Discussion Paper No. 651

APPLICATIONS OF A CRISIS INSTABILITY INDEX: ARMS CONTROL AGREEMENTS AND SPACE-BASED MISSILE DEFENSES

By

Barry O'Neill

March 1985

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, Illinois 60201

Abstract

Crisis instability is the danger of preemptive war due to each government's fear that the other is about to attack. A previous paper developed an index of crisis instability that depended on the relative costs of striking first versus second and gave a unique ranking of situations by degree of instability.

Using a simplified model of a nuclear war this paper derives the costs of striking first versus second as a function of the weapons holdings of the two governments and compares the stability consequences of various bilateral arms control agreements. Treaties to decrease accuracies, numbers of MIRVs and reliabilities are most stabilizing, while limitations on civil defense or numbers of weapons have little effect.

The model also implies that building anti-missile weapons in space is destabilizing unless the systems are both very effective and cannot be attacked with relative advantage to the first-striker. Space-based defenses help the first-striker eliminate the other's retaliating missiles, and this source of instability usually overwhelms the benefit from degrading the first-striker's counterforce. The model implies that space-based weapons are especially harmful if they can be used against otherwise invulnerable submarine missiles.

Acknowledgement: The author would like to thank James Morrow for helpful comments.

1. The Crisis Stability Index

In a previous paper (O'Neill 1985) we developed a measure of crisis instability based on the costs of attacking first versus retaliating. Here we will use the measure to analyze the effects of various arms control agreements and of ballistic missile defenses in space.

We will outline the crisis instability index briefly but for complete details and rationale the reader should refer to the original paper. Two governments, 1 and 2, were assumed to assign costs a_1, r_1, a_2 and r_2 as shown in Matrix 1.

Gov't 2

		Decide to Refrain	Decide to Attack
Gov't 1	Decide to Refrain	0, 0	$-r_1$, $-a_2$
	Decide to Attack	$-a_1$, $-r_2$	$\frac{a-r_1}{2}$, $\frac{-a_2-r_2}{2}$

Matrix 1. The payoffs in general form.

It was assumed that $r_i > a_i > 0$, so that playing the role of retaliator is more disastrous than being the attacker, which in turn is worse than enjoying peace. The values in the lower right cell are the expected costs if both decide to attack. In that instance some chance event determines which of the two has the ability to strike first. In specific terms, one government or the other starts its deliberations or implements its decision to strike first more quickly than the other.

Sample costs with this ordering of preference are shown in Matrix 2.

		Gov't 2	
		Decide to Refrain	Decide to Attack
Gov't 1	Decide to Refrain	0, 0	-45, -12
	Decide to Attack	-15, -20	-30, -16

Matrix 2. Sample values.

A series of axioms was stated that led to the following index of instability, which was interpreted as an ordinal measure of the probability of a war. This probability was construed as that assessed by an outside observer who knows only the costs appearing in Matrix 1.

$$\text{Crisis Instability Index (CII)} = (r_1/a_1 - 1)(r_2/a_2 - 1).$$

The value of CII for Matrix 2 for example would be $(45/15 - 1)(20/12 - 1) = 1.33$.

To say that this index is an "ordinal measure" means that it may be different from the probability of war but that for two situations the one with the higher CII will have the higher likelihood. If the costs in the matrix change, the CII tells us whether the danger of war goes up or down, although it does not specify by how much.

2. The Nuclear Exchange Model

In this paper we discuss how to improve crisis instability by limiting weapons. The above index gives the crisis instability as a function of the war costs, so now we must connect the physical characteristics of the weapons to the war costs.

A change in weapons holdings can alter crisis instability by altering the damage done in a war, and also more subtly by changing the way the

		Gov't 2: Decide to Refrain	Gov't 2: Decide to Attack
Gov't 1	Decide to Refrain	0, 0	-45, -12
	Decide to Attack	-15, -20	-30, -16

Matrix 2. Sample values.

A series of axioms was stated that led to the following index of instability, which was interpreted as an ordinal measure of the probability of a war. This probability was construed as that assessed by an outside observer who knows only the costs appearing in Matrix 1.

$$\text{Crisis Instability Index (CII)} = (r_1/a_1 - 1)(r_2/a_2 - 1).$$

The value of CII for Matrix 2 for example would be $(45/15 - 1)(20/12 - 1) = 1.33$.

To say that this index is an "ordinal measure" means that it may be different from the probability of war but that for two situations the one with the higher CII will have the higher likelihood. If the costs in the matrix change, the CII tells us whether the danger of war goes up or down, although it does not specify by how much.

2. The Nuclear Exchange Model

In this paper we discuss how to improve crisis instability by limiting weapons. The above index gives the crisis instability as a function of the war costs, so now we must connect the physical characteristics of the weapons to the war costs.

A change in weapons holdings can alter crisis instability by altering the damage done in a war, and also more subtly by changing the way the attacker would use the weapons. For example hardening of silos may induce the opposing side to attack cities more heavily. To capture both effects, we will use a "nuclear exchange model" which prescribes how a government would target its weapons to maximize some objective in the war, and gives the resulting costs.

The world depicted in nuclear exchange models is largely unreal. Absent are hard-to-quantify factors such as the longer term effects of a nuclear war, unreliabilities of groups of weapons and difficulties in command and control. However, using a model may prove useful to understanding crisis instability by making the workings of the Crisis Instability Index clearer, and showing the rough pattern of how one would calculate instability given more adequate models. Our view is that the numbers we get are not precise predictions for the real world, but the comparative sizes or directions of change are worth attention especially until better models are constructed.

The nuclear exchange model used here is a free adaptation of one described by Grotte (1982). Details are given in Appendix A. A nuclear war is assumed to have two stages: one government strikes at the other's value and missiles, then the latter government retaliates at the former's value. The retaliator has no reason to attack the former's missiles since all the silos are now empty.

Some strategists have contemplated the possibility of wars that continue to several exchanges. However this assumption of a two-strike model seems appropriate for the specific type of crisis situation outlined here: the governments believe a war will probably occur and in desperation try to limit damage to themselves. It is also consistent with the fragile state of each side's command and control system.

Each government's cost is proportional to the fraction of its value destroyed in the war. Each hopes mainly to preserve its own value, but has an additional motive, a desire to harm the other, that will determine its execution of the war. We postulate that government i, if it strikes first against j, allocates its weapons to minimize $2a_i - r_j$ where r_j is the level of j's value given i strikes first, and a_i is the level of i's value after j's retaliation. This objective function for war plans means that i would sacrifice one unit of its value to destroy two units of the other side's. Government i wants to damage j's value, perhaps to prevent recovery, to take retribution, or because weapons systems had been set up that way before the war for purposes of deterrence.

We assume very simple weapons holdings, where each side has a single type of strategic missile in hardened silos, with these parameters:

number of missiles	1000 missiles
warheads/missile	5 warheads
yield of each warhead	500 KT
inaccuracy (CEP)	.13 nmi
warhead/missile reliability	80%
hardness of silos	2000 psi
invulnerability of value	577 warheads

Table 1. Base parameters.

The warhead/missile reliability is the likelihood that the warhead reaches the vicinity of target and explodes given that the missile is launched, while the CEP describes the accuracy of those missiles that operate reliably. Our basic model will not require us to specify how much of the failure rate is situated in the launch, the booster, the warhead, etc. -- all of these can be summed up in one value of reliability. The invulnerability of value is the number of detonating 500 kiloton warheads that would destroy one half of the nation's value. The values in Table 1 were chosen to be simple yet not too far off the parameters of weapons in used in current quantitative strategic analysis.

Using a computer algorithm to apply the nuclear exchange model, we find that the attacker will direct 80% of its warheads against the other's missiles and expect 275 warheads in return. Assuming the value of each to be 1, the resulting payoff matrix is shown in Matrix 3, which has CII = 2.75.

		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

<u>Matrix</u> <u>3.</u> Costs using the base parameters.

With this example as a baseline we can alter the weapons characteristics, use the nuclear exchange model to determine how the weapons would be used, what the effects would be and calculate a new Crisis Instability Index.

<u>3.</u> <u>Some</u> <u>Simple</u> <u>Bilateral</u> <u>Agreements</u>

What arms control agreements decrease crisis instability, and how much do they help relative to one another? We will analyze the type of agreement in which one feature of a weapon is altered and others are fixed at their current levels. Other types of treaties have been outright prohibitions on classes of armaments, such as weapons of mass destruction in space, and weapons that injure by fragments invisible to X-rays, or, like the SALT agreements, they have placed ceilings on combinations of weapons features and left each side free to choose a force within the constraints. However here we will examine only treaties that give the two governments no further choice.

Since the numerical values of CII have only comparative meaning, it will be enough to state which agreements give equal instability. The changes in Table 2 decrease crisis instability to the same level, from 2.75 to .556. (The latter value was chosen indirectly since it arose from lowering the warheads per missile from 5 to 4.) The changes are listed in order of the percentage difference in the objective features.

Agreement		% change in parameter	% missiles against silos
increase inaccuracy	.13 nmi to .145 nmi CEP	+ 12%	(80%)
decrease reliability	80% to 69.6%	- 13%	(75%)
decrease warheads/missile	5 to 4 warheads	- 20%	(75%)
decrease yield	500 kt to 351 KT	- 30%	(80%)
increase silo hardness	2000 psi to 2750 psi	+ 38%	(80%)
decrease value invulnerability	577 to 185 warheads	- 61%	(88%)
increase number of missiles	1000 to 3120 missiles	+312%	(87%)

Table 2. Parameter changes producing equal improvements in the Crisis Instability Index.

The third column in Table 2 shows the proportion of weapons the government striking first would allocate to silos.

The most effective remedy for instability is to limit accuracy. Accuracies of both sides could be lowered from what they would have become by a mutual agreement to limit testing, or by one side relocating its missiles further away from the opponent's borders.

Reliability is the second most crucial factor, again one influenced by banning tests. Number of MIRVs, yield and silo hardness affect instability in the direction expected, but an odd result appears: increasing the size of the two missile forces increases stability. The buildup postulated would be a joint one, of course, and a great increment in the missile force would be necessary for the given benefit in stability.

Increasing the vulnerability of the nations' values helps. How this might be done is not clear, but we must remember that instability is a matter of perception, so a discovery that nuclear weapons are more damaging to value than previously believed would have the same effect as an objective change. As with an increase in the number of missiles, a large increase is required to get a small benefit.

Many authors have observed that the arms race is like a repeated Prisoners' Dilemma game in that each side tries to maximize its payoff in "security" and by both doing so they make the situation worse. We can look at those weapons changes that would make a government more effective in a war _if_ the other's weapons were held constant, but worsen crisis stability when both sides implement the change. They are: increasing accuracy, reliability, warheads/missile or yield. These are the entrapping features of the weapons. The prisoners can solve their dilemma if they transform a game played without cooperation into one with communication and agreements, and likewise the governments should make these four features subjects of arms limitations agreements, since stability will deteriorate if they are left to follow their natural courses.

4. Ballistic Missile Defenses

In March of 1983, President Reagan announced his view on how to free the nuclear powers from the threat of mutual destruction. He called on scientists to invent ways to destroy missiles before they reach their targets and suggested that America might share this technology with the Soviet Union.

Administration-sponsored studies later emphasized the stationing of weapons in space to destroy ICBMs during their boost phase. Boost interception has the advantage that the missile is attacked while it is a large, visible, slow-moving target and before it releases warheads and penetration aids. It has another feature attractive to the defender: the opponent cannot explode a lattice of high-altitude bursts over the defender's territory to confuse home-based defenses.

The American administration has frequently stated that space-based missile defenses would improve crisis stability by making it more difficult for a first striker to eliminate almost all the other's silos. But assuming that both sides deploy space defenses, the counterclaim is that these systems help eliminate the retaliating missiles and thus give confidence to the first striker. Many Western critics of space-based defenses and the Soviet leadership have stated this sceptical position.[1]

There are several reasons to believe that the government striking first could use a space-based BMD system more effectively, which implies that they are an incentive to be the first-striker:

1) The first striker will face fewer incoming missiles, so the stress on its system due to exhaustion of interceptors or shortness of time will be less.

2) The first striker will have greater foreknowledge of time of the war and can plan the management of its antimissile defenses better.

3) BMD systems are themselves open to an attack. Both the first and second striker can exploit this vulnerability, but the first striker is in a better position to do so.

One reason for 3), the first striker's advantage in attacking the other's BMD system, is the analogue of 1), that the first striker would have better weapons management. A second reason is that in many cases the proposed systems are effective weapons against themselves.[2] A logical opening move would be for the first striker to use its BMD system against the other's, reducing the latter's ability to attack missiles and space defenses. In the form of satellites the stations are beyond the cloak of the atmosphere, in the vacuum of space where most of their proposed methods of destruction operate best. They are difficult to harden and predictable in their location, thus effectively stationary targets like land-based ICBMs. They are also "MIRVed" in the sense that each can attack several of the opponent's weapons.

We wish to determine which effect of space defenses, the lessening of the counterforce potential of the attacker's ICBMs or the reduced dependability of retaliator's surviving missiles, has greater impact effect on instability. We will add a pair of BMD systems to the nuclear exchange model and calculate the consequent instability. Details are given in Appendix B, but we summarize the modification here.

Considerations 2) and 3) above are introduced by an exogenously given parameter, the relative invulnerability RI of the second striker's BMD system. The model then determines how consideration 1), the larger number of ICBMs facing the second striker's defenses, balances off against the reduction of the first-striker's counterforce to affect instability.

The sequence of events would be as follows, calling the first-striker F and the second-striker S.

F attacks S:

1) F uses its anti-missile defenses against S's anti-missile defenses, and launches its missiles against S's silos and value.
2) S uses its remaining anti-missile defenses against F's missiles.

S attacks F:

3) S uses its remaining anti-missile defenses against F's defenses, and launches its remaining missiles against F's value.
4) F uses its remaining anti-missile defenses against S's missiles.

Space-based systems fall into two categories. Some can be approximated by assuming that a single shot either hits and eliminates the booster or does no damage at all. This type includes X-ray lasers and self-propelled or magnetically accelerated projectiles. The other class causes damage to the booster that accumulates during the time the BMD is on target, and includes continuous lasers and particle beams. (For a description of the various types see Weiner 1983, Carter 1984 and Robinson 1983.) The model we give here is more appropriate for systems in the all-or-nothing class rather than the cumulative damage class.

We define the effectiveness E_F of the first striker's BMD system as (the rate of fire in units per second) times (the probability of destruction by a single unit) times (the duration the boosters are vulnerable or the duration before the defense's fire is exhausted, whichever is less). The measurement is to be taken on F's BMD system after S has attacked it.

We make these assumptions:

1) The boosters are launched simultaneously.

2) The BMD system cannot determine when a booster has been destroyed.

3) There is no time cost to retargetting the BMD stations.

4) The BMD's fire is infinitely divisible among the ICBM targets.

Assumption 2) would be accurate especially for particle beams methods, which aim to destroy the electronics of the missile. The other assumptions are regarded as approximations to at least some possible events.

Given these assumptions F's best plan is to divide its fire evenly among the boosters.

By the definition of effectiveness we continue to count hits that would destroy the booster even after the booster has been destroyed, and thus are able to measure effectiveness independently of the number of boosters launched. Another appealing feature of the definition is that if the effect of an attack is to reduce a system's numbers, a system that is twice as big will have twice the value of E_F.

The two sides are assumed to have identical BMD systems. We define the <u>relative invulnerability</u> RI such that the effectiveness E_S of S's BMD system is RI x E_F. The relative invulnerability RI reflects S's disarray from the surprise strike and F's relatively more successful attack on the S's BMD. That is, RI embodies factors 2) and 3) listed above. Using a system to attack the other's might reduce the attacking system's size but we neglect this factor on the assumption that the reduction would be approximately the same for both.

It is shown in Appendix B that the number of boosters remaining after passing through the other's defenses will be

$B_F\ e^{-E_S/B_F}$ for the first striker, and

$B_S\ e^{-E_F/B_S}$ for the second striker,

where B_F and B_S are the numbers of boosters of the respective sides that reach intercept.

For illustration we might suppose that both have deployed 600 satellite stations each with 200 rocket projectiles all capable of firing within the time the opposing boosters are vulnerable, and that the absentee ratio of the satellites is 80%. Assume also that F's attack on S reduces the system down to 30% of its original size and S's attack on F reduces the system down to 60% and that the kill probability of a single rocket is 1/6. Optimistically for the systems' effectiveness we assume that their use in the attack on the other's BMD reduces their size negligibly. We then calculate E_F as 60% x t sec of duration of fire x 200 rockets per t sec of fire per satellite x 20% x 600 satellites x 1/6 prob of kill = 2400. Relative invulnerability RI = 30%/60% = .5 and E_S = .5 x 2400 = 1200. We can substitute these parameters in the two functions above and determine the number of boosters intercepted as a function of the number launched as shown in Figure 1.

FIGURE 1 HERE

Given that a booster reaches intercept, its probability of being intercepted is the height of the functions in Figure 1 relative to the 45^{o}-line. Figure 1 shows that this probability is less for the first striker's boosters and declines with the number of missiles launched. To add space defenses to the nuclear exchange model, we use this probability to adjust the missile unreliability, with each side's unreliability depending on whether it has the role of the first or second striker. We then use the other base parameters of Table 1 to calculate the costs of attacking and retaliating and then the Crisis Instability Index.

The effect on stability of course depends on the effectiveness and relative invulnerability of the BMD system. Figure 2 shows the results for several values of RI and for continuously varying E_F, drawn horizontally.

FIGURE 2 HERE

The vertical scale shows the degree of stability. If this scale depicted the value of CII directly the height of the curve would have no more than ordinal meaning. To make instability more imaginable we determine a world that matches the given value of CII, that has no missile defenses but has a silo hardness different from the baseline value of 2000 psi. The silo hardness of this equivalent system becomes the vertical scale and an indirect measure of CII.

According to Figure 2, with total relative invulnerability (RI=1), as BMD becomes more effective the system approaches total stability. However if the systems have significant relative vulnerability they can be disastrously unstable. For example BMD systems with the parameters in

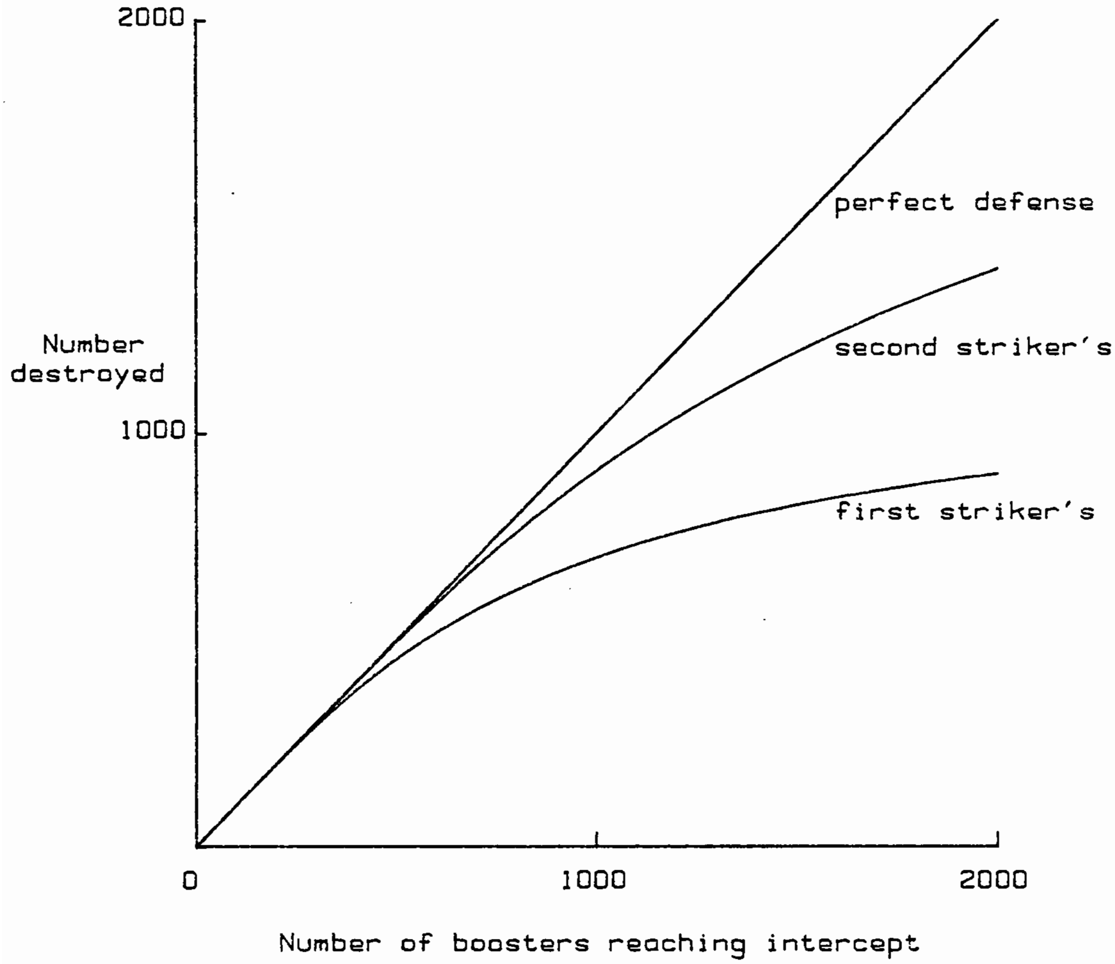

Figure 1: Number of boosters destroyed as a function of number reaching intercept for effectiveness E_F = 2400 and relative invulnerability RI = .5.

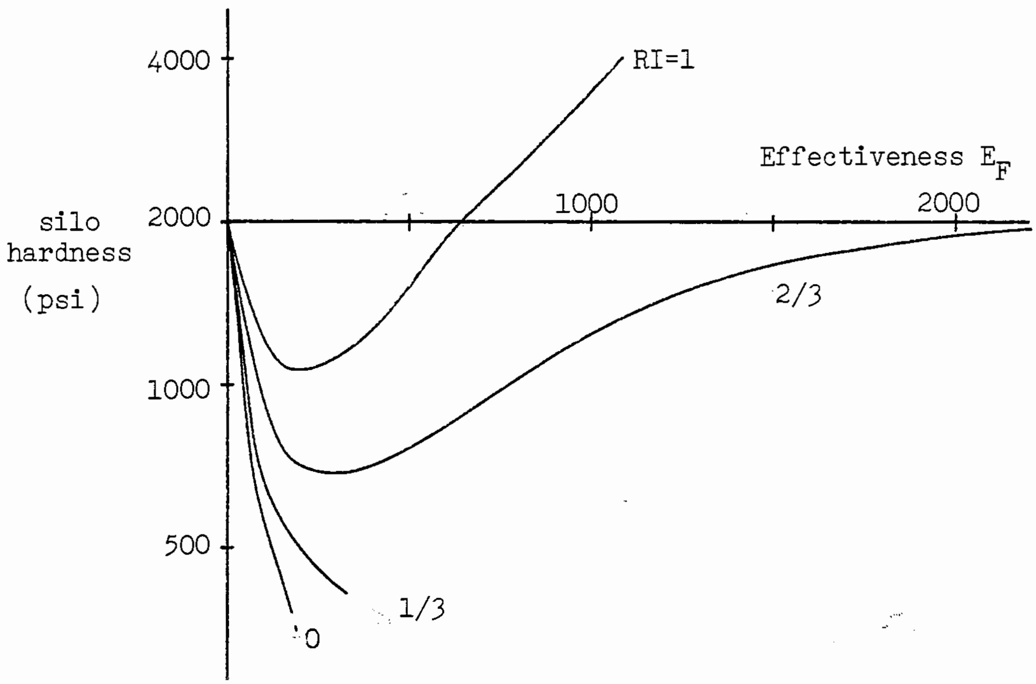

Figure 2: Stability as a function of effectiveness E_F and relative invulnerability RI of BMD. Stability is measured by silo hardness in an equally stable defenseless system.

Figure 1 have lower stability, equivalent to a world in which missile silo hardness was reduced from 2000 psi to 1100 psi.

Usually the effect on stability depends on the joint values of effectiveness and relative invulnerability, but we can make one general statement: any systems with RI less than approximately 2/3 will be destabilizing no matter how effective they are.

In the debate on missile defenses supporters have emphasized the power of technology to come up with unforeseen solutions while critics have cited cost, technical infeasibility and possibility of passive countermeasures. These factors translate here into estimates of effectiveness without regard to the value of RI. Our results suggest that relative invulnerability is also very important.

Introducing a BMD system shifts the question of an offensive advantage one step backwards from a battle of ICBMs to one of BMD systems, but adds a penalty: even an invulnerable BMD system is destabilizing if it is not sufficiently effective. For example a system with RI = 1 will be destabilizing if its effectiveness is below approximately 635.

The reasons behind this dip in stability at low values of effectiveness are revealed by examining the way wars would be fought as E_F increases with RI=1 (and thus $E_S = E_F$). There is almost no change in the number of missiles that F allocates to S's silos between $E_F = 0$ and $E_F = 635$ so allocation of missiles is not an important factor in the dip. The number of missiles surviving F's attack rises gradually from 55 to 386, which restores stability as the system is built up. The decisive factor in the sharp fall is the low number of S's retaliating missiles that get through. At low values of E_F (and therefore E_S) the number of retaliating missiles from S is so small that F's BMD eliminates most of them. Thus consideration 1) causes instability for BMD systems that are not relatively vulnerable to attack by the first striker.

Some side effects of BMD systems should be noted. (These can be calculated from the model but are not shown in Figure 2.) Above E_F values of 3000 or 4000 F diverts its warheads entirely to countervalue and relies entirely on its BMD to defend against retaliation.

If defenses make the price of eliminating an opposing silo too high, a government maintains its deterrent by redirecting its missiles to the other side's vulnerable spots.

For most joint values of effectiveness and relative effectiveness, a joint BMD system decreases the average damage to value should a war occur. (This statistic is determined by the payoffs in the fourth cell in the matrices corresponding to Matrix 1). However when low relative invulnerability combines with moderate to high effectiveness, the damage increases. In our model this occurs when both E_F is above roughly 400 and RI is below 1/4. The reason seems to be that F relies on its defenses to deal with retaliation and increases its proportion of countervalue missiles which do their damage relatively unhindered by S's weakened defenses.

5. Space-based Ballistic Missile Defenses and Submarine Invulnerability

Currently the survivable submarine-based missile forces possessed by the two superpowsers increase crisis stability. Methods to locate all submarines hiding in the oceans seem unlikely in the near future and thus submarines guarantee each goverment the ability to retaliate.

The invention of space-based weapons systems that could destroy sub-launched missiles would be equivalent to a breakthrough in anti-submarine warfare. They would nullify submarine invulnerability since it is strategically irrelevant whether the missiles are destroyed before or after they leave their launching tubes.

The strength of their impact on the stability provided by submarines can be judged by adding submarine missiles to the 1000 land-based missiles already postulated by the model. We suppose each side also possesses 250 submarine missiles with features identical to the land-based missiles both in the missiles' destructive power and their liability to interception by the BMD system.[3] The trend of stability as E_F grows has the same overall shape as Figure 2, the curves are lower, i.e., stability decreases with the addition of BMD for more pairs of values of effectiveness E_F and relative invulnerabilty RI.

The pairs of values of E_F and RI for which stability is improved with and without SLBMs are shown in Figure 3. For example with RI = 1, in a world without sub missiles any BMD system with $E_F < 635$ will hurt stability, but any BMD system with $E_F < 1250$ will hurt, given sub missiles.

FIGURE 3 HERE

---

6. Discussion

The basic result is that space defenses will be destabilizing if they are less than very effective or if they can be degraded with a relative advantage to the first striker. Also destabilization is stronger when one side has a near first-strike potential and more likely when stability is bolstered by submarine missiles.

One might seek to challenge this conclusion by questioning the adequacy of our particular crisis instability measure. However the formula for CII strikes us as supported by the plausibility of its axioms and the intuitive acceptability of the results when it investigated the effect of increases of MIRVs, accuracy and reliability. Perhaps equally justified measures of crisis instability could be designed that would give a different picture of

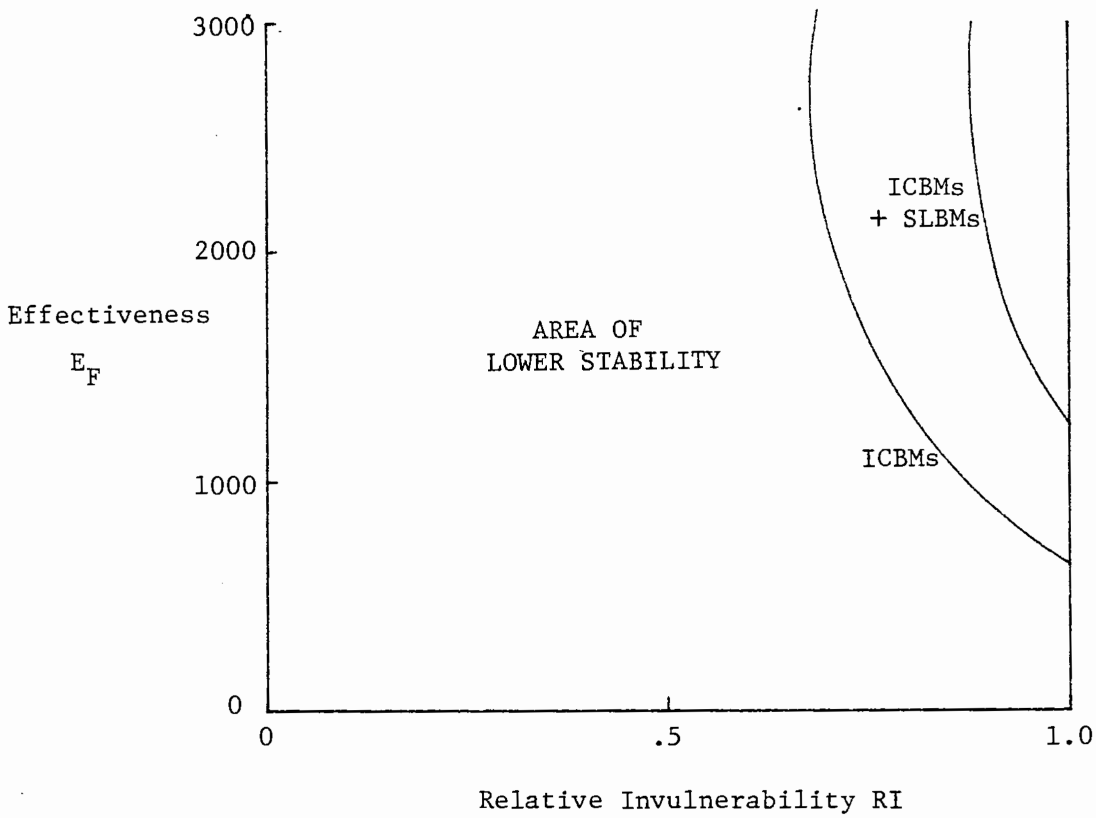

Figure 3: For given E_F and RI of a BMD system stability decreases if the pair of parameters lies to the left of the appropriate curve. "ICBMs" is the curve for weapons holdings of 1000 ICBMs; "ICBMs + SLBMs" for 1000 ICBMs plus 250 SLBMs.

space defenses but we know of none at present.

A different objection to these results might be based on the adequacy of the nuclear war model. Many factors are left out: command and control strikes; fratricide; uncertainty of groups of components; human unpredictability; the fact that most proposed missile defense systems are multilayered; and that each side would not keep its weapons holdings constant as the other built defenses. No doubt there are other varibales overlooked.

Most of these could be added without much difficulty but we judged it best to keep the model relatively simple. A model including them all would be confusingly complicated and perhaps would not convince a sceptic any more than the present one, as the details of the new these factors are open to debate.

Some of the considerations omitted, such as increased weapons holdings or additional layers of defense, involve changes that would probably arise with space defenses. However we have followed the practice common in science and informal argument: isolate the effect of one factor by changing while holding the rest constant. The risk of error here is that the effect of a change of a variable may reverse <u>direction</u> when other variable levels are changed, but we see no positive reason to believe this will happen here.

An example of an additional factor is post-boostphase intercept defenses, either midcourse or terminal. Midcourse interception systems attack all warheads whether they are targeted at force or value and would thus aggravate any effect of space-based defenses if both are relatively vulnerable. Terminal layer defenses designed solely to protect missile silos would contribute to stability. This change would mitigate the effects of space-defenses but could not be expected to reverse their direction. So the policy decisions on the two systems seem to be independent as far as crisis instability is concerned: if space-defenses would worsen the situation without terminal defenses, it seems likely that they so with them.

Constructing models can help us understand the interaction of mechanisms even if we do not derive precise predictions. In the present case, there seems to be a general mechanism at work that would reappear in more complete models of space defenses. That mechanism is the increased effectiveness of a system that has to deal with small numbers of returning missiles compared to one that is near saturation. The condition that there are few returning missiles depends on the counterforce effectiveness of the first strike. Many writers have argued that the feasibility of a first strike is exaggerated, while others regard it as a serious threat and have advocated BMD defenses as an antidote to the vulnerability of land-based missiles. Our model follows the latters' premise through to its ironic conclusion that in situations where first strike potential is greatest, space defenses may increase the danger.

Appendix A. The Basic Nuclear Exchange Model

Justifications for most of the following equations are given by Grotte (1982) and some further descriptions by Bennett (1982).

These equations deal with the effects of an attack of government i against government j. Variables resulting from decisions at the time of war are designated by small letters, and previously determined values by capitals.

M_i : number of i's missiles.

W_i : number of warheads on each of i's missiles.

x_i : number of i's warheads targeted at j's silos.

H_j : hardness of j's silos (pounds per square inch)

$LR_j = .1649[3.48/H_j + \mathrm{sqrt}(12+3.3H_j)/H_j]^{2/3}$
: lethal radius for j's silos, that is, distance at which the missile would be destroyed by a 1 MT explosion.

YD_i : yield of i's warhead (kilotons)

CEP_i : i's circular error probable (nautical miles), defined as the radius of a circle around the target within which one-half of i's missiles will fall.

This definition of CEP assumes that i's missiles land in a distribution centered on j's silos, with no systematic bias due to uncertainty of silo location, or systematic guidance errors during the flight. Warheads detonating within a CEP radius are regarded as sure to destroy the silo. Uncertainty in the model comes not from the properties of the silo or warhead, but from aiming accuracy and an ICBM reliability factor.

$PK_{ij} = 1 - .5 \exp[(LR_j\ YD_i\ /CEP_i)^j]$
: single shot kill probability, likelihood that a single functioning i warhead will destroy its target missile.

R_i : reliability of an i warhead, including failure of boosters and bus.

Government i divides its anti-silo warheads as evenly as possible among j's silos, and therefore some silos will receive exactly one more than the others. (Our model is different from Grotte's in this aspect which allows non-integral number of warheads assigned to silos and can handle larger problems involving different types of weapons.)

wl_i = greatest integer less than x_i/M_i
: number of warheads targeted at some silos.

$wg_i = wl_i + 1$: number of warheads targeted at all other silos.

$sg_j = x_i - wl_i\ M_j$: number of warheads against silos receiving greater number of warheads.

$sl_j = M_j - sg_j$: number of warheads against silos receiving lesser number of warheads.

$y_j = W_j sl_j(1 - R_i PK_{ij})^{wl_i} + W_j sg_j(1 - R_i PK_{ij})^{wg_i}$
: number of j warheads surviving

We have also assumed the effects of successive warheads are independent. This is not true in fact due to the phenomenon of fratricide, in which one explosion degrades the accuracy of or destroys later warheads. If we were to take account of fratricide, the crisis instability would improve and the number of missiles targeted against silos would decrease.

$AEMT_i = YD_i^{.4}$: the adjusted equivalent megatons of i's warhead.

This variable measures the damage from a single explosion against value taking account that damage is circular and cannot be adjusted to the size and shape of the target (Downey, 1976). This formula was assumed to be the same for both first and second strikes, implying that a second striker is able to target its anti-value warheads with the same efficiency as the first striker.

V_i : i's total value.

VI_i : value invulnerability of country i.

Both nations are assumed to have total value 1.

$U_i = .0577/VI_i$: i's rate of value damage, increment in value damage per small increment in adjusted equivalent megatons. This was estimated to be .0001.

$d_{j,ij} = V_j(1 - e^{[R_i U_j (M_i W_i - x_i) AEMT_i]})$
: damage to j's value when i strikes first against j.

$d_{i,ij} = V_i(1 - e^{(-U_i R_j y_j AEMT_j)})$
: damage to i's value when i strikes first against j.

K_i : i's tradeoff of j's value for its own.

Government i's war plans are set up to trade a decrement of K_i units of j's value for $1-K_i$ of its own. This parameter is assumed to be 1/3 for both.

Government i controls the variable x_i, the number of warheads directed against j's silos. The two damages are determined as functions of x_i and i can approximate an optimal x_i by solving the nonlinear program for the proportion of warheads allocated to silos

$$\max_{p_i \text{ in } [0,1]} [K_i d_{j,ij}(p_i) - (1-K_i) d_{i,ij}(p_i)]$$

where and setting $x_i = p_i M_i W_i$. Since this objective function is one-dimensional, and in all cases we calculated was verified to be a single-peaked function of p, the program was solved using a simple line search routine, in which four points were place along the line at intervals 0, .4,

.6 and 1. If the function evaluated at the second point was greater than the third, then the third point became the new end of the line, otherwise the second point became the new beginning of the line. This procedure was repeated until the possible interval of the maximum was acceptably narrow.

Appendix B. The Nuclear Exchange Model including Space-based Defenses.

We assume that neither country can retarget other missiles to replace losses due to the opponent's BMD system. The introduction of a BMD system by i is then equivalent to a reduction in j's missile reliability, which for large number of missiles can be approximated by a decrease in warhead reliability, R_j. (If we allowed retargetting, i's BMD system would be equivalent to a decrease in j's missiles M_j.)

Some of the factors causing the failure of j's warhead can become manifest before the intercept point, and some can afterwards. We must add an assumption of how much occurs in each phase but the exact proportion chosen will not change the results greatly. We assume that half the decrement from perfect reliability occurs before, and that any boosters that fail then are not targeted by i's BMD system.

$R_j' = 1 - (1-R_j)/2$: probability that j's warhead arrives at intercept point.

To find the likelihood of an intercept we assume that the rate of fire of i's entire BMD system is F_i units per second. The attacking units are spread evenly over j's boosters, independent of the number of boosters as justified in the text.

The boosters are vulnerable for T_j seconds and a single BMD unit has a single-shot kill probability of PD_i. We define the effectiveness of i's system when i takes the role of the first striker to be:

$EF_j = F_i PD_i T_j$: effectiveness of i's BMD when i has the role of the first striker.

$RS_j = e^{-EF_i/R_j'M_j}$: approximate probability that j's warhead passes the intercept point given that it arrives there, when j is the second striker.

Government i as the first striker will attack certain of j's anti-BMD weapons and degrade j's anti-BMD capacity. This decline can result in a decrease in the size of j's system (F_j) or in the accuracy of the system (PD_j), or speed of locating the boosters T_i, but since these occur as multiplicative factors in the formula for j's effectiveness, we can represent the effect by a single overall relative effectiveness.

RI_j : j's relative invulnerability when j is the second striker.

Then $E_{Sj} = RI_j F_j PD_j T_i$: effectiveness of j's BMD system when j has the role of the second striker,

and $RF_i = e^{-ES_j/R_i M_i}$: approximate probability of i's warhead passing intercept given it arrives there, when i is the first striker.

Finally we take account of unreliability in the rest of the journey:

$R_j'' = 1 - (1-R_j)/2$: probability that j's warhead detonates given that it passes the intercept point.

The overall reliability is then the product of these components, R_i' RF_i R_i'' for a first striker or R_j' RS_j R_j'' for a second striker. We assume that the attack on the other side's BMD system is made with one's own system, and that the part allocated to this attack is somehow fixed, and not counted in the effectiveness used in calculating the destruction of boosters. The nuclear exchange model with BMD is then applied by substituting these products for R_j in the model of Appendix B.

References

Baugh, W.H. The Politics of the Nuclear Balance. New York: Longman. 1984.

Bennett, B.W. Assessing the capability of strategic nuclear forces: limits of current methods. RAND N-1441. Santa Monica, Ca. 1980.

Carter, Ashton. Directed Energy Weapons in Space Report for the Office of Technology Assessment, U.S. Congress, April 1984.

Davis, W.A. Nine personal views. in A. Carter and D. Schwartz, eds., Ballistic Missile Defense. Washington: The Brookings Institution. 1984.

Downey, T. How to avoid Monad and disaster. Foreign Policy. 1976. 24.

Drell, S., P. Farley and D. Holloway. A critique of the Reagan Strategic Defense Initiative. International Security. 1984, 51-86.

Glaser, C. Why even good defenses may be bad. International Security 1984, 92-123.

Grotte, J.H. Further development of a strategic weapons exchange model. P-1322 Institute for Defense Analyses, Arlington, Va. 1978.

Grotte, J.H. Measuring strategic stability with a two-sided nuclear exchange model. Journal of Conflict Resolution, 24, 1980, 213-239.

Grotte, J.H. An optimizing nuclear exchange model for the analysis of nuclear war and deterrence. Operations Research 1982, 30, 428-445.

Hoeber, F. Strategic nuclear exchange models (Ch.7) in F. Hoeber, Military Applications of Modeling. New York: Gordon. 1981.

Payne, K. Strategic defense and stability. Orbis Summer 1984. 216-229.

Robinson, C. Panel urges boost-phase intercepts. Aviation Week and Space Technology Dec. 5, 1983, 50-57.

Weiner, S. Systems and technology. in A. Carter and D. Schwartz, eds. Ballistic Missile Defense. The Brookings Institution. Washington, D.C. 1983.

Footnotes:

1: For a typical statement of the "deterrence enhancement" argument see the White House issued pamphlet The President's Strategic Defense Initiative (New York Times, Jan.4, 1985): "By significantly reducing or eliminating the ability of ballistic missiles to attack military forces effectively and thereby rendering them impotent and obsolete as a means of supporting aggression, advanced defenses could remove this potential source of instability."

The counterclaim was made by Premier Andropov immediately after President Reagan's speech "... the strategic offensive force of the US will continue to be developed and upgraded at full tilt and along quite a definite line at that, namely that of acquiring a first nuclear strike capability. Under these conditions the intention to secure itself the possibility of destroying with the help of the ABM defenses the corresponding strategic systems of the other side that is of rendering it unable of dealing a retaliatory strike is a bid to disarm the Soviet Union in the face of the US nuclear threat.", (Pravda Mar.27, 1983, cited in Drell et. al. p.105.)

For the position that instability considerations are irrelevant to space defenses or are correctable see Davis (1983) and Glaser (1984).

2: Points 1) and 2) previously made refer to BMD systems attacking ICBMs. The "second reason" given here is the analogue of 2) when one views the conflict as BMD attacking BMD, just as the first reason is the analogue of 1).

3: The assumption of equal holdings by the two sides is made only for conceptual simplicity -- the nuclear exchange model and the crisis stability index are able to handle assymetrical force structures without difficulty.