

ECONOMIC RESEARCH REPORTS

DETERRENCE VERSUS DEFENSE:
A GAME-THEORETIC MODEL OF STAR WARS

by

Steven J. Brams

and

D. Marc Kilgour

R.R. #86-35

December 1986

**C. V. STARR CENTER
FOR APPLIED ECONOMICS**



**NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003**

DETERRENCE VERSUS DEFENSE: A GAME-THEORETIC MODEL OF STAR WARS

Steven J. Brams*
Department of Politics
New York University
New York, NY 10003

D. Marc Kilgour**
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario
Canada N2L 3C5

*Steven J. Brams gratefully acknowledges the financial support of the Ford Foundation under Grant No. 845-0354 and the National Science Foundation under Grant No. SES84-08505. Valuable support from the C.V. Starr Center for Applied Economics at New York University is also acknowledged.

**D. Marc Kilgour gratefully acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada under Grant No. A8974.

ABSTRACT

A game-theoretic model of the Strategic Defense Initiative, or "Star Wars," is developed based on a deterrence model founded on the game of Chicken. In this model, two players can choose any level of preemption, and threaten any level of retaliation against preemption, whereas in the Star Wars Game they are constrained in these choices by the defensive capabilities of an opponent.

Nash equilibria, or stable outcomes, are derived in this game and illustrated for three different scenarios involving various postulated relationships between the first-strike and second-strike defenses of the players. Unlike the deterrence model, mutual preemption emerges as an equilibrium in the Star Wars Game, underscoring the problem--particularly if defensive capabilities are unbalanced--of deterrence's being subverted by the development of Star Wars. Ramifications of this model for avoiding preemption and preserving crisis stability, especially in superpower relations, are discussed.

DETERRENCE VERSUS DEFENSE: A GAME-THEORETIC MODEL OF STAR WARS

Steven J. Brams and D. Marc Kilgour

1. Introduction

The furor caused by President Reagan's Strategic Defense Initiative (SDI)--or, more popularly, Star Wars--has gone unabated since his speech of March 23, 1983, announcing the initiative.¹ We suggest there are two main reasons for the continuing controversy. First, the initiative took the nation and the world, including the defense community, by surprise; even today, more than three years later, questions about SDI's feasibility and effectiveness sharply divide proponents and opponents.

Second, an extraordinary amount of money over a short period of time has been allocated to research and testing of various aspects of SDI; this has created a growing pro-SDI constituency in addition to instigating a strong anti-SDI reaction. Moreover, given the possibility of much more money to come for further research and development and eventual deployment of the system, the forces on both sides are likely to expand as competition for SDI funds intensifies.

Our concern in this paper is not with this domestic political competition but rather with assessing SDI's strategic effects on nuclear deterrence. Our approach is unorthodox. We shall ignore all issues connected with the technological feasibility of SDI, which have dominated debate about Star Wars and even questions of whether a viable defense against nuclear weapons is possible, much less cost-effective. Neither

shall we try to analyze SDI's international political effects, including its impact on Soviet decision making, the behavior of allies, or the prospects of negotiating arms-control agreements. Rather, our focus will be on modeling the stability of nuclear deterrence as Star Wars becomes a more and more effective defensive system.

To be sure, the "stability question" has been discussed in the literature, but there has been almost no formal modeling of this question to try to ascertain the precise trade-offs between deterrence and defense.² Ostensibly, as each side develops measures that could blunt a first strike against it and enhance its ability to retaliate, a potential opponent's uncertainties about the putative advantages of attacking first would seem to increase, thereby strengthening deterrence. On the other hand, because this same defensive system can be used to attenuate a second strike, it could undermine the threat of a retaliatory counterattack meant to deter the first strike. Thus, the effects of Star Wars may pull in different directions, helping and hurting deterrence at the same time.

In the next section we shall outline a game-theoretic model of deterrence, based on Chicken, that has been used to analyze the stability of nuclear deterrence between the superpowers and the optimality of different kinds of threats (Brams, 1985; Brams and Kilgour, 1985a, 1985b, 1986a, 1986b, 1986c, 1986d, 1986e). In section 3 we shall modify the rules of the "Deterrence Game" to allow for the possibility of varying levels of defense against first and second strikes in a "Star Wars Game." We will then demonstrate in section 4 the existence of different kinds of Nash (1951) equilibria in this game.

In section 5 we posit three different scenarios for the Star Wars Game and show graphically the regions in which various equilibria, often overlapping, exist. In the first scenario, for example, we assume that each side's first-strike and second-strike defenses are the same and show that, in addition to the deterrence and unilateral preemption equilibria of the Deterrence Game, new and disturbing equilibria crop up in the Star Wars Game--namely, equilibria of mutual preemption--especially as Star Wars defenses become better and better.

The general picture that emerges from the different scenarios--putting aside the question of the cost of building a Star Wars system--is not encouraging. Basically, at low levels of defense Star Wars leaves deterrence intact, even when only one side has it, chiefly because deterrence via a retaliatory second strike is still possible. At higher levels, an incentive for mutual preemption develops, especially when one side is substantially ahead of the other and is therefore not deterred to the extent it would be if it were facing an equal foe, which may in turn induce the weaker side also to preempt.

A symmetry in Star Wars capabilities sometimes offers help, staying preemption and prolonging deterrence as defenses improve. Yet, each side may still be motivated to preempt unilaterally at low levels of defense and --more dangerous--at high levels mutual preemption is an equilibrium. Fortunately, the stability of mutual preemption in the Star Wars Game is counterbalanced by the fact that deterrence equilibria often coexist with the mutual-preemption equilibria when the Star Wars capabilities of the two sides are relatively high and more or less equal--and, in most plausible situations, the deterrence equilibria are better for both players.

In a crisis, however, the notion of preemption may come to the fore and, especially if it is suspected that an opponent may preempt, one does better also by preempting than by absorbing a first strike and then retaliating. This incentive toward mutual preemption never exists in the Deterrence Game, which suggests that Star Wars, paradoxically, may be compatible with deterrence (especially at low levels of defense) yet undermine crisis stability (by making mutual preemption an equilibrium at higher levels).

We shall explore these contradictory effects of Star Wars in the final section of the paper and indicate the tricky time path, or trajectory, that the superpowers must adhere to if they are to avoid a possible cataclysm, particularly in the transition period from principally deterrent to principally defensive postures. We conclude with recommendations on controls to strengthen crisis stability, the Achilles heel of Star Wars if its development by the superpowers is uncoordinated and one side should get a defensive edge on the other.

2. The Deterrence Game³

The Deterrence Game is based on the two-person game of Chicken, which we shall first describe and briefly analyze. In Chicken, each player can choose between two strategies: cooperate (C) and not cooperate (\bar{C}), which in the context of deterrence may be thought of as "not attack" and "attack," respectively. These strategies lead to four possible outcomes, which the players are assumed to rank from best (4) to worst(1). These rankings

are shown as ordered pairs in the outcome matrix of Figure 1, with the first number indicating the rank assigned by the row player (called "Row"), and the second number indicating the rank assigned by the column player (called

Figure 1 about here

"Column"). Chicken is defined by the following outcome rankings of the two players:

1. Both players cooperate (CC)--next-best outcome for both players: (3, 3).
2. One player cooperates and the other does not ($\bar{C}C$ and $C\bar{C}$) --best outcome for the player who does not cooperate and next-worst outcome for the player who does: (2, 4) and (4, 2).
3. Both players do not cooperate ($\bar{C}\bar{C}$)--worst outcome for both players: (1, 1).

Outcomes (2, 4) and (4, 2) in Figure 1 are circled to indicate that they are Nash equilibria: neither player (Row or Column) would have an incentive unilaterally to depart from these outcomes because he would do worse if he did. For example, from (2, 4) Row would do worse if he moved to (1, 1), and Column would do worse if he moved to (3, 3). By contrast, from (3, 3) Row would do better if he moved to (4, 2), and Column would do better if he moved to (2, 4).⁴

The shorthand verbal descriptions given for each outcome in Figure 1 suggest the vexing problem the players confront in choosing between C and \bar{C} : by choosing \bar{C} , each can "win" but risks disaster; by choosing C, each might benefit from compromise but could also "lose." Each of the Nash equilibria

FIGURE 1

OUTCOME MATRIX OF CHICKEN

		<u>Column</u>	
		Cooperate (C)	Do not cooperate (\bar{C})
<u>Row</u>	Cooperate (C)	(3, 3) Compromise	(2, 4) Column "wins," Row "loses"
	Do not cooperate (\bar{C})	(4, 2) Row "wins," Column "loses"	(1, 1) Disaster

Key: (x,y) = (rank of Row, rank of Column)

4 = best; 3 = next best; 2 = next worst; 1 = worst

Circled outcomes are Nash equilibria

shown in Figure 1 favors one player over the other, and the stability of these equilibria as such says nothing about which of the two--if either--will be chosen.

Other concepts of equilibrium distinguish (3, 3) as a stable outcome, but the rules of play that render compromise stable presume that the players (i) act nonmyopically or farsightedly and (ii) cannot threaten each other (Brams and Wittman, 1981; Kilgour, 1984, 1985; Zagare, 1984). If threats are possible in repeated play of Chicken under still different rules, the stability of (3, 3) is undermined (Brams and Hessel, 1984).

The effect that threats may have in Chicken is not hard to grasp. If one player (say, Row) threatens the other player (Column) with the choice of \bar{C} , and this threat is regarded as credible, Column's best response is C, leading to (4, 2).

Clearly, the player with the credible threat--if there is one--can force the other player to back down in order to avoid (1, 1). Although Row would "win" in this case by getting his best outcome, Column would not "lose" in the usual sense by getting his worst outcome but rather his next-worst.

This illustrates that Chicken is not a constant-sum game, in which what one player wins the other player loses. That is why we have put "win" and "lose" in quotation marks here and in Figure 1. In variable-sum games like Chicken, the sum of the players' payoffs at each outcome (if measured cardinally by utilities rather than ordinally by ranks) is not constant but variable. This means that both players may do better at some outcomes [e.g., (3, 3)] than at others [e.g., (1, 1)]. Outcomes, such as (1, 1) in Chicken, which are inferior for both players to some other outcome(s) in

a game, are called Pareto-inferior; those outcomes which are not Pareto-inferior are Pareto-superior, as are the other three outcomes in Chicken.

The Deterrence Game is based on Chicken but adds two refinements: (i) the players can make quantitative choices of levels of cooperation (or noncooperation), not just qualitative choices of C or \bar{C} ; (ii) once these initial choices, which we shall interpret as levels of nonpreemption (or preemption) are made, each player's choice of a subsequent level of retaliation comes into play, provided his opponent's initial choice was regarded as preemptive and his own was not. More formally, the Deterrence Game is defined by the following rules:

1. The final outcome will be one of the four outcomes of Chicken. The payoffs are the same as those of Chicken, except that cardinal utilities replace ordinal rankings. Thus r_n and c_n signify the highest payoffs for Row and Column, respectively, r_1 and c_1 the lowest, etc.
2. The players do not choose initially between C and \bar{C} , as in Chicken, but instead choose (unspecified) actions that have associated a nonpreemption probability (s for Row and t for Column) and a complementary preemption probability ($1 - s$ for Row and $1 - t$ for Column). With these probabilities, the actions will be interpreted as cooperative (C) and non-cooperative (\bar{C}) strategy choices, respectively.
3. If both players' initial choices are perceived as the same, the game ends at that position (i.e., CC or $\bar{C}\bar{C}$). If one player's choice is perceived as C and the other's as \bar{C} , the

former player chooses a subsequent action with an associated nonretaliation probability (p for Column and q for Row) and a complementary retaliation probability ($1 - p$ for Column and $1 - q$ for Row). With the retaliation probability, the conflict is escalated to the final outcome \overline{CC} ; otherwise it remains (at \overline{CC} or \overline{CC}).

4. The players choose their preemption probabilities and retaliation probabilities before play of the game. Play commences when each player simultaneously chooses initial actions that may be interpreted as either C or \overline{C} , with associated preemption probabilities. One player may then choose subsequent actions, according to rule 3, with the associated retaliation probability specified at the beginning of play.

The Deterrence Game is represented in Figure 2. Note that besides

Figure 2 about here

the fact that the initial strategy choices of the two players are probabilities (with assumed underlying actions), rather than actions (C and \overline{C}) themselves, this representation differs from the Figure 1 outcome matrix in having expected payoffs rather than (certain) payoffs in its off-diagonal entries. This is because we assume that if one player is perceived to preempt, the other player's (probabilistic) retaliation will be virtually instantaneous, so it is proper to include in the off-diagonal entries a combination of payoffs--reflecting both possible retaliation and possible nonretaliation--by means of an expected value.

FIGURE 2

DETERRENCE GAME

		<u>Column</u>	
		t	1 - t
<u>Row</u>	s	(r_3, c_3)	$q(r_2, c_4) + (1 - q)(r_1, c_1)$ $= (qr_2, q)$
	1 - s	$p(r_4, c_2) + (1 - p)(r_1, c_1)$ $= (p, pc_2)$	$(r_1, c_1) = (0, 0)$

Key: $(r_i, c_j) = (\text{payoff to Row, payoff to Column})$

$r_4, c_4 = \text{best}; r_3, c_3 = \text{next best}; r_2, c_2 = \text{next worst};$

$r_1, c_1 = \text{worst}$

$s, t = \text{probabilities of nonpreemption}; p, q = \text{probabilities of nonretaliation}$

Normalization: $0 = r_1 < r_2 < r_3 < r_4 = 1;$

$0 = c_1 < c_2 < c_3 < c_4 = 1$

We assume, of course, that $0 \leq s, t, p, q \leq 1$ because they represent probabilities. To simplify subsequent calculations, we normalize the payoffs of the players so that the best and worst payoffs are 1 and 0, respectively. Hence,

$$0 = r_1 < r_2 < r_3 < r_4 = 1;$$

$$0 = c_1 < c_2 < c_3 < c_4 = 1.$$

Because we assume the preemption and retaliation probabilities are chosen independently by the players, the expected payoffs for Row and Column are simply the sums of the four payoffs (or expected payoffs) in the Figure 2 matrix, each multiplied by the probability of its occurrence:

$$E_R(s, q; t, p) = str_3 + (1 - s)tp + s(1 - t)qr_2; \quad (1)$$

$$E_C(t, p; s, q) = stc_3 + s(1 - t)q + (1 - s)tpc_2. \quad (2)$$

Besides an equilibrium in mixed strategies (analogous to the one in Chicken described in note 4), which we called "naive" (Brams and Kilgour, 1985a) because it is Pareto-inferior to the Deterrence Equilibrium and hence not likely to arise in practice, there are effectively three Nash equilibria in the Deterrence Game. They can be grouped into two classes:

I. Deterrence Equilibrium: $s = 1, q \leq c_3; t = 1, p \leq r_3$. This equilibrium is one in which the players never preempt ($s = t = 1$), but Row retaliates with probability $1 - p \geq 1 - r_3$ and Column retaliates with probability $1 - q \geq 1 - c_3$. Essentially, these inequalities ensure that a player's expected payoff as the sole preemptor-- p for Row and q for Column, as shown in the off-diagonal entries in Figure 2--is not greater than what

is obtained from the cooperative outcome of the underlying Chicken game, with payoffs (r_3, c_3) .

II. Preemption Equilibria: (1) $s = 1, q = 1; t = 0, p$ arbitrary; (2) $s = 0, q$ arbitrary; $t = 1, p = 1$. The first equilibrium is certain preemption by Column and no retaliation by Row; because Row is deterred by Column's initiative, Column's retaliation probability is arbitrary since it never comes into play. The second equilibrium is analogous, with the roles of Column and Row switched. At these equilibria, the outcomes of the Deterrence Game are the outcomes of the underlying Chicken game associated with wins for Column and Row, with payoffs $(r_2, 1)$ and $(1, c_2)$, respectively.

Of the three Pareto-superior Nash equilibria, only the Deterrence Equilibrium (Class I) depends on the possibility of retaliation-- specifically, precommitted threats to respond (at least probabilistically) to a provocation when it is viewed as equivalent to the choice of \bar{C} . Such threats distinguish the Deterrence Game from the underlying game of Chicken, in which retaliation against the choice of \bar{C} is not permitted.

Note that the two Preemption Equilibria in Class II occur only when retaliatory threats are never used ($p = 1$ or $q = 1$ or both). They correspond precisely to the two pure-strategy Nash equilibria in Chicken and so introduce no new element into the analysis of deterrence beyond what was earlier provided by Chicken. However, when a threat structure is added to Chicken to yield the Deterrence Game, a qualitatively different equilibrium (the Deterrence Equilibrium) emerges. This

equilibrium demonstrates how threats can work to the advantage of both players to stabilize the Pareto-superior cooperative outcome (r_3, c_3) , which is unstable in Chicken without the possibility of retaliation. Moreover, not only is this outcome a Nash equilibrium but it is also a dominant-strategy Nash equilibrium when precommitted retaliation probabilities are fixed, since $r_3 > p$ and $qr_2 > 0$, so that an initial cooperative choice is unconditionally best for Row, and similarly for Column. This dominance would seem a strong argument for choosing the Deterrence Equilibrium over the Preemption Equilibria.

Because the Deterrence Equilibrium depends fundamentally on threats that would be costly for the threatener to implement, it is neither perfect nor subgame-perfect in the sense of Selten (1975).⁵ Nevertheless, the Deterrence Equilibrium possesses a dynamic-stability property that should, once the equilibrium forms, contribute to its persistence in repeated play. That is to say, given that the players are at the Deterrence Equilibrium, if one player (say, Column) for any reason suspects that the other player (Row) may contemplate preemption, thereby rendering $s < 1$, Row can still do no better than continue to choose $t = 1$.

In other words, even should Row think he might be preempted, he should nonetheless continue to refuse to preempt in order to keep his expected payoff at its maximum. This obviates the problem that Schelling (1960, ch 7) called "the reciprocal fear of surprise attack" that leads inexorably to preemption.

Our proof (Brams and Kilgour, 1985a, Appendix) of this dynamic-stability property of the Deterrence Equilibrium shows, in effect,

that any perceived departures from 1 of s or t will not initiate an escalatory process whereby the players are motivated to move closer and closer to preemption for certain. The fact that the Deterrence Equilibrium is impervious to any perturbation in s or t means that the players, instead of being induced to move up the escalation ladder, will have an incentive to move down should one player deviate from $s = t = 1$.

The restoration of the Deterrence Equilibrium depends on probabilistic threats of retaliation that satisfy

$$0 < q < c_3; \quad 0 < p < r_3.$$

But note that if deterrence for any reason should fail, it is irrational to retaliate, even on a probabilistic basis, because retaliation leads to a worse outcome for the retaliator, as well as for the player who preempted and thereby provoked retaliation.

The apparent irrationality of retaliating in the Deterrence Game is precisely what makes the Deterrence Equilibrium imperfect. Despite its imperfectness, it, like the Preemption Equilibria, has the essential equilibrium property of "mutually fulfilled expectations": if either player anticipates that his opponent will choose an equilibrium strategy--associated with one of the three equilibria--he can never do better than choose his corresponding equilibrium strategy (associated with this same equilibrium) to maximize his payoff.⁶

We have also explored the question of whether, if the players are at one of the two Preemption Equilibria, the preempted player can change his strategy in such a way as to induce the preemptor to switch his own strategy to the Deterrence Equilibrium (Brams and Kilgour, 1985b). In fact,

if the preempted player threatens to move from his next-worst to his worst outcome unless the preemptor cooperates initially, he makes it irrational for the preemptor to ignore his threat. Thereby he induces the preemptor to shift to (r_3, c_3) , the Deterrence Equilibrium.

The questions we address next are the possible effects of a Star Wars defense on Nash equilibria in the Deterrence Game.

3. The Basic Star Wars Game

Recall that, in the Deterrence Game, Row could choose any nonpreemption probability s and nonretaliation probability q , and Column could choose nonpreemption and nonretaliation probabilities of t and p , respectively. In the Star Wars Game, we restrict these choices by assuming that the players' defensive capabilities put upper bounds on the amounts of preemption $(1 - s, 1 - t)$ and retaliation $(1 - q, 1 - p)$ that they can choose or, equivalently, lower bounds on $s, t, q,$ and p .

To be more precise, we assume that the partial shield of Column's Star Wars defense implies that Row's first strike must satisfy $v_C \leq s \leq 1$, where $0 \leq v_C < 1$; similarly, Row's defense will force Column's first-strike choice to satisfy $v_R \leq t \leq 1$, where $0 \leq v_R < 1$. In each case, the parameters v_C and v_R constrain the range of the players' choices, giving minimum levels of cooperativeness (or maximum levels of noncooperativeness) of a first strike. Put another way, because v_C and v_R measure Column and Row's defensive capabilities against first strikes by their opponents, $1 - v_C$ and $1 - v_R$

measure their opponents' maximal attack capabilities and may be thought of as ceilings on the magnitude of the players' first strikes: Row and Column's choices of $1 - s$ and $1 - t$ may not exceed these ceilings ($1 - s \leq 1 - v_C$; $1 - t \leq 1 - v_R$), which are equivalent to the upper bounds of v_C for s , and v_R for t , given earlier.

We use additional parameters w_C and w_R to denote the restrictions on each player's second strike imposed by Column and Row's defense capabilities. Specifically, these restrictions are

$$w_R \leq p < 1 \text{ and } w_C \leq q < 1.$$

For example, $1 - w_R$ is a ceiling on the magnitude of a second strike by Column, in response to a first strike by Row.

It is reasonable to assume that the parameters w_R and w_C are functions of v_R and v_C (and, possibly, of other factors as well). Generally it is to be expected that w_R is an increasing function of v_R (because Row's first-strike defense would help against second strikes as well) and a decreasing function of v_C (because the better Column's defense against Row's first strike, the stronger Column's second strike can be). In fact, to illustrate specific results of our basic model in section 5, we postulate in one scenario

$$w_R = v_R(1 - v_C); w_C = v_C(1 - v_R).$$

Before investigating the implications of this and other functional relationships, it is useful to clarify how the Star Wars constraints, v_C , v_R , w_C , and w_R , affect the payoffs of the players and ultimately their rational strategy choices in the Star Wars Game. To begin with, assume there is no first strike by Column (i.e., $t = 1$). If Row also chooses

no first strike (i.e., $s = 1$), then the payoffs to the players are (r_3, c_3) . On the other hand, if Row's actions are considered to be a first strike (with probability $1 - s > 0$), and there is no Star Wars defense by Column, the payoffs to the players are $(1, c_2)$, at least temporarily (since Column might retaliate).

Now assume that Column has some defensive capability, so the maximal first strike Row can inflict on him is $s = v_C$, somewhere between the extremes of $s = 0$ and $s = 1$. This means that the payoff Row derives from a maximal first strike is equivalent to a mixture of no first strike (r_3) and a maximal first strike against no defense (1) in the proportions of v_C to $1 - v_C$.

Thus, the greater v_C , the more Row is forced to partake of r_3 , rather than this maximal payoff of 1 after he preempts. The reason, of course, is that Column's Star Wars defense prevents Row from moving beyond v_C toward $s = 0$. By contrast, when Column has no defense $v_C = 0$, allowing for $s = 0$ and payoffs of $(1, c_2)$; when Column's defense is total ("leakproof") $v_C = 1$, ensuring that $s = 1$ so that Row's choice of a first strike is payoff-equivalent to the choice of no first strike, and payoffs are (r_3, c_3) . It is worth noting that the $v_C:1 - v_C$ payoff ratio we assume of Row when he strikes first simultaneously applies to the mixture between c_3 and c_2 for Column.

Now consider possible second strikes. Assume that Column's actions are believed to constitute a first strike against Row (with probability $1 - t > 0$). If Row does not retaliate, the players' payoffs are $(r_2, 1)$, whereas maximal retaliation by Row would give payoffs of $(0,0)$ if Column had no defense.

If Column has some defensive capability, however, the maximal retaliation, w_C , that Row can inflict on him falls between the extremes of $q = 0$

and $q = 1$. Analogous to our assumptions about first strikes, we interpret Row's maximal second strike against Column's defense as payoff-equivalent to a mixture in the proportion $w_C:1 - w_C$ of no retaliation (payoff of r_2 to Row) and maximal retaliation against no defense (payoff of 0).

As with v_C in the case of first strikes, $w_C = 0$ means that Row can retaliate to any extent he chooses, and $w_C = 1$ means that Row has no retaliation option. As before, we assume that the payoff mixture for Column is in the same ratio as for Row.

These assumptions for the players fix the payoffs they derive in the Deterrence Game when Star Wars defenses are erected. Expected payoffs in the resulting Star Wars Game are the same as those given by (1) and (2) in the Deterrence Game, except that there are now the previously given constraints on the values that the players' strategy choices-- s and q for Row, t and p for Column--can assume in the Star Wars Game.

4. Nash Equilibria

The Nash equilibria of the Star Wars Game are derived in the Appendix. Some of these equilibria closely resemble those in the Deterrence Game, so we use the same terminology as in the latter game:

I. Deterrence Equilibrium (DE): $s = 1, w_C \leq q \leq c_3; t = 1, w_R \leq p \leq r_3$. As in the Deterrence Game, the players never preempt but threaten to retaliate if preempted with probabilities sufficient to deter their opponents but within

their retaliation thresholds. The difference in the Star Wars Game is that, to make a first strike unprofitable, the players must have enough second-strike capability to retaliate against each other's defenses. This capability is ensured if Row can choose a retaliation probability $1 - q$ less than the ceiling of $1 - w_C$ imposed by Column's second-strike Star Wars defense (i.e., $1 - q \leq 1 - w_C$), and Column can choose an analogous probability not unduly constrained by Row's second-strike Star Wars defense. As before, the payoffs to the players at DE are (r_3, c_3) .

II. Preemption Equilibria (PE): (1) PE_C : $s = 1, q = 1; t = v_R,$
 $w_R \leq p \leq [v_R r_3 + (1 - v_R)r_2]/v_R$; (2) PE_R : $t = 1, p = 1; s = v_C,$
 $w_C \leq q \leq [v_C c_3 + (1 - v_C)c_2]/v_C$.

As in the Deterrence Game, the Preemption Equilibrium for Column (PE_C) is maximal preemption by Column ($t = v_R$) and no preemption ($s = 1$) or retaliation ($q = 1$) by Row. Unlike the Deterrence Game, however, the p in the Star Wars Game that supports PE_C is not arbitrary but instead is constrained by the bounds given above. These say that Column's threat of retaliation, $1 - p$, if attacked (actually he never is attacked at PE_C) will be (i) not more than that allowed by Row's second-strike Star Wars defense (which may be less than, equal to, or greater than v_R , Row's first-strike Star Wars defense), and (ii) not less than that allowed by the upper bound given above, which is a function of $r_2, r_3,$ and v_R .⁷ At PE_C , the payoffs to the players are $(v_R r_3 + [1 - v_R]r_2, 1 - v_R[1 - c_3])$; at PE_R the payoffs are $(v_C c_3 + [1 - v_C]c_2, 1 - v_C[1 - r_3])$.

To interpret the PEs, note that the upper bound on p for PE_C is always greater than r_3 , which means that Column's threat of retaliation does not have to be as great as that required to sustain DE, except when $v_R = 1$. At this extreme, Row has a perfect first-strike Star Wars defense, so preemption by Column is harmless; in fact, it is as if Column never preempted. Hence, Column's threat of retaliation must be the same as for DE to prevent Row from doing better by preempting than by capitulating. On the other hand, if Row has a less-than-perfect defense ($v_R < 1$), the bound on Column's threat of retaliation can be relaxed from $p \leq r_3$ to $p \leq r_3 + [(1 - v_R)/v_R]r_2$ and still ensure PE_C .

Thus, Column can afford to be less threatening to deter Row's preemption, and stabilize his own, the weaker Row's first-strike Star Wars defenses are. This means that stronger threats are needed against stronger defenses to make unilateral preemption a Nash equilibrium.

What all this amounts to is that unilateral preemption is rational if two conditions are met:

1. One player's preemption can be deterred by the threat of retaliation of the other player.
2. It is optimal for the other player (the preemptor) to choose full preemption (to the extent permitted by the nonpreemptor's first-strike Star Wars defense).

Furthermore, the better the nonpreemptor's first-strike defense, the more retaliation the preemptor must threaten in order to deter the nonpreemptor's own first strike. In other words, harsher threats of retaliation are needed against more formidable defenses to ensure that preemption stays one-sided--

that is, the preemptor himself is not preempted simultaneously.

Such threat escalation may not always be possible. For example, there is an inconsistency between assuming Row has a perfect defense against first and second strikes and that Column is able to choose a p according to the conditions for PE_C . Clearly, if $v_R = w_R = 1$, the condition $1 \leq p \leq r_3$ for PE_C cannot be satisfied, meaning that nothing Column can do will sustain PE_C . Instead, an impregnable Row can do better by striking back unless, of course, Column is also impregnable, in which case threats of retaliation are not needed to sustain (r_3, c_3) .

In fact, the no-preemption outcome can be stabilized either by retaliatory threats of both players (DE), by both players' perfect first-strike Star Wars defenses, or by one player's perfect first-strike Star Wars defense coupled with a retaliatory threat by the other player that deters the first from attacking. In the last case, observe that when either $v_R = 1$ or $v_C = 1$, and either a p or a q is selected that satisfies the conditions for PE_C or PE_R , the payoffs to the players given in the description of Preemption Equilibria above reduce to (r_3, c_3) .

The next kind of equilibrium is not duplicated in the Deterrence Game:

III. Mutual Preemption Equilibrium (MPE): $s = v_C, q = 1; t = v_R, p = 1$. Each player preempts to the maximum level permitted by the other's Star Wars defense, and neither retaliates. This equilibrium requires that $v_C \geq c^*$ and $v_R \geq r^*$, where

$$c^* = c_2 / (1 - c_3 + c_2); \quad r^* = r_2 / (1 - r_3 + r_2).$$

The payoffs to the players at this equilibrium are $(v_R v_C r_3 + v_R [1 - v_C] + [1 - v_R] v_C r_2, v_R v_C c_3 + [1 - v_R] v_C + v_R [1 - v_C] c_2)$.

Recall that in the definition of PE_C , $v_R \leq r^*$, the reversal of the above inequality for MPE, was the condition that made the upper bound on p innocuous--Column could preempt with impunity without threatening retaliation ($p = 1$) if attacked himself. The MPE condition, however, turns this situation around: both players' first-strike Star Wars defenses are now so strong as to blunt preemption significantly; in this case, neither player can sustain a PE without also threatening retaliation.⁸

Given the MPE condition, each player has a first-strike defense at or above the thresholds of c^* and r^* . With these stronger defenses, first strikes are less damaging to the struck parties than threatening--and possibly carrying out--second strikes [with $p < 1$ or $q < 1$ sufficient to deter the first strikes and ensure (r_3, c_3)] to deter preemption by an opponent.

Each player may therefore prefer MPE to DE if the two equilibria coexist, provided his own defenses are sufficiently strong. In the Appendix we indicate how these preferences may create a zone near $(v_R, v_C) = (1, 1)$ where MPE is Pareto-superior to DE (if they coexist), and always creates a larger zone around $(v_R, v_C) = (0, 0)$ where DE is Pareto-superior to MPE. Notice that MPE and DE coincide at $v_R = v_C = 1$ in the sense that the players always receive (r_3, c_3) . In other words, with a perfect first-strike Star Wars defense, nothing changes after mutual preemption (except each player's stock of arms, which is not reflected in his payoff), so the payoffs remain (r_3, c_3) .⁹

There are two other kinds of Nash equilibria, derived in the Appendix, that we shall not discuss in detail here because they are always Pareto-inferior to the other Nash equilibria in the Star Wars Game. The first kind are Line Equilibria (LE), and they are similar in nature to the Preemption Equilibria. For Column, LE_C requires that Row's first-strike Star Wars defense parameter v_R exactly equal r^* (defined earlier), and that Column preempt maximally by choosing $t = v_R = r^*$; Row may choose any level of preemption s that satisfies $c^* \leq s$ and $v_C < s < 1$. Neither player ever retaliates (i.e., $p = q = 1$).

The main difference between MPE and LE_C is that Row, instead of preempting fully, chooses a level of preemption less than the maximum permitted by Column's first-strike defense. Although Row's level of preemption does not affect his own payoff, Column would obviously prefer less preemption (a higher s). When $s = 1$, LE_C becomes PE_C at $v_R = r^*$; this PE_C outcome gives Column a higher payoff, and Row not a lower payoff, making it Pareto-superior to any LE_C outcome on the line $v_R = r^*$ (with $s < 1$) in a $v_R - v_C$ coordinate system (to be illustrated later).

The other kind of Pareto-inferior Nash equilibrium in the Star Wars Game is analogous to the Naive Equilibrium (NE) in the Deterrence Game, in which both players partially preempt and neither retaliates. In the Star Wars Game, this equilibrium is characterized by $s = c^*$, $q = 1$; $t = r^*$, $p = 1$.

As in the Deterrence Game, NE in the Star Wars Game is a mixed-strategy equilibrium for both players that renders them indifferent to each other's strategy choices. It is always Pareto-inferior to DE and

hence unlikely to be chosen since DE, if it exists, can not only make both players better off but also is a dominant-strategy Nash equilibrium when a sufficient probability of retaliation is guaranteed. Moreover, it completely overlaps NE in the two most realistic scenarios that we shall discuss in section 5.

The different kinds of equilibria we have described in this section may arise in all Star Wars Games, provided that the players can make strategic choices within the constraints imposed by the Star Wars defenses given for each equilibrium. We shall not consider further Line or Naive Equilibria because they are always Pareto-inferior to the other equilibria (DE, PE, or MPE) and would therefore be eschewed by rational players.¹⁰

5. Pareto-Superior Equilibria in Different Situations

To render more concrete the abstract definitions of the Pareto-superior equilibria, we shall make different assumptions about (i) possible relationships between first-strike and second-strike Star Wars defenses and (ii) maximum values that these defenses may assume. Altogether, we shall posit three scenarios, beginning with the simple case in which

1. First-strike and second-strike defenses are exactly the same:

$$\underline{w_C = v_C} \ ; \ \underline{w_R = v_R}.$$

In this scenario, each player's Star Wars defense is independent of whether he (say, Column) is struck first and can limit Row's first strike to $s \geq v_C$, or Column strikes first himself and then must defend against

Row's retaliation, which he can limit to $q \geq w_C = v_C$. This equality assumption is sensible if it is correct to think of Star Wars as simply putting a cap on the number of missiles that can penetrate each player's defenses, either in a first or second strike. Although a player's first strike before he has incurred any losses may be more devastating than his second strike after he has, the cap in either case is the same in this scenario. (In the next scenario, each player's second-strike defense is assumed to be at least as good, and generally better, than his first-strike defense.)

In Figure 3, we have depicted the regions in which DE, MPE, and

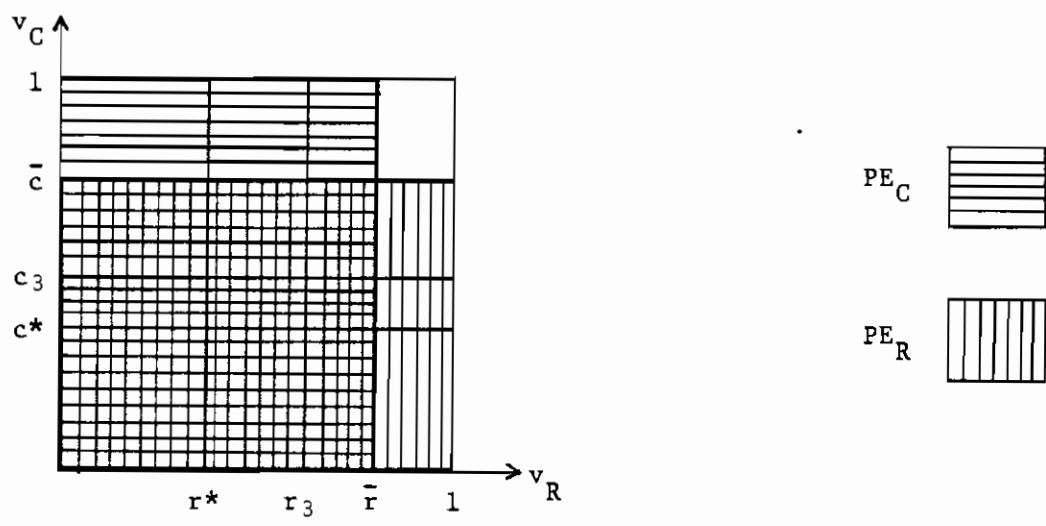
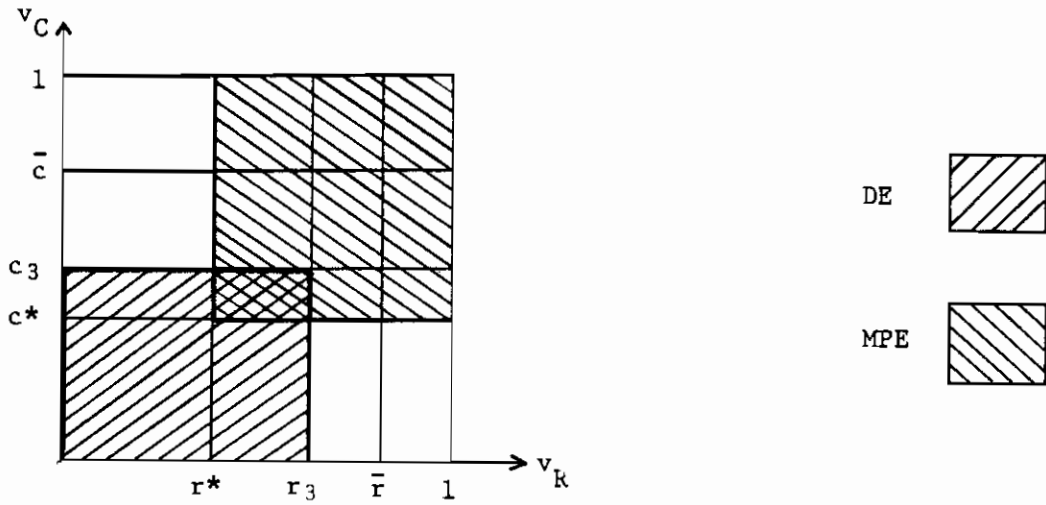
Figure 3 about here

PE_C and PE_R exist in terms of v_C and v_R . The first graph shows overlapping regions containing the former two equilibria, and the second graph overlapping regions containing the latter two equilibria, in the (v_R, v_C) plane. (With color overlaps, all these could be shown in one graph, but in their absence points are given on the two axes to describe precisely where these equilibria do and do not coexist.)

In the first graph, DE exists for lower levels of Star Wars defenses (up to $v_C = c_3$ and $v_R = r_3$), and MPE for higher levels (at or above $v_C = c^*$ and $v_R = r^*$). In the DE - MPE overlap region between c^* and c_3 on the vertical axis and r^* and r_3 on the horizontal axis, the DE payoff to each player is greater than the MPE payoff. This is because, as indicated in the Appendix, Column prefers DE to MPE whenever they coexist and $v_C \leq c_3$, and similarly for Row. Hence, DE dominates MPE whenever they overlap in this scenario.

FIGURE 3

REGIONS OF EXISTENCE OF NASH EQUILIBRIA IN SCENARIO (1): $w_R = v_R, w_C = v_C$



- Key:
- v_C = level of Column's first-strike defense
 - v_R = level of Row's first-strike defense
 - DE = Deterrence Equilibrium
 - MPE = Mutual Preemption Equilibrium
 - PE_C = Preemption Equilibrium for Column
 - PE_R = Preemption Equilibrium for Row

Note: DE dominates MPE in their region of overlap.

So far, so good if the Star Wars defenses of the two players are not above the threshold values of c_3 and r_3 : excluding the PEs for now, either DE is the unique (Pareto-superior) Nash equilibrium, or it dominates the (Pareto-inferior) MPE equilibrium. However, above these thresholds MPE excludes DE, because each player's best response to the other player's preemption--if it is not accompanied by the threat of retaliation, which we consider later--is to attack himself. Mutual preemption would appear not to be a problem if each's Star Wars defense is perfect, because nothing penetrates it, but what if it is not?

To illustrate the effects of a good but not perfect defense, let $v_C = v_R = 1 - \epsilon$, where ϵ is a small positive number such that $v_C > c_3$ and $v_R > r_3$. Since their defenses are equal, the players are on the diagonal in the upper right part of the MPE region. After rearranging terms, Column's MPE payoff can be shown to be

$$E_C = c_3 + \epsilon[(1 - \epsilon) - c_3(2 - \epsilon) + c_2(1 - \epsilon)],$$

which may be greater or less than c_3 , according as the quantity in brackets is positive or negative.

In general, as shown in the Appendix, both players can improve on their (r_3, c_3) payoffs at DE if their Star Wars defenses are (i) sufficiently good to put them in the MPE region but (ii) not perfect. This phenomenon also depends on the values of the players' utilities, as specified in the Appendix and illustrated above. Nevertheless, as their defenses approach perfection, the players' payoffs invariably approach (r_3, c_3) , which can be seen in the previous example by letting $\epsilon \rightarrow 0$, in the above expression for E_C . Then $v_C = v_R \rightarrow 1$, because $v_R = v_C = 1 - \epsilon$ and $E_C \rightarrow c_3$.

Analogously, $E_R \rightarrow r_3$, so the players can actually do worse in the MPE region when they can stop all, rather than most, incoming missiles launched in a first (or second) strike.

It may seem astounding that mutual preemption, when Star Wars defenses are good but not perfect, is not only a Nash equilibrium but can also lead to higher payoffs than DE, or than MPE when the Star Wars defenses of both players are perfect. In effect, with good defenses the players can "afford" to be less cautious in the Star Wars Game, but--more to the point--their deterrent capability is undermined. Thereby, mutual preemption may become their jointly preferred equilibrium.

But observe that the exclusive MPE region requires that the defenses of the players be very strong--above \bar{c} on the horizontal axis and above \bar{r} on the vertical axis, as defined in the Appendix and illustrated in Figure 3. It is by no means clear that Star Wars technologies will ever permit such near-perfect defenses, so apprehensions about being in the exclusive MPE region are probably ill-founded, at least in the near future. Nevertheless, both the stability and the Pareto-superiority of mutual preemption are disconcerting if Star Wars should ever approach perfection.

Figure 3 also shows that unilateral preemption by Column or Row give equilibria that overlap both DE and MPE. The payoff to the preemptor increases the weaker his opponent's first-strike defense, becoming 1 when his opponent has no Star Wars defense. Thus, each player has an incentive to be the sole preemptor--but only when the opponent's defense parameter, v , is less than \bar{c} or \bar{r} --after which MPE becomes the unique Nash equilibrium.

Although PE_C and PE_R cover much of the Figure 3 square, we think they do not pose a serious threat now for two reasons. First, in the foreseeable future when Star Wars is likely to be quite imperfect (well below c_3 and r_3 in Figure 3), retaliatory threats will almost surely be a sufficient deterrent. Second, although the PEs presume no retaliatory response on the part of the preempted player, in fact the command and control procedures of the superpowers make a response almost automatic, suggesting the acronym PDM, or "probabilistic doomsday machine" (Brams and Kilgour, 1986a) as a concomitant to MAD, or "mutual assured destruction."

Superpower intentions reinforce their deterrent capabilities and virtually dictate the choice of DE over PE_C and PE_R in their regions of overlap. More worrisome is preemption--either unilateral or mutual--past the c_3 and r_3 points in Figure 1. In fact, past these points and until \bar{c} and \bar{r} , when both sides' defenses are moderately strong, each player prefers his PE to MPE, as shown in the Appendix: each side has not only the wherewithal to deter the other side, but does better when it preempts, and threatens retaliation, than when both sides preempt but neither threatens retaliation.

Beyond \bar{c} and \bar{r} , when both sides' defenses are very strong, deterrence breaks down completely. Neither side can deter the other by threats, and MPE is the unique Nash equilibrium and may even be Pareto-superior to (r_3, c_3) , except when each side's Star Wars defenses are perfect so that the players' payoffs are exactly (r_3, c_3) .

We believe it highly unlikely that the two superpowers will ever find themselves either in the exclusive MPE region, or the region where the PEs

and MPE overlap and both sides prefer a different PE (of course, Column always prefers PE_C to PE_R , and Row prefers PE_C to PE_R in this region). Nevertheless, unilateral preemption is stable, and better for the preemptor, even in the DE region, which might conceivably threaten deterrence, especially in a crisis, if one side thinks the other is bluffing with its threats of retaliation. PDMS that support a retaliatory policy based on MAD or a similar doctrine, while frightening, probably remain the best means of preserving DE--at least in the early and middle stages of the development of a Star Wars system.

2. Second-strike defenses are perfect, first-strike defenses variable:

$$\underline{v_C \leq 1, v_R \leq 1; w_C = w_R = 1.}$$

When each side's defense against a second strike is perfect, as assumed in this scenario, DE is eliminated altogether. The reason, of course, is that neither side can credibly threaten the other if it has no capability to retaliate against a first strike.

Thus, only preemption equilibria show up in Figure 4, with PE_C occurring

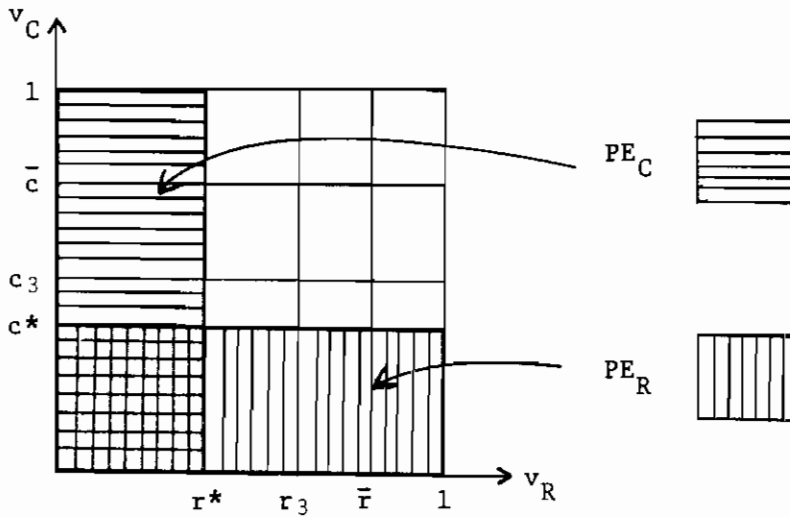
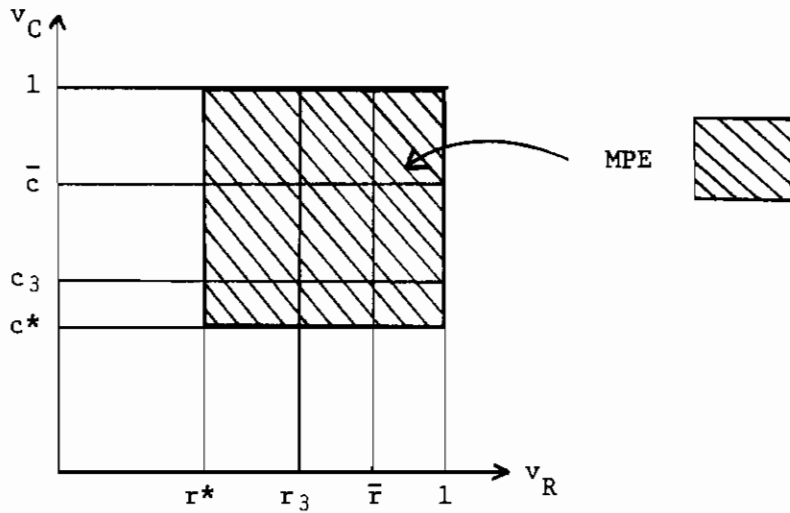
Figure 4 about here

when Row's Star Wars defense is relatively weak, PE_R when Column's defense is relatively weak. MPE covers the rest of the area in Figure 2, where both players' defenses are stronger.

There is no overlap between these equilibria, as in the previous scenario, because threats of retaliation are meaningless against a perfect second-strike defense. Without the ability to threaten, no PE can coexist with MPE: if one side preempts, it is better for the other side to capitulate or to respond in kind.¹¹

FIGURE 4

REGIONS OF EXISTENCE OF NASH EQUILIBRIA IN SCENARIO (2): $w_R = w_C = 1$



Key: v_C = level of Column's first-strike defense
 v_R = level of Row's first-strike defense
MPE = Mutual Preemption Equilibrium
 PE_C = Preemption Equilibrium for Column
 PE_R = Preemption Equilibrium for Row

More precisely, if one player preempts at levels of defense below c^* for Column and r^* for Row, the other player is deterred because he would do worse by moving the outcome toward the disastrous (0, 0) outcome in the Star Wars Game (to the degree permitted by his opponent's defense). By comparison, above these thresholds deterrence fails, because the preempted player can do better by preempting also, which leads to MPE. It is shown in the Appendix that the PEs and MPE in this scenario are all Pareto-superior.

Plainly, deterrence is undermined by a perfect second-strike defense, which would seem easier to approach than a perfect first-strike defense, the ultimate raison d'être of Star Wars. The principal reason is that after a first strike, the ability of the attacked party to retaliate will be impaired, making it more difficult for this party to penetrate the attacker's second-strike defense rather than its first-strike defense.

Although an absolutely perfect Star Wars second-strike defense is probably an impossibility, it is useful to posit, nevertheless, in order to dramatize the extreme dependence of DE on the lack of a second-strike Star Wars defense. Certainly it is reasonable to assume that each side's first-strike defense will never be more effective than its second-strike defense, as in the present scenario. Furthermore, the factor of surprise lends additional plausibility to this assumption, for a retaliatory second strike can be better anticipated than a first strike, magnifying the effectiveness of a player's second-strike defense.

3. Second-strike defenses are never greater than first-strike defenses:

$$w_C = v_C(1 - v_R); w_R = v_R(1 - v_C).$$

In one sense, this scenario is the opposite of scenario (2) above,

wherein second-strike defenses were perfect and so could never be topped by first-strike defenses. Whereas scenario (2) wiped out the possibility of DE, our third scenario gives a substantial boost to DE, with scenario (1), in which first-strike and second strike defenses are equal, an intermediate case.

Before offering a specific interpretation of the functional relationship between first-strike and second-strike defenses postulated in the present scenario, consider the circumstances under which first-strike defenses might actually exceed second-strike defenses. If space-based defenses are essentially invulnerable to a first strike, which they are meant to stymie, then these defenses would presumably be equally effective against a second strike, as assumed in scenario (1).

But now suppose that an attacker's first strike is largely counterforce. Then the attacked party's rational response will be largely countervalue since it would be foolish for it to attack empty missile silos. Because "soft" cities and industrial targets that are hit in a countervalue strike are harder to defend than "hard" missile sites and other protected defense targets hit in a counterforce strike, Star Wars defenses against even a diminished (countervalue) second strike--albeit of a different character than the (counterforce) first strike--may actually be less effective than against the first strike.

It is possible, therefore, to imagine that second-strike defenses may in fact be weaker than first-strike defenses, for strategic (and perhaps ethical) considerations may dictate different kinds of first and second strikes. One must be careful, however, in drawing any such comparison, because Star Wars defenses, as measured by v's and w's, are based on different scales. First-strike defenses are anchored to a scale ranging

from one's next-best to next-worst outcome, whereas second-strike defenses refer to a scale from one's best to worst outcome.

Our specific assumption in this scenario is that each player's second-strike defense is directly related to his first-strike defense (the better Star Wars is against first strikes, the better it is against second strikes) and inversely related to his opponent's first-strike defense (the more an opponent is able to stop a first strike, the harder he can hit back and penetrate one's second-strike defense). These assumptions are embodied in the simple functional relationship postulated above, which, we think, accords well with intuition, especially at the extremes when defenses are all or nothing.

For example, $w_C = 0$ (Column has no second-strike defense) if either $v_C = 0$ (Column has no first-strike defense) or $v_R = 1$ (Row has a perfect first-strike defense and hence can launch an unimpeded second strike). On the other hand, $w_C = 1$ (Column has a perfect second-strike defense) if both $v_C = 1$ (he has a perfect first-strike defense) and $v_R = 0$ (Row has no first-strike defense and hence will absorb the full force of Column's first strike).

Less defensible, perhaps, are the inverse relationships, which negatively link second-strike defenses of one player to the first-strike defenses of the other. Admittedly, one's second-strike defenses are not necessarily helped as an opponent's first-strike defenses deteriorate, and vice versa. Yet the connection, even if indirect, is probably there: if an opponent cannot defend himself well when hit first, his retaliation will be degraded, thereby enhancing the attacker's second-strike defense.

This third scenario is the most complex and, not surprisingly, the relationship among Nash equilibria the most subtle. As shown in Figure 5,

Figure 5 about here

DE, MPE, and both PE_C and PE_R overlap with each other, and occasionally all four overlap simultaneously, as when the Star Wars defenses of both players are at or above point (r^*, c^*) and also more or less equal.

A symmetry in Star Wars defenses is good in this scenario, because DE not only always exists when the first-strike defenses of the players are reasonably close (i.e., near the $v_R = v_C$ diagonal in Figure 5), but MPE never dominates DE in the region of DE - MPE overlap provided a certain condition, derived in the Appendix, is met:

$$(1 - c_3)(1 - r_3) \leq (c_3 - c_2)(r_3 - r_2). \quad (3)$$

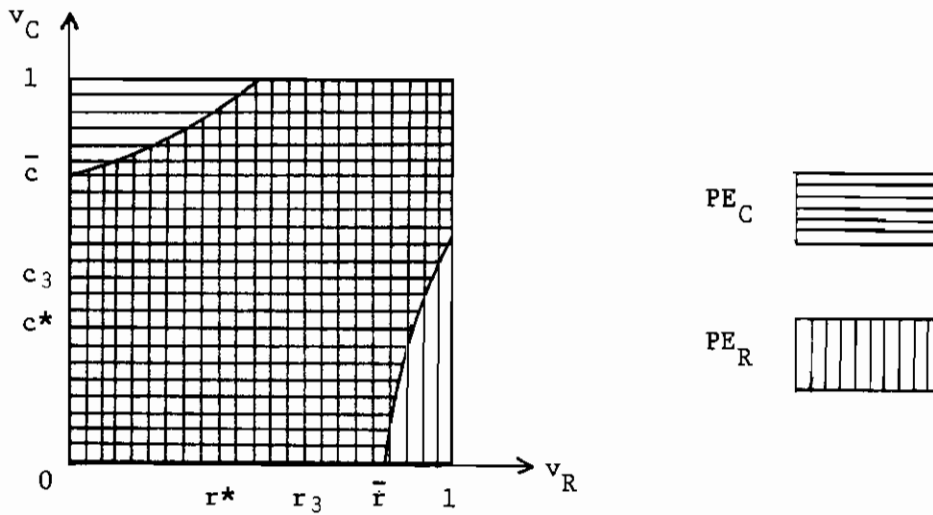
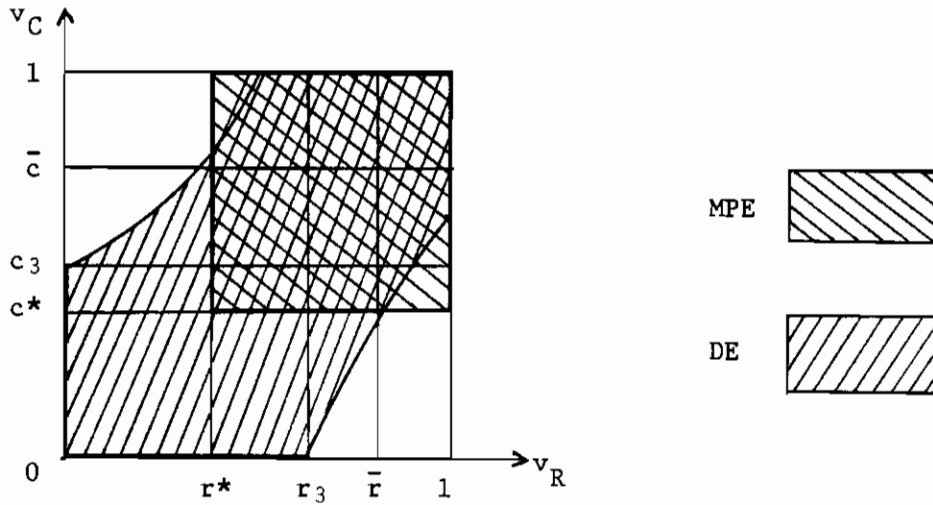
This condition may be interpreted to say that the advantages of unilateral preemption for the players, as embodied in the differences on the left side of (3), are less than the damages the players would suffer if preempted (on the right). A further consequence of (3) is that the region in which DE dominates MPE, located near the $v_R = v_C$ diagonal, extends all the way to $v_R = v_C = 1$.

We think this relationship between payoffs characterizes almost any conceivable nuclear conflict between the superpowers. Hence, it is virtually certain that DE will dominate MPE in this important region should the Star Wars defenses of the superpowers ever approach perfection.

As in scenario (1), if the defenses of the players are weaker (Row below r^* and Column not too strong, or Column below c^* and Row not too

FIGURE 5

REGIONS OF EXISTENCE OF NASH EQUILIBRIA IN SCENARIO (3): $w_R = v_R(1 - v_C)$, $w_C = v_C(1 - v_R)$



Key: v_C = level of Column's first-strike defense
 v_R = level of Row's first-strike defense
MPE = Mutual Preemption Equilibrium
 PE_C = Preemption Equilibrium for Column
 PE_R = Preemption Equilibrium for Row
DE = Deterrence Equilibrium

strong), MPE does not exist when DE does. This salutary finding is counter-balanced, however, by the more ominous result that the upper left and lower right parts of the rectangle (within the unit square), defined by corners at (r^*, c^*) and $(1, 1)$, contain MPE but not DE.

In these regions, both players have moderately good Star Wars defenses, but one player's is significantly better than the other's. In such a situation, each player's best response to maximum possible preemption by his opponent is maximum preemption himself, leading to MPE, and neither can deter the other from preemption.

Slipping momentarily into one of these MPE regions may not be disastrous if the inferior player can quickly attain a better defense. Then the players' more symmetrical defenses would move them into a region that also admits DE, which would come to dominate MPE, subject to inequality (3) above. DE is able to reassert itself over MPE in this case by the equalization of the retaliatory capabilities of the players, with each being strong enough to deter the other when their v 's (and therefore w 's in this scenario) are more or less equal.

If DE is generally preferred to MPE where they overlap, what about PE_C and PE_R , which together completely cover the unit square and therefore always overlap with DE? Indeed, the two PEs overlap except when one player's Star Wars defense substantially exceeds the other's; then, as in the earlier scenarios, the stronger player can deter the weaker and especially benefit from unilateral preemption. Observe that as the inferior player catches up, however, the players move from exclusive PE regions to regions that admit both MPE and DE, but no PE is ever dominated by MPE or DE when they overlap.

Clearly, PE_C and PE_R are a terrible problem when one or both are the only Nash equilibria in a region. But it takes a comparatively large asymmetry in defense capabilities to trigger unilateral preemption, which seems most likely when one of the PEs is the unique equilibrium.

We suspect such an asymmetry is unlikely to persist as soon as the inferior side realizes how precarious its position is--or might become--and takes steps to catch up. In this process of catching up, first both PE_C and PE_R appear (replacing just one), followed either by DE (at lower levels of defense), MPE (at higher levels), and both DE and MPE (at the highest levels). In the short term, DE will almost surely dominate MPE and, given PDMs, deflect players from the choice of unilateral preemption, but DE neither dominates nor is dominated by the PEs generally.

All scenarios we have described probably capture some part of the truth; they need to be compared and the plausibility of their conclusions assessed. Accordingly, we shall offer in the final section a composite picture of Star Wars and its potential strategic effects.

6. Conclusions

The Deterrence Game is founded on Chicken but allows players to choose any level of preemption at the start and threaten any level of retaliation if preempted. The Star Wars Game puts constraints on both the preemption (first strike) and retaliation (second strike) levels that the players can choose in the Deterrence Game, wherein preemption by one player (PE_C and PE_R) and deterrence by both (DE) are the only Pareto-superior Nash equilibria.

The constraints generally preserve the PEs and DE, but the regions in which they exist are different in each of the scenarios we postulate. An exception occurs in our second scenario, when both players have a perfect second-strike Star Wars defense. With retaliation against a first strike ruled out, deterrence becomes impossible and DE is wiped out.

In this scenario, mutual preemption (MPE) emerges as an equilibrium when the first-strike Star Wars defenses of both players exceed a particular threshold. At this equilibrium, because the players cannot be deterred by retaliatory threats, they do not make them. Instead, they find it better to attack rather than absorb a first strike, given their defenses are above threshold values.

We regard scenario (2) as the most unrealistic of the three, for perfect defenses--either first-strike or second strike--simply do not seem in the cards. But scenario (2) does underscore the fragility of deterrence if Star Wars should ever approach perfection; in fact, the goal of SDI now seems to be to shore up deterrence rather than provide a perfect defense (Mohr, 1986, p. E4).

Yet deterrence could be undercut if an all-out first strike, with increasingly accurate nuclear weapons, could cripple an opponent's retaliatory ability. This possible negative relationship between one player's first-strike defense and his opponent's second-strike defense is explicitly modeled in our third scenario. Specifically, we assume that the stronger a player's first-strike defense (permitting him to hit back harder if attacked first), the weaker will be his opponent's second-strike defense (because he will have to suffer greater retaliation if he attacks first). This assumption, coupled

with the assumption that each player's first-strike and second-strike defenses are positively related, leads to the substantial overlap among all equilibria in scenario (3).

Such overlap was anticipated in part from our first scenario, wherein we assumed strict equality between the players' first and second-strike defense parameters. In this scenario, however, DE is Pareto-superior to MPE when the two equilibria overlap, but MPE ultimately displaces even the PEs as the Star Wars defenses approach leakproofness.

In scenario (1), DE vanishes precisely when both players' Star Wars defenses exceed the c_3 and r_3 thresholds necessary to deter, yielding to PEs and MPEs at higher levels of defense. Fortunately, though, these levels will probably be difficult to achieve, even in the distant future, so seem not a present danger if this scenario is an appropriate model.

The overlap of equilibria in our third scenario preserved DE all the way up to both players' having perfect first-strike Star War defenses--as long as these defenses remain more or less equal (how near to equality depends on the players' payoffs). The most dangerous zones in this scenario occur when one player gains a substantial defensive lead over the other, and when, at higher levels of defense, MPE may actually be Pareto-superior to DE. The latter possibility, however, seems remote, because near-perfect defenses are unlikely and, even if attained, a necessary condition for MPE to dominate DE at this level is that a first strike produce unrealistically high rewards for the players.

All in all, our scenarios suggest that deterrence will remain viable as long as Star Wars defenses are primitive enough that each side retains

a substantial second-strike capability. As Star Wars becomes better and erodes these defenses, however, the primary danger lurks in one side's developing considerably stronger defensive capabilities than the other and thereby finding it rational to attack, secure that either it can deter retaliation or survive a ragged response and even be comparatively better off.

Preemption would be most advantageous for the superior player if he has a substantial defensive lead over the inferior player. Then, with DE eliminated, either PE (for the superior player) exists by itself, or both PE and MPE exist.

Perhaps the greatest peril occurs when there is no Deterrence Equilibrium. Then a severe form of crisis instability may grip the players and lead them to an abyss. More probable in superpower relations, though, is that deterrence will remain reasonably secure, mainly because both sides have largely invulnerable second-strike capabilities (principally, submarine-launched ballistic missiles and cruise missiles) that Star Wars will have no effect on, at least presently.

Our major concern is that, short of being leakproof, Star Wars is probably more destabilizing than stabilizing. For one thing, it inevitably introduces MPE into the Star Wars Game, which did not crop up in the Deterrence Game. For another, it shows, especially in the third scenario, that the unbalanced development of Star Wars capabilities by both sides is preemption-inducing, and becomes more so as the Star Wars defense of the superior player approaches perfection. And, of course, Star Wars will be a horrendously costly venture, especially for such dubious returns.

If there is to be a full-fledged Star Wars Game, a time path for its development that keeps the two sides' defenses balanced is surely the best way to preserve deterrence. Balance can probably most readily be achieved through the sharing of defensive technologies by the superpowers, as President Reagan recommended in his 1983 speech on the subject but others have ridiculed as naive in the extreme (Ignatius, 1985, p. 15). Also, more frequent and open communication, and better coordination of activities (e.g., military exercises), are obviously desirable to squelch fears of "breakout."

At some point, however, perhaps in a severe crisis, crisis stability could be upset and preemption, perhaps even mutual preemption, might appear attractive (Brams and Kilgour, 1986d). This has occurred at lower levels of conflict between the superpowers in different parts of the world. If we are to steer clear of nuclear preemption as a rational option, it is imperative that the superpowers recognize that they must carefully chart a course of balanced development--and, preferably, reach a verifiable agreement on no deployment--if Star Wars is not to jeopardize deterrence.

APPENDIX

The basic Star Wars Game is described in section 3, based on the Deterrence Game described in section 2. For future reference, we repeat the expected payoffs to Row and Column given in section 2:

$$E_R(s,q; t,p) = str_3 + (1 - s)tp + s(1 - t)qr_2; \quad (1)$$

$$E_C(t,p; s,q) = stc_3 + s(1 - t)q + (1 - s)tpc_2. \quad (2)$$

For the Star Wars Game, the strategic variables s , t , q , and p satisfy:

$$\begin{aligned} v_C \leq s \leq 1, \quad w_C \leq q \leq 1 & \quad \text{for Row} \\ v_R \leq t \leq 1, \quad w_R \leq p \leq 1 & \quad \text{for Column.} \end{aligned}$$

Throughout this Appendix, we will assume that the defense parameters satisfy

$$0 < v_R, v_C < 1; \quad 0 < w_R, w_C \leq 1,$$

but our results do in fact apply when v_R and/or v_C is zero or one, or when w_R and/or w_C is zero. Our assumption is for convenience only, to avoid many tedious special cases.

The following Lemma is easily verified. It will prove useful in deriving Nash equilibria:

Lemma. Fix v and w such that $0 < v, w < 1$ and consider

$$f(x, y) = H + xK + xyL,$$

where H , K , and L are constants with $L > 0$. Suppose that x and y must satisfy

$$v \leq x \leq 1, \quad w \leq y \leq 1.$$

Then

- (i) $y = 1$ is necessary for a maximum of f ;
- (ii) $x = 1, y = 1$ maximizes f iff (if and only if) $K + L \geq 0$;

(iii) $x = v, y = 1$ maximizes f iff $K + L \leq 0$;

(iv) for any x' satisfying $v < x' < 1$, $(x, y) = (x', 1)$ maximizes f iff $K + L = 0$.

The following definitions of r^* and \bar{r} are based on Row's payoffs of r_2 and r_3 . Analogous definitions of c^* and \bar{c} will be assumed; of course, their properties are analogous as well. Recall that

$$0 < r_2 < r_3 < 1.$$

Set

$$r^* = \frac{r_2}{1 - r_3 + r_2}.$$

It is easy to verify that

$$r_2 < r^* < r_3.$$

Now let

$$Q(v) = v^2 - v(r_3 - r_2) - r_2.$$

Because Q is quadratic with $Q(0) = -r_2 < 0$ and $Q(1) = 1 - r_3 > 0$, $Q(v)$ must have one zero in $(0, 1)$ -- denote this zero by \bar{r} .

Since $Q(r_3) = r_2(r_3 - 1) < 0$, it is clear that

$$0 < r_2 < r^* < r_3 < \bar{r} < 1.$$

Furthermore, if $0 \leq v \leq 1$, then $Q(v) \leq 0$ iff $v \leq \bar{r}$, and $Q(v) = 0$ iff $v = \bar{r}$.

First we will identify all Nash (1951) equilibria at which $s = 1$. If $s = 1$, (2) shows that

$$E_C(t, p; 1, q) = q + t(c_3 - q), \quad (4)$$

so Column maximizes his expected payoff by choosing $t = 1$ iff $c_3 \geq q$. Since $q \geq w_C$, this is possible iff $w_C \leq c_3$. An analogous calculation for Row, and a straightforward verification, yields the Deterrence Equilibrium (DE):

I. There exists an equilibrium with $s = 1$ and $t = 1$ iff $w_C \leq c_3$ and $w_R \leq r_3$. In this case, all such equilibria are given by

$$s = 1, w_C \leq q \leq c_3; \quad t = 1, w_R \leq p \leq r_3.$$

At any DE, the expected payoffs are $E_R = r_3$ and $E_C = c_3$.

Now observe from (4) and the fact that $t \geq v_R$ that, if $s = 1$ and $q \geq c_3$, Column maximizes his expected payoff by choosing $t = v_R$. By (1),

$$E_R(s, q; v_R, p) = v_R p + s[v_R r_3 - v_R p] + sq[(1 - v_R)r_2].$$

The Lemma now shows (i) that $q = 1$ is necessary to maximize E_R , and (ii) that $s = 1, q = 1$ maximizes Row's expected payoff iff

$$v_R r_3 - v_R p + (1 - v_R)r_2 \geq 0,$$

which is equivalent to

$$p \leq \frac{v_R r_3 + (1 - v_R)r_2}{v_R}.$$

Since $w_R \leq p$, this means that

$$w_R \leq \frac{v_R r_3 + (1 - v_R)r_2}{v_R}. \quad (5)$$

Inequality (5) is not an effective restriction on w_R if the right side is at least 1; it is easy to verify that this occurs iff $v_R \leq r^*$. This leads to Column's Preemption Equilibrium (PE_C):

II. There exists an equilibrium with $s = 1$ and $t = v_R$ iff either $v_R \leq r^*$, or $v_R > r^*$ and $w_R \leq \frac{v_R r_3 + (1 - v_R)r_2}{v_R}$. In this case, all such equilibria are given by

$$s = 1, q = 1; \quad t = v_R, w_R \leq p \leq \frac{v_R r_3 + (1 - v_R)r_2}{v_R}.$$

At any PE_C , the expected payoffs are $E_R = v_R r_3 + (1 - v_R)r_2$ and $E_C = 1 - v_R(1 - c_3)$.

It is instructive to consider the implications of (II) in the case of the functional relationship in scenario (3), $w_R = v_R(1 - v_C)$. Substitution in (5) and simplification leads to the condition

$$v_C \geq \frac{Q(v_R)}{v_R^2},$$

where $Q(v_R)$ is as defined above. It follows that, if $w_R = v_R(1 - v_C)$, there is PE_C if either $v_R \leq \bar{r}$ or, if $v_R > \bar{r}$, then $v_C \geq \frac{Q(v_R)}{v_R^2}$.

We now show that (I) and (II) define the only equilibria with $s = 1$. For an equilibrium with $v_R < t < 1$, (2) implies that

$$\frac{\partial E_C}{\partial t}(t, p; 1, q) = c_3 - q = 0$$

is necessary, so $q = c_3$ at an equilibrium with $s = 1$ and $v_R < t < 1$. But (1) yields

$$\frac{\partial E_R}{\partial q}(s, q; t, p) = s(1 - t)r_2 > 0$$

since $t < 1$, so $q = 1$ at equilibrium. This contradiction shows that there are no equilibria with $s = 1$ other than (I) and (II).

By symmetry, we conclude that the only equilibria with $t = 1$ are DE and PE_R ; the latter exists iff either $v_C \leq c^*$ or $v_C > c^*$ and

$$w_C \leq \frac{v_C c_3 + (1 - v_C)c_2}{v_C}. \text{ It consists of all strategy combinations } s = v_C, w_C \leq q \leq \frac{v_C c_3 + (1 - v_C)c_2}{v_C}; t = 1, p = 1.$$

At any PE_R , the expected payoffs are $E_R = 1 - v_C(1 - r_3)$ and $E_C = v_C c_3 + (1 - v_C)c_2$.

We now search for equilibria with $s = v_C$ and $t = v_R$. By (1),

$$E_R(s, q; v_R, p) = v_R p + s[v_R r_3 - v_R p] + sq[(1 - v_R)r_2].$$

The Lemma shows that (i) $q = 1$ is necessary to maximize E_R , (iii) $s = v_C$, $q = 1$ maximizes E_R iff

$$v_R r_3 - v_R p + (1 - v_R)r_2 \leq 0.$$

This is equivalent to

$$p \geq \frac{v_R r_3 + (1 - v_R)r_2}{v_R}.$$

Because $p \leq 1$, a necessary condition for equilibrium is

$v_R r_3 + (1 - v_R) r_2 \leq v_R$, which is easily seen to be equivalent to $v_R \geq r^*$. An analogous calculation for Column's expected payoff, and a simple verification, give the Mutual Prremption Equilibrium (MPE):

III. There is an equilibrium with $s = v_C$ and $t = v_R$ iff $v_C \geq c^*$ and $v_R \geq r^*$.

In this case, the only equilibrium is

$$s = v_C, q = 1; \quad t = v_R, p = 1.$$

At any MPE, the expected payoffs are $E_R = v_R v_C r_3 + v_R (1 - v_C) + (1 - v_R) v_C r_2$ and $E_C = v_R v_C c_3 + (1 - v_R) v_C + v_R (1 - v_C) c_2$.

We now search for equilibria with $s = v_C$ and $v_R < t < 1$. By (2)

$$E_C(t, p; v_C, q) = v_C q + t[v_C c_3 - v_C q] + t_p [(1 - v_C) c_2].$$

The Lemma shows that (i) $p = 1$ at any equilibrium, and (iv) $v_R < t < 1$ at an equilibrium iff

$$v_C c_3 - v_C q + (1 - v_C) c_3 = 0,$$

which is equivalent to

$$q = \frac{v_C c_3 + (1 - v_C) c_2}{v_C}. \quad (6)$$

But,

$$\frac{\partial E_R}{\partial q}(s, q; t, p) = s(1 - t)r_2,$$

which implies that $q = 1$ at any equilibrium with $t < 1$. Combining with (6) gives the necessary condition,

$$v_C = \frac{c_2}{1 - c_3 + c_2} = c^*.$$

Applying the Lemma to E_R , under the conditions $v_C = c^*$, $v_R < t < 1$, and $p = 1$, shows that $t \geq r^*$ is necessary at any equilibrium with $s = v_C$. It is straightforward to verify Row's Line Equilibrium (LE_R):

IV. There is an equilibrium with $s = v_C$ and $v_R < t < 1$ iff $v_C = c^*$. In this case, all such equilibria are given by

$$s = v_C = c^*, q = 1; v_R < t < 1 \text{ and } t \geq r^*, p = 1.$$

At any LE_R , the expected payoffs are $E_R = t - (t - r^*)(1 - r_3 + r_2)c^*$ and $E_C = c^*$. Between two different LE_R equilibria, Row prefers the one with the larger value of t , and Column is indifferent.

As before, we note that there is an equilibrium LE_C analogous to LE_R , with $v_C < s < 1$ and $t = v_R$ iff $v_R = r^*$. All such equilibria are then given by

$$v_C < s < 1 \text{ and } s \geq c^*, q = 1; t = v_R = r^*, p = 1.$$

The expected payoffs at LE_C are analogous to those at LE_R .

Finally, we search for equilibria with $v_C < s < 1$ and $v_R < t < 1$. By differentiating (2), it is easy to see that $p = 1$ at such an equilibrium, giving

$$E_R(s, q; t, 1) = t + s[tr_3 - t] + sq[(1 - t)r_2].$$

The Lemma shows (iv) that Row's best response to $v_R < t < 1$ and $p = 1$ is $v_C < s < 1$ and $q = 1$ iff

$$tr_3 - t + (1 - t)r_2 = 0,$$

which is equivalent to $t = r^*$. It follows from $v_R \leq t$ that $v_R \leq r^*$ is necessary for this equilibrium. A similar calculation for Column's expected payoff, and a simple verification, yields the Naive Equilibrium (NE):

V. There is an equilibrium with $v_C < s < 1$ and $v_R < t < 1$ iff $v_R \leq r^*$ and $v_C \leq c^*$. In this case, the unique such equilibrium is

$$s = c^*, q = 1; t = r^*, p = 1.$$

The expected payoffs are $E_R = r^*$ and $E_C = c^*$.

Based on the definitions of equilibria (I) - (III), the drawing of Figures 3 and 4 for scenarios (1) and (2) in section 5 is straightforward.

In the case of scenario (3), where $w_C = v_C(1 - v_R)$ and $w_R = v_R(1 - v_C)$, the construction of the curves in Figure 5 may require some explanation. As a consequence of (I), the region of existence of DE in Figure 5 is bounded by the curves $v_C = c_3/(1 - v_R)$ (upper left) and $v_R = r_3/(1 - v_C)$ (lower right). As indicated above, the boundary of the PE_R region in Figure 5 is $v_C = Q(v_R)/(v_R)^2$, and the boundary of the PE_C region is $v_R = Q(v_C)/(v_C)^2$.

We now turn to the question of the Pareto-superiority of the different Nash equilibria. It is easy to show that Row always prefers PE_R to DE to PE_C , and Column always prefers PE_C to DE to PE_R , so none of these three equilibria ever dominates any other. It can also be readily verified that DE is better for both players than NE, and that all line equilibria are always dominated by the preemption equilibrium, which is their limiting case. (For example, LE_R is dominated by PE_C .) Thus the Pareto-superior equilibria are always to be found among DE, PE_R , PE_C , and MPE.

The status of MPE is especially interesting. Straightforward calculations give

$$E_R(PE_R) - E_R(MPE) = (1 - v_R)[1 - v_C(1 - r_3 + r_2)] > 0$$

$$E_C(MPE) - E_C(PE_R) = (1 - v_R)(1 - c_3 + c_2)(v_C - c^*) \geq 0$$

since MPE does not exist unless $v_C \geq c^*$. Similar computations for PE_C yield that none of PE_R , PE_C , and MPE is Pareto-inferior to any of the others.

The comparison of MPE and DE is more complex. It is easy to see that

$$E_R(MPE) > E_R(DE) \text{ iff } v_R > f_C(v_C) = \frac{r_3 - v_C r_2}{1 - v_C(1 - r_3 + r_2)}.$$

Now $f_C(v_C)$ is a smoothly increasing function satisfying $f_C(0) = r_3$ and

and $f_C(1) = 1$. It follows that $E_R(\text{DE}) \geq E_R(\text{MPE})$ if $v_R \leq r_3$; Row prefers MPE to DE only if v_R is large enough. Analogous conclusions can be drawn for Column's payoffs.

It follows that MPE dominates DE exactly when $v_R > f_C(v_C)$ and $v_C > f_R(v_R)$ hold simultaneously, and DE dominates MPE when both inequalities fail simultaneously. Consideration of the derivatives $f'_C(1)$ and $f'_R(1)$ shows that MPE dominates DE iff the two equilibria coexist in a region near $(v_R, v_C) = (1,1)$, and $(1 - r_3)(1 - c_3) > (r_3 - r_2)(c_3 - r_2)$. The region in which DE dominates MPE includes $v_R \leq r_3$, $v_C \leq c_3$, as noted above. It intersects every neighborhood of $(v_R, v_C) = (1,1)$ iff

$$(1 - r_3)(1 - c_3) \leq (r_3 - r_2)(c_3 - c_2).$$

NOTES

1. A sophisticated discussion of the issues SDI has raised, and a careful weighing of the arguments of both proponents and opponents, can be found in Office of Technology Assessment (1985). Stability questions about SDI particularly relevant to this paper are discussed in Marsh (1985).

2. There are some recent exceptions, including Chrzanowski (1985a, 1985b) and Bracken (1986). Bracken shows, among other things, how the one-sided deployment of SDI can create an incentive for the side deploying to strike first; his analysis anticipates some of our conclusions, including the problem of avoiding preemption in the transition to full deployment. While Bracken's model is not game-theoretic, O'Neill (1985a, 1985b) defines axiomatically a game-theoretic index of crisis instability and applies it to SDI.

3. This section is based in part on Brams and Kilgour (1985b).

4. There is a third Nash equilibrium in Chicken, but it is not in pure strategies, or specific strategies that players would choose with certainty. Rather, it is in mixed strategies, which are defined as probability distributions over a player's pure strategies. The calculation of equilibria involving mixed strategies requires that payoffs be given in cardinal utilities--not just ordinal ranks--which we shall introduce when we define the Deterrence Game late in this section.

5. See also Shubik (1982, pp. 265-270) for a recent discussion of perfect equilibria and related concepts.

6. We underscore "corresponding" because equilibrium strategies in the Deterrence Game are not interchangeable (Luce and Raiffa, p. 106). Specifically, one player's choice of his Deterrence Equilibrium strategy combined with the other player's choice of his Preemption Equilibrium strategy does not constitute an equilibrium.

7. This upper bound is not an effective restriction if $v_R \leq r^*$, where $r^* = r_2 / (1 - r_3 + r_2)$, because then it is at least 1. In specifying the upper bound, we assume that it is not less than the lower bound (so that there exists a p that satisfies the double inequality for p given in the definition of PE_C), and that $v_R > r^*$. (It can be shown that $r_2 < r^* < r_3$.) Details are given in the Appendix.

8. Note that when $v_R = 0$ and Row has no Star Wars defense, PE_C in the Star Wars Game is PE_C in the Deterrence Game.

9. Presumably, if one has arms that are totally ineffective against an opponent, one would be foolish to use them. Instead, it is reasonable to suppose that a player would save them for use against another opponent.

10. In the Appendix we show that neither PE nor MPE may dominate the other. If they exist, DE and MPE may both be Pareto-superior, or either may dominate the other. In any case, there is always at least one Pareto-superior equilibrium, and all Pareto-superior equilibria are neither DE, PE_R , PE_C , or MPE.

11. In the first scenario, "response in kind" could sometimes be deterred by threats. The possibility of retaliation permits PEs to coexist with MPE, which never involves threats.

REFERENCES

- Bracken, Jerome (1986). "Stability, SDI, Air Defense and Deep Cuts." In Modeling and Analysis of Arms-Control Problems, ed. Rudolf Avenhaus and Reiner K. Huber. Berlin: Springer-Verlag.
- Brams, Steven J. (1985). Superpower Games: Applying Game Theory to Superpower Conflict. New Haven, CT: Yale University Press.
- Brams, Steven J., and Marek P. Hessel (1984). "Threat Power in Sequential Games." International Studies Quarterly 28, no. 1 (March): 15-36.
- Brams, Steven J., and D. Marc Kilgour (1985a). "Optimal Deterrence." Social Philosophy and Policy 3, no. 1 (Autumn): 118-135; reprinted in Nuclear Rights/Nuclear Wrongs, ed. Ellen Frankel Paul et al. London: Basil Blackwell, 1986, pp. 118-135.
- Brams, Steven J., and D. Marc Kilgour (1985b). "The Path to Stable Deterrence." In Dynamic Models of International Conflict, ed. Urs Luterbacher and Michael D. Ward. Boulder, CO: Lynne Rienner, pp. 11-25.
- Brams, Steven J., and D. Marc Kilgour (1986a). "Is Nuclear Deterrence Rational?" PS 19, no. 2 (Summer): 645-651.
- Brams, Steven J., and D. Marc Kilgour (1986b). "Optimal Threats." Mimeographed.
- Brams, Steven J., and D. Marc Kilgour (1986c). "Rational Deescalation." In Evolution, Games, and Learning: Models for Adaptation in Machines and Nature, Physica 16D, ed. Burton Wendroff et al. Amsterdam: North Holland.
- Brams, Steven J., and D. Marc Kilgour (1986d). "Threat Escalation and Crisis Stability: A Game-Theoretic Analysis." Mimeographed.

- Brams, Steven J., and D. Marc Kilgour (1986e). "Winding Down if Preemption or Escalation Occurs: A Game-Theoretic Analysis." Mimeographed.
- Brams, Steven J., and Donald Wittman (1981). "Nonmyopic Equilibria in 2×2 Games." Conflict Management and Peace Science 6, no. 1 (Fall): 36-62.
- Chrzanowski, Paul (1985a). "Crisis Stability during a Transition to a Deterrence Posture Reliant on Defenses." Lawrence Livermore National Laboratory, UCID 20590 (October).
- Chrzanowski, Paul (1985b). "Strategic Defense and Crisis Stability." Lawrence Livermore National Laboratory, UCID 20699 (December).
- Ignatius, David (1985). "In Star Wars Debate, Tactical Issues Nearly Get Lost in Shuffle." Wall Street Journal (October 15): 1, 15.
- Kilgour, D. Marc (1985). "Anticipation and Stability in Two-Person Non-cooperative Games." In Dynamic Models of International Conflict, ed. Urs Luterbacher and Michael D. Ward. Boulder, CO: Lynne Rienner, pp. 26-51.
- Kilgour, D. Marc (1984). "Equilibria for Far-sighted Players." Theory and Decision 16, no. 2 (March): 135-157.
- Luce, R. Duncan, and Howard Raiffa (1957). Games and Decisions: A Critical Survey. New York: Wiley.
- Marsh, Gerald E. (1985). "SDI: The Stability Question." Bulletin of the Atomic Scientists 41, no. 9 (October): 23-24.
- Mohr, Charles (1986). "Star Wars Planners Are Digging Themselves In." New York Times (April 20), p. E4.
- Nash, John (1951). "Non-cooperative Games." Annals of Mathematics 54: 286-295.
- Office of Technology Assessment (1985). "Ballistic Missile Defense Technologies: Summary." Washington, DC: Office of Technology Assessment (September).

- O'Neill, Barry (1985a). "Applications of a Crisis Instability Index: Arms Control Agreements and Spaced-Based Missile Defenses." Discussion Paper No. 651, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
- O'Neill, Barry (1985b). "A Measure for Crisis Instability." Discussion Paper No. 652, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
- Schelling, Thomas C. (1960). The Strategy of Conflict. Cambridge, MA: Harvard University Press.
- Selten, Reinhard (1975). "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." International Journal of Game Theory 4, no. 1: 25-55.
- Shubik, Martin (1982). Game Theory in the Social Sciences: Concepts and Solutions. Cambridge, MA: MIT Press.
- Zagare, Frank C. (1984). "Limited-Move Equilibria in 2×2 Games." Theory and Decision 16, no. 1 (January): 1-19.
- Zagare, Frank, C. (1985). "Toward a Reformulation of the Theory of Mutual Deterrence." International Studies Quarterly 29, no. 2 (June): 155-169.