.[12] One partly generalized version of the Lanchester equations has the following form

$$\frac{dA}{dt} = -K_d A^r D^s \qquad \frac{dD_i}{dt} = -K_a D^t A^u \tag{A.1}$$

in which the attrition rates and exponents are time-independent parameters. Sometimes the equations are extended to include constant reinforcement-rate terms.

Most authors doing analytical work (as distinct from computer simulations) have focused on one of two special cases: the "square law" corresponds to s=u=1 and r=t=0; the "linear law" corresponds to r=s=t=u=1.

$$dA / dt = -K_d D \qquad dD / dt = -K_a A \qquad \text{square law} \tag{A.2}$$

$$dA / dt = -\tilde{K}_d AD \qquad dD / dt = -\tilde{K}_a AD \qquad \text{linear law} \tag{A.3}$$

It is usually said that the square law applies to "aimed fire" (e.g., tank versus tank) and the linear law to "unaimed fire" (e.g., artillery barraging an area without precise knowledge of target locations). Alternatively, it is sometimes said that the key feature of the square law is that it describes concentration of fire.

Although the simple Lanchester equations with constant coefficients remain useful for demonstrating some features of combat (e.g., the value of concentrating effort

---

[12]For extensive discussion of Lanchester equations see the treatises by Taylor (1980, 1983) and more recent work by Lepingwell (1987), Homer-Dixon (1987) and Epstein (1990), all of which describe the shortcomings of Lanchester theory. See also Wise (1991), which discusses effects of maneuver and command-control, and Helmbold (1993, 1994), which discuss alternative formulations useful for examining empirical data and appreciating some of the more subtle implications of the formulation. The Lanchester equations were discovered simultaneously and independently by the Russian scientist Osipov.

and the associated penalty for breaking up one's forces), especially when it is desirable to do so analytically, they are a poor basis for describing most combat situations. Computer simulations may use Lanchester expressions "locally" (i.e., for attrition estimates within a given time interval), but the coefficients of those equations change from time step to time step as conditions of terrain, defender preparations, and many other factors change. Good computer simulations recognize that the losing side may choose to break off battle rather than be annihilated. Some use equations in which the exponents are much smaller than called for in the square law and in which there are some differences in exponent between attacker and defender (e.g., to reflect the different mix of aimed and unaimed fire that might result from the defender having better cover and the attacker relying more heavily on artillery preparation).[13] Most computer simulations deal separately with different classes of weapon-on-weapon interactions and treat maneuver as fundamental, not an annoying complication. Unfortunately, such computer simulations are then more complicated to understand and discuss. Hence, Lanchester equations continue to have a place in explaining simple points.

For readers interested in understanding the relationship between Lanchester equations and "physics-level calculations," a recent study may be illuminating (Hillestad, Owen, and Blumenthal, 1993). It illustrates how a Lanchester square law can—in simple cases—be a reasonable approximation of events when the opponents approach each other frontally. The authors began with item-level simulations with individual shooters (e.g., tanks) and kill-per-shot probabilities dependent on range. They assumed flat, featureless, terrain. Even in this case, moving to and understanding the Lanchester representation was nontrivial and, in practice, was informed by theory and experimentation with the higher-resolution simulations.

ESTIMATING THE STRENGTHS OR SCORES USED WITHIN LANCHESTER FORMULATIONS

Lanchester equations assume that the sides' strengths can be characterized by scalar quantities that are usually called scores. In practice, estimating appropriate scores can be very troublesome, especially when the sides each have a mix of equipment and especially when the opponents have different equipment, organization, and doctrine. The most important considerations are accounting for the number of items of relevant equipment and gross features of context (type terrain, type battle, and whether

[13]See, e.g., Allen (1992, p. 41) for the expressions used in RAND's RSAS and JICM models. $(dA/dt)/A$ goes as $1/F^{.93}$; $(dD/dt)/D$ goes as $F^{.64}$; the ratio of fractional loss rates goes as $1/F^{1.6}$. These were based on loose fits to historical data as well as approximate theoretical arguments.

there is a serious mismatch of capabilities). Early scoring methods, known as static methods, did not reflect context, but a newer situational scoring method does so, albeit in a way dependent on expert judgment for correction factors (Allen, 1992). The situational scoring method is used by RAND in the RSAS and JICM theater-level models. It has been used in Germany for NATO-sponsored work on multipolar stability concepts (Huber and Helling, 1995).

Another subtle problem in using scores involves the treatment of qualitative factors (e.g., the effects of terrain or the differences in competence between equally sized and equipped forces of different nations). Lanchester intended that A and D measure numbers of entities (e.g., people or tanks). Applying Lanchester laws to force *strength* (i.e., scores reflecting both numbers and qualitative features of combatant entities) requires great care to avoid logical inconsistencies (Lepingwell, 1987; Homer-Dixon, 1987).[14] It is mathematically cleaner to treat qualitative effects by modifying the attrition coefficients rather than the scores. In this report I assume that appropriate scores can be constructed.

THE 3:1 RULE

The 3:1 rule in ground combat has been discussed for centuries, but it is difficult to find authoritative sources justifying it in any detail. For discussion and some citations, see Mearsheimer (1989). For rejoinders see Epstein (1989) and Dupuy (1989). My own view (consistent, I believe, with Mearsheimer's intended message) is that for modern mechanized combat the 3:1 rule applies approximately, when applied to scores such as WEI/WUV or equivalent-division scores determined largely by the number of pieces of major equipment such as tanks. It applies only to equally competent opponents when one of them is fighting from prepared positions in good defensive terrain and the other is conducting a frontal attack. In other situations, the defender advantage is normally less. Dupuy (1987) deals with this by assessing a "combat power," which is something like a WEI/WUV score modified by a series of correction factors for terrain, defensive preparations, surprise effects, and so on. After making such corrections, Dupuy treats break-even as a ratio of 1:1—in "combat power." Models like RAND's RSAS or JICM treat the same effects in somewhat different ways.

---

[14]For example, models such as the RSAS and JICM that use qualitatively adjusted scores to compute attrition must calibrate the scores so that the results are the same as if the situation-dependent effects had been included in the attrition coefficients. Further, they must keep separate track of the unadjusted and adjusted force levels, because the ratios of loss rates are different for these quantities. See Allen (1992, p. 41).

BIBLIOGRAPHY

Allen, Patrick (1992), *Situational Force Scoring: Accounting for Combined Arms Effects in Aggregate Combat Models*, RAND, N-3423-NA, Santa Monica, CA.

Davis, Paul K., and Richard Hillestad (eds.) (1992), *Proceedings of Conference on Variable-Resolution Modeling, Washington, D.C., 5-6 May 1992*, RAND, CF-103-DARPA, Santa Monica, CA.

Davis, Paul K., and Richard J. Hillestad (1993), "Families of Models That Cross Levels of Resolution: Issues for Design, Calibration, and Management," in *Proceedings of the 1993 Winter Simulation Conference, December, 1993*, Society for Computer Simulation, San Diego, CA.

Davis, Paul K. (1990), "Central Region Stability in a Deep-Cuts Regime," in Reiner Huber (ed.), *Military Stability: Prerequisites and Analysis Requirements for Conventional Stability in Europe*, Nomos-Verlagsgesellschaft, Baden-Baden, Germany.

Davis, Paul K. (1995), "Distributed Interactive Simulation in the Evolution of Warfare Modeling," *Proceedings of the IEEE*, Vol.83, No.8, August.

Dupuy, Trevor N. (1987), *Understanding War*, Paragon House Publishers, New York.

Dupuy, Trevor N. (1989), "Combat Data and the 3:1 Rule," *International Security*, Vol. 14, No. 1.

Epstein, Joshua M. (1990), *Conventional Force Reductions: A Dynamic Assessment*, Brookings Institution, Washington, D.C.

Epstein, Joshua M. (1989), "The 3:1 Rule, the Adaptive Dynamic Model, and the Future of Security Studies," *International Security*, Vol.13, No.4.

Helmbold, Robert (1993), "Combat Analysis," *Phalanx*, Vol. 26, No.3, September, Military Operations Research Society, Alexandria, VA.

Helmbold, Robert (1994), "Combat Analysis," *Phalanx*, Vol. 27, No.4, December, Military Operations Research Society, Alexandria, VA..

Hillestad, Richard J., John G. Owen, and Donald Blumenthal (1993), *Experiments in Variable-Resolution Combat Modeling*, RAND, N-3631-DARPA, included in Davis and Hillestad (1992).

Hines, John (1990), "The Soviet Correlation of Forces Method," in Huber (1990).

Hoffmann, Hans, Robert Schnurer, and Andreas Tolk (1995), *On the Impact of Stochastic Modeling in a Rule Oriented Combat Simulation Model on Division/Corps Level*, technical report, University of the Bundeswehr, Munich, Neubiberg. To be included also in the proceedings of the NATO-sponsored "Symposium on Coping with Uncertainty in Defence Decision Making," held at The Hague, 16-18 January, 1995.

Homer-Dixon, Thomas (1987), "A Common Misapplication of the Lanchester Square Law," *International Security*, Vol. 12, No. 1, Summer.

Horrigan, Timothy (1992), "The 'Configuration Problem' and Challenges for Aggregation," in Davis and Hillestad (1992).

Huber, Reiner (ed.) (1990), *Military Stability: Prerequisites and Analysis Requirements for Conventional Stability in Europe*, Nomos-Verlagsgesellschaft, Baden-Baden, Germany.

Huber, Reiner, and Otto Schindler (1993), "Military Stability of Multipolar Power Systems: An Analytical Concept for Its Assessment, Exemplified for the Case of Poland, Byelarus, the Ukraine and Russia," in Reiner Huber and Rudolf Avenhaus (eds.), *International Stability in a Multipolar World: Issues and Models for Analysis*, Nomos-Verlagsgesellschaft, Baden-Baden, Germany.

Huber, Reiner, and Klaus Helling (1995), "Military Stability and Uncertainty About Intentions, Risk Attitudes, Operational Objectives, and Combat Performance of Regional Actors," University of the Bundeswehr, Fakultät fur Informatik, Institut für Angewandte Systemforschung und Operations Research, Neubiberg, Germany.

Lepingwell, John W.R. (1987), "The Laws of Combat? Lanchester Reexamined," *International Security*, Vol. 12, No. 1.

Mearsheimer, John (1989), "Assessing the Conventional Balance: The 3:1 Rule and Its Critics," *International Security*, Vol. 13, No. 4.

RDA (1990), *Battle Command Training Program (BCTP) Opposing FOrce (OPFOR) Command and Staff Handbook*, Book 2, Chapter 5, R&D Associates/Logicon, Los Angeles, CA.

Taylor, James G. (1980), *Force on Force Attrition Modeling*, Military Operations Research Society of America, Alexandria, VA (originally published for the society by Ketron Inc.).

Taylor, James G. (1983), *Lanchester Models of Warfare*, Ketron Inc., Arlington, VA.

Wise, Ben (1991), "Lanchester Equations for Automated Planning," *Phalanx*, Vol. 24, No. 1, March, Military Operations Research Society, Alexandria, VA.