

ORIGINAL COPY

AD-E501 079
Copy 14 of 75 copies

2

IDA PAPER P-2145

A COMPARISON AND ANALYSIS OF STRATEGIC DEFENSE TRANSITION STABILITY MODELS

AD-A298 033

Ivan Oelrich
Jerome Bracken

December 1988

DTIC
ELECTE
MAY 15 1989
S H D

Prepared for
U.S. Arms Control and Disarmament Agency

DISTRIBUTION STATEMENT A
Approved for public release
Distribution is unlimited



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

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, or (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.

Papers

Papers normally address relatively restricted technical or policy issues. They communicate the results of special analyses, interim reports or phases of a task, ad hoc or quick reaction work. Papers are reviewed to ensure that they meet standards similar to those expected of refereed papers in professional journals.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts to record substantive work done in quick reaction studies and major interactive technical support activities; to make available preliminary and tentative results of analyses or of working group and panel activities; to forward information that is essentially unanalyzed and unevaluated; or to make a record of conferences, meetings, or briefings, or of data developed in the course of an investigation. Review of Documents is suited to their content and intended use.

The results of IDA work are also conveyed by briefings and informal memoranda to sponsors and others designated by the sponsors, when appropriate.

The work reported in this document was conducted under contract MDA 903 84 C 0031 for the Department of Defense. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that agency.

This paper has been reviewed by IDA to assure that it meets high standards of thoroughness, objectivity, and sound analytical methodology and that the conclusions stem from the methodology.

Approved for public release, unlimited distribution. Unclassified.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE				
1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) IDA Paper P-2145			5. MONITORING ORGANIZATION REPORT NUMBER (S)	
6a. NAME OF PERFORMING ORGANIZATION Institute for Defense Analyses		6b. OFFICE SYMBOL (if applicable)		7a. NAME OF MONITORING ORGANIZATION U.S. Arms Control and Disarmament Agency
6b. ADDRESS (CITY, STATE, AND ZIP CODE) 1801 North Beauregard Street Alexandria, Virginia 22311			7b. ADDRESS (CITY, STATE, AND ZIP CODE) 320 21st Street, NW Washington, DC 20451	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Arms Control and Disarmament Agency		8b. OFFICE SYMBOL		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AC88SD401
8c. ADDRESS (City, State, and Zip Code) 320 21st Street, NW Washington, DC 20451			10. SOURCE OF FUNDING NUMBERS	
			PROGRAM ELEMENT	PROJECT NO. TASK NO. ACCESSION NO. WORK UNIT
11. TITLE (Include Security Classification) A COMPARISON AND ANALYSIS OF STRATEGIC DEFENSE TRANSITION STABILITY MODELS				
12. PERSONAL AUTHOR(S) Ivan Oelrich, Jerome Bracken				
13. TYPE OF REPORT Final	13b. TIME COVERED FROM TO		14. DATE OF REPORT (Year, Month, Day) December 1988	15. PAGE COUNT 116
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	Strategic stability/instability, transition stability models, stability measures, strategic defense, ballistic missiles.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This paper discusses the nature and causes of strategic stability and instability, how defenses affect the stability equation, problems of partial defenses and various attempts to model stability during the transition to effective defense. All of the models reviewed demonstrate that there is some stable path from no defenses to near-perfect (99 percent effective) defenses, and the paper presents an integrated theory that shows the common nature of all of the stable paths. This paper demonstrates that the results of the various transition stability differ not in numerical results but in the measures of stability used. Recommendations for future transition stability studies are included, based on the review of the various models. The appendices contain detailed descriptions of models used in the various transition stability studies reviewed.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS REPORT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL			22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD FORM 1473 84 MAR

BF 3/10/89

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

IDA PAPER P-2145

A COMPARISON AND ANALYSIS OF STRATEGIC DEFENSE
TRANSITION STABILITY MODELS

Ivan Oelrich
Jerome Bracken

December 1988



INSTITUTE FOR DEFENSE ANALYSES

Contract AC88SD401

PREFACE

This paper has been prepared by the Institute for Defense Analyses (IDA) for the U.S. Arms Control and Disarmament Agency (ACDA), under contract number AC 88 SD 401.

Reviews of this paper were conducted by Dr. Bruce Anderson of the Strategy, Forces and Resources Division (SF&RD) and Dr. Tony Hagar of the System Evaluation Division (SED) of IDA. Mr. Barry Pavel of SF&RD collected documents and provided bibliographical assistance.

Accession For	
NTIS SP&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A-1	



CONTENTS

A. INTRODUCTION	1
B. THE NATURE AND CAUSES OF STRATEGIC STABILITY AND INSTABILITY	2
1. Stability Without Defenses.....	3
2. Stability With High Levels of Defenses.....	4
3. Stability With Partial Defenses	5
4. Transition Stability Models	6
5. Offensive Forces.....	12
6. Defensive Forces.....	13
7. Stability Measures	15
C. INTEGRATING THEORY OF STABLE TRANSITION	17
D. SUMMARY AND CONCLUSIONS.....	19
E. RECOMMENDATIONS	18

APPENDICES

A. Comparison of Stability Measures	
B. Summary of Bracken Paper	
C. Exploration of Canavan Paper	
D. Summary of Chrzanowski Paper	
E. Summary of Kent and Devalk Paper	
F. Summary of O'Neill Paper	
G. Summary of Wilkening and Watman Paper	
H. Bibliography	

FIGURES

1. Percentage Defenses	7
2. Threshold Defenses	9
3. First Strike Payoff With Simple Symmetric Threshold Model.....	9

A. INTRODUCTION

President Reagan, in his March 1983 speech, announced his intention to investigate the possibility of defending the Nation against Soviet ballistic missiles. The speech was the start of the Strategic Defense Initiative and it has affected all strategic debate since. The program as originally set out was to be a *research* program to evaluate the possibility of defense against missiles, but clearly--at least early on--the ultimate defenses that were contemplated, and which created public support for the program, were robust, nation-wide defenses of population.

Few doubt the fundamental stability of the current strategic relationship between the superpowers based on deterring attack through the mutual threat of assured counter-strike and retaliation. This stability requires a certain level of survivability of retaliatory forces but with submarines, bombers, and large numbers of ICBMs which could in theory be launched under attack, both sides could certainly mount devastating retaliations. Therefore, either side could destroy the other whether it struck first or second so there is no incentive to strike first. Similarly, perfect nuclear defenses on both sides would be stable--at least considering only the offensive nuclear part of the relationship--because nuclear attack would have been made irrelevant and there would be no incentive to strike first or second.

Even though the two end-points may be stable, there was some worry, even among many of the proponents of strategic defenses, about stability during the transition to robust defenses. Very early on in the debate, general worries about arms races and more specific worries about crisis stability appeared. Partly, these concerns resulted from a cautious and perfectly reasonable conservatism regarding a nuclear relationship that is theoretically very dangerous but most agree is quite stable in its current state. Why sail unnecessarily into uncharted waters?¹

Most transition studies have assumed implicitly that the ultimate goal of near-perfect defenses is feasible, the question has been how to get there safely. The near unanimous current judgment within the technical community is that near-perfect (say, 99% effective) defenses are not now practical (there is still no consensus on the utility of lesser degrees of effectiveness). The implication for this study is that the transition--that is, the period leading up to perfect defenses--may never reach its end-point. The "transition" and the associated stability problems may be permanent, so by studying the stability of the transition, we may be studying the nature of the strategic superpower relationship for the

¹ See James Schlesinger, "Reykjavik and Revelations: A Turn of the Tide?" *Foreign Affairs* 65 (3) page 447 1987.

foreseeable future if defenses are deployed.² In any case, we should not expect the transition to be fast. It has taken forty years to get from no nuclear weapons to the strategic relationship of today and it may take as long to work our way back.

In response to concerns about transition stability, many studies of the problem were begun. These studies ranged from purely qualitative discussions to detailed quantitative analyses using game theoretic and operations research approaches. The results of several of these studies were presented at two conferences sponsored, in part, by ACDA. Some seeming contradictions among various models were noted at the conferences. Even if all of the models used were identical, the similarity may have been obscured because there was no agreed terminology or notation, no common measures of effectiveness, and no standard offensive or defensive forces. We hope to explain the differences between the models and to demonstrate that some of the difference are more apparent than real. Finally, most of the models of the transition demonstrate that there is some stable path from no defenses to perfect defenses. We believe that we have developed an integrated theory that shows the common nature of all of the apparently different stable paths.

In the following sections we discuss first the nature of strategic stability, then how defenses affect the stability equation, the particular problems of partial defenses, and the various attempts to model stability during the transition. We conclude with a general explanation of instability during the transition and an explanation of why some models find a stable transition easy while others find it hard. We have not had access to the computer model codes themselves nor the resources to explore deeply their behavior even if we did. Therefore, this analysis uses more of a case study approach rather than the quantitative sensitivity analysis approach more familiar to most modelers. The second part of the paper is made up of appendices describing the various models in some detail.

B. THE NATURE AND CAUSES OF STRATEGIC STABILITY AND INSTABILITY

Much has been written on strategic stability and only the barest outline of the subject can be reviewed here. We use as a definition of "stability" that implied by its mechanical analog, that is, if a system is perturbed by some arbitrary outside influence, the system is called "stable" if it tends to return to its original condition. If a perturbation starts a process of increasing changes in the system, then the system is unstable.

² "Transition of What?", Albert Carnesale, *Strategic Defenses and Soviet-American Relations*, page 175. Samuel Wells and Robert Litwak, Eds, Ballinger, 1987.

Many discussions of stability identify three different types of stability, arms race stability, crisis stability, and first strike stability. A relationship is arms race stable if changes in the military forces of one side do not require comparably large changes in the military forces of the other side. A relationship is crisis stable if at any level of tension there is no incentive to escalate to the next highest level; ideally, there would be incentive to de-escalate to the next lower level. A relationship is first strike stable if, during an extreme crisis in which either or both sides consider war to be a real possibility, neither side has an incentive to strike first. The division of stability into these categories is somewhat arbitrary. One should also note that the last step in escalation in a crisis is to strike first, so first strike stability is a subset of crisis stability.

The three types of stability may appear to be very different in some ways but they have important underlying similarities. The different types of stability differ primarily in their time scales. A useful rule-of-thumb approach to defining "stability" is to consider how each side, when considering what it knows about the other side and--just as importantly--keeping in mind what it does not know about the other side, would respond to the question, "Can I afford to wait?" If the answer is *Yes*, then the relationship tends to be stable and if it is *No*, then it tends to be unstable.

For example, if the first *visible* sign of a breakout by one side of the ABM treaty could appear only months before the breakout would occur and countermeasures would take years, then the other may feel compelled to plan responsive actions even without clear evidence of a planned breakout. Each side, knowing the same thing about the other, might make similar preparations which could reinforce each others' worries. This is a type of arms race instability. If strategic forces' survival depend on their alert status and if putting them on alert takes longer than the flight time of the weapons used to attack them, then each side has an incentive to go to higher states of alert in a time of crisis even without clear signs that an attack is being planned. Furthermore, if higher alert also makes the weapons appear more threatening (for example, airborne alert bombers or surged submarines) then one side's increasing alert may force the other side to respond similarly. This is a type of crisis instability. If either side can survive by striking first but will be destroyed if it waits and the other side strikes first, a type of first strike instability occurs.

1. Stability without Defenses

As unsatisfying as the strategic relationship between the superpowers may be, it is now quite stable and this stability is due largely to a single cause: the very large number of survivable nuclear weapons on both sides. Each side has essentially covered, one might

say "saturated," the important targets of the other side. If one side starts to build a few more weapons, there is little incentive for the other side to follow suit so little arms race pressure occurs. Even with forces at normal levels of alert, the essential targets can be hit, reducing any military incentive to go to higher states of nuclear alert (the situation may be very different with regard to general purpose forces) which dampens crisis instabilities. Even after absorbing a first strike, surviving bombers, mobile ICBMs, SLCMs, and SLBMs--and potentially ICBMs launched under attack--can hit a crushing number of important targets and this limits first strike instabilities.

2. Stability with High Levels of Defenses

High levels of defenses, that is defenses able to cope with the entire arsenal of the opponent, change the stability equations. If the defenses are easy to counter simply by building more or better offensive weapons, then the defenses should not be built in the first place.³ If a large investment in offenses can be offset by a small investment in defenses, then there should be no offense-defense arms race instability because of the hopelessness of building more offensive weapons. On the other hand, if the cost of defenses are roughly comparable to the cost of the countered offenses, then the potential for an arms race occurs.

If the defenses can be attacked, either by offensive weapons (an ICBM could be used as an anti-satellite weapon) or the other side's defensive weapons (for example, space-based weapons could attack each other), then strong arms race instabilities can develop. Without defenses, the size of the survivable, secure offensive arsenal of one side is properly compared to the number of *targets* on the other side. The number of offensive weapons on the other side is of secondary importance (except as they are targets), so any tendency toward an arms race is strongly dampened. However, with vulnerable defenses, arms race incentives occur because the two weapon types can shoot at each other. If they can shoot at each other, then the number of weapons in one side must be compared to the number of *weapons* on the other side and increases in the number of weapons on one side could be offset by increases on the other which can cause further increases and so on causing a classic action-reaction arms race.

Some types of defenses could contribute to crisis instabilities. If, for example, the effectiveness or the survivability of the defenses depended on changing the deployment or lofting additional assets into orbit, or if the defenses could be countered by preparations of

³ Herbert York argues that qualitative improvements are a more likely response than quantitative improvements. See Herbert York, Does Strategic Defense Breed Offense? CSIA Occasional Paper, University Press of America, Lanham, Maryland, 1987.

the other side, then actions on one side could induce the other side to counteract in a way that would appear to be a further escalation resulting in a worsening of the crisis.

High levels of defenses should not contribute to first strike instabilities if they are invulnerable. If the defenses are able to stop all offensive weapons, then neither side should have any incentive to strike first or second. If the defenses themselves are vulnerable to a first strike, then striking first against high levels of defenses could mean the difference between being untouched and being destroyed and severe first strike incentives could occur.

None of the defensive models considered here consider the vulnerability of the defenses to a first strike attack or defense suppression in general. Vulnerability will magnify any instability caused by defenses. Even if offensive forces are perfectly survivable, if the defensive forces protecting value targets are vulnerable to whichever side goes first, important first strike incentives will occur. If the defenses and the retaliatory forces are vulnerable, the destabilizing effect of both can be, in some cases, greater than the sum of the individual effects.

Because all of the instabilities described above are undesirable, one would hope that future defensive systems will be designed and built to minimize or eliminate them. Indeed, if the causes of the instabilities--for example, the vulnerability of defensive satellites--can not be removed, that may be the reason that the defenses are not built. However, even with a stable end-point, there is the potential for instabilities arising during the transition to high levels of defense.

3. Stability with Partial Defenses

Regardless of the degree of stability of complete defenses, special stability problems will arise during the transition to complete defenses. If defenses are effective, they will be able to block retaliatory strikes, the one thing that now counters the use of offensive nuclear weapons. If just one side developed defenses, the defenses would make its *offensive* weapons usable. Neither side could afford to allow that situation to develop, so the deployment of defenses by either side will lead inevitably to some sort of action-reaction cycle. This process may be a carefully planned cooperative transition or it may have the characteristics of an arms race.

During periods of partial defenses, there may appear critical problems of crisis instability. If the defenses can stop an offensive attack larger than that presented by the day-to-day alert forces but smaller than that presented by more fully alert forces, then going

on alert makes a potentially enormous difference in the damage the other side will suffer. For example, if one side has defenses that can just barely stop the number of normally alerted nuclear-armed missiles, then it may see the other side's surging of its SSBN fleet as very threatening and provocative. If both sides have these partial defenses, then neither will be willing to risk falling behind in a mobilization race. Nor will either side, once both have alerted forces above the defense threshold level, be eager to be the first to de-escalate below the threshold.

In an extreme form, these crisis instabilities could lead to first strike instabilities if either side believed that war was very likely. If one side takes actions that soon would put it over the defense threshold (again, for example, starting to surge the SSBNs or disperse rail mobile MX) then the other side may calculate that it is better to strike first now and suffer no damage rather than wait and risk losing the ability to overcome the threshold in the near future.

The purer first strike incentive problem occurs due to the "ragged second strike" effect which was identified early on in the strategic defense debate. If partially built defenses are unable to stop a well coordinated first strike, but can stop a degraded second strike, then there may be very strong incentives for both sides not to wait but to strike first to limit damage to themselves.

4. Transition Stability Models

Quantitative models of the transition can allow greater understanding of the instabilities described above and can potentially provide insight into how some of the problems may be avoided. Most of the models of the transition, and almost all of the models presented at the ACDA-sponsored conferences, were models of first strike instability. Some game theoretic models of crisis instabilities have been created. Although the potential arms race instabilities have inspired a great deal of qualitative discussion, very little quantitative modeling of this problem has been reported. This paper, like the presentations at the ACDA-sponsored conferences, concentrates primarily on studies of first strike instabilities. Some work relating to crisis instabilities is briefly mentioned at the end.

Mathematical models of first strike instability with defenses typically assume two sides (referred to as "One" and "Two", "Prime" and "Unprime", and so on. We will use "Red" and "Blue.") and assume that attack by one or the other is inevitable. The payoffs for the two sides if one side strikes first are calculated and then the payoffs if the other side

strikes first are calculated. The difference in the payoffs that each side sees between the two cases is the difference between going first and second. If going first is better, then this is a measure of the incentive to go first⁴. The models typically assume some sort of allocation of first strike forces among value and force targets. The retaliator typically attacks only value targets. A slightly more sophisticated approach is to allow an optimization routine to determine this allocation. In this case, the measure of goodness for each side becomes very important. Not only does it potentially bias the description of the outcome, but because the model, in optimizing, tries to maximize "goodness" by that particular measure, it *determines*, in part, the outcome.

Defenses are then added to the calculation. The first strike incentive is calculated at increasing levels of defenses up to levels high enough to deal with the highest level of offensive forces available. How the defenses are modeled is important and can determine whether a situation is stable or not. Some assumptions made in modeling the defenses result because they make the model easier to execute, which is typical of any modeling effort. Other assumptions in the model reflect differences in how the actual defenses are expected to operate.

Early studies of the transition modeled defenses in one of two general ways, which we call "percentage" defenses and "threshold" defenses⁵. How percentage defenses work is illustrated in Figure 1. One side strikes first with some or all of his forces against the forces and value of the other side. The defenses of the second side intercept some fixed percent of the first strike forces. The second side retaliates with all of his surviving forces against the value of the first striker. The defenses of the first striker intercept some other fixed percentage of the retaliatory forces, *independent of the size of the attack*.

With a percentage defense, first strike incentive decreases smoothly when going from no defenses to perfect defenses. There really is no "transition" problem and the more defenses both sides have, the better. Because of this attractive property and perhaps because models of this type are very easy to execute, many early defense models investigated percentage type defenses. Hypothetical defensive systems might have these

4 In a real confrontation, neither side can really choose to go second, it can only wait and allow the other side the option of going first. In models of first strike instability described here, waiting is not an option and that is a limitation that is best treated using a game theoretic approach. In a game model, the "value" of waiting is the current value of staving off retaliation--a response of the other side--minus the potential cost in the future of losing the initiative and allowing the other side the option of going first.

5 "Percentage defenses" is our term, we have not found any other explicit term used to describe this type of defense. "Threshold" defenses are also referred to as "minimum buy-in defenses" and "cost-of-entry" defenses.

properties, for example: A defense made up of a quadrillion ball bearings in orbit might have a probability of about 0.01 of intercepting any particular ballistic missile in an attack regardless of the number of missiles in an attack. A defense made up of essentially infinite numbers of interceptors that could get in, for whatever reason, only one shot per target, would have an overall intercept probability equal to the intercept probability of the individual interceptors regardless of attack size. Unfortunately, no real, plausible defensive system will have properties and firing doctrines very accurately described by percentage defense models. As a result, by the second ACDA-sponsored conference, the percentage models had fallen out.

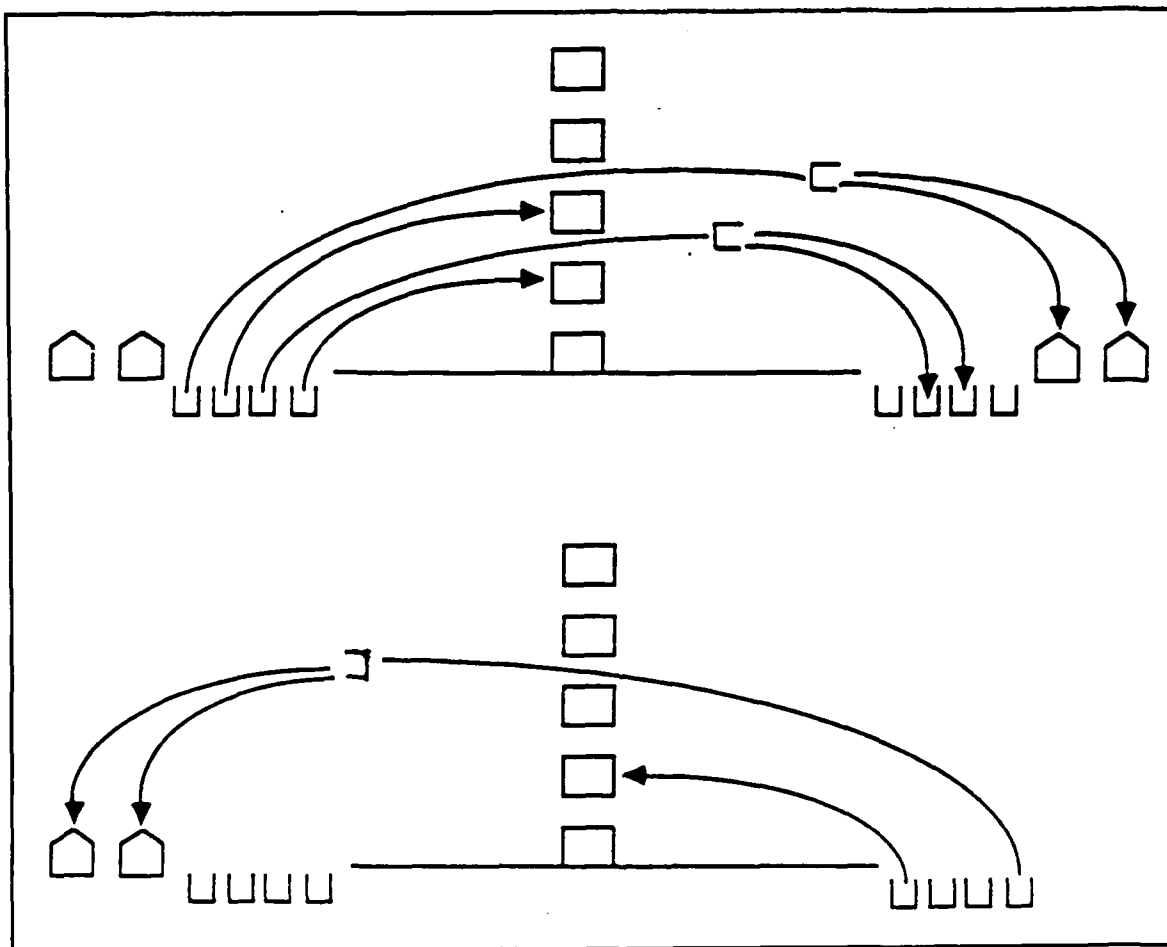


Figure 1. Percentage Defenses

Any real system will demonstrate threshold effects, that is, the system will be able to cope--with some degree of effectiveness--with attacks up to a certain size and any part of the attack greater than that size will overwhelm the defenses and gets through unscathed. This is clearly true for ammunition limited defenses, for example, an attack against a limited

number of space-based interceptors. Threshold effects also appear for systems that at first seem strictly to be rate-of-fire limited, for example, ground-based lasers attacking boosters via space-based mirrors. In this case, the shooter can fire essentially indefinitely but the finite engagement time (we assume that the enemy does not cooperate by attacking slowly) multiplied by the finite rate of engagement yields the number of targets that can be attacked.

Figure 2 illustrates how an engagement between two sides with threshold defenses might look. The vertical barriers in Figure 2 symbolize the threshold defenses of the opposite side. One side strikes first. Some number of his weapons is destroyed but the total number is over the threshold so some are able to penetrate to execute countervalue and--very importantly--counterforce attacks. After the counterforce attacks, the second striker does not have enough forces even to get over the threshold and his retaliatory strike is shut out completely.

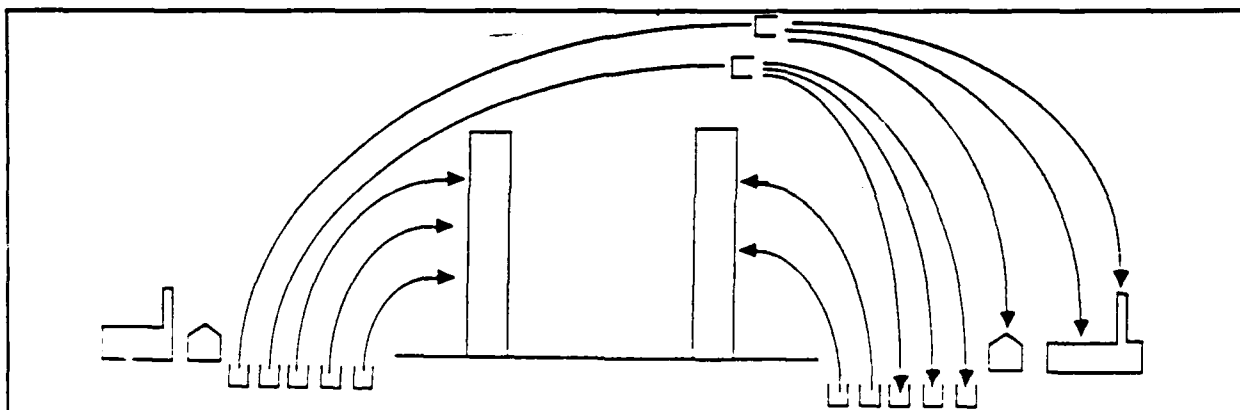


Figure 2. Threshold Defenses

Figure 3 shows the results of an extremely simple threshold model. The model was developed strictly for purposes of illustration here but, although the model is extremely simplified, it shows many of the same effects as more complex models. To simplify the model, the same offensive forces are assumed for both sides, specifically five thousand warheads on each side, two thousand of which are invulnerable to preemptive attack. The same MIRV numbers, probabilities of kill of offensive weapons and interceptors, and so forth are assumed for the two sides, and the defenses levels for the two sides are assumed to be symmetric. One side goes first (since the two sides are identical, it does not matter which one) and allocates his offensive weapons against the other to maximize the payoff function, which is damage to the other side minus retaliatory damage to himself assuming

that the second striker sends all of his retaliatory strike against the first striker's value targets. "Stability" is the negative of the first strike payoff, that is, the system is stable when the payoff is small and unstable when it is large. The angles in the plot occur because the first strike incentive measure was calculated at just a few points. The target sets on both sides can be saturated, that is, increasing the number of warheads has a diminishing return in damage.

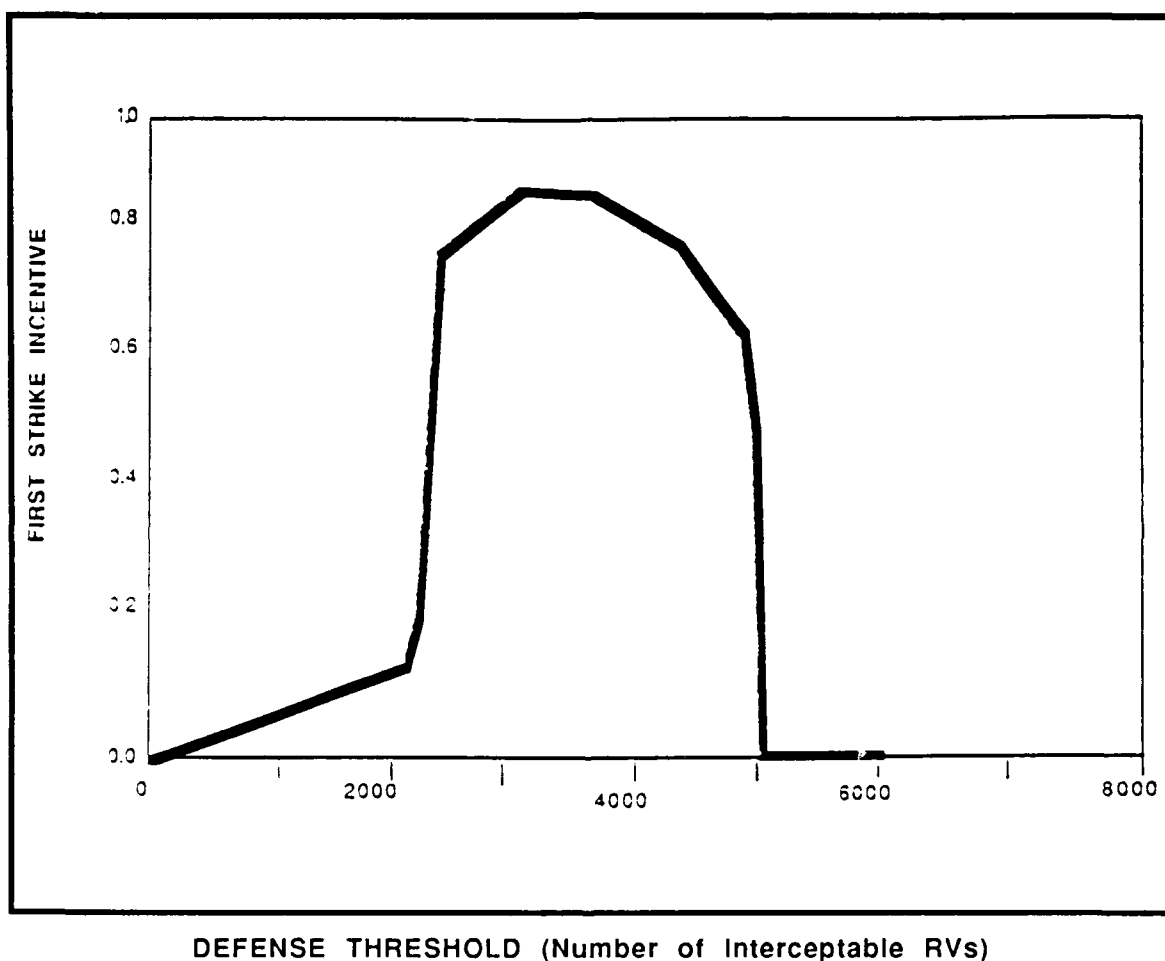


Figure 3. First Strike Payoff With Simple Symmetric Threshold Model

We can see from Figure 3 what happens to stability as defense levels are increased. With no or low defenses, the two thousand untargetable offensive weapons provide a retaliatory force adequate to saturate the other side's value targets. As a result, both the first striker and the second striker are destroyed and there is no incentive to go first. Once the number of interceptors reaches two thousand, however, there is a big payoff in going first because now the five thousand warheads of the first strike can easily get over the two thousand warhead threshold, execute an effective counter strike, have enough warheads left

over for a counter value strike, and the two thousand survivable warheads are then stopped by the defenses. Defenses have created first strike incentives for exactly the same reason that vulnerable retaliatory forces cause them: the first strike can limit damage to the attacker. Indeed, defenses have made all forces vulnerable, some are vulnerable to preemptive attack and the balance are vulnerable to defenses.

As the defense level increases, the first strike incentive goes back to zero because the defenses are high enough to block out even the first strike. The first strike is so weakened that counter force attacks are impractical and the first striker goes for pure counter value attacks, just as the other side goes for pure counter value retaliation attacks. At this point there is no incentive to go first. Eventually, the defenses are high enough to stop all missiles and there is at that point, of course, no difference between first and second strikes, both sides are throwing weapons away.

Although the first strike incentive at low levels of defenses and at high levels of defenses is the same, that is, zero, note that the situation is in fact very different in the two cases. At low levels of defense, both sides are destroyed and the damage *difference* between 100% destruction on both sides is zero. At high levels of defenses, the damage difference between no damage on either side is also zero. In other words, at low levels of defenses, there is mutual assured destruction, at high levels, there is mutual assured survival which is exactly where we want to be. The problem is getting past the instability barrier in between.

More complex models relax most of the constraints imposed by the very simple illustrative model laid out above. The first constraint to go is the symmetry of defense force levels. Displaying results then requires two axes showing defense levels, a third perpendicular axis is usually depicted with contour lines of varying degrees of stability. All of the models reviewed here contain additional details to more closely represent the behavior of real defenses. Some allow preferential rather than simple subtractive defenses, some include the effects of multiple defense layers, and some model air-breathers. Only very brief descriptions are given here; the models are examined in much more detail in the appendices.

The purpose of this study is to compare the various models that were presented at the ACDA-sponsored conference and, whenever the conclusions of the modeling investigations differed, determine the causes of the differences. To that end, we studied in detail the models of Bracken, Canavan, Chrzanowski, Kent and DeValk, and Wilkening

and Watman.⁶ All investigations found that stable transitions were possible but they fell into two broad categories: Kent and DeValk and Canavan found that stability during the transition was fairly robust, the transition was fairly easy and the transition path broad for most cases considered; the rest--Bracken, Chrzanowski, and Wilkening and Watman -- found that stable transitions were possible but only if very particular conditions were met for both the offensive and defensive forces. We also studied a paper by O'Neill, oriented toward developing and explaining a new stability measure.

We compared each of the analyses on the basis of the offensive forces assumed, the way that defensive forces were modeled, and the stability measure that was used. One important criterion for being chosen for review is that the authors provide enough documentation to make a review possible.

5. Offensive Forces

All of the models used as a base case the normally expected (that is, not defense-responsive) offensive forces of the turn of the century. Canavan and Kent and DeValk did not consider air breathers. The other models include air breathers to various degrees.

All of the investigations considered substantial variations from the offensive force base case. These variations usually included altering the distribution of the forces among the legs of the triad--for example, moving forces from ICBMs to SLBMs--and changing the survivability of the forces. Changes in survivability can be effected in either of two ways: directly, by deploying the forces in a less vulnerable way--for example, moving silo-based ICBM warheads onto mobile ICBMs, or indirectly, by moving forces to systems with less prompt hard target kill capability--for example, by moving warheads from silo-based ICBMs to bombers--which reduces the vulnerability of the *enemy* forces.

In general, all of the model results are in agreement that increased retaliatory force survivability increases stability and allows a stable transition under a broader range of conditions. This is easy to understand if the instability is caused by the ragged second strike effect. The threshold character of the defenses, which leads to the first strike incentives, is pronounced only when the threshold is lower than the full first strike but higher than the reduced second strike. In the extreme case, if offensive forces were perfectly survivable, then the second strike of either side would be just as big as its first strike and the defense thresholds--although still there--would not lead to first strike incentives. Similarly, if one side moves forces onto less threatening weapons, for

⁶ This work was not presented at the ACDA-sponsored conferences but is included here because it has been published in a major open literature journal.

example, bombers, then the other side will see its own forces as less vulnerable and will see less risk in waiting because there will be a smaller chance of having so many of its forces destroyed on the ground that they can no longer get over the opponent's defense threshold.

Specifically, one or another of the models considered variations in hardness of silos, moving a greater fraction of forces from ICBMs to SLBMs or mobile ICBMs, and moving from ICBMs to bombers. (Several models increased survivability of offensive forces by defending them but that will be considered next under defense assumptions.)

The offensive forces are sometimes used differently in the different models. Bracken assumes that all offensive missiles can either shoot at value targets to cause damage to the other side or shoot at retaliatory force targets to reduce damage *from* the other side. The model employs an allocation optimization to maximize the difference between the two sides' surviving value. The other models assume counterforce-capable missiles which all shoot at retaliatory forces and counterforce-incapable missiles, which shoot just at value. Canavan and Kent and DeValk do not include air breathers. The other models include them but they are all assumed to be counterforce-incapable and participate only in the countervalue attack. The bombers are considered to be survivable to the extent that they are on alert but some number of SLCMs are considered untargetable.

6. Defensive Forces

As we pointed out above, between the first and second ACDA-sponsored conference, all of the percentage models dropped out. All of the models considered here use some type of threshold defense.

Several types of defensive tactics are conceivable. The simplest is simple subtractive defense wherein attacking RVs are attacked in a statistically random way. A discriminating defense cannot determine which particular target an RV is heading for but it can determine the type of RV (from the booster or launch area, for example) so the hard-target-capable RVs can be attacked preferentially. Simple preferential defenses used in these models concentrate the defensive effort on some predetermined subset of the retaliatory forces to be defended; if the attacker wants to attack all of the targets but does not know which targets are defended, then he must attack as though all of them are defended although only some are. With the adaptive preferential defense used here, the defender does not decide ahead of time which retaliatory forces to defend but adjusts the defense as the progress of the attack unfolds. If a target is destroyed or if all of the RVs sent against a

target are destroyed, then the remaining defensive assets for that target can be freed up to defend other targets. The defenses can be layered and the various layers can use different strategies, for example, a boost phase defense may be subtractive, the mid-course could be discriminating, and finally, a terminal layer could be adaptive.

Bracken (Appendix B) uses a simple single layer subtractive threshold. No actual defense will have razor-sharp thresholds and Bracken models this by including a variable leakage under the threshold. The model treats distinct missile and air defenses.

Canavan (Appendix C) explicitly models the layers in the defense. He treats first a subtractive boost-phase defense, then an adaptive preferential defense (defending retaliatory forces), and finally adds in a percentage-like mid-course defense (which we believe is not handled correctly). Since the model does not include air breathers, it naturally does not include air defenses.

Chrzanowski (Appendix D) considers both the case of simple subtractive missile defense and the case of preferential defense of retaliatory assets. Although the model contains air breathers as counter value and retaliatory forces, it does not include air defenses.

Kent and DeValk (Appendix E) consider cases that they call "discriminating" and "non-discriminating" defenses. Non-discriminating defenses are simple subtractive defenses. Discriminating defenses are defenses that can identify the *type* of RV, specifically whether or not it is hard target capable, and preferentially defend against the hard target capable RVs, but can not identify the RVs' specific targets. Discriminating defense is intermediate in its effects between subtractive defense and preferential defense. Because the model does not include air breathers, air defenses are not included.

Wilkening and Watman (Appendix G) assume simple subtractive boost-phase defenses backed up by preferential terminal defenses of retaliatory forces. They also include air defenses and combine air and missile defenses into an overall defense potential. They describe the defense potential in terms of percent; however, it should be noted that this is not a percentage defense model, rather the defensive threshold is normalized to the size of the opponent's offensive forces such that a defense just able to handle all of the possible offensive forces has a defense potential of 100%.

In general, the model results show that the more favorable the assumptions that one makes about the defense, the more stable the transition. In increasing order of effectiveness, the types of defenses are simple subtractive, discriminating, preferential, and adaptive. A multi-layer defense consisting of a boost-phase layer removing the bulk of the

attacking RVs in a simple subtractive manner backed up by an adaptive terminal layer is the most powerful defense configuration. The correlation between defense effectiveness and stability occurs because the defenses become more effective by more efficiently concentrating their defensive capability, specifically, they can concentrate it on defending retaliatory assets. This reduces the "raggedness" of any second strike and reduces the incentive to go first.

7. Stability Measures

The five different models used three different stability measures. We call these measures minimum mutual deterrence, damage (or surviving value) differences, and mutual expectations.

The minimum mutual deterrence measure of stability is used by Kent and DeValk and Canavan. With this measure, the strategic relationship is "stable" if (1) both sides are deterred, (2) either side dominates the other because of high levels of defenses, or (3) both sides are safe behind high levels of defenses. A situation is "unstable" only if one or both sides find that they are vulnerable when the other side goes first but they are invulnerable when they go first themselves. (This unstable situation is called by Kent "conditional survival"). There is no allowance for an incentive for inflicting damage on the other side and, hence, no difference between inflicting 10% damage and 90%. Furthermore if the enemy can get one or a few warheads through, then the putative first striker is deterred. As a result, there is effectively no incentive to strike first to reduce potential damage from, say, 90% to 10% if even 10% is over the level that deters.

Stability by this measure is fairly easy to achieve and both Canavan and Kent and DeValk conclude that the transition should present no major stability problems. This measure is perfectly consistent with what Kent (who developed the measure) considers to be the justification of strategic defenses, namely, eliminating or reducing to a bare minimum the damage that potential enemies can inflict on us.⁷ Since only a few score to a few hundred warheads on urban centers could threaten our existence as a society, this criterion places extraordinarily high demands on the technical performance of any future defense system. Indeed, this criterion makes the end-point perhaps impossible to achieve but, somewhat ironically, it also makes stability on the way easy to achieve.

A somewhat more complex stability measure, used by Bracken and Chrzanowski, uses first strike incentives defined as the differences in surviving value seen by either side

⁷ "Toward Damage Limitation", Glenn Kent, page 179, *Strategic Defenses and Soviet American Relations*, Samuel Wells and Robert Litwak, Eds. Ballinger 1987

when going first and when going second. If either or both sides have first strike incentives, then the situation is unstable. With this type of measure, used by Bracken and Chrzanowski, there is some incentive to accept damage in exchange for inflicting damage on the other side [although typically there is some weighting factor to put a greater value on own value preserved than on enemy value destroyed]. There is also some incentive to go first to reduce damage to oneself, for example, either side may be tempted to go first to accept for certain the 10% damage from the opponent's certain retaliation rather than wait and risk having the other side go first and suffer 90% damage from his first strike.

Simple, linear difference models suffer from some weaknesses. For example, they do not distinguish between the difference between 100% and 50% damage and the difference between 50% and 0% damage. Also, because of the value placed on doing damage to the other side, these measures suggest that if one side is destroyed, whether it goes first or second, then it may still go first to destroy the other side--knowing it was committing suicide--rather than risk allowing the other side to go first to escape damage.

It turns out that stability by this measure is much harder to achieve than by the mutual deterrence measure and, in general, models that use this measure find that stable transitions are possible only if careful attention is given to the detailed nature of the defenses, the order of deployment of defense layers, and the survivability of retaliatory forces.

Finally, there is the mutual expectation measure of stability used by Wilkening and Watman. With this measure, one starts out by calculating first strike incentives as in the previous measure. However, the measure takes into account the expectations that each side has about the other. If either side were certain that the other side were going to strike, then it would be better to preempt; if either side were certain that the other side were not going to strike, then it would almost always be better to wait. Recognizing that the choice is not between striking first and striking second (striking first would almost always be better) but between striking first and waiting and thereby *risking* being second, each side must estimate the probability that the other side will strike first given the chance.

The mutual expectation measure accounts for these expectations on each side. This measure has a feedback mechanism that tends to push the first strike incentives to their extremes, either no incentive to strike or no incentive to wait.

In its simplest form, the measure implies that, if the first strike incentive is over a certain level, then each side has a certain basic incentive to strike first but also a real fear that the other side will strike; the other side knows that the first side has that expectation,

so the other side will expect him to strike and the other side will want to preempt, so whenever the other side is *expected* to strike, it *ought* to strike to avoid the preemption. The first side does the same calculation and the expectations of the two sides reenforce until the probability of one side's striking is near certainty. Below the level, the low expectations of each side dampen the incentives with a similar feedback mechanism until the probability of either side striking is near zero. The weakness of the measure as it is implemented by Wilkening and Watman, however, is that the threshold levels are arbitrarily supplied by the analysts as 0.3.

Clearly, the choice of the incentive level that separates the regions of upward spiraling and downward spiraling expectations is critical. The mutual expectation measure incorporates attitudes that each side has toward the other and, in that way, is more complete than the simple surviving value difference measures. However, in practice, because of the sensitive, self-reinforcing nature of the measure, it is little better than picking a particular first strike incentive from the surviving value difference measure and defining all incentives above that as unstable and all below as stable. Depending on the level chosen, the mutual expectation measure can make the transition appear easy or hard.

C. INTEGRATING THEORY OF STABLE TRANSITION

An examination of all of the models reviewed here show that all of the transition instabilities have similar causes and when stable transitions are found they have certain common traits. The instabilities are due to a combination of two factors: the threshold saturation characteristic of the defenses and the vulnerability of offensive forces to a counterforce first strike. These two things cause the "ragged second strike" problem, namely, that the first strike will be large and well organized and will be able to overwhelm partially-built defenses to a great enough degree to carry out both countervalue attacks and just enough counterforce attack to bring the second strike below the threshold of what the first striker's defenses can handle. This creates powerful incentives to go first. The side going first can spare himself destruction while assuring defeat of the other side. Heightened sensitivity can be expected since each side will realize that the other is making the same calculation.

Note that *both* offensive vulnerability *and* a threshold effect are required for the instabilities to appear. All of the models that show stable paths do so by effectively removing one or the other of these requirements. There are several possibilities.

One approach is to reduce the vulnerability of offensive forces. The first strike incentive occurs because of the ragged second strike. If the second strike is not so ragged, then the incentive to strike first is reduced, that is, if the second strike of one side were just as big as its potential first strike, then there is no incentive for the other side to strike first to limit damage to himself. There may still be an incentive to strike first if the putative first striker's own forces are vulnerable; he may then strike first to destroy the other side knowing that if he strikes second, the other side will remain unharmed. This ability to inflict damage on the other side may create some first strike incentive but it is certainly weaker than that created by damage limitation to oneself.

All of the studies that investigated this case found that increasing offensive force survivability reduced first strike incentives. This might have been accomplished directly by hardening silos or increasing submarine alert rates. Either side can indirectly but effectively increase the other side's survivability by reducing the prompt hard target kill of its own forces. For example, by putting a greater fraction of the total number of warheads on presumably survivable and presumably first strike incapable bombers, one reduces the vulnerability of the other side's forces and reduces his incentive to go first.

Offensive forces' survivability can also be increased by defending them, an option that is consistent with a hypothetical world with strategic defenses. The models that investigated this case found that building terminal defenses of retaliatory forces first, and only then area defenses, reduced the extent of instability during the transition.

Some of the earlier modeling efforts used what we call percentage defenses. These models were later dropped because they are unrealistic but they did have the attractive mathematical property that instabilities did not appear during the defense build-up. For instabilities to appear, the defense must show threshold effects. However, the percentage models suggest how to suppress the bad effects of defense thresholds. The threshold effects would not manifest themselves if the threshold were larger even than the largest first strike. If we could create a "leaky" defense with thresholds that high, then it would appear to be a percentage defense. Therefore, one stable route through the transition is to build leaky defenses until the thresholds are very high, letting the leakage provide deterrence during the transition, and then turn down the leakage only at the end.

The Bracken model, which explicitly allows for ballistic missile defense leakage has shown that this approach can lead to a stable transition. In practice, this could be achieved by building first a boost-phase intercept system--which would inevitably let some missiles through--and only when that is complete, build the ground-based terminal interceptor

system that would mop up leakers attacking value targets (a short-range terminal system that defended ICBMs could be included in the earlier stage and not upset stability).

Another approach is to use airbreathers to provide the "leakage." In this case, missile defenses are built up while air defenses are deliberately left weak. The air defenses provide deterrence during the transition and when the missile defenses are strong enough, the air defenses would be built up. Either of these approaches may require substantial cooperation to be effective.

D. SUMMARY AND CONCLUSIONS

The question that originally motivated this study, "What are the causes of the apparent disagreements between the several transition models?" can now be answered. All of the models used comparable offensive forces and made similar types of excursions from their base case assumptions so these assumptions cannot account for the differences. The models used different assumptions about the *type* of defense, for example, whether it is simple subtractive or adaptive. These assumptions have effects but the effects are not dominant and are well understood.

Where the models do differ importantly is in their measures of stability. In general, mutual deterrence measures imply that stability is easy to achieve and damage difference measures imply that it will be much harder to achieve (but not impossible). So we see that the models do not vary so much in their basic, numerical results as they may at first have seemed, but the modelers have very different ideas of what these results imply for stability and these differences are reflected in the quantitative stability measures that they choose.

E. RECOMMENDATIONS

The following should, perhaps, not be called "recommendations" since most reports already incorporate most of them but they do provide a checklist for future work in the area. The strongest recommendation to come from this study is to make very clear what kind of numerical stability measure is being used and to ***make clear what the stability measure implies about the assumed behavior of nations.*** If the model contains an optimization routine or allocation algorithm, this too will contain assumptions about the behavior of the players and these assumptions always should be made explicit. Perhaps the numerical results should not be carried to a final measure of stability, rather the results should be presented at a level where each reader is able to apply the stability measure he feels most appropriate. In particular, it will sometimes be better to disaggregate the first

strike incentives of each side, that is, to express the incentive of either side separately rather than adding them into a single stability index.

Since strategic defenses do not yet exist, it is perfectly plausible to make various assumptions about their capabilities and how the defense might be organized. However, each assumption of the nature of the defense implies certain technical capabilities. The modeler presumably picks one type of defense model over another for some reason, either for ease of modeling or because one particular defense is considered more plausible than another. What drives the choice should always be made clear. *If the modeler believes that one type of defense is more likely than another, then he has an obligation to outline the technical implications.* For example, an adaptive defense implies impact point prediction, robust communication, and a very flexible firing system. It also implies that the enemy *does not have* MaRVs that can defeat impact point prediction. (Indeed, if adaptive defense is technically feasible, then one can almost as easily envisage a communication system among mid-course MaRVs that would allow adaptive *attack*. To the best of our knowledge, no one has considered this case.)

We believe that *the assumptions about offensive forces should be kept very general.* A world with strategic defenses would be very different from the world today. We should not assume that the defensive environment will be radically changed while the offensive environment remains much what we predict it to be. Perhaps investigations of strategic defenses should use broad categories of types of future offensive arsenals to discover trends and general relationships and not worry overly much about the fine detail.

Next steps? Possible future works could include: (1) The specific effects of specific early deployment defense systems. (2) Detailed investigation of various quantitative stability measures. (3) Analysis of the dynamics of defense, specifically, the interaction of defenders and weapons that could attack the defenses or the interaction of defenses that could attack each other. This would be a much more complex modeling task than any presented here. (4) Many game theoretic models of crisis stability predict escalation incentives based on a comparison of the potential risk and benefit from moving up one escalation step versus the risk and benefit of waiting and allowing the other player to take the initiative. These risks and benefits are usually left as variables. Models of the sort discussed in this paper could be used to quantify these values. An interaction between the two approaches may yield useful results.

APPENDIX A

COMPARISON OF STABILITY MEASURES

A. STABILITY MEASURES

There are two types of stability which are of primary concern -- arms race stability and crisis stability, of which first strike stability is an important part.

Arms race stability is the simpler of the two concepts. It exists whenever, if one side adds forces or improves forces, the other side does not have to respond by adding forces or improving forces. First strike stability is a more complex concept. It exists when there is no incentive for either side to strike first in a crisis situation.

B. DAMAGE DIFFERENCE STABILITY MEASURES

The minimum mutual deterrence measure of stability is fairly simple but the damage difference measure used by Bracken and Chrzanowski is somewhat more complex so it is described in more detail in this section.

Let the initial levels of value of Blue and Red be denoted by B and R . Let a Blue first strike followed by a Red second strike be denoted by br and a Red first strike followed by a Blue second strike be denoted by rb . Let R_{k1} and B_{k2} denote the Red value killed in the Blue first strike and the Blue value killed in the Red second strike, and B_{k1} and R_{k2} denote the Blue value killed in the Red first strike and the Red value killed in the Blue second strike.

Bracken and Chrzanowski define Blue first strike payoff as:

$$F^B = \underset{\text{all allowable } br}{\text{maximum}} \quad (B - B_{k2}) - (R - R_{k1})$$

and Red first strike payoff as:

$$F^R = \underset{\text{all allowable } rb}{\text{maximum}} \quad (R - R_{k2}) - (B - B_{k1})$$

These definitions state that the first striker's surviving value minus the second striker's surviving value, after an optimized centerforce/countervalue allocation, is the first strike payoff. If the initial values, B and R , are 1, then F^B and F^R lie between -1 and 1.

The lower bound of -1 corresponds to the first striker's being destroyed by the second strike after causing no damage in the first strike; the upper bound of 1 corresponds to the first striker suffering no damage from the second strike after causing complete damage in the first strike.

Bracken and Chrzanowski now introduce the measure of first strike instability:

$$G = FB + FR.$$

Since FB and FR take on values between -1 and 1, G takes on values between -2 and 2.

Figures A-1, A-2, A-3, and A-4 give four generic cases of first strike payoff optimization outcomes. These figures show starting and ending value for the two sides when one or the other goes first. They will be interpreted in terms of first strike instability.

Figure A-1 is a simple situation of mutual assured destruction. There is no first strike payoff ($FB = 0$ and $FR = 0$) and no incentive to preempt ($G = 0$). The end states are identical regardless of who goes first. Again B and R are the surviving values.

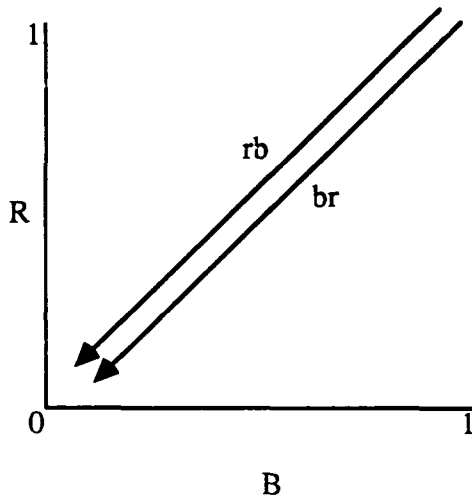


Figure A-1: $FB = 0$, $FR = 0$, $G = 0$

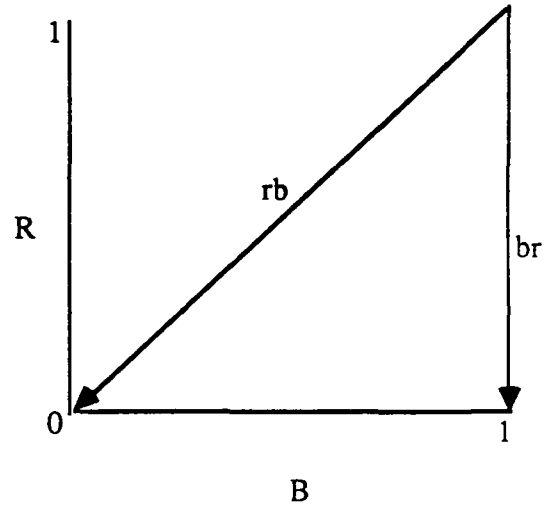


Figure A-2: $FB = 1$, $FR = 0$, $G = 1$

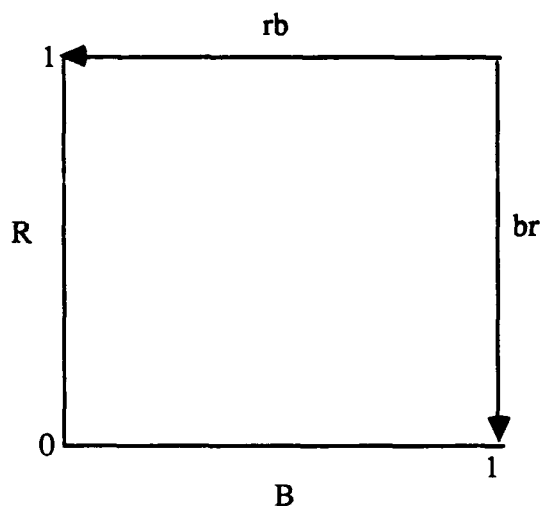


Figure A-3: $F^B = 1, F^R = 1, G = 2$

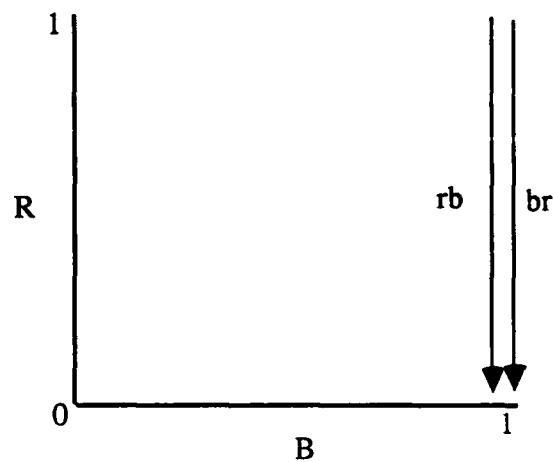


Figure A-4: $F^B = 1, F^R = -1, G = 0$

Figure A-2 is straightforward from a first strike payoff point of view, for Blue can prevail and Red can't ($F^B = 1$ and $F^R = 0$). From a first strike instability point of view, however, this is a disturbing case. The measure $G=1$ states that both sides would be motivated to strike first, Blue to prevail unharmed and Red to prevent that from happening, even if it means its destruction. Some analyses would not call this situation unstable, on the theory that Red would never be induced to commit suicide to prevent Blue from "winning"; we will return to this later. (Note that only the illustration of Figure A-2 involves any analytical disagreement; the other three are unambiguous.). Red would find this situation intolerable from an arms race instability point of view, and would attempt to at least restore the situation, of Figure A-1.

Figure A-3 is an illustrative case of extreme first strike instability. If Blue strikes first, $F^B = 1$ while if Red strikes first $F^R = 1$. Both could not be more highly motivated to strike first and $G = 2$.

It is worth discussing an example to illustrate when such a situation might occur. It is particularly interesting in the case of two-sided offensive and defensive force structures. Let Blue have 1000 ICBMs with 3000 RVs and 3000 deployed SLBM RVs. Let Red have 1400 ICBMs with 5000 RVs and 1000 deployed SLBM RVs. Let both sides proceed to build defenses at equal levels until each has 4000. At this point, if Blue strikes first, the 2000 of its 6000 RVs which exceed the Red defensive threshold can be targeted to the 1400 Red ICBMs and to value targets. Blue's 4000 defenders can ward off the surviving Red

ICBM RVs and the 1000 Red SLBM RVs. Thus, it is possible that $F^B = 1$. If Red strikes first, the 2000 of its 6000 RVs which exceed the Blue defense threshold can be targeted to the 1000 Blue ICBMs and to value targets. Red's 4000 defenders can ward off the surviving Blue ICBM RVs and the 3000 Blue SLBM RVs. Thus, it is possible that $F^R = 1$. This example could match Figure A-3 quite well, and illustrates the concept of substantial first strike instability in the offense-defense context.

Figure A-4 illustrates a case that is unstable from an arms race point of view and stable from a crisis point of view. Red will not tolerate force structures leading to $F^B = 1$, so will build up forces to eliminate this situation. However, since $F^R = -1$, there is no incentive for either side to preempt in a crisis, since the outcome is the same regardless of who goes first. (It can be argued, however, that Blue may wish to go first to "collect its winnings" in a crisis; there is no net benefit to either side from going first.)

The above discussion of Figure A-4 also implicitly raises an issue which shows up in many analyses of first strike instability and changing force structures. If Red would build up forces to improve upon $F^B = 1$, $F^R = -1$, $G = 0$, the outcomes might well pass through the region where $F^B = 1$, $F^R = 0$, $G = 1$ enroute to $F^B = 0$, $F^R = 0$, $G = 0$. Thus, eliminating arms control instabilities involves transitioning from $G = 0$ through $G = 1$ to $G = 0$ and includes first strike instabilities during the transition (e.g., the transition proceeds from Figure A-4 to Figure A-2 to Figure A-1.)

C. COMPARISON OF CRISIS STABILITY MEASURES

In this section, we compare three sets of stability measures. All three are similar in most cases. However, two are the same while one differs in the type of case illustrated in Figure A-2 above.

For convenience of this discussion, we define some new symbols for surviving value C, D, E, and F:

	<u>Side 1</u>	<u>Side 2</u>
Value at Beginning	B	R
Value Surviving when Side 1 First	$C = B - B_{K2}$	$D = R - R_{K1}$
Value Surviving when Side 2 First	$E = B - B_{K1}$	$F = R - R_{K2}$
Value Destroyed when Side 1 First	B_{K2}	R_{K1}
Value Destroyed when Side 2 First	B_{K1}	R_{K2}

The first strike instability indexes compared are as follows:

Bracken (1987)

$$\text{Payoff to Side 1} = C-D$$

$$\text{Payoff to Side 2} = F-E$$

$$\text{Crisis Instability} = (C-D) + (F-E) \text{ with range } [-2,2]$$

Wilkening, et.al (1986)

Assuming that the first striker is indifferent between the damage received and the damage inflicted, first strike incentive for Side 1 I_1 and first strike incentive for Side 2 I_2 are as follows:

$$\begin{aligned} I_1 &= \left[\frac{1}{2}R_{K1} - \frac{1}{2}B_{K2} \right] - \left[\frac{1}{2}R_{K2} - \frac{1}{2}B_{K1} \right] \\ &= \left[\frac{1}{2}(R-D) - \frac{1}{2}(B-C) \right] - \left[\frac{1}{2}(R-F) - \frac{1}{2}(B-E) \right] \\ &= \frac{1}{2}(C-D) + \frac{1}{2}(F-E) \quad \text{with range } [0,1] \end{aligned}$$

$$\begin{aligned} I_2 &= \left[\frac{1}{2}B_{K1} - \frac{1}{2}R_{K2} \right] - \left[\frac{1}{2}B_{K2} - \frac{1}{2}R_{K1} \right] \\ &= \left[\frac{1}{2}(B-E) - \frac{1}{2}(R-F) \right] - \left[\frac{1}{2}(B-C) - \frac{1}{2}(R-D) \right] \\ &= \frac{1}{2}(F-E) + \frac{1}{2}(C-D) \quad \text{with range } [0,1] \end{aligned}$$

Note that the sum of the Wilkening incentive measures for each side is one-half of the Bracken measure.

O'Neill (1985)¹

Crisis instability index

$$= \left(\frac{E}{C} - 1 \right) \left(\frac{D}{F} - 1 \right) \quad \text{with range } [0, \infty]$$

Table 1 shows six representative situations, 1 through 6, with end states C, D and E, F in the second column. The three measures are given in the third, fourth and fifth columns.

¹ Appendix F presents the more general measure of O'Neill (1987). The measure treated here of O'Neill (1985) is, however, a special case, and is that applied in the example of O'Neill (1987).

Table A-1. Three Sets of Measures for Six Representative Situations

		Range [-2,2]	Ranges [0,1], [0,1]	Range [0,1]
<u>Situation</u>	<u>C,D</u> <u>E,F</u>	<u>Bracken</u>	<u>Wilkening. et.al</u>	<u>O'Neill</u>
1	1,0 0,1	$G = 1+1= 2$	$I_1= 1, I_2 = 1$	$(-1)(-1) = 1$
2	1,0 0,0	$G = 1+0= 1$	$I_1= .5, I_2 = .5$	$(-1)(?) = ?$
3	1,.1 .1,.1	$G = .9+0= .9$	$I_1= .45, I_2 = .45$	$(-.9)(0) = 0$
4 .45	1,.1 .1,.2	$G = .9+.1= 1$	$I_1= .50, I_2 = .50$	$(-.9)(-.5) =$
5 .603	1,.1 .1,.3	$G = .9+.2= 1.1$	$I_1= .55, I_2 = .55$	$(-.9)(-.67) =$
6 .16	.9,.4 .2,.5	$G = .5+.3= .8$	$I_1= .4, I_2 = .4$	$(-.78)(-.2) =$

Situation 1 is the very crisis unstable situation of Figure A-2 above. All three measures take on their maximums, which is a desirable characteristic of the measures.

Situation 2 is the situation of Figure A-2 above, while situations 3, 4, and 5 are variants thereof. The first two measures are completely consistent for all four situations. The third measure is undefined for situation 2. For the small variation of situation 2 illustrated in situation 3, however, it states that there is stability. Since end states E and F are the same, at .1,.1, there is no incentive for Side 2 to strike and since this is the case, there is stability. The first two measures add the great incentive of Side 1 to the zero incentive of Side 2, while the third measure multiplies the two incentives. Thus situation 3 illustrates very well the significant differences in the measures. It alerts the analyst to the importance of understanding the measure.

Situations 4 and 5 of Table 1 are minor variations of situation 3. The third measure is very sensitive to minor perturbations, caused by the effects of ratios of end states rather than differences.

Situation 6 of Table 1 is an example where both sides gain marginally by striking first, Side 1 by .5 and Side 2 by .3. The first and second measures yield incentives nearer to the maximums, again due to the effect of a ratio of Side 2 end states in the case of the third measure rather than differences in Side 1 and Side 2 end states in the case of the first and second measures.

APPENDIX B

SUMMARY OF BRACKEN PAPER¹

¹. Jerome Bracken, *Stable Transitions from Mutual Assured Destruction to Mutual Assured Survival*, P.O. Box 151048, Chevy Chase, MD 20815, May 1988.

A. SUMMARY EXTRACTED FROM BRACKEN PAPER

This paper treats the strategic offensive and defensive forces of the United States and the Soviet Union. It addresses the problem of transitioning from the current state of mutual assured destruction to a future state of mutual assured survival. The paper identifies sequences of two-sided force structures which possess the properties that (1) neither side has positive first strike payoff and (2) there is no incentive to strike first in a crisis. These properties correspond, respectively, to arms race stability and crisis stability.

The paper treats both ballistic missile and air-breathing offensive forces and missile defenses and air defenses against them.

A straightforward mathematical model is presented which includes an optimized counterforce/countervalue first strike followed by a payoff and crisis instability measures are defined. Data on offensive forces and air defenses are presented. Assumptions are made on counterforce probabilities of kill and on numbers of weapons required to achieve destruction of value targets. Current offensive forces and "deep cut" (half of current) offensive forces are treated, along with four levels of air defenses on both sides. Missile defense forces on both sides are varied over a wide range.

Subject to all of the assumptions given in the paper, the results are as follows:

First, starting from the current force structures there exist several stable transitions from mutual assured destruction to mutual assured survival. These transitions involve maintaining assured destruction with air-breathing systems, joint deployment of missile defenses, and finally, joint deployment of air defenses. The current forces permit a limited initial deployment of missile defenses without inducing crisis instability.

Second, deployment by the U.S. of the Trident D-5 RV makes the limited initial deployment of SDI by the U.S. unstable. This is because the Soviet ICBMs and bombers can be destroyed in a first strike and even a modest U.S. SDI can block the second strike by the surviving, deployed, Soviet SLBMs. The Soviets would have to adopt some ICBM survivability measure, such as launch on warning or mobility, to avoid destruction by a U.S. first strike.

Third, a very interesting stable transition from mutual assured destruction to mutual assured survival is identified. It consists of the following steps: (1) a limited initial deployment of missile defenses by both sides, (2) deep cuts in missiles only, and (3) joint deployment of increased air defenses.

Finally, brief discussions of the effects of changes in the assumptions of the model and the data, and of the limitations of the analysis, are presented.

B. OBJECTIVE AND SCOPE EXTRACTED FROM BRACKEN PAPER

The purpose of this investigation is to identify sequences of Blue and Red strategic offensive and defensive force structures which possess the properties that

- (1) neither side has positive first strike payoff,
- (2) there is no motivation to strike first in a crisis, and
- (3) a stable transition can be made from mutual assured destruction to mutual assured survival.

It is of particular interest that property (2) above be identified and avoided. The analysis highlights the identification of those combinations of forces for which both sides have strong incentives to strike first in a crisis.

First strike payoff is defined as first striker fraction survival minus second striker fraction survival. Crisis instability is defined as the sum of Blue and Red first strike payoff. First strike payoff lies between -1 and 1 , the former being when the first striker survivors are 0 and the second striker survivors are 1 and the latter being when the first striker survivors are 1 and the second striker survivors are 0 . Crisis instability lies between -2 and 2 , the former characterizing situations where both sides have first strike payoff of -1 and the latter characterizing situations where both sides have first strike payoff of 1 .

Some combinations of force structures will have first strike payoff of 1 for one side and 0 for the other side. These combinations are unstable from an arms race point of view in that the second side will attempt to build up to disallow the first side the first strike payoff of 1 , converting it to 0 and thus achieving assured destruction. These combinations are also unstable from a crisis instability point of view in that in a crisis both sides might wish to strike first. The first side will wish to achieve his payoff of 1 ; the second side will wish to deny the first side his payoff of 1 , and perhaps will prefer the payoff of 0 (mutual suicide) to the first side's winning. This particular case is of great interest and will be discussed in detail.

The offensive forces treated in the paper are Blue and Red ICBMs, SLBMs, bombers and cruise missiles, including Blue and Red RVs per ICBM and SLBM and Blue and Red air breathing weapons per bomber and cruise missile. The defensive forces treated

in the paper are Blue and Red SDI and air defense characterized by thresholds over which all attacking RVs and air breathers reach their targets and by percentages of RVs and air breathers below the thresholds which reach their targets.

The abstracted situation in which the analysis is performed is a two-strike exchange. The first striker allocates all of his ICBMs and SLBMs to the attack of the second striker's ICBMs, bombers and value targets, and all of his bombers and cruise missiles to the attack of the second striker's value targets. The first striker's allocation of ICBMs and SLBMs is optimized (within a broad set of allocation options) to maximize first striker's value surviving minus second striker's value surviving, which lies between -1 and 1. The second striker does not have an optimal allocation problem since he uses all of his surviving weapons in the countervalue second strike.

In the above exchange, the counterforce RVs which exhaust the SDI constitute the first strike against ICBMs and bomber bases. The countervalue RVs and air breathing weapons which exhaust the SDI and air defense, and the countervalue RVs and air breathing weapons which leak through the SDI and air defense, constitute the first and second strikes against value targets. No credit is given for leakage in the counterforce attack since the attacker would probably not plan on the basis of leakage. If leakage is small (say 10 percent) the counterforce damage will not add greatly to the first striker's payoff unless the exhaustion threshold is very high. If leakage is large, however (say 25 percent), there might be significant damage to forces when there are medium to large exhaustion thresholds. Thus the present analysis may underestimate first strike payoff in the presence of leakage.

In the two-strike exchange, the counterforce RVs are characterized by their probabilities of kill against ICBMs and bomber bases. The countervalue RVs and air breathers are characterized by the percent of the value target complex expected to be killed by each arriving weapon, which depends on the makeup of the target complex and on the terminal defenses.

Force structures for current offensive forces and current air defense forces are taken from *The Military Balance, 1985-1986* (Reference [9]). Effectiveness parameters for offensive and defensive forces are assumed.

The overall framework of the analysis of stable transitions considers eight cases of joint Blue and Red offensive forces and air defense forces. Offensive forces are (1) current and (2) deep cuts. Air defense forces are (1) none, (2) current, (3) medium (double of

current), and (4) large (quadruple of current). For each of the eight combinations of offensive forces and air defense forces, Blue and Red SDI capabilities are varied over a wide range of SDI thresholds.

C. FORMAT OF PRESENTATION OF RESULTS

Figure B-1 is extracted from the Bracken paper. It shows regions of positive Blue first strike payoffs ($FB > 0$); positive Red first strike payoff ($FR > 0$) and regions where they overlap and have large crisis instability ($G > 1$).

From a starting point of current Blue and Red air defenses and no Blue and Red missile defenses, it shows three stable transitions from the current state of mutual assured destruction to mutual assured survival. Path A involves (1) reducing air defenses to zero, (2) deploying missile defenses, (3) redeploying air defenses, (4) deploying more air defenses, (5) performing deep cuts in offensive forces, and, (6) reducing air defenses and missile defenses.

Path B consists of adjustments in Blue and Red air breathing forces, followed by joint deployment of missile defenses, followed by (4) through (6) above.

Path C allows an interim deployment of missile defenses first, followed by increases in Blue air breathers and/or decreases in Red air defenses, followed by further joint deployment of missile defenses, followed by (4) through (6) above.

1. Exchange Model

The exchange model used by Bracken is completely documented in the paper. The steps are as follows, summarized in terms of a Blue first strike on Red.

Red ICBMs killed in the first strike by Blue ICBM and SLBM RVs are a function of attacking Blue RVs exceeding the Red SDI threshold which are allocated to Red ICBMs, the number of Red ICBMs, and the Blue ICBM and SLBM RV single-shot probability of kill.

Red bombers killed in the first strike by Red ICBM and SLBM RVs are a function of attacking Blue RVs exceeding the Red missile defenses threshold which are allocated to Red bomber bases, the number of Red bomber bases, the number of Red bombers and the RV single-shot probability of kill.

Red value killed in the first strike by Blue ICBMs, SLBMs, bombers, and cruise missiles is a function of Blue RVs exceeding the Red SDI which are allocated to value targets, Blue RVs leaking through the Red SDI which are allocated to value targets, Blue air breathing exceeding the Red air defense, and Blue air breathers leaking through the Red air defense. The fraction of value killed per weapon is a parameter of the model.

Blue value killed on the second strike is a fraction of surviving Red ICBM RVs and SLBM RVs exceeding the missile defense threshold and those leaking through the defenses, and surviving Red bomber delivered weapons and cruise missiles exceeding the air defense threshold and leaking through the air defense.

The optimal first striker counterforce/countervalue allocation is determined for each case being analyzed from among a set of 15 possible combinations of allocations of weapons to ICBMs, bomber bases and value targets.

2. Scope of Cases Considered

In addition to the base case, the principal cases analyzed are as follows:

- (1) effect of deploying Trident D-5 RV's,
- (2) effect of deep cuts in missiles only,
- (3) effect of assuming 1500 weapons to destroy all of Red value and 500 weapons required to destroy all of Blue value,
- (4) effect of air defense leakage,
- (5) effect of "generated" forces,
- (6) effect of adding tactical air-delivered bombs to Blue first strike.

Subsequent work by Bracken addresses (1) reducing the number of weapons required to completely destroy Blue and Red to 100, (2) increasing that number to 2500, and (3) assuming differing effects of Blue and Red air breathing weapons due to Red surface to air missile defenses.

APPENDIX C

EXPLORATION OF CANAVAN PAPER¹

¹ Gregory H. Canavan, Simple Discussion of the Stability of Strategic Defense, LA-UR-85-1377, Los Alamos National Laboratory, April 1985.

A. INTRODUCTION

This analysis builds on the simple and elegant treatment of strategic stability contained in Gregory H. Canavan, *Simple Discussion of the Stability of Strategic Defense*, LA-UR-85-1377, Los Alamos National Laboratory, April 1985.

Canavan's model is summarized and the data utilized in his paper are illustrated in Cases 1 and 2. A variation of Case 1 is given as Case 3.

A variation of Canavan's model which explicitly defines assured destruction is presented. It is applied in Cases 4 and 5 with more realistic forces than those of Cases 1-3. While Cases 1-3 imply that deployment of interceptors is stable under a wide range of assumptions, Cases 4 and 5 suggest the conclusion that it will be very difficult to perform a stable deployment if SLBMs can kill ICBMs.

Finally, two-sided aspects of midcourse defense are explored.

B. CANAVAN'S MODEL

We adopt Canavan's notation completely. Let the two sides be called unprime and prime. Define:

L, L' = number of launchers (ICBMs)

M, M' = number of RVs per launcher

B, B' = number of RVs on untargetable systems (SLBM RVs)

I, I' = number of interceptors

Assume that the defense of launchers by interceptors is adaptive preferential. In this case the allocation of attacking RVs to launchers can be observed by the defender and then interceptors can be assigned to protect a subset of the launchers. This is equivalent to assuming both impact-point prediction by the defense (no MARVs in the attack) and the capability to act on it. The attack must be essentially simultaneous in order that the defense can be utilized to protect the least-attacked launchers. In this situation, it is well-known that the optimal attack should be uniformly allocated over the launchers and the optimal defense should be allocated so as to save as large a subset as possible of the launchers, after observing the attack. (One further assumption is that all attacking RVs and all interceptors are perfect).

For illustration, if prime's R' RVs attack launchers protected by unprime's I interceptors, the number of unprime's launchers surviving is $L\left(\frac{I}{R'}\right)$. Expressing this in RVs, if all of prime's ICBMs attack all of unprime's ICBMs the number of surviving unprime ICBM RVs is $ML\left(\frac{I}{M'L'}\right)$. If all of prime's ICBMs and SLBMs attack unprime's ICBMs the number of surviving unprime ICBM RVs is $ML\left(\frac{I}{M'L'+B'}\right)$.

Canavan defines assured survival of unprime and prime in his equations (14) and (15) as:

$$\text{AS: } I \geq B' + M'L' \quad (1)$$

$$\text{AS': } I' \geq B + ML \quad (2)$$

The interpretation of these equations is that the defense can counter all RVs of the other side. Mutual assured survival (MAS) is defined as the region where both (1) and (2) hold.

Canavan defines conditional survival of unprime and prime in his equations (17) and (16), assuming only ICBMs attack ICBMs, as:

$$\text{CS: } I \geq B' + M'L' \left(\frac{I'}{ML} \right) \quad (3)$$

$$\text{CS': } I' \geq B + ML \left(\frac{I}{M'L'} \right) \quad (4)$$

Conditional survival means that one side or the other can survive only by going first and is not a good situation. If CS holds, for instance, then unprime's ML ICBM RVs can attack prime's M'L' ICBM RVs protected by I' interceptors, after which unprime will have sufficient interceptors to ward off the counterattack by prime's B' SLBM RVs and $M'L' \left(\frac{I'}{ML} \right)$ surviving ICBM RVs.

Canavan essentially extends the above equations for conditional survival in his equations (20) and (19), allowing SLBMs to join in the attack on ICBMs, resulting in what we denote as:

$$\text{DS: } I \geq B' + M'L' \left(\frac{I'}{ML+B} \right) \quad (5)$$

$$\text{DS: } I' \geq B + ML \left(\frac{I}{M'L'+B'} \right) \quad (6)$$

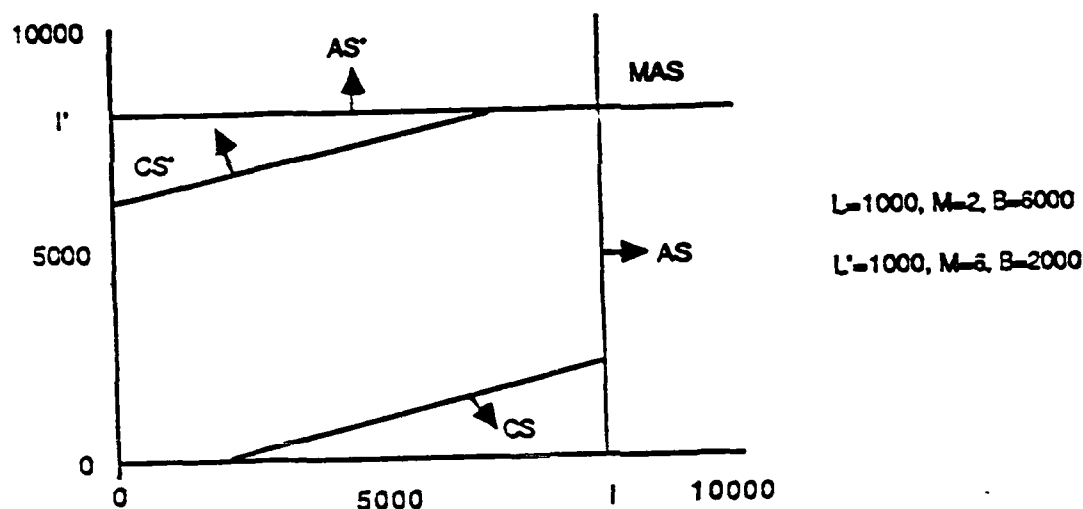
C. APPLICATION WITH CANAVAN'S DATA

The following data, from Canavan's paper, are assumed as Case 1 for application of the model, where unprime corresponds to U.S. forces and prime corresponds to Soviet forces:

$$L = 1000, M=2, B = 6000$$

$$L' = 1000, M'=6, B'= 2000$$

For the above data, Figure C-1 (Canavan's Figure C-2) shows AS, AS', CS, CS' and MAS. There are large numbers of mutual deployments of I and I' which do not come close to CS or CS', although if deployment is along the diagonal it is somewhat close to CS when it begins and somewhat close to CS' when it enters MAS. (A reverse S-shaped path could preserve the largest distance from CS and CS' during the deployment.)



$$AS: I \geq 2000 + 6 \times 1000$$

$$AS': I' \geq 6000 + 2 \times 1000$$

$$CS: I \geq 2000 + 6 \times 1000 \left(\frac{I'}{2 \times 1000} \right)$$

$$CS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{6 \times 1000} \right)$$

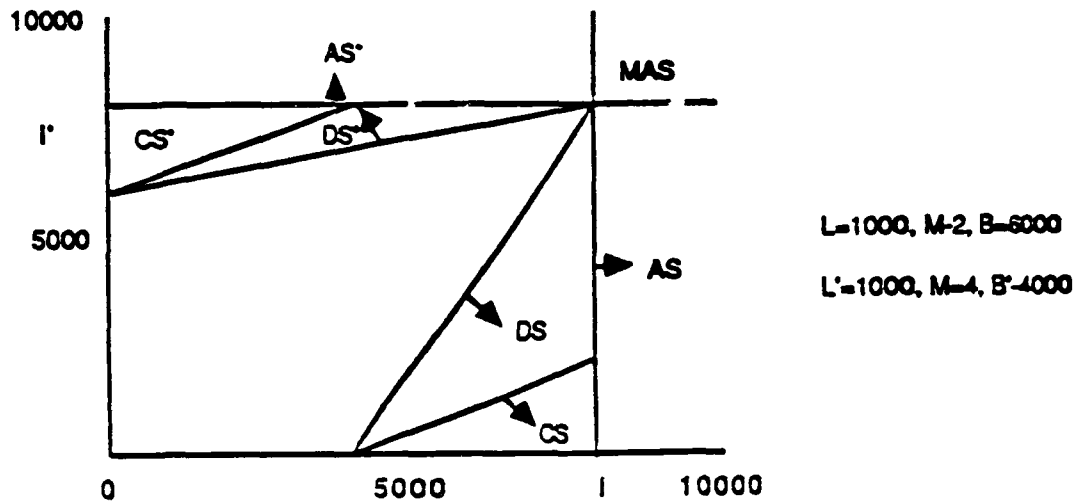
Figure C-1. Results for Case 1

Prior to including SLBMs in the attack, Canavan changes the data of the problem. He assumes for what we term Case 2 that for prime 2000 RVs are moved from targetable to untarvable forces, as follows:

$$L = 1000, M=2, B=6000$$

$$L' = 1000, M'=4, B'=4000$$

For the above data, Figure C-2 (Canavan's Figure C-4) shows AS, AS', CS, CS', DS and DS', as well as MAS.



$$AS: I \geq 4000 + 4 \times 1000$$

$$AS': I' \geq 6000 + 2 \times 1000$$

$$CS: I \geq 4000 + 4 \times 1000 \left(\frac{I'}{2 \times 1000} \right)$$

$$CS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{4 \times 1000} \right)$$

$$DS: I \geq 4000 + 4 \times 1000 \left(\frac{I'}{2 \times 1000 + 6000} \right)$$

$$DS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{4 \times 1000 + 4000} \right)$$

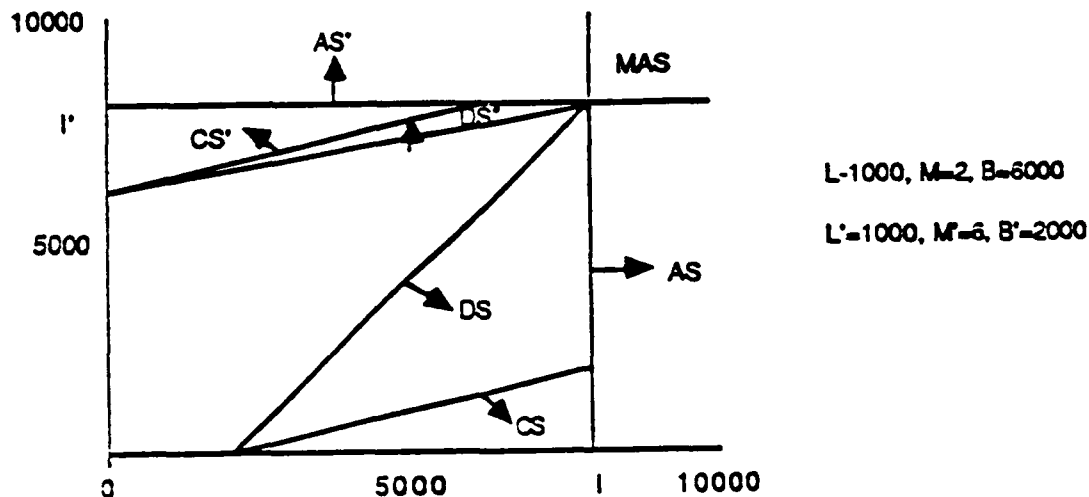
Figure C-2. Results for Case 2

The CS region is pushed to the right, which is desirable, and the CS' region is pushed to the top, which is also desirable. The channel of DS and DS' to $I=8000$ and

$I'=8000$ is very wide at the beginning, allowing initial deployments without coming near to them. The channel narrows at the end but is still feasible.

Shifting forces from targetable to nontargetable forces is stabilizing in terms of reducing CS and CS'. Allowing SLBMs to target ICBMs, and thus producing DS and DS' as well as CS and CS', makes deployment more difficult, but in this case it is still feasible without entering the DS as DS' regions.

Case 3 below is not shown by Canavan, though it is the direct extension of Case 1. Case 3, in which submarines in the current force structure are permitted to attack ICBMs, implies that to maintain stability while staying as far as possible from DS and DS' interceptor deployments should be asymmetric, with significantly more for the Soviets than the U.S. at the lower levels of defenses.



$$AS: I \geq 2000 + 6 \times 1000$$

$$AS': I' \geq 6000 + 2 \times 1000$$

$$CS: I \geq 2000 + 6 \times 1000 \left(\frac{I'}{2 \times 1000} \right)$$

$$CS': I \geq 6000 + 2 \times 1000 \left(\frac{I}{6 \times 1000} \right)$$

$$DS: I \geq 2000 + 6 \times 1000 \left(\frac{I'}{2 \times 1000 + 6000} \right)$$

$$DS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{6 \times 1000 + 2000} \right)$$

Figure C-3. Results for Case 3

D. A SLIGHT VARIATION OF CANAVAN'S MODEL

The standard concept of assured destruction can be very naturally represented by a variation of Canavan's model. Two levels of assured destruction are examined here. The first is based on the assumption that the delivery of a single RV in a second strike constitutes finite deterrence and is sufficient. This is analogous to Canavan's conditional stability. The second requires that 1000 or more RVs be delivered in the second strike.

For unprime and prime, assuming that SLBMs can attack ICBMs, the first type of assured destruction is present when the following is true:

$$AD_1: I' < B + ML \left(\frac{I}{M'L' + B'} \right) \quad (7)$$

$$AD_1': I < B' + M'L' \left(\frac{I'}{ML + B} \right) \quad (8)$$

That is, when interceptors are less than the sum of nontargetable RVs plus protected targetable RVs, assured destruction is present. Equivalently, AD_1 means that unprime possesses assured destruction when prime has fewer interceptors than unprime has retaliating RVs. It follows directly that for I, I' pairs where both (7) and (8) hold, mutual assured destruction MAD_1 is present.

The second type of assured destruction is present when the following is true:

$$AD_2: I' < B + ML \left(\frac{I}{M'L' + B} \right) - 1000 \quad (9)$$

$$AD_2': I < B' + M'L' \left(\frac{I'}{ML + B} \right) - 1000 \quad (10)$$

That is, each side possesses assured destruction when the two sides have fewer interceptors than those required to reduce the second strike to less than 1000 RVs. Mutual assured destruction MAD_2 is present when both (9) and (10) hold.

E. APPLICATION WITH REVISED DATA

To come closer to a realistic case, consider the following offensive force structures:

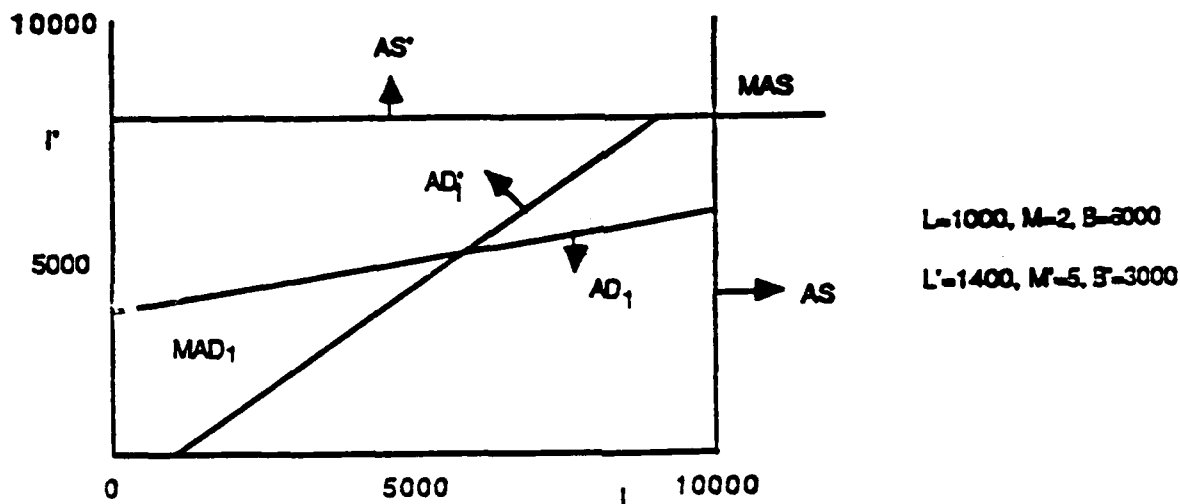
$$L = 1000, M = 2, B = 6000$$

$$L' = 1400, M' = 5, B' = 3000$$

This corresponds to the current U.S. and Soviet force structures. For Case 4, assume a steady state deployment of U.S. and Soviet submarines consisting of 2/3 of U.S.

submarines and 1/3 of Soviet submarines. Assume that all SLBMs can take part in the first strike but only deployed SLBMs can take part in the second strike.

Figure C-4 presents AS, AS', MAS, AD₁, AD'₁, and MAD₁ for Case 4. Note that MAD₁ is present only in a region in the lower left corner.



$$AS: \quad I \geq 3000 + 5 \times 1400$$

$$AS: \quad I' \geq 6000 + 2 \times 1000$$

$$AD_1: \quad I' < 4000 + 2 \times 1000 \left(\frac{I}{5 \times 1400 + 3000} \right)$$

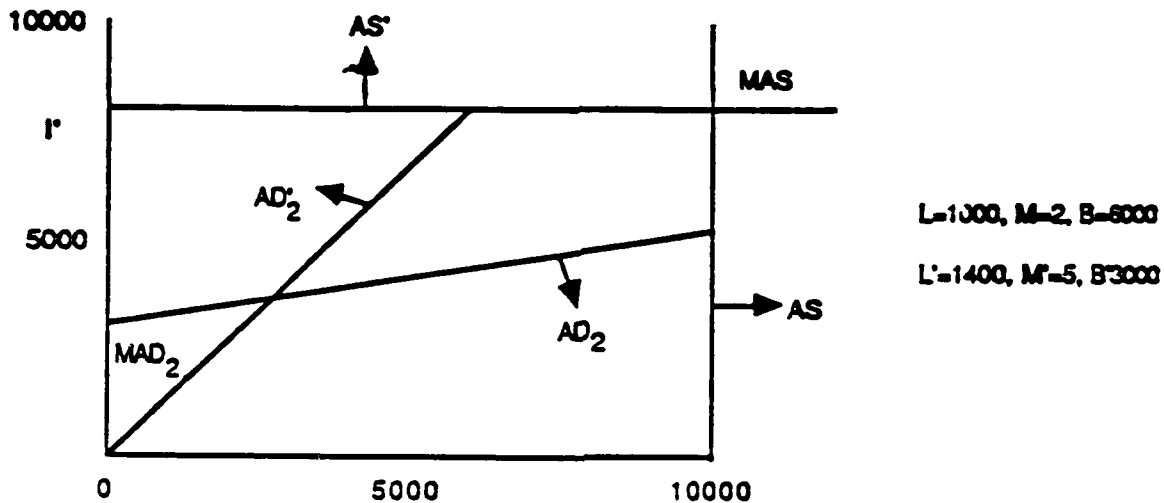
$$AD_1': \quad I < 1000 + 5 \times 1400 \left(\frac{I'}{2 \times 1000 + 6000} \right)$$

Figure C-4. Results for Case 4

A deployment of interceptors up to the level $I \cong 5000$, $I' \cong 5000$ can be permitted and still retain mutual assured destruction. To remain in the center of the MAD₁ region involves initial interceptor deployments slightly in favor of the Soviets. Beyond this point either side could carry out a first strike on ICBMs and completely ward off the second strike by the deployed SLBMs and surviving ICBMs.

Figure C-5 presents Case 5, the same results for AD₂ and AD'₂. The more stringent specification of assured destruction substantially shrinks the region of MAD₂ with respect to the space of permissible interceptor deployments. The conclusion to be drawn from Cases 4 and 5 is that by adopting standard views of mutual assured destruction,

making reasonable assumptions on submarine deployments, and allowing SLBMs the capability to destroy ICBMs the problem of two-sided deployment of interceptors is severely complicated.



$$AS: \quad I \geq 3000 + 5 \times 1400$$

$$AS': \quad I' \geq 6000 + 2 \times 1000$$

$$AD_2: \quad I' < 4000 + 2 \times 1000 \left(\frac{I}{5 \times 1400 + 3000} \right) - 1000$$

$$AD'_2: \quad I < 1000 + 5 \times 1400 \left(\frac{I'}{2 \times 1000 + 6000} \right) - 1000$$

Figure C-5. Results for Case 5

F. NOTE ON CANAVAN'S ANALYSIS OF MIDCOURSE AND BOOST-PHASE DEFENSES

Returning to Canavan's model, let us consider the analysis of midcourse and boost-phase defenses.

First, consider midcourse defense. Canavan postulates that unprime has N midcourse interceptors equal to a specified fraction of the threat, or

$$N = f(m'L' + B').$$

He also postulates on analogous midcourse intercept force for prime, or

$$N' = f'(mL + B).$$

When either N or N' midcourse interceptors are included, conditional survival is as follows:

$$DS = I \geq B' + M'L' \left(\frac{I'}{ML + B - N'} \right)$$

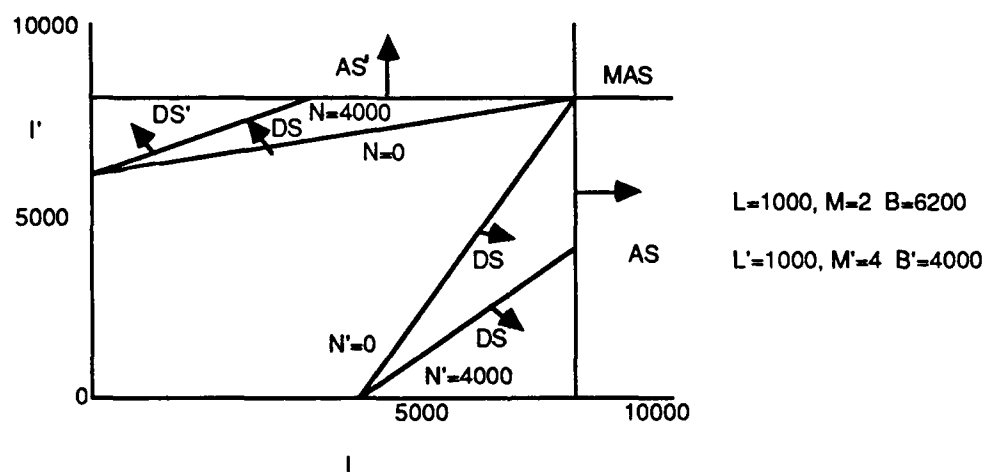
$$DS' : I' \geq B + ML \left(\frac{I}{M'L' + B' - N} \right)$$

Figure C-6 displays the effect of midcourse interceptors for N=0 and N=4000 (Canavan's f=0 and f=.5) and for N'=0 and N'=4000 (Canavan's f'=0 and f'=.5). Figure C-6 is analogous to Canavan's Figure C-7. Canavan points out that an addition of N=4000 or N'=4000 interceptors opens up the region of stable transitions.

However, the above argument is incomplete. If both sides add midcourse interceptors, the equations for conditional survival are:

$$ES = I + N \geq B' + M'L' \left(\frac{I'}{ML + B - N'} \right)$$

$$ES' = I' + N' \geq B + ML \left(\frac{I}{M'L' + B' - N} \right)$$



$$AS: I \geq 4000 + 4 \times 1000$$

$$AS': I' \geq 6000 + 2 \times 1000$$

$$DS: I \geq 4000 + 4 \times 1000 \left(\frac{I'}{2 \times 1000 + 6000 - N'} \right)$$

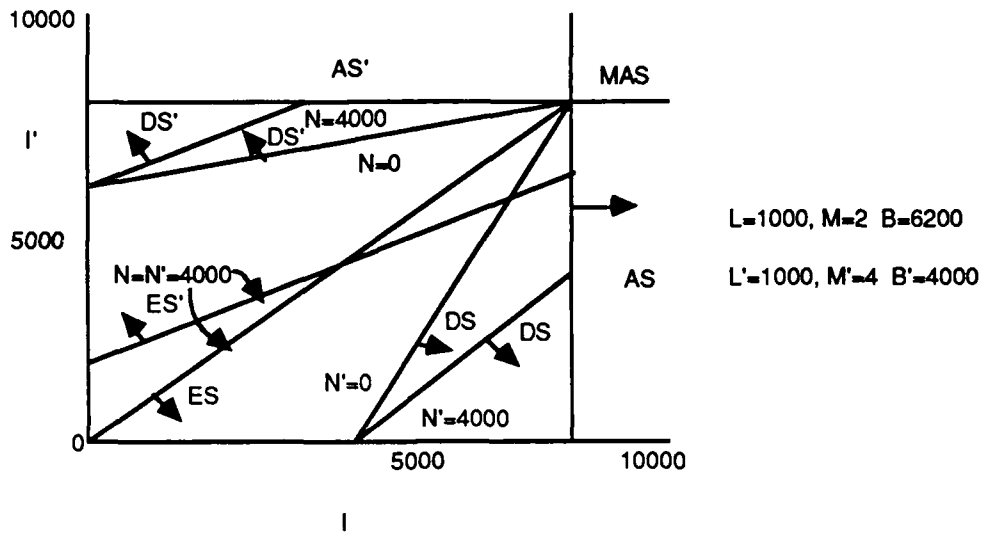
$$DS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{4 \times 1000 + 4000 - N} \right)$$

Figure C-6. Results For One-Sided Deployment of Midcourse Defense

In addition to midcourse interceptors being introduced on one side to subtract RVs from the first strike, midcourse interceptors are introduced on the other side to ward off the second strike.

Figure C-7 builds upon Figure C-6 to illustrate the effect of both sides adding 4000 interceptors. The areas of conditional survivability increase dramatically. There is only a small area where neither side has conditional survivability. It is easy to understand why this occurs, namely, adaptive preferential use of I for protecting land-based missiles is more efficient than subtractive use of N, but both I and N have equal ability to ward off a counterattack. Therefore, addition midcourse interceptors on both sides is destabilizing as compared to adding terminal interceptors on both sides.

Boost-phase defense operating in a subtractive mode would have exactly the same characteristics as shown above in the analysis of midcourse defense. Boost-phase defense is less efficient than adaptive preferential defense of land-based interceptors while having the same capabilities to ward off a second strike.



$$AS: I \geq 4000 + 4 \times 1000$$

$$AS': I' \geq 6000 + 2 \times 1000$$

$$DS: I \geq 4000 + 4 \times 1000 \left(\frac{I'}{2 \times 1000 + 6000 - N'} \right)$$

$$DS': I' \geq 6000 + 2 \times 1000 \left(\frac{I}{4 \times 1000 + 4000 - N} \right)$$

$$ES: I + N \geq 4000 + 4 \times 1000 \left(\frac{I'}{2 \times 1000 + 6000 - N'} \right)$$

$$ES': I' + N' \geq 6000 + 2 \times 1000 \left(\frac{I}{4 \times 1000 + 4000 - N} \right)$$

Figure C-7. Results For Two-Sided Deployment Of Midcourse Defense

APPENDIX D

SUMMARY OF CHRZANOWSKI PAPER¹

¹ Paul L. Chrzanowski, Transition to Deterrence Based on Strategic Defense, E&TR, January-February 1987.

A. EXCHANGE MODEL

An exchange model is adopted to compute surviving value targets for both sides. Each side has untargetable offensive missiles, untargetable air-carried weapons, targetable missiles and targetable value targets. Offensive forces are used against the opposing side's targetable offensive missiles and targetable value targets. Each side has defensive weapons to intercept ballistic missiles. In the first stage of the exchange, one side uses all of its weapons to attack the opponent's silos and value targets. The opponent uses all of its interceptors in defense against ballistic missiles. Air-carried weapons are used against value targets only. (There are no defenses against air-carried weapons). In the second stage of the exchange the side attacked retaliates by sending all of its surviving weapons against the aggressor's value targets. The aggressor uses its interceptors to defend its value targets against the ballistic missile weapons. (There are again no defenses against the air-carried weapons).

The value functions, used to optimize attack and defense tactics, depend only on number of surviving value targets. Either the difference in surviving targets or the ratio of surviving targets measures the value of the exchange.

B. MEASURES

The results of the exchange are characterized by value functions calculated for Red's going first and for Blue's going first. The measures are normalized such that the best possible outcome for Red is +1 and the best outcome for Blue is -1.

The crisis stability index is defined as $V_{Rfirst} - V_{Bfirst}$. It quantifies the comparative regrets that each side would have if it chose to wait and had to retaliate rather than to attack. The crisis stability index varies between -2 and 2, the latter corresponding to the most unstable case. (Since the value increases with greater instability, it would have perhaps been better called an "instability index").

C. PRESENTATION SCHEME

Most of the presentations show the crisis stability index and the value function of the first strike as a function of the ratio of Blue interceptors to Red attackers and of Red interceptors to Blue attackers. Figure D-1 is a typical presentation of results. The value function of the first strike is the magnitude of the value function V that one of the two sides can achieve by striking first. It is generated by combining the positive regions of V_R first

with the positive regions of $-V_B$ first. When this magnitude is small neither side can "win" a strategic exchange.

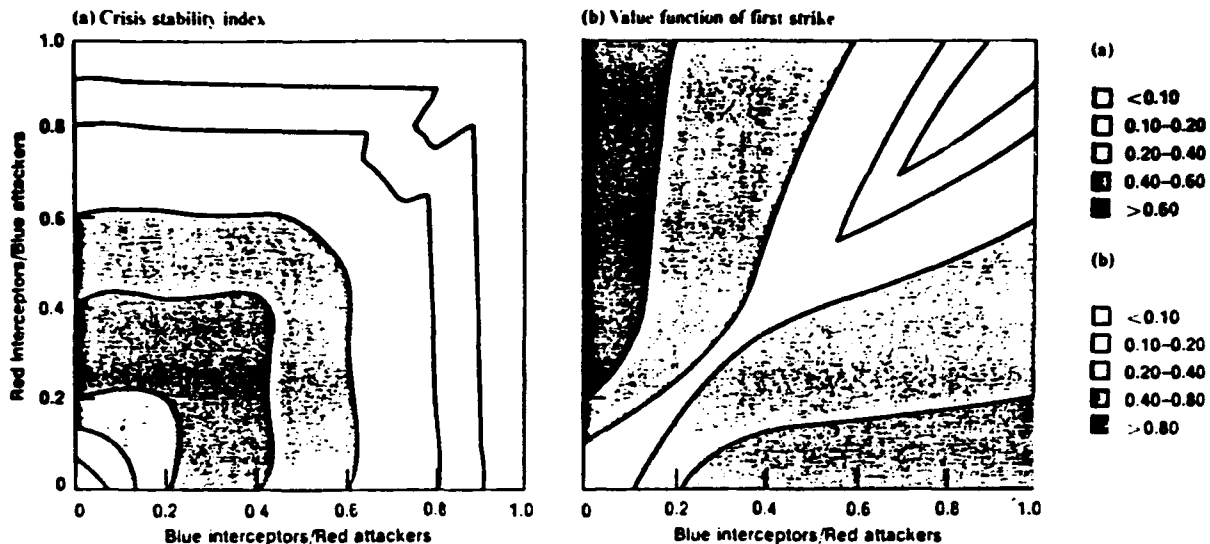


Figure 6. Crisis stability (a) and force balance (b) for the case where 75% of each side's forces are based in vulnerable silos and preferential defenses are being deployed. These graphs are analogous to those in Figure 2. The difference is that more of the offensive forces are vulnerable to attack, which causes a large regret for not striking first (a large crisis stability index value) when I/A is about 0.30 for either side. Also, the channel of force balance is narrow to nonexistent, depending on the criterion considered.

Figure D-1. Figure 6 of Chrzanowski Paper--75% of Forces in Vulnerable Silos and Preferential Defenses

D. TYPES OF DEFENSE

Preferential defense and randomly subtractive defense are analyzed. With a preferential defense, the defender is able to identify the aimpoints of attacking RVs and protect the largest possible subset of the targets, leaving the others undefended. When aimpoints cannot be identified and RVs are intercepted at random the defense is randomly subtractive.

E. SURVIVABILITY OF FORCES

Survivability of forces is highlighted in the analysis. Figure D-2 shows the case where 25% of the forces are based in vulnerable silos and preferential defense is deployed.

The crisis stability index and value functions of the first strike take on much lower values than in Figure D-1.

Comparison of Figures D-1 and D-2 leads Chrzanowski to suggest that a possible defensive buildup which would be crisis-stable and maintain the strategic balance is to first make offensive forces as survivable as possible and then deploy defenses so that the ratio of interceptors to attackers is the same for both sides. In terms of Figures D-1 and D-2 this transition is to (1) move from the lower left corner Figure D-1 to the lower left corner of Figure D-2, and (2) deploy defenses symmetrically to the upper right corner of Figure D-2.

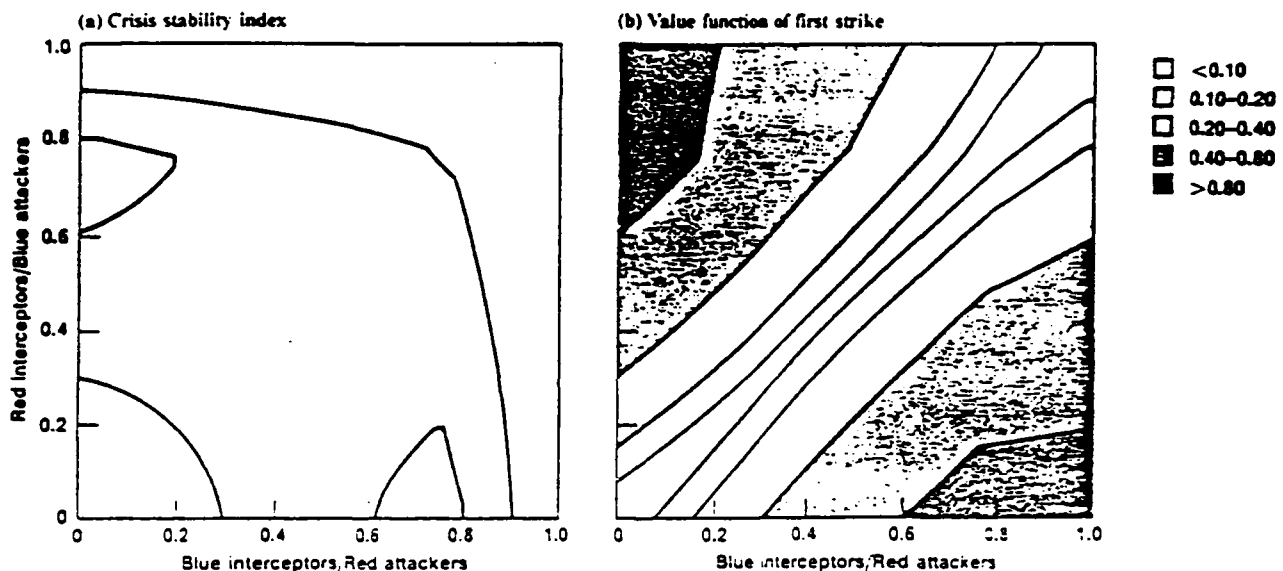


Figure 2. Crisis stability and force balance for the case where 25% of each side's forces are based in vulnerable silos and preferential defenses are being deployed. (a) A contour plot of the crisis stability index ($V_{\text{Red first}} - V_{\text{Blue first}}$), a measure of how much worse off either side would be in choosing not to strike first in a crisis situation (and, instead, having to retaliate). (b) A contour plot depicting the magnitude of the value function (V) that one of the two sides can achieve by striking first. It is generated by combining the positive value regions of $V_{\text{Red first}}$ with the positive value regions of $-V_{\text{Blue first}}$. Where this magnitude is small, neither side can "win" a strategic exchange because there is a balance of forces. Notice in this example that there is crisis stability with most any path from no defenses to mutual, highly competent defenses, and only along a channel near the diagonal is there strategic force balance. Where there is stability but no balance, one of the sides can be subjected to coercion.

Figure D-2. Figure 2 of Chrzanowski Paper--25% of Forces in Vulnerable Silos and Preferential Defense

F. TERMINAL DEFENSE OF SILOS AND ALERT RATE OF SLBMS

It is pointed out that terminal defense of silos is a reasonable first step in making them survivable, but if defenses can be used to protect other targets they may be valuable to the first striker in protecting against the counterstrike against value targets. Thus terminal defense of ICBMs must be restricted to that purpose.

Increasing the alert rate of SLBMs might significantly affect the ratio of interceptors to attackers during a deployment of defenses. This issue is discussed briefly.

G. ROLE OF AIR-DELIVERED WEAPONS

The difficulties of making bomber/missile tradeoffs are discussed, there being many attributes of the two systems which are very different--the number of weapons, uncertain defenses not limited by agreement and reusability of bombers. As ballistic missile defenses were being introduced, the capabilities and characteristics of the bomber force would become increasingly important.

H. IMPACT OF UNCERTAINTIES

"Conservative" planners can assign high capabilities to opposing forces (offensive and defensive) and low capabilities to their own forces. When this is done, the channels of stable deployment look different to Blue and Red. The intersection of the channels may involve much smaller, if any, regions of acceptability.

Another aspect of uncertainty, however, is the reduction in optimism of any first striker. During a crisis, a risk-averse aggressor may be more deterred, thus making the situation more stable. (It should be noted that his argument may not go far enough. Though the putative first striker may calculate his first strike payoff as lower, he may calculate the opponent's first strike payoff as higher and be led to strike to deny the opponent the larger payoff).

APPENDIX E

SUMMARY OF KENT AND DEVALK PAPER*

* Glenn A. Kent and Randall J. DeValk, *Strategic Defenses and the Transition to Assured Survival*, R-3369-AF, The Rand Corporation, October 1986.

A. INTRODUCTION

This summary contains as an Annex the full Summary extracted from the Kent and DeValk paper.

B. BASE CASE DATA AND FORMAT OF RESULTS

The base case data and format of results are shown in Figure E-1.

Soviet assured survival exists when Soviet defense potential is equal to a greater than 5,000 (the number of as RVs) and U.S. assured survival exists when U.S. defense potential is equal to or greater than 7,000 (the number of Soviet RVs).

The data identify number of "killer RVs" within the total number of ICBM RVs (1,500 of 2,000 U.S. ICBM RVs, with $P_k=.4$ and 5,000 of 6,000 Soviet ICBM RVs, with $P_k=.7$).

A significant asymmetry of killer RVs versus ICBMs exists, namely, 1,500 U.S. killer RVs ($P_k=.4$) vs 1,400 Soviet ICBMs and 5,000 Soviet killer RVs ($P_k=.7$) vs 1,000 U.S. ICBMs. Also, a significant asymmetry of nontargetable RVs (deployed SLBM RVs) exists, namely, 3,000 U.S. nontargetable RVs versus 1,000 Soviet nontargetable RVs.

C. ANALYSIS OF BASE CASES

The key concept of *conditional survival* is fundamental to the Kent and DeValk work. It is defined as the region of two-sided defense deployments where one side can strike first and protect itself completely against the second strike. Note that *conditional survival* is equivalent to *assured destruction denial* if assured destruction is defined to be the ability to deliver one or more weapons in a second strike. That is, for a particular two-sided defense deployment, one side's possession of conditional survival is equivalent to denial of the other side's possession of assured destruction.

Figure E-2 presents Kent and DeValk's Base Case I, with discriminating strategic defenses. Discriminating strategic defenses engage only killer RVs in a random subtractive manner, with ordered fire (uniform allocation of defending shots to RVs).

Note that a symmetric deployment of defenses intersects the Soviet conditional survival region when U.S. and Soviet defense potentials equal 4,000 or so. Note also that there is a wide channel between the Soviet conditional survival and U.S. conditional

survival region. Non-symmetric deployments of nationwide defenses could easily be designed to avoid the areas of conditional survival.

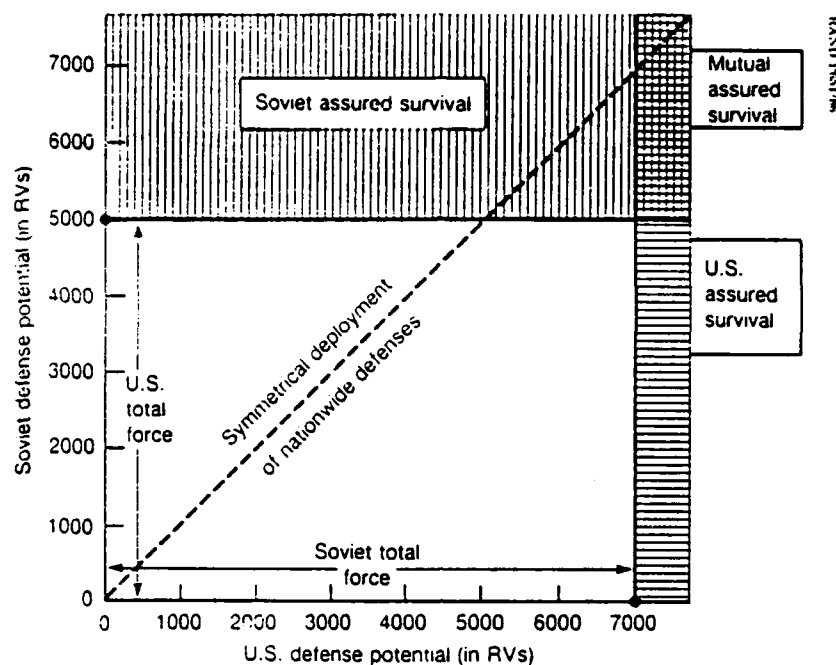


Figure E-1. Figure 1 of Kent and DeValk -- Base Case Data and Format of Results

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

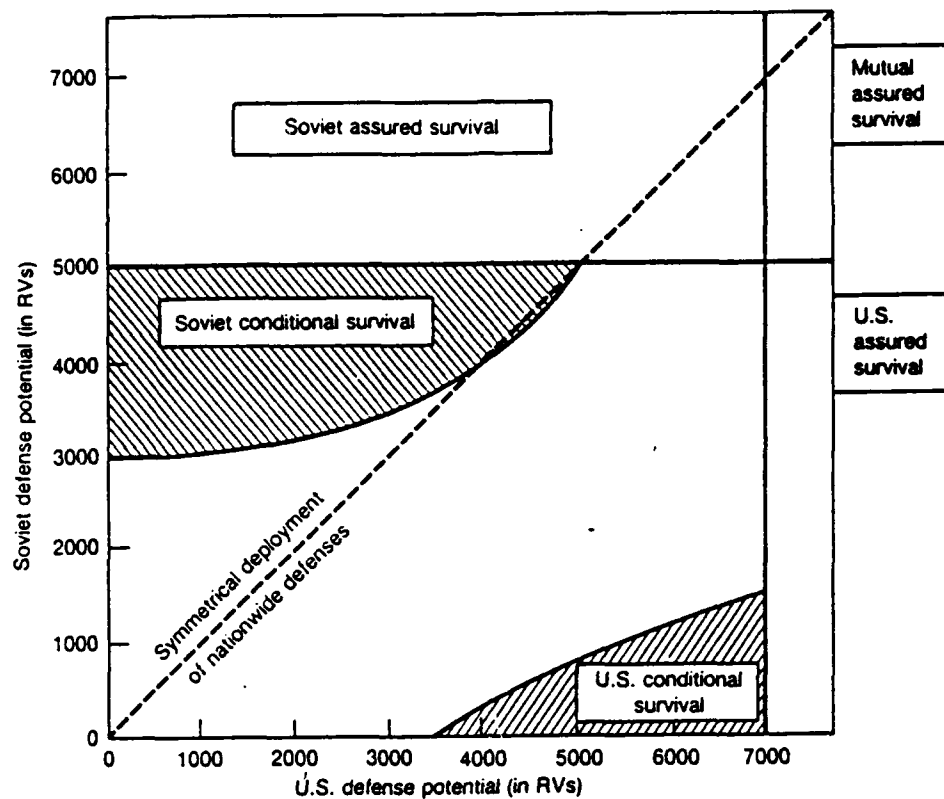


Figure E-2. Figure 4 of Kent and DeValk -- Results of Base Case I with Discriminating Strategic Defenses

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

Figure E-3 presents Kent and DeValk's Base Case II, with nondiscriminating strategic defenses. Nondiscriminating strategic defenses engage all RVs in a random subtractive manner, with ordered fire.

Symmetric deployments of nationwide defense in this case encounters Soviet conditional survival when U.S. and Soviet defense potential reaches just over 3,000. Furthermore, there is no wide channel between Soviet conditional survival and U.S. conditional survival.

1. Analysis of Selected Excursions

Four of Kent and DeValk's excursions from Base Case I will be presented here, as follows:

1. U.S. deployment of 100 MX in Minutemen silos
2. U.S. deployment of 1,000 small ICBMs on hardened mobile launchers
3. U.S. deployment of local defenses of ICBMs
4. U.S. deployment of 1,000 small ICBMs and local ICBM defense

Figure E-4 shows the effect of the U.S. deploying 1,000 more killer RVs with P_k of .7 and 500 fewer killer RVs with P_k of .4. The region of U.S. conditional survival increases significantly. The region of Soviet conditional survival decreases slightly. A symmetric deployment of nationwide defenses does not enter the region of Soviet conditional survival, though it comes close.

Figure E-5 shows the effect of the U.S. deploying 1,000 small ICBMs on hardened mobile launchers. Assuming that the deployment area is 10,000 square miles and the effective kill bombardment area per Soviet RV is 4 square miles, there is an effective addition of 2,500 aimpoints to the previous 1,000 aimpoints. The region of Soviet conditional survival shrinks appreciably because of this. Addition of 1,000 RVs increases the region of U.S. conditional survival slightly more than in Figure E-4 because there is a net addition of 1,000 rather than 700 U.S. RVs. There is a wide channel between Soviet unconditional survival and U.S. conditional survival.

Figure E-6 shows the effect of U.S. deployment of 1,000 and 2,000 units of local defense in what is termed a semipreferential mode. The semipreferential mode is equivalent to preallocated preferential defense when the attack size is known and defenses are perfect (see the Appendix of the Kent and DeValk paper). When U.S. defense potential is between 2,000 and 4,000 RVs the effect U.S. local defense of ICBMs is fairly substantial, significantly shrinking the region of Soviet conditional survival. However, when the U.S. and Soviet defense deployments reach 5,000 each, the symmetric deployment line comes very close to the region of Soviet conditional survival. Thus Kent and DeValk conclude

that local defenses do not by themselves provide a safe transition (unlike, for instance, the excursion of deploying 1,000 small ICBMs in hardened mobile launchers presented in Figure E-5).

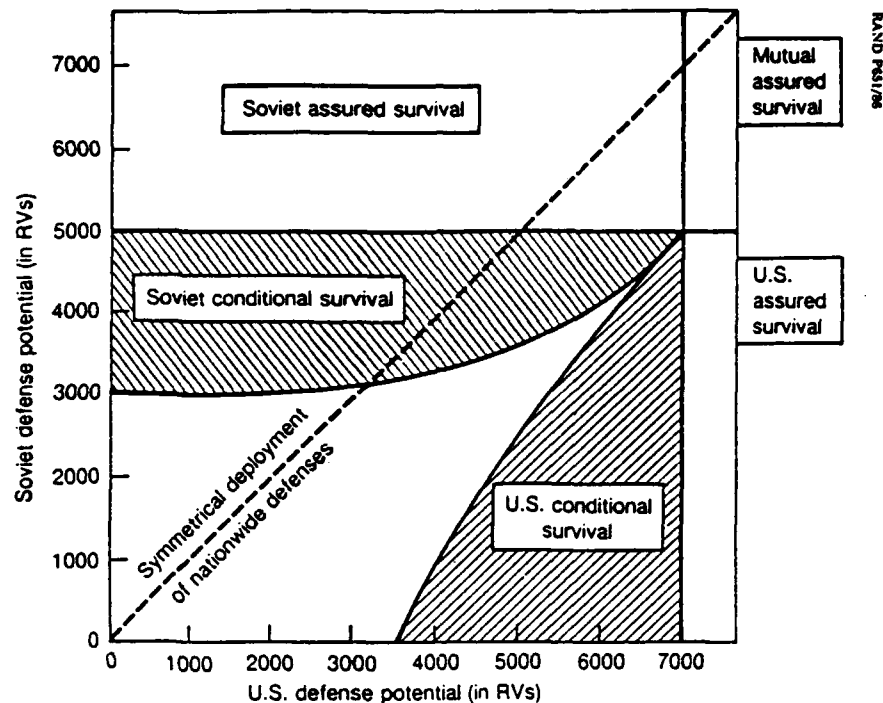


Figure E-3. Figure 5 of Kent and DeValk – Results of Base Case II with Nondiscriminating Strategic Defenses

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

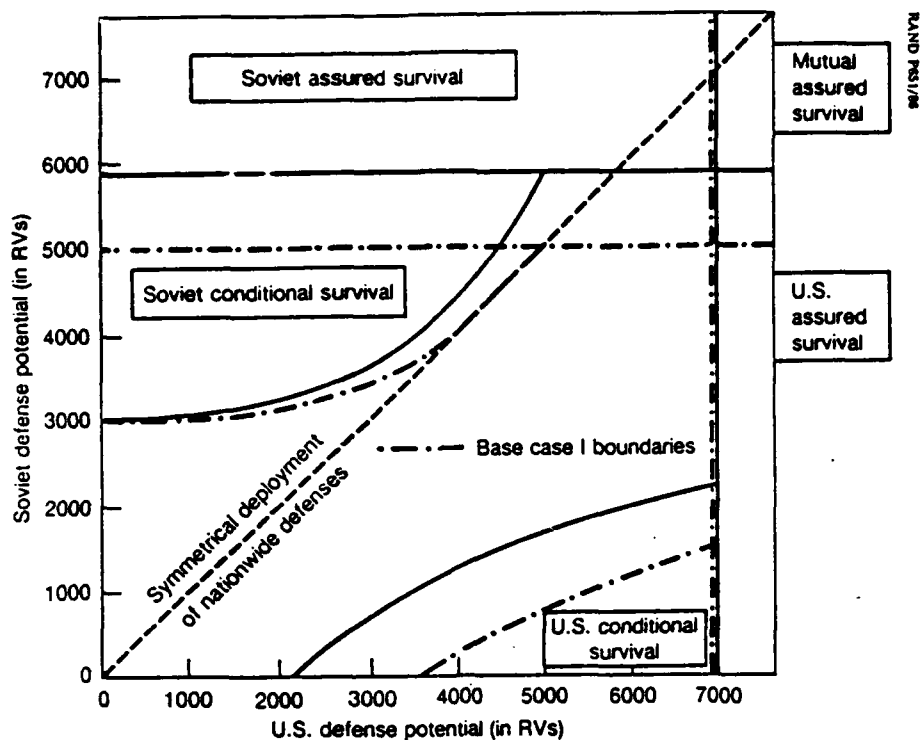


Figure E-4. Figure 6 of Kent and DeValk -- Results for Excursion with U.S. Deployment of 100 MX in Minutemen Silos

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,700 on station RVs, including
 - 2,700 ICBM RVs in 1,000 silos, of which 2,200 RVs are killers, 1,200 with $0.4 P_k$ and 1,000 with $0.7 P_k$ against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with $0.7 P_k$ against U.S. silos
 - 1,000 nontargetable RVs

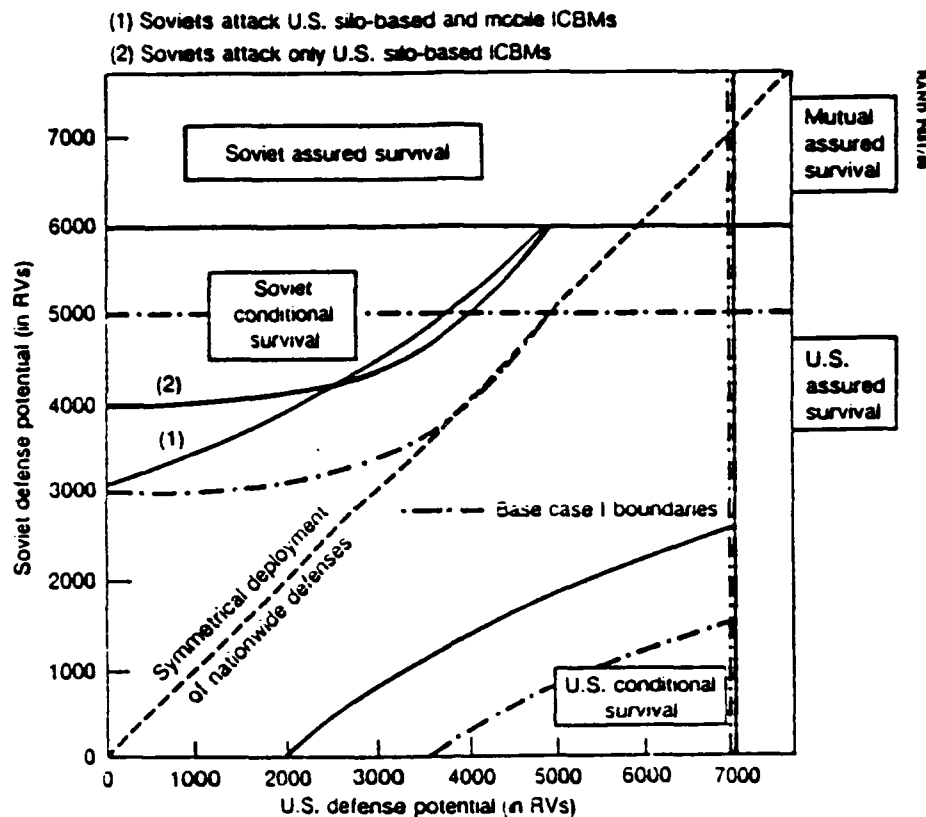


Figure E-5. Figure 7 of Kent and DeValk -- Results for Excursion with U.S. Deployment of 1,000 Small ICBMs on Hardened Mobile Launchers

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 6,000 on station RVs, including
 - 3,000 ICBM RVs, of which 2,500 RVs are killers, 1,500 with 0.4 P_k and 1,000 with 0.7 P_k against Soviet silos; the 3,000 include
 - 2,000 RVs in 1,000 silos
 - 1,000 RVs on transporters deployed randomly over 10,000 nm²
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

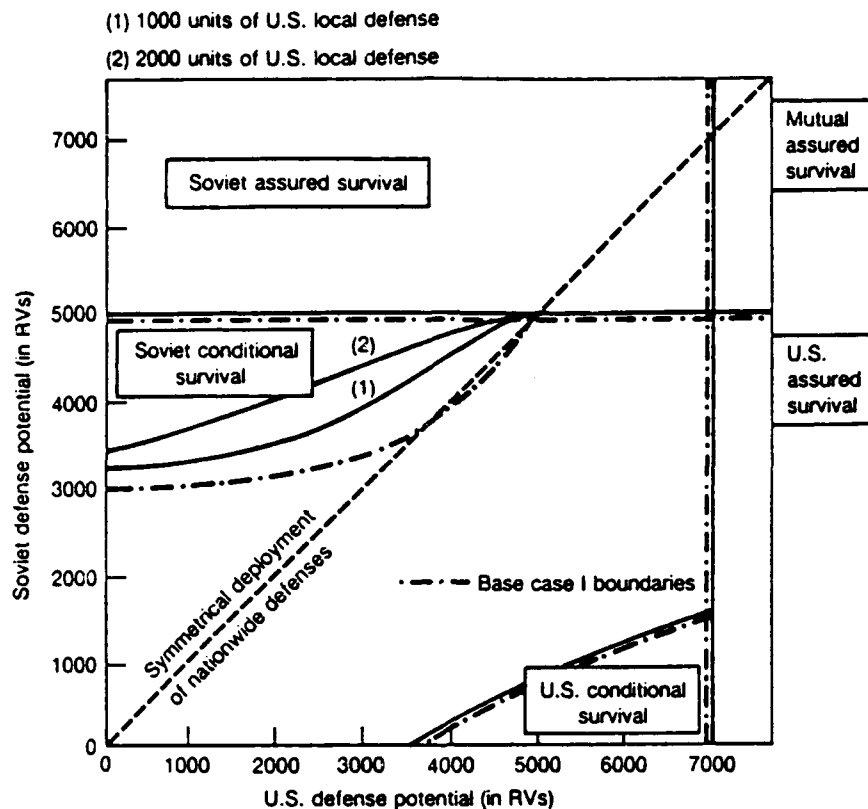


Figure E-6. Figure 8 of Kent and DeValk -- Results for Excursion with U.S. Deployment of Local Defenses of ICBM

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

Figure E-7 shows the combined effect of the measures of Figures E-1 and E-6. The Soviet conditional survival region shrinks significantly.

D. EXTENSIONS OF CONDITIONAL SURVIVAL REGIONS

Conditional survival as discussed thus far has been defined as the region where one side can strike the other side's forces and protect himself against a second strike such that no RVs penetrate. We have also termed this assured destruction denial.

Kent and DeValk present regions of what they term ballistic missile retaliatory gradients. Figures E-8 and E-9 give these results.

If one defines 2,000 RVs surviving and penetrating after a first strike as constituting assured destruction for both the U.S. and Soviets, then there is a narrow channel near the 2:1 ratio of U.S. to Soviet defense, specifically to 5,000 U.S. and 2,500 Soviet defenders. After that the two regions overlap and there is no stable transition. Relaxing that requirement to 1,000 RVs, and further to "few" RVs, as in the previous definition of conditional survival, makes the stable transitions far more feasible.

E. GENERAL REMARKS

The main focus of this paper is on the identification of ballistic missile offense and defense force structures on both sides which are stable, with particular emphasis on the transition from no defenses on both sides to defenses on both sides which provide mutual assured survival.

Along the way conditional survival provides the unifying concept, where conditional survival is defined as the ability to carry out a first strike and defend completely against the second strike. Conditional survival is thus a bad characteristic. It is equivalent to assured destruction denial, if assured destruction is defined as being achieved by delivering a "few" weapons. Kent and DeValk's retaliatory gradients analysis is equivalent to defining assured destruction in arms of 1,0000, 2,0000, etc. weapons in the second strike.

The Kent and DeValk paper does address a number of interesting and important physical parameters. It shows many cases where the channel of stable transitions is fairly wide. It also indicates which factors might make the channel very narrow or nonexistent.

A limitation of the paper is the fact that it does not address air-breathing offensive forces and defenses against them. It may be the case that for many sets of assumptions, there is no stable transition involving ballistic missile offenses and defenses, but a stable transition can be attained when second strike capabilities of air-breathing systems are taken into account.

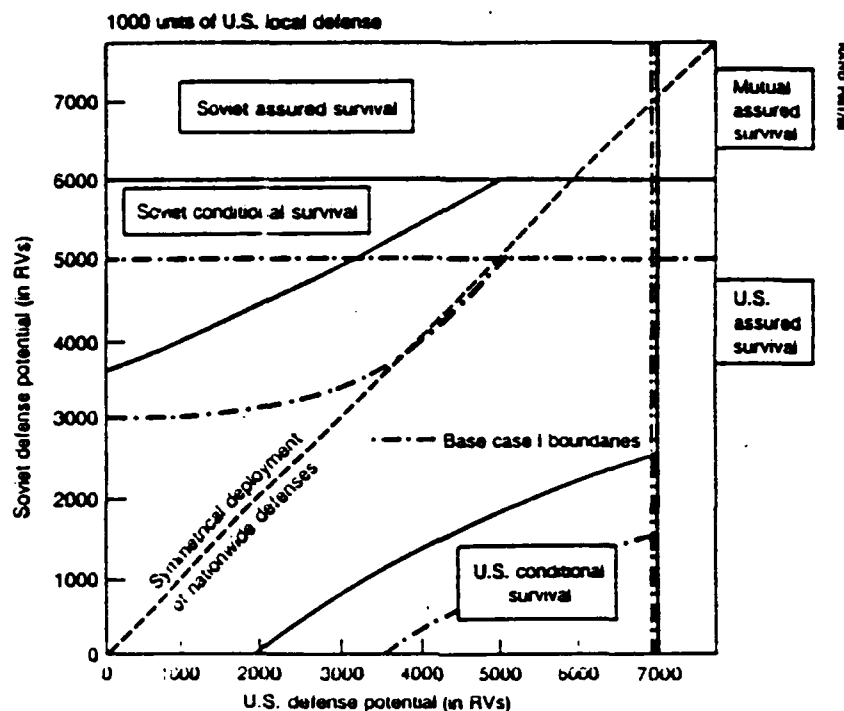


Figure E-7. Figure 9 of Kent and DeValk -- Results for Excursion with U.S. Deployment of 1,000 Small ICBMs as Hardened Missile Launchers and Local Defenses of ICBMs

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 6,000 on station RVs, including
 - 3,000 ICBM RVs, of which 2,500 are killers, 1,500 with 0.4 P_k and 1,000 with 0.7 P_k against Soviet silos; the 3,000 RVs include
 - 2,000 RVs in 1,000 silos
 - 1,000 RVs on transporters deployed randomly over 10,000 nm²
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos and 4 nm² bombardment area against mobile launchers
 - 1,000 nontargetable RVs

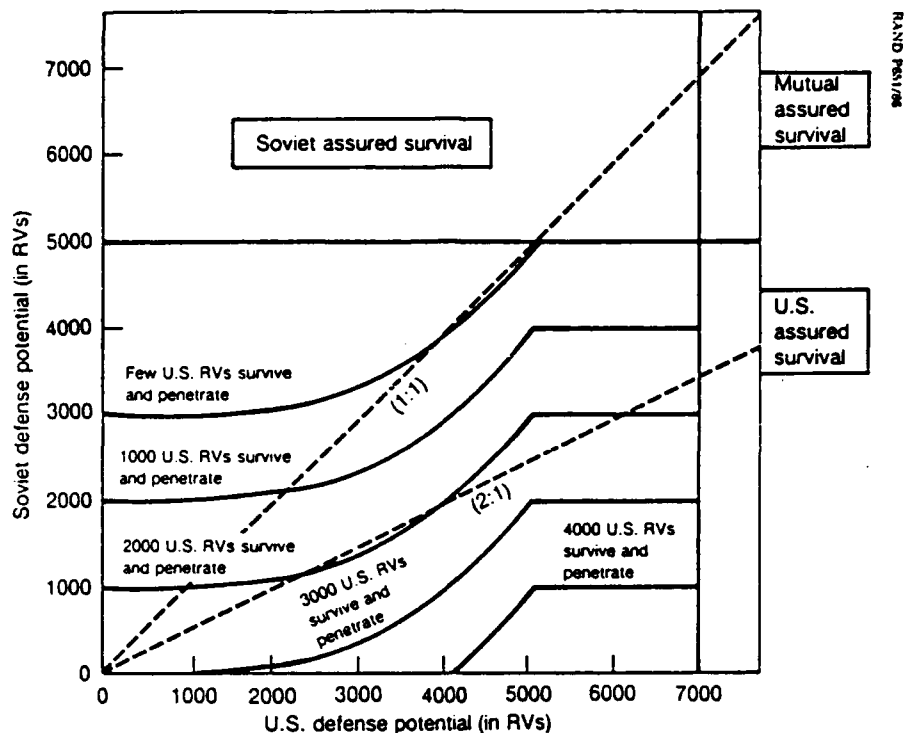


Figure E-8. Figure 12 of Kent and DeValck -- Extensions of Soviet Conditional Survival Regions

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

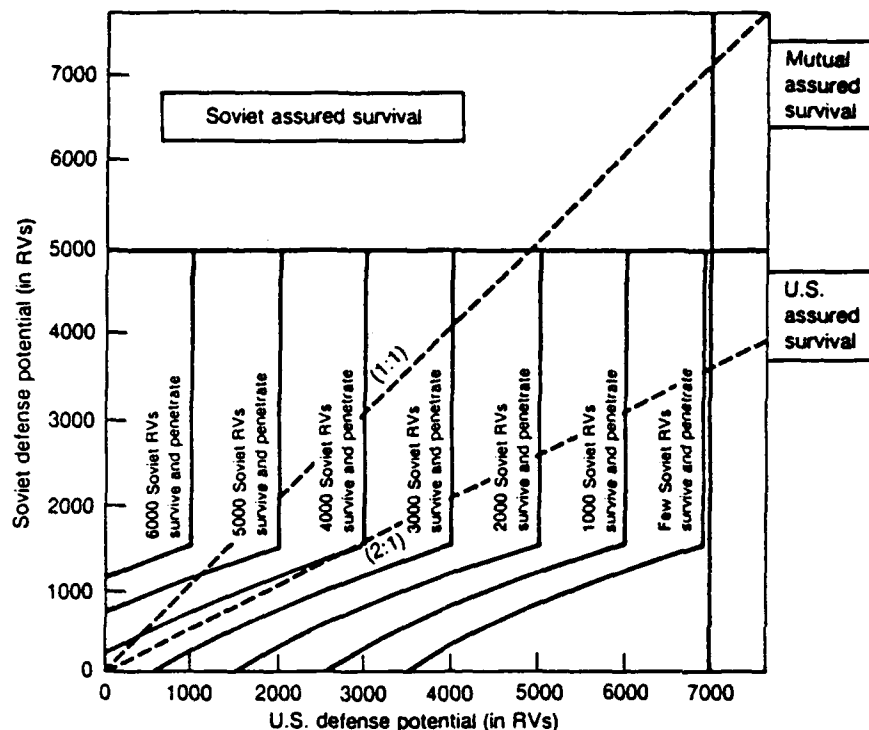


Figure E-9. Figure 13 of Kent and DeValk -- Extensions of U.S. Conditional Survival Regions

DATA (Notional)

- Nationwide defenses operate in discriminating random subtractive mode and are invulnerable to suppression
- U.S. Force: 5,000 on station RVs, including
 - 2,000 ICBM RVs in 1,000 silos, of which 1,500 RVs are killers with 0.4 P_k against Soviet silos
 - 3,000 nontargetable RVs
- Soviet Force: 7,000 on station RVs, including
 - 6,000 ICBM RVs in 1,400 silos, of which 5,000 RVs are killers with 0.7 P_k against U.S. silos
 - 1,000 nontargetable RVs

F. SPECIFIC DISCUSSION OF STABILITY MEASURE

First strike incentive for one side is typically measured in terms of payoff achieved by striking first. First strike instability, or crisis instability, is typically measured by combining in some way the first strike incentive of both sides.

The conditional survival region used in the Kent and DeValk paper is a region whose boundaries have no first strike payoff but whose interiors have positive first strike payoff. When the conditional survival regions are thought of as assured destruction denial regions, and expanded as assured destruction is defined as consisting of 1,000, 2,000 etc. weapons in the second strike, these "first-strike incentive" regions become very large (in the example cases).

Thus an interesting problem is to understand the similarities and differences of the conditional survival (or assured destruction denial) measures used in the Kent and DeValk paper, and the first-strike incentive, first-strike instability and crisis instability measures used in other papers.

ANNEX 1

Summary Extracted From Kent and DeValk Paper

This report details the anatomy and calculus of the ballistic missile portion of the transition to a robust nationwide strategic defense posture, as proposed by President Reagan on March 23, 1983.¹ To provide insight into the policy issues surrounding the transition, we develop an analytic format based on ballistic missile "defense potential." We then use the defense-potential format to demonstrate how various postures of strategic offensive, defensive, and defense-suppression forces might interact to provide or deny each superpower certain strategic capabilities.

Strategic capabilities include the capabilities for "assured survival" and "conditional survival" from ballistic missile attack.

- Assured survival implies the capability to survive as a nation under all circumstances, including an enemy first strike. To attain this posture, a nation would have to deploy highly survivable defenses that could intercept nearly all of the weapons in the enemy's arsenal of strategic ballistic missiles.
- Conditional survival implies the capability to survive the ragged ballistic missile retaliatory attack of the enemy after one's own first strike against the enemy's offensive, defensive, and defense-suppression forces. Unlike assured survival, conditional survival would not require invulnerable defenses.

The report concludes that only the following two postures would offer both first-strike stability² and arms-race stability:

¹ The term *nationwide strategic defenses* refers to ballistic missile defenses that would consist primarily, but not solely, of space-based components; we assumed that when on station these defenses would be capable of intercepting any reentry vehicle (RV) in a given enemy attack involving ballistic missiles. We make no judgment as to whether the United States or the Soviet Union is technically capable of deploying ballistic missile defenses robust enough to provide the capabilities discussed in the report. The analysis simply examines how various changes in the U.S. and Soviet postures of offensive and defensive forces would affect a transition to robust nationwide defenses.

² First-strike stability exists when neither side has the incentive to launch a disarming first strike on the other's strategic forces: That is, neither side calculates that it would be considerably better off, in relative or absolute terms, after launching a would-be disarming first strike against the other and neither side feels pressed to launch a first strike in order to avoid the far worse consequences of going second.

- Mutual assured retaliation, in which both countries possessed-and could continue to possess, regardless of an adversary's actions-the capability to retaliate and, in so doing, to inflict massive, unacceptable damage on the attacker.
- Mutual assured survival, in which both countries possessed-and could continue to possess, regardless of an enemy's actions-the capability to survive as a nation under all circumstances.

The unilateral U.S. or Soviet possession of the capability for assured retaliation, conditional survival, or assured survival would trigger an effort by the other to attain this same capability for itself; hence, it would cause arms-race instability. Mutual conditional survival, a posture in which each superpower had the capability to execute a disarming first strike, would lead to extreme first-strike instability.

The defense-potential format demonstrates that if highly survivable strategic defenses were deployed as an adjunct to current superpower ballistic missile forces, the United States could make the transition to the President's goal of assured survival from ballistic missile attack without having to pass through a destabilizing period during which either country possessed the capability for conditional survival from ballistic missile attack.

The avenue along which a stable transition would be possible is fairly wide.³ If, however, both the United States and the Soviet Union continued to deploy reentry vehicles (RVs) capable of destroying hard targets but failed to adopt corresponding offensive force survivability measures, the avenue to stable transition would close. The absence of offensive force survivability measures would lead to a destabilizing posture of mutual conditional survival from ballistic missile attack, even if the strategic defenses deployed by the superpowers were totally invulnerable.⁴

In general, a safe transition would be possible if both sides deployed many weapons at sea and if the number of RVs available for effective attack on the other's land-based forces did not greatly exceed the number of aim points on which those RVs were

³ A wide avenue for transition in this context means that each country could ultimately attain an assured survival capability while preventing the other from acquiring a capability for conditional survival.

⁴ To the degree that U.S. and Soviet defenses were vulnerable to suppression efforts, a world of mutual conditional survival would become more likely. The extreme case would occur, of course, if both superpowers deployed highly vulnerable defenses. In this situation, neither U.S. nor Soviet assured survival would be possible.

based.⁵ It is demonstrated that the most important actions that the superpowers could take to facilitate a stable transition include:

- Placing sustained and comprehensive constraints on ballistic missile RVs and throwweight in an effort to limit or reduce each side's counterforce capability.
- Unilaterally implementing force survivability measures to increase the number of aim points, for example, by deploying ICBM RVs in redundant silos and/or on hardened mobile launchers that could move about in large basing areas.
- Increasing the number of RVs at sea.

The defense-potential analysis of the ballistic missile portion of the transition to nationwide strategic defenses suggests the following conclusions:

- The United States should not seek to amend the 1972 antiballistic missile (ABM) treaty with the aim of deploying local defenses to increase the number of RVs on intercontinental ballistic missiles (ICBMs) likely to survive a Soviet attack. The deployment of local defenses to protect U.S. ICBM sites would contribute only marginally to a stable transition to assured survival.

The current Soviet ballistic missile force contains a relatively large number of RVs capable of hard target kill, while the current U.S. ballistic missile force has a relatively small number of aim points. Thus, even fairly large deployments of local defenses in the absence of nationwide strategic defenses would not significantly increase the number of U.S. ICBM RVs likely to survive a Soviet first strike.

- Given current U.S. and Soviet ballistic missile forces, the symmetrical deployment of intermediate levels of strategic ballistic missile defenses would erode the U.S. ballistic missile deterrent and decrease first-strike stability.

From the U.S. perspective, intermediate levels of symmetrical ballistic missile defenses would deny the United States a ballistic missile counterforce option without really protecting the U.S. population or strategic forces from a Soviet ballistic missile attack. From the Soviet perspective, modest levels of symmetrical superpower ballistic missile defenses would allow the USSR to deny the United States a ballistic missile counterforce option while not significantly detracting from the Soviet counterforce option.

⁵ The Soviets currently deploy approximately 6,000 RVs on intercontinental ballistic missiles (ICBMs), 5,000 of which are capable of effectively attacking the 1,000 U.S. ICBM silos. The Soviets thus enjoy an overkill capability with respect to U.S. ICBM silos. The United States, in contrast, has barely one effective RV for each Soviet ICBM silo. The United States today has a total of roughly 2,000 ICBM RVs, of which approximately 1,500 can be considered hard target killers capable of effectively attacking some 1,400 Soviet ICBM silos.

As a net result of a symmetrical deployment of intermediate levels of defense, the Soviets would dangerously approach a capability to draw down with their own offenses U.S. weapons on ICBMs and submarine-launched ballistic missiles (SLBMs) in port. The Soviets could then use their defenses to stop surviving U.S. ballistic missile RVs launched in retaliation. They might even be able to limit the damage to the Soviet Union from a U.S. retaliatory attack involving only ballistic missiles to such an extent that, at least in the eyes of U.S. strategists, they might deem the damage acceptable.

- During the period when U.S. ballistic missile retaliatory capability was eroding, the United States would have to depend heavily on strategic bombers and bomber weapons (gravity bombs and cruise missiles) to deter the Soviets.
- Given present-day U.S. and Soviet ballistic missile forces, the United States would have to deploy ballistic missile defense capability at a much faster pace than the Soviet Union to guarantee a stable transition to assured survival from ballistic missile attack.
- Finally, the United States must redress the existing asymmetry in ballistic missile force capability through arms control and/or the modernization of basing modes of existing forces, or it must prepare to build and deploy strategic ballistic missile defense capability nearly twice as fast as the Soviet Union builds and deploys its strategic defenses.

APPENDIX F

SUMMARY OF O'NEILL PAPER¹

¹. Barry O'Neill. A Measure for Crisis Instability with an Application to Space-Based Antimissile Systems, *Journal of Conflict Resolution*, Volume 31, Number 4, December 1987, pages 631-672.

A. INTRODUCTION

The O'Neill paper is organized as follows:

1. Introduction
2. The Crisis Instability Index and Its Adequacy
 - a. Model of the Crisis Decision
 - b. Definition of the Crisis Instability Measure
 - c. Adequacy and Uniqueness of CRNSTB
3. Past Studies of Crisis Instability
 - a. Recurrent Problems in Defining Stability:
The Mutual Attack Cell
 - b. Recurrent Problems in Defining Stability:
Confounding Deterrence Stability with Crisis Stability
4. A Simplified Nuclear Exchange Model
5. Some Simple Bilateral Agreements
6. Space-Based Ballistic Missile Defenses
7. (missing)
8. Space-Based Ballistic Missile Defenses and Submarine Invulnerability
9. Discussion

Appendix A: Axioms for the Crisis Instability Index

Appendix B: The Basic Nuclear Exchange Model

Appendix C: The Nuclear Exchange Model Including Space-Based Defenses

References

It contains a rigorous axiomatic description of a crisis instability index, a discussion of past work and of concepts of instability, an analysis of bilateral force structure changes resulting in reduced instability and an analysis of spaced-based missile defenses.

B. THE CRISIS INSTABILITY INDEX

A non-zero sum game is defined in Matrix 1, as follows:

Matrix 1.

		Gov't 2's policy	
		Would refrain	Would attack
Gov't 1's policy	Would refrain	p ₁ , p ₂	s ₁ , f ₂
	Would attack	f ₁ , s ₂	b ₁ , b ₂

The payoffs to 1 and 2 are associated with peace (p₁ and p₂); first strike (f₁ and f₂), both strike simultaneously (b₁ and b₂) (O'Neill discusses the circumstances of simultaneous attack). In practice, this result could come about, presumably, from some sort of launch under attack strategy rather than from a truly simultaneous attack and second strike (s₁ and s₂). O'Neill assumes that the natural ordering is p_i > f_i > b_i > s_i for i=1 and i=2. An example having this characteristic is shown in Matrix 2, as follows:

Matrix 2.

		Gov't 2's policy	
		Would refrain	Would attack
Gov't 1's policy	Would refrain	4,4	1,3
	Would refrain	3,1	2,2

The following crisis instability index (CRNSTB) is proposed:

$$(1) \quad \text{CRNSTB} = \left(\frac{b_1 - s_1}{p_1 - f_1} \right) \left(\frac{b_2 - s_2}{p_2 - f_2} \right),$$

if $p_i > f_i > b_i > s_i$ for $i = 1$ and $i = 2$,

- (2) CRNSTB = infinity, if either denominator is zero and both numerators are positive, or if the payoffs for one side make attacking preferable no matter what the other side does.
- (3) CRNSTB = 0, if refraining is preferable for both no matter what the other side does.
- (4) CRNSTB = undefined, for other matrices.

CRNSTB takes values from zero to infinity, inclusive. It is proposed as an ordinal measure meaning that only the comparative ranking of two CRNSTB values, not their specific magnitudes, is meaningful.

O'Neill states:

"The formula for CRNSTB is the product of two factors, each of which depends on a single player's payoffs. A factor is the player's ultimate incentive which will depend on the other player's incentive, as well, since side 1 is induced to launch by side 2's incentive, and by fear that side 2 fears side 1's incentive, and so forth.

O'Neill further states: "The components within a factor can also be interpreted. The numerator $b_i - s_i$ is the payoff lost from holding back given the other side's policy to attack, termed the restraint regret, and the denominator $p_i - f_i$ is the loss from an attack policy when the other's decision is to hold back, the attack regret. Thus, the crisis instability idea CRNSTB is the product of ratios of regrets."

Later, in discussions of criteria of adequacy, O'Neill states:

"CA3. When one government's payoffs are ordered $p_i > f_i = s_i$ and the others are $p_j > f_j > s_j$, a measure should take its minimum value."

and

"CA4. When one government's payoffs are ordered $p_i = f_i > s_i$ and the others are $p_j > f_j > s_j$, a measure should take its maximum value."

C. COMPARISONS WITH PAST STUDIES OF CRISIS INSTABILITY

O'Neill presents an interesting and reasonably comprehensive discussion of previous work, focusing on policy issues, measures of value, names of concepts and functional forms of measures. The comparison does not include numerical investigations of stability measures for different cases.

O'Neill discusses the mutual attack cell which Ellsberg's analysis (starting with the same non-zero sum game as O'Neill) left blank.

O'Neill then discusses distinctions between deterrence instability and crisis instability. He presents matrices 3(a) and 3(b) below:

Matrices 3(a) and (b).

100,100	40, 60	100,100	40, 60
90, 30	50, 40	110, 30	50, 40

He states: "Matrix 3(a) represents an unstable situation since the payoff from striking first makes it close to Matrix 3(b) in which the row-chooser prefers to attack regardless of what the column-choosers do. A technological development of change in goals and attitudes might alter the row-chooser's payoffs and trigger a war. In Matrix 3(a) deterrence is precarious, a condition some have termed "crisis instability." However, this concept is clearly different from the classical statements of crisis instability exemplified in the introduction. We prefer to call it deterrence instability. It is a type that depends on change in the payoff matrix, as opposed to crisis instability, in which the matrix is fixed and two players' choices shift from peace to war."

D. DATA FOR BASE CASE

O'Neill assumes the following set of parameters on both sides

number of missiles	1,000 missiles
warheads/missiles	5 warheads
yield of each warhead	500 kilotons
inaccuracy (CEP)	.13 natural miles
warhead/missile reliability	80%
hardness of silos	2,000 pounds per square inch
invulnerability of resources	577 warheads

He assigns to peace in Matrix 1 the payoff of both players. Attacking first has payoff $-a_i$ for government i and retaliation has payoff $-v_i$ for government i . The payoff associated with both attacking is $\frac{-(a_i + r_i)}{2}$, the average of the first and second strike payoffs. In this notation, the formula for crisis instability reduces to

$$CRNSTB = 4 \left(\frac{a_1}{r_1} - 1 \right) \left(\frac{a_2}{r_2} - 1 \right)$$

O'Neill computes values for the payoff matrix utilizing a two strike warfare model documented in the paper. He assumes that the first striker will allocate its resources to minimize $2a_i - r_j$ where r_j is the damage to j 's resources given that i strikes first and a_i is the damage to i 's resources from the counterstrike by j .

The result is given in Matrix 5, as follows:

Matrix 5. PAYOFFS USING THE BASE PARAMETERS

		Gov't 2	
		Decide to Refrain	Decide to Attack
Gov't 1	Decide to Refrain	0, 0	-.618, -.232
	Decide to Attack	-.232, -.618	-.425, -.425

This matrix has CRNSTB = 11.

E. SOME SIMPLE BILATERAL AGREEMENTS

O'Neill presents a set of alternative parameter changes which decrease instability to the same level, namely, from 11 to 2.2. These parameter changes are listed in the following table:

Agreement		% Change in parameters	Missiles Against Silos
incr inaccuracy	.13 nmi to .145 nmi CEP	+12 (+.11)	80
decr reliability	80% to 69.6%	-13 (-.14)	75
decr warheads/missile	5 to 4 warheads	-20 (-.22)	75
incr silo hardness	2,000 psi to 2,750 psi	+38 (+.32)	80
decr yield	500 KT to 351 KT	-30 (-.35)	80
decr value invulnerability	577 tp 185 warheads	-61 (1-1.14)	88
incr number of missile	1,000 to 3,120 missiles	+312 (+1.14)	87

F. SPACE-BASED BALLISTIC MISSILE DEFENSES

O'Neill then performs calculations based on a number of detailed assumptions about space-based missile defenses and presents various results. These results are not in

formats which emphasize two-sided stable transitions, but are focused on stability aspects of specific weapon system characteristics.

G. COMPARISONS OF CRNSTB WITH ANOTHER MEASURE

In Chapter III of the main report, comparisons are made of O'Neill's earlier crisis stability index with that of Bracken and Wilkening.

In this section, we compare the new measure, CRNSTB, with the measure used by Bracken for several cases, as follows:

		$G = (f_1 - s_1) + (f_2 - s_2)$		$CRNSTB = \left(\frac{b_1 - s_1}{p_1 - f_1} \right) \left(\frac{b_2 - s_2}{p_2 - f_2} \right)$	
		$p_1 \ f_1 \ s_1 \ b_1$			
		$p_2 \ f_2 \ s_2 \ b_2$			
Situation		Bracken		O'Neill	
A	1 .9 .1 .2	$G = .8 + .8 = 1.6$		$CRNSTB = \left(\frac{.1}{.1} \right) \left(\frac{.1}{.1} \right) = 1$	
	1 .9 .1 .2				
B	1 .9 .1 .2	$G = .8 + .1 = .9$		$CRNSTB = \left(\frac{.1}{.1} \right) \left(\frac{.1}{.1} \right) = .125$	
	1 .2 .1 .2				
C	1 .9 .1 .2	$G = .8 + 0 = .8$		$CRNSTB = \left(\frac{.1}{.1} \right) \left(\frac{.1}{.9} \right) = .111$	
	1 .1 .1 .2				
D	1 .9 .1 .1	$G = .8 + 0 = .8$		$CRNSTB = \left(\frac{0}{.1} \right) \left(\frac{0}{.9} \right) = 0$	
	1 .1 .1 .1				

Situation A is highly unstable by the Bracken measure (which ranges from -2 to 2) but not by the O'Neill measure (which ranges from 0 to infinity). There is little restraint regret $b_i - s_i$ or attack regret $p_i - f_i$ on either side.

Situation B is quite unstable by the Bracken measure but very stable by the O'Neill measure. This is close to the Side 1 prevail and Side 2 commit suicide case. Both of the O'Neill measures, the one discussed in Chapter III and the one discussed in this summary, do not ascribe significant instability to this case.

Note that in the discussion of CA3 and CA4, the Side 1 prevail/Side 2 commit suicide case is raised by implication. CA3 says when $p_i > f_i = s_i$ and $p_j > f_j > s_j$ the measure should take on its minimum value and CA4 says when $p_i = f_i > c_i$ and $p_j > f_j > s_j$ the measure should take on its maximum value. The case of interest is equivalent to $p_i = f_i > s_i$ and $p_j > f_j = s_j$. What should the measure be here?

APPENDIX G

SUMMARY OF WILKENING AND WATMAN PAPER¹

¹ Dean Wilkening and Kenneth Watman, Strategic Defenses and First-Strike Stability, R-3412-FE RS, The Rand Corporation, November 1986.

A. INTRODUCTION

This summary contains in an annex the summary extracted from the Wilkening and Watman paper. It concisely conveys (1) the scope of the paper, (2) the definition of first-strike stability (3) the motivation for studying first-strike stability in the context of defenses, which at some levels may enable a first-strike (exceeding the defense capabilities) on forces and value targets, followed by a second strike (significantly less than the defense capabilities) on value targets--the outcome being a "successful" first-strike when measured by difference in surviving value targets or something similar, (4) the general conclusions of the paper with respect to the factors which most influence stable transitions, and (5) various methodological considerations of the work.

The present summary discusses three of the principal figures presenting results, the behavior of the measures of stability and the assessment mechanism.

B. PRINCIPAL RESULTS

Figures 1 and 2 present the key results of the Wilkening and Watman paper. Figure G-1 shows regions of Soviet first-strike incentive. Regions of U.S. first-strike incentive and regions of first-strike instability for the base case, plotted as a function of U.S. defense potential (%) and Soviet defense potential (%). Defense potential measures the percent relative to total force of opposing ballistic missile RVs and air-delivered weapons which can be stopped.² Note that the BMD and air defense are coupled in this presentation.

Figure G-1 shows no stable transition from mutual assured destruction in the bottom left corner to mutual assured survival in the top right corner. Much of the entire space of defenses involves a region of first-strike instability.

Figure G-2 arrives at a completely different conclusion. If there are 6,500 terminal interceptors deployed, with 4,000 defending preferentially the ICBMs and 2,500 defending preferentially the value targets, there is a stable transition, with no region of first-strike instability.

² Although the axes are labeled a defense potential in units of percent. The Wilkening & Watman model is a threshold model, not a "percentage" model which we describe elsewhere in the paper.

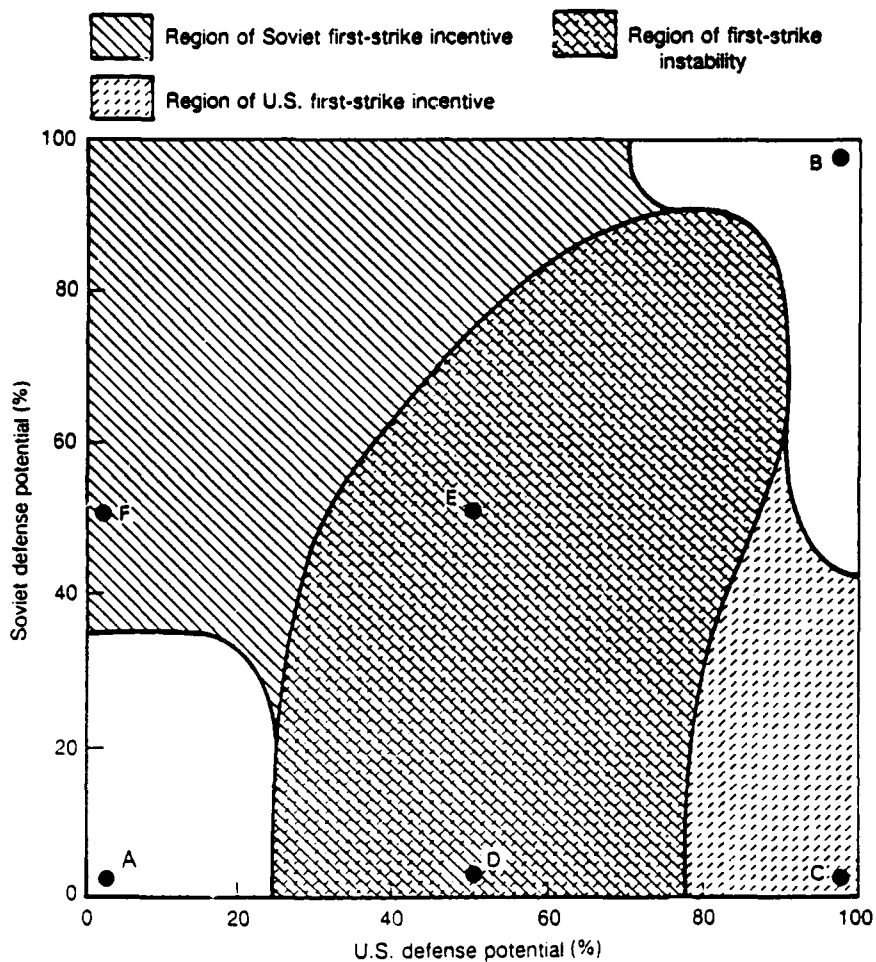


Fig. 2—Regions of high U.S. and Soviet first-strike incentive
(no terminal BMD)

**Figure G-1. Figure 2 of Wilkening and Watman Paper--First-Strike Incentives
and Region of First-Strike Instability with No Terminal Interceptors**

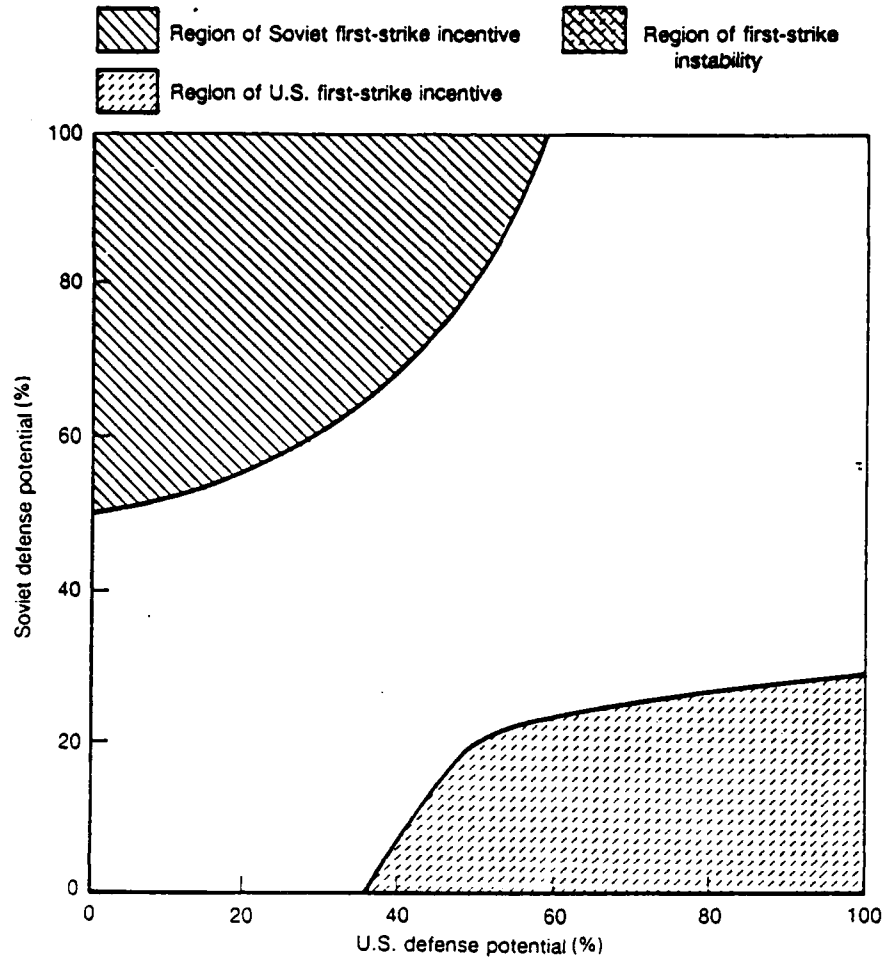


Fig. 6—Regions of high U.S. and Soviet first-strike incentive
(6,500 terminal BMD interceptors deployed)

Figure G-2. Figure 6 of Wilkening and Watman Paper--First-Strike Incentives
with 6500 Terminal Interceptors

C. MEASURES OF FIRST-STRIKE INCENTIVE AND REGIONS OF INSTABILITY

The measurement scheme of Wilkening and Watman is discussed in some detail in Appendix B of their paper, but its essence can be understood to a great extent by means of an example.

Figure G-3 presents the results of exchange calculations in which the first strike simultaneously attacks his adversary's strategic forces and other military targets (holding

back sufficient forces to significantly damage 200 urban/industrial targets), and the adversary responds with his remaining forces on other military targets (also holding back sufficient forces to significantly damage 200 urban/industrial targets). There is no statement in the paper about what is optimized in the nuclear exchange, if anything; from the tone of the discussion one assumes that the weapon allocations are "roughly right" with some attempt by the first-striker to maximize, and second striker to minimize, damage achieved by first-striker minus damage achieved by second striker, perhaps weighted to place more value on self-preservation by the first striker.

Figure G-3 is tied to Figure G-1 above. Case (a) in Figure G-3 corresponds to point A in Figure G-1, and so on.

Consider (a) and (b) of Figure G-3, the first corresponding to mutual assured destruction and the second corresponding to mutual assured survival. The measurement scheme yields negative first-strike incentives in the first case and zero first-strike incentives in the second case.

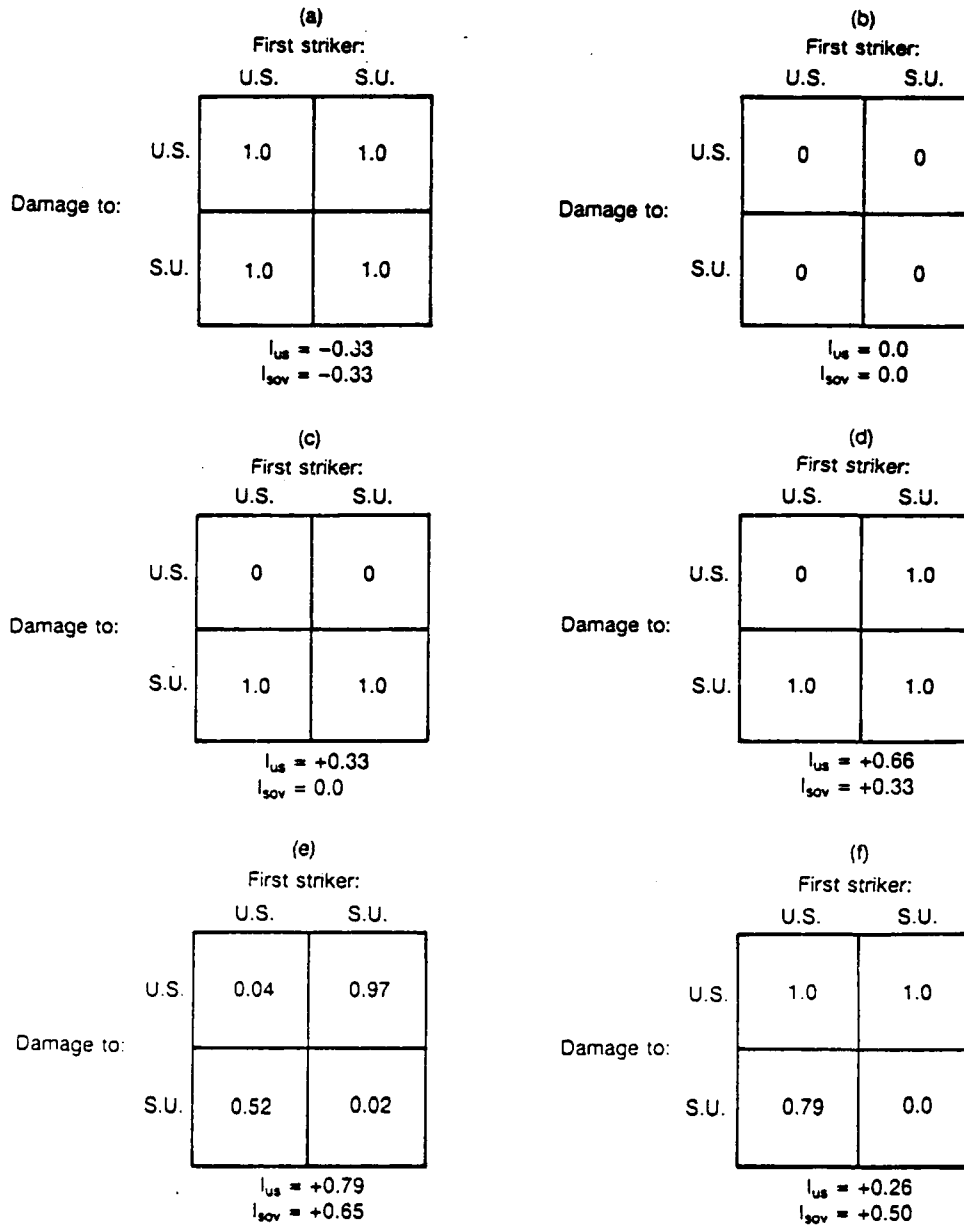
Figure G-1 displays the regions throughout which U.S. and Soviet first-strike incentives exceed $+0.30$ and the overlap region where both the U.S. and Soviet first-strike incentives exceed $+0.30$. The latter is called the region of first-strike instability.

Case (d) is of particular interest. In case (d) $I_{US}=+0.66$ and $I_{SU}=0.33$. The United States "wins" if it strikes first and the Soviet Union "commits suicide" if it strikes first. This is defined to be a situation of first-strike instability.

Case (c), which involves dominance by the United States either in a first-strike mode or second strike mode, has $I_{US}=+0.33$ and $I_{SU}=0$. The U.S. has a first-strike incentive exceeding the threshold.

Case (f) is not the mirror image of case (d) though it is located on the corresponding part of the figure. The Soviet Union can strike first with impunity to "win" and has $I_{US}=+0.50$. The United States striking first, however, does not result in "mutual suicide" but damage to the Soviet Union of 0.79 and to the United States 1.0 . Thus the incentive to the U.S. is $+0.26$, under the threshold of 0.33 , and the region is declared to be one of Red first-strike incentive.

Case (e), in the middle of the region, involves significant advantages to both sides of going first. It is an example of the typical first-strike-advantage-for-both-sides case, and



(Numbers in the boxes represent damage expectancy to other military targets)

Fig. 3—Examples of exchange outcomes and first-strike incentives

Figure G-3. Figure 3 of Wilkening and Watman Paper--Examples of Exchange Outcomes and First-Strike Incentives

very clearly represents the type of situation which should be avoided. The point of going through all of these combinatorial examples is to show the importance of the definition of the measures. There are two measures in this illustration which are particularly evocative. Should case (d) be one of first-strike instability--would the Soviet Union commit suicide here as the measure assumes? Should case (f) be one of first-strike instability--would the United States commit suicide here, as the measure does not assume? Note that if one were to add the first-strike incentive measures rather than look at thresholds, the results would be somewhat different.

D. EXCHANGE MODEL

Appendix B of the paper contains a through discussion of the measures of first-strike incentive and regions of instability. However, the discussions of exchange calculations is not as thorough.

Rather than present example calculations or equations describing each step, descriptive statements about the various levels of the assessment process are made. From the discussion it is clear that the exchange model optimizes some objective function. Figure 20 of the paper shows allocations to counterforce in a U.S. first-strike as a function of U.S. defense potential and Soviet defense potential.

There are no citations to the model or models used. The reader is thus left with the impression that there is no documentation of the exchange model.

ANNEX 1

SUMMARY EXTRACTED FROM THE WILKENING AND WATMAN PAPER

The impact of strategic defenses on stability is a central theme in the Strategic Defense Initiative (SDI) debate. This report examines the effects of defenses on first-strike stability. (Since the methodology we have developed is sufficiently general to account for decision-making under conditions where the likelihood of an enemy attack is low, as well as when it is high, we use the term "first-strike stability" instead of the more common "crisis stability.")

First-strike instability refers to a situation in which both the United States and the Soviet Union have high incentives to strike first, resulting in a strong pressure to preempt. In this study, the incentive to strike first is viewed solely as arising from the offensive and defensive central strategic balance. This formulation obviously ignores many political and circumstantial factors which influence so fateful a decision, not to mention the correlation of conventional and theater nuclear forces. Nevertheless, it captures the effect of defenses on that part of the decision influenced by the central strategic balance. This report is principally concerned with assessing first-strike stability during the transition from an offense-dominated balance. The implications of various offensive and defensive force structures are also examined.

Typically, first-strike instability does not arise when neither side possesses strategic defense, provided the offensive forces are sufficiently survivable. However, instability can arise if both sides acquire moderately effective nationwide defenses. This condition represents the so-called "ragged retaliation" problem. That is, with moderate levels of defense and partially vulnerable offensive forces, both sides may have an incentive to strike first in the expectation that the adversary's retaliation, already impaired by the counterforce attack, will be unable to penetrate the defenses. In general, any increase in the vulnerability of the strategic offensive forces-e.g., increased countersilo capability, more capable antisubmarine warfare, an effective barrage of bomber flyout corridors-exacerbates this instability. This strongly suggests that first-strike instabilities are likely to occur during the

defense transition if either side neglects to ensure the survivability of its strategic offensive forces. It also suggests that, to the extent that defenses help limit damage if one strikes first, a greater emphasis will be placed on offensive counterforce capabilities during the defense transition--thus exacerbating the opponent's offensive force survivability problem. Not surprisingly, this instability can be diminished by increasing strategic force survivability. This can be accomplished actively (with terminal ballistic missile defense (BMD)) or passively (with silo hardening, proliferating ICBM aimpoints, increased bomber and submarine alert rates, quieter submarines, etc.).

Therefore, a phased transition in which terminal BMD, defending the strategic nuclear forces, precedes nationwide BMD minimizes first-strike instability. Passive measures could be used as well to ensure the survivability of the strategic nuclear forces. The choice between active or passive defense of the strategic offensive forces is largely one of cost effectiveness. It should also be noted that a launch-on-warning or launch-under-attack policy increases stability by ensuring the survivability of silo-based ICBMs, although the specter of an accidental nuclear war triggered by faulty warning makes this option less attractive.

An important conclusion of this work is that air defenses play a significant role in generating first-strike instabilities. If air defenses are weak, instabilities are less apt to occur, since each side's bomber force can then inflict devastating retaliation (unless effective barrage attacks can be made against the bomber flyout corridors). Consequently, a phased transition in which the BMD transition is completed before significant air defenses are deployed should minimize first-strike instabilities. Under these circumstances, deterrence would rest on only one leg of the Triad. After a nearly perfect BMD is in place, the air defense transition is stable because air-breathing forces presumable are not effective counterforce weapons, due to the significant tactical warning associated with bomber attacks. Thus, by completing the BMD transition first and then embarking on the air defense transition, it might be possible to arrive at a world with "perfect" defenses in a stable manner (provided, of course, these hypothetically "perfect" defenses are technically achievable and survivable). Therefore, to minimize instabilities during the defense transition, terminal BMD, nationwide BMD, and nationwide air defenses should be deployed sequentially. As one might expect, the cooperation of both sides will be required to effect this phased transition.

Instabilities can also rise if strategic defenses are vulnerable to suppression. If both sides' defenses and defense-suppression forces are vulnerable to a first-strike, instability is

likely to occur. If both sides' defenses are vulnerable but the defense-suppression forces are not, instability is less likely. This highlights the importance of the survivability of the defenses and the defense-suppression forces, in addition to that of the strategic offensive forces.

Arms control measures are frequently proposed as a way to reduce instabilities during the defense transition. However, merely reducing the number of warheads does not necessarily reduce instabilities, as illustrated by the notional arms control force posture used in this study. This force posture assumes across-the-board reductions to approximately 9,300 weapons on each side. However, no specific attempt is made to reduce the counterforce capability of each side's forces. Without terminal BMD interceptors deployed, the impact of this force-structure change on first-strike instabilities is small. This result suggests that, unless an arms control regime is designed specifically to reduce the number of counterforce warheads relative to the opponent's counterforce targets, stability will not be greatly enhanced.

The above results are predicated on a world in which each side has perfect knowledge of the other's offensive and defensive capabilities and the resulting first-strike incentives. In reality, perceptual distortions and biases invariably influence each side's assessments of the costs and benefits of a first-strike. Though many kinds of perceptual biases are possible, we examine the effects of only two; "defense conservative planning" and "worst-case planning." In both cases, each side plans as though it is weaker and its adversary stronger than is really the case. The difference between defense conservative and worst-case planning arises from the assumptions each side makes about how his adversary calculates his first-strike incentives. If side A is defense conservative, it believes that side B underestimates its own capabilities and exaggerates those of A. On the other hand, if side A proceeds from worst-case assumptions, it believes that side B overestimates its own capabilities and underestimates those of A. The effect of defense conservative planning is to diminish both U.S. and Soviet first-strike incentives, thereby diminishing any first-strike instability. The effect of worst-case planning is quite different. When both sides view the world from this perspective, first-strike instability can increase. These results illustrate how the defense transition can be made more or less stable, depending on what each side assumes about the opponent's first-strike calculations. This is important, because perceptions can change much more rapidly than shifts in the nuclear balance brought about by actual force deployments.

Finally, it is often argued that defenses enhance first-strike stability by increasing the uncertainty in the first striker's attack calculations. This is only true if the number of warheads employed in the first strike remains constant as the level of defense effectiveness increases. If the first strike responds to an opponent's defenses by proliferating warheads, he can meet his attack objectives with any degree of confidence he desires. Thus, before it is possible to conclude that defenses enhance stability by increasing uncertainty, it is necessary to determine the likely arms competition that will ensue after defenses are deployed. To the extent that the first striker responds with greater offensive forces, uncertainty will be no more of a problem than it was before the defenses were deployed, assuming the first strike correctly gauges the size of the defense. If the first striker deploys a comparable level of defense, mutual defenses will probably undermine confidence in one's ability to retaliate to a greater degree than it will confidence in one's ability to strike first. If anything, this undermines first-strike stability.

The methodology that produces these results defines first-strike incentives in terms of the outcomes of strategic attacks. The influence that one side's incentive to strike has on the opponent's incentive to strike has been accounted for explicitly. This coupling of incentives reflects the fact that when the Soviet incentive to strike is sufficiently high, the United States will assign a high value to the probability of a Soviet attack. This increases the U.S. incentive to strike. However, as the U.S. incentive increases, the probability of a U.S. first-strike increases, which in turn influences the Soviet incentive. The intent here is not to arrive at a predictive model for the actual probability that the United States or the Soviet Union would launch a nuclear strike, but rather to build a model that reflects the essential qualitative features of the incentive to strike first, based on the damage that can be inflicted by each side's strategic nuclear forces (i.e., based on the correlation of nuclear forces.)

The magnitude of each side's incentive depends on various assumptions concerning the scenario, the offensive force postures, and the defensive force postures used in the exchange calculation. The scenario we have used involves massive counter-military attacks directed at the homeland of the opponent. That is, we have assumed that deterrence rests on the threat of retaliation against the opponent's military forces (nuclear and conventional). Urban/industrial targets are not directly attacked in this scenario, although we assume that each side withholds a secure reserve for threatening these targets.

More explicitly, each side's first-strike incentive is calculated by comparing the following quantities: the expected damage to military targets from a first-strike on the

adversary and from his retaliatory strike, the expected damage from a first-strike by the adversary and a retaliatory strike against his military targets, and the value of each side of having no nuclear strike (i.e., preservation of the status quo for each side). Incorporating these factors permits the model to represent nuclear decision-making under conditions ranging from belief by each side that the likelihood of a first-strike by the adversary is very small to belief by each side that the likelihood of a preemptive strike is very high. To the extent that the nuclear balance (including the effects of defense) affects the outcomes, a link is established between the correlation of nuclear forces and each side's incentive for launching a nuclear attack. Situations in which both sides have a strong incentive to strike first are deemed to be unstable.

The offensive force postures used in this study represent notional forces for the year 2000. They come from a Congressional Budget Office study entitled *Modernizing U.S. Strategic Offensive Forces: The Administration's Program and Alternatives*. This notional force posture reflects the long-standing weapon preferences of both sides and continues the existing trend of increasing hard-target kill capability. The U.S. and Soviet arsenals consist of approximately 13,600 and 18,500 weapons, respectively. The U.S. force posture is a balanced Triad with sufficient hard-target kill capability to destroy the Soviet silo-based ICBM force. The Soviet force consists predominantly of ICBMs, with the capability to hold the U.S. silo-based ICBM force at risk.

The strategic defenses examined in the study are of three types: a nationwide BMD that destroys incoming warheads until the capacity of the system is exceeded or the warheads are all destroyed, terminal BMD interceptors that operate preferentially to protect strategic and other military targets, and nationwide air defenses that destroy air-delivered weapons until the air defense system is saturated. The level of defense effectiveness for each type is varied parametrically to examine the impact these defenses have on a country's incentive to strike first.

The exchange-model methodology has the advantages of explicitness, replicability, and flexibility, and the disadvantages of artificiality and incompleteness. Because the model is designed to explore the influence of offensive and defensive forces on first-strike stability, other factors have not been included. While the regions of instability produced with such a model do not exactly correspond with reality, we believe the relative effects of various offensive and defensive force deployments are accurately portrayed. To some extent, the methodology could be extended to incorporate other assumptions, for example, basing first-strike incentives on the damage to urban/industrial targets (instead of military

targets), changing the magnitude of the incentive assumed necessary to trigger a first-strike, and elaborating the nature of the perceptual biases. This report simply introduces the methodology; it makes no attempt to exhaust the many variations that could be invented.

APPENDIX H
BIBLIOGRAPHY

BIBLIOGRAPHY

- Bella, David A., "Nuclear Deterrence: An Alternative Model," *IEEE Technology and Society Magazine* 6, 2: pp. 18-23, 1987.
- Bracken, Jerome., *Stable Transitions from Mutual Assured Destruction to Mutual Assured Survival*, (mimeo), 1987.
- Brams, Steven J. and D. Marc Kilgour, "Winding Down If Preemption or Escalation Occurs: A Game-Theoretic Analysis". *Journal of Conflict Resolution* 31, 4: pp. 547-572, 1987.
- Brams, Steven J. and D. Marc Kilgour, *Deterrence versus Defense: A Game-Theoretic Model of Star Wars*, (mimeo).
- Brams, Steven J., "Deterrence versus Defense: A Game-Theoretic Model of Star Wars" *International Studies Quarterly* 32, 1: pp. 3-28, 1988.
- Brown, Thomas A., *Some Stability Questions Associated with Defense Against Ballistic Missiles*, The RAND Corporation, Working Draft WD-2003-USDP, 1983.
- Bunn, Matthew G., *Strategic Stability: Theory and Measurement for Arms Control*, Master's Thesis, Massachusetts Institute of Technology, Department of Political Science, 1985.
- Canavan, G., et al., *Comparison of Analyses of Strategic Defense*, Draft, Los Alamos National Laboratory, LA-UR-85-754, 1985.
- Canavan, Gregory H., *Simple Discussion of the Stability of Strategic Defense*, Los Alamos National Laboratory, LA-UR-85-1377, 1985.
- Canavan, Gregory H., *An Assessment of Strategic Defense*, Los Alamos National Laboratory, LA-UR-87-520, 1987.
- Canavan, Gregory H., *An Assessment of Strategic Defense: Part II*, Los Alamos National Laboratory, LA-UR-87-520, 1987.
- Canavan, Gregory H., *Interaction Between Strategic Defenses and Arms Control*, Los Alamos National Laboratory, Center for National Security Studies Brief, 1987.
- Chrzanowski, Paul L., *Ballistic Missile Defense and Crisis Stability*, Lawrence Livermore National Laboratory, Draft, 1985.
- Chrzanowski, Paul L., *Crisis Stability During a Transition to a Deterrence Posture Reliant on Defenses*, Lawrence Livermore National Laboratory, UCID-20590, 1985.
- Chrzanowski, Paul L., *Strategic Defense and Crisis Stability*, Lawrence Livermore National Laboratory, UCID-20699, 1985.
- Chrzanowski, Paul L., *Factors Affecting a Transition to a Deterrence Posture More Reliant on Strategic Defense*, University of California, Lawrence Livermore National Laboratory, DDV-86-0020, 1986.
- Chrzanowski, Paul L., "Transition to Deterrence Based on Strategic Defense," E&TR: pp. 31-45, 1987.

- DeNardo, James, *Are Strategic Defenses Strategically Defensible?*, Princeton University, Department of Politics, (mimeo).
- De Santis, Hugh, "An Anti-Tactical Missile Defense for Europe," *SAIS Review* 6, 2: pp. 99-116, 1986.
- Diaz, O., et al., *Evaluation of Strategic Missile Defenses and Crisis Stability -- Developing an Analytic Framework*, ANSER, MPDN 86-2, 1986.
- Ellsberg, Daniel, *The Crude Analysis of Strategic Choices*, The RAND Corporation, P-2183, 1960.
- Fought, Lt. Col. Stephen O., "SDI: A Policy Analysis," *Naval War College Review* 38: pp. 59-95, 1985.
- Glaser, Charles L., "Do We Want the Missile Defenses We Can Build?" *International Security* 10, 1: pp. 25-57, 1985.
- Glaser, Charles L., *Managing the Transition*, in Samuel F. Wells, Jr. and Robert S. Litwak (eds.) *Strategic Defenses and Soviet-American Relations*, Cambridge, MA: Ballinger Publishing Company, 1987.
- Hopkins, Kevin R., *SDI and Crisis Stability: A Force Exchange Model Approach*, Hudson Institute, HI-3789, 1985.
- Intriligator, Michael D. and Dagobert L. Brito, "Can Arms Races Lead to the Outbreak of War?" *Journal of Conflict Resolution* 28, 1: pp. 63-84, 1984.
- Intriligator, Michael D. and Dagobert L. Brito, "Mayer's Alternative to the I-B Model," *Journal of Conflict Resolution* 30, 1: pp. 29-31, 1986.
- Kent, Glenn A., et al., *A Calculus of First-Strike Stability (A Criterion for Evaluating Strategic Forces)*, The Rand Corporation, N-2526-AF, 1988.
- Kent, Glenn A. and DeVal, Randall J., *Strategic Defenses and the Transition to Assured Survival*, The Rand Corporation, R-3369-AF, 1986.
- Kerby, William, "The Impact of Space Weapons on Strategic Stability and the Prospects for Disarmament: A Quantitative Analysis," *Hamburger Beitrage* 6: pp. 1-42, 1986.
- Max, C., et al., *Deployment Stability of Strategic Defenses*, JASON, The Mitre Corporation, JSR-85-926, 1986.
- Mayer, Thomas F., "Arms Races and War Initiation: Some Alternatives to the Intriligator-Brito Model," *Journal of Conflict Resolution* 30, 1: pp. 3-28, 1986.
- Mizrahi, Maurice M., *On Defenses and Stability*, (mimeo), 1985.
- O'Neill, Barry, *A Measure for Crisis Instability*, Second Draft, Northwestern University, Department of Industrial Engineering and Management Sciences, (mimeo), 1985.
- O'Neill, Barry, *Applications of a Crisis Instability Index: Arms Control Agreements and Space-Based Missile Defenses*, Second Draft, Northwestern University, Department of Industrial Engineering and Management Sciences, (mimeo), 1985.
- O'Neill, Barry, "A Measure for Crisis Instability with an Application to Space-Based Antimissile Systems," *Journal of Conflict Resolution*, Vol. 31, pp. 631-672, 1987.
- Payne, Keith B. and Colin S. Gray, "Nuclear Policy and the Defensive Transition," *Foreign Affairs* 62, 4: pp. 820-842, 1984.

- Powell, Robert, "Crisis Bargaining, Escalation, and MAD," *American Political Science Review* 81, 3: pp. 717-735, 1987.
- Powell, Robert, "Nuclear Brinkmanship with Two-Sided Incomplete Information," *American Political Science Review* 82, 1: pp. 155-178, 1988.
- Powell, Robert, *Crisis Stability in the Nuclear Age*, Harvard University, Department of Government and the Center for International Affairs, (mimeo), 1988.
- Radner, Roy, *A Model of Defense-Protected Build-Down*, in Alvin M. Weinberg and Jack N. Barkenbus (eds.) *Strategic Defenses and Arms Control*, New York: Paragon House Publishers, 1988.
- Rathjens, George and Jack Ruina, *BMD and Strategic Instability*, in Franklin A. Long, et al. (eds.) *Weapons in Space*, New York: W.W. Norton & Company, 1986.
- Saperstein, Alvin M. and Gottfried Mayer-Kress, *A Nonlinear Dynamical Model of the Impact of S.D.I. on the Armsrace*, (mimeo).
- U.S. Congress, Office of Technology Assessment, *Ballistic Missile Defense Technologies*, OTA-ISC-254. Washington, DC: U.S. Government Printing Office, 1985.
- Wilkening, Dean and Kenneth Watman, *Strategic Defenses and First-Strike Stability*, The Rand Corporation, R-3412-FF/RC, 1986.
- Wilkening, Dean, Kenneth Watman, Michael Kennedy and Richard Darilel, "Strategic Defenses and First-Strike Stability," *Survival*, March-April, pp. 137-165, 1987.