"PRISONERS' DILEMMA" AND "CHICKEN" MODELS IN INTERNATIONAL POLITICS

By Glenn H. Snyder

State University of New York at Buffalo

#1

July 7, 1969

"PRISONERS' DILEMMA" AND "CHICKEN" MODELS IN INTERNATIONAL POLITICS

Two conflict models, the "prisoners' dilemma" and the "game of chicken", are generally assumed to have considerable social relevance. Some such assumption would seem to be implied, at least, in the large amount of recent laboratory experimentation based on these models, and in the frequency with which they are invoked metaphorically in the verbal analysis of conflict situations. Little has been done, however, by way of clarifying and spelling out the various kinds of social situations which are modeled by these games, or in relating them to other theoretical ideas. Nor have the two models themselves been exhaustively compared to clarify differences in their logic and social implications. This paper will attempt an inquiry into these matters in the context of international politics.

The Prisoners' Dilemma

A typical prisoners' dilemma matrix is the following:[1]

		B: 1	B: 2
A	1	5, 5	-10, 10
	2	10, -10	-5, -5

Fig. 1

The central characteristic of this game is that although the parties could enjoy mutual benefits by cooperating, they are forced into conflict and mutual losses by the logic of their situation. A and B could achieve the best mutual benefit by both playing strategy 1. However, if either thinks the other will play this cooperative strategy, he has two kinds of incentives to "doublecross" the other by playing strategy 2. The first, which might be called the "offensive"

incentive, is that each can increase his gains (from 5 to 10) by playing 2 when the other stays at 1. The second, the "defensive" incentive, springs from the fear that the other may doublecross by playing 2, producing the maximum loss for oneself; consequently there is an incentive to doublecross preemptively--i.e., play strategy 2--in order to limit one's losses (to -5 rather than -10). If both parties act on either of these incentives, they end up in the lower right-hand box, which, although not the worst possible outcome, is definitely worse than could have been achieved by mutual cooperation. In strict logic, assuming both players are rational and act only according to self-interest, there is no way of escaping this outcome. In the language of game theory, strategy 2 "dominates" strategy 1 for both players: whatever one assumes about the other party's choice, the best result is obtained by playing strategy 2. Even though both parties might want to cooperate with strategy 1, they are prevented from doing so by the structure of the game.[2]

The Prisoners' Dilemma and International Politics

The prisoners' dilemma is a paradigm of many social situations in which there are rewards for cooperation and penalties for mutual non-cooperation, but in which the reward for unilateral non-cooperation exceeds both the benefit from mutual cooperation and the cost of mutual conflict. Some examples are business firms in an oligopolistic market, newspaper reporters pledged to hold a story until some future release date and organizations trying to collect dues from their memberships to pay for some collective good.[3] In such situations, there are strong incentives for each party, individually, to defect from agreed or potential collaboration (e.g., reduce prices, publish the story prematurely, or fail to pay dues) in order to gain a greater benefit, but the outcome of all-around defection is an all-around loss which could have been avoided by cooperation.

The most obvious example in international politics is that of arms competition and its obverse, disarmament. Let us assume that in the long run all

states would be best off if none armed at all. But if one state thinks another will not arm, it is tempted to arm itself in order to make gains by coercion or war against the disarmed state. Or, one state, fearing another will reason this way, is motivated to arm to protect itself against the other's possible armament. When both states act on either of these incentives, the result is mutual armament with no greater security than before plus the economic burden of the arms. Conversely, if both start in a condition of mutual armament, and seek to negotiate disarmament or arms reduction, they find it impossible to do so because each fears that if it disarms the other may cheat on the agreement. The states are trapped in the double-defection box of a prisoners' dilemma and cannot escape from it into the cooperative box without a super-ordinate "government" which can enforce cooperation.

Alliance competition is the political counterpart of arms competition and may also be rooted in a prisoners' dilemma. Adversary states, unable to achieve mutual cooperation at 1, 1, are driven to seek protection from each other by acquiring allies, thus incurring costs in the form of new political commitments and risks and reduced flexibility of policy. The prisoners' dilemma aspect of alliances is seen most clearly in the "preclusive" alliance. For example, in the 1870's the most prominent international conflict in Europe was that between Germany and France. Bismarck, fearing an alliance between Austria and France, made an alliance with Austria in 1879, in part, to "dig a ditch", as he put it, between the two countries. Ultimately, this increased the insecurity of both France and Russia and drew them together in an alliance. The end result was no greater security for either Germany or France than they had enjoyed prior to 1879, but with the added burdens of the alliance ties. The "self-defeating" aspect of the prisoners' dilemma is plain here, roughly parallel to the arms race version. From Germany's point of view the "best" outcome (1, 1) would have been no alliances, the "worst" (2, 1 or 1, 2) would have been an alliance between France and Austria and perhaps Russia as well, the "second-best" or

"second-worst" was German alliance with Austria. For France, the obverse was true. The alliances prevented the worst possible outcomes for both but they entailed additional burdens, costs and risks with little or no ultimate increase in security for either.[4]

Several other phenomena in international relations represent some variant of the prisoners' dilemma or contain elements of it. "Collective security", e.g., as envisaged under the League of Nations, is an attempt to organize universal cooperation in defeating or deterring "aggression." In the long run, it might well be to all states' advantage to participate in such an enterprise, regardless of the identity of the aggressor or victim. With the entire international community taking part, each state's costs would be small, most aggression probably would be deterred, and each would benefit from the general guarantee.[5] But with each particular case of aggression, each participant in the scheme is tempted to stand aside and let the other participants do the job. Or a participant may defect because it fears inaction by the others, which would leave it to face the aggressor alone. It turns out to be impossible to organize international community action against international "crimes" because of such self-centered temptations and fears. Collective security thus degenerates into a "balance of power" in which each state is willing to fight only in situations where its own "vital interests" are at stake.

Another example might be the imperialistic competition which took place in the latter part of the 19th century. Free trade and free investment in Asia and Africa, without political control, would have provided maximum economic benefits for all; it has been conclusively shown that in most cases the costs of administering and defending colonies outstripped the economic benefits of political domination. But even if this were perceived at the time, it was overwhelmed by another consideration: the <u>worst</u> possible outcome would have been for one's own state to refrain from acquiring colonies while others did not. Then one would face the prospect of being shut out entirely from trade

with the colonial areas. If the stronger motivation for acquiring colonies were prestige rather than economic gain, the same reasoning applied. When one or a few European countries began acquiring colonies, others had to join the race to avoid slipping lower in the prestige ranking. On both counts, the worst possible outcome was to have no colonies while others had some. But the all around race for prestige and economic gain through colonies wound up in little prestige or gain for anyone, and all the colonial powers found themselves with costly burdens which would not have been incurred had they not started the race in the first place.[6]

World War I has been described as "a war which nobody wanted." The statesmen appear to have become trapped in a spiral of action and reaction which led inexorably from a single assassination to general war. This spiral was fundamentally the result of two interlocking prisoners' dilemmas. The first was the product of the power configuration: a virtual equilibrium between two opposing alliances, each alliance being composed of countries of roughly comparable strength, so that the continued allegiance of each member of each alliance was essential to the equilibrium. Hence, in terms of the prisoners' dilemma, the defeat or alienation of an ally was the "worst possible outcome." Allies had to be supported, even at the cost of war. If Germany and France could have cooperated in restraining Austria and Russia the war probably would have been avoided. But both feared that withholding support from their allies would result in either the alienation or defeat of the ally. Consequently both Germany and France made firm commitments to Austria and Russia (i.e., both "defected") which encouraged the latter countries to take actions which made war inevitable. Reinforcing this logic was another prisoners' dilemma which was inherent in the lead times for mobilization and in the prevalent belief at the time that mobilization meant war. Once one country started to mobilize, its rival feared the outbreak of war before it was ready. Therefore, it was driven to mobilize and attack first to avoid being caught at a disadvantage.

When Russia, acting on rumors of German mobilization, began a general mobilization, Germany responded with mobilization and an ultimatum, followed by a declaration of war. Fearing the "worst possible outcome" of being attacked before they were ready, the two countries chose the "second worst"--war with some semblance of preparedness.

As Thomas Schelling has pointed out, the mobilization race of World War I has its counterpart in the "reciprocal fear of surprise attack" in the nuclear age.[7] Two nuclear powers, each possessing a first strike capability, are under strong pressures to attack pre-emptively, in order to forestall a feared attack by the other. Fortunately, technology has not yet allowed the development of first strike capabilities by the United States and the Soviet Union, so this particular prisoners' dilemma is only hypothetical. However, even though there have been no incentives to pre-emptive attack at the action level, the prisoners' dilemma has nevertheless operated at the preparedness level in the nuclear arms competition. Consider the following quotation from former Secretary of Defense Robert S. McNamara:

> In 1961, when I became Secretary of Defense, the Soviet Union possessed a very small operational arsenal of intercontinental missiles. However, they did possess the technological and industrial capacity to enlarge that arsenal very substantially over the succeeding several years.
>
> Now, we had no evidence that the Soviets did in fact plan to fully use that capability. But as I have pointed out, a strategic planner must be "conservative" in his calculations; that is, he must prepare for the worst plausible case and not be content to hope and prepare merely for the most probable.
>
> Since we could not be certain of Soviet intentions--since we could not be sure that they would not undertake a massive buildup--we had to insure against such an eventuality by undertaking ourselves a major buildup of the Minuteman and Polaris forces. . .
>
> Clearly, the Soviet buildup is in part a reaction to our own buildup since the beginning of this decade. Soviet strategic planners undoubtedly reasoned that if our buildup were to continue at its accelerated pace, we might conceivably reach, in time, a credible first-strike capability against the Soviet Union.

> This was not in fact our intention. Our intention was to assure that they--with their theoretical capacity to reach such a first-strike capability--would not in fact outdistance us.
>
> But they could not read our intentions with any greater accuracy than we could read theirs. And thus the result has been that we have both built up our forces to a point that far exceeds a credible second-strike capability against the forces we each started with. . .
>
> It is futile for each of us to spend $4 billion, $40 billion, or $400 billion--and at the end of all the spending, and at the end of all the deployment, and at the end of all the effort, to be relatively at the same point of balance on the security scale that we are now.[8]

Thus, the Secretary of Defense, with remarkable clarity, and in a tone which can only be described as wistful frustration, expounded the essence of the prisoners' dilemma in the nuclear age. The dilemma functioned in McNamara's time despite technological inhibitions against it--i.e., a considerable gap between a first-strike and a second-strike capability. These inhibitions may disappear if the Soviets and the United States begin competitive development of anti-missile missiles and multiple warhead offensive missiles. For then the difference between numbers of offensive missiles needed for a first-strike and those needed for a retaliatory blow may narrow or disappear; an intended retaliatory capacity might easily double as a first-strike force. Uncertainties and fears about the opponent's intentions would multiply. The situation would be much less stable than presently, i.e., a much more virulent "prisoners' dilemma" could emerge.

Some Related Theories: Security Dilemma, Mirror Image and Deterrence

In the International relations literature, the theory most closely resembling the prisoners' dilemma is that of the "security dilemma." This theory has a venerable history, beginning at least as early as Hobbes and elaborated in the modern international context by John Herz, Herbert Butterfield, Arnold Wolfers and others. The dilemma is said to arise inevitably out of the fundamental structure of the international system--a "state of nature", or a

system of decentralized power and multiple sovereignties. Lacking any powerful central authority which can regulate conflict, states are under continual apprehension of attack by other states, and their relations take on the character of a continuous struggle for security in the shadow of war. The dilemma arises because states can never be sure that the security measures of others are intended only for security and not for aggression. Consequently, each state's efforts to gain security through power accumulation tend to increase the insecurity of other states, stimulating them to enhance their power, which then leads to further apprehension and power accumulation by the first state, and so on. Thus the very existence of states in a condition of anarchy produces a competition for security which is objectively "unnecessary" and ultimately futile.[9]

The British historian, Herbert Butterfield, has given a particularly succinct statement of the predicament:

> It is the peculiar characteristic of the situation I am describing--the situation of what I should call Hobbesian fear--that you yourself may vividly feel the terrible fear that you have of the other party, but you cannot enter into the other man's counter-fear, or even understand why he should be particularly nervous. For you know that you yourself mean him no harm, and that you want nothing from him save guarantees for your own safety; and it is never possible for you to realize or remember properly that since he cannot see the inside of your mind, he can never have the same assurance of your intentions as you have. As this operates on both sides the Chinese puzzle is complete in all its interlockings--and neither party sees the nature of the predicament he is in, for each only imagines that the other party is being hostile and unreasonable.[10]

Thus Butterfield stresses the core of the dilemma: the inability ever to be sure of the other party's intentions. Kenneth Waltz, in his "third image" of the causes of war (the nature of the system, as opposed to the nature of man and the nature of the state) shows how this uncertainty, as in the prisoners' dilemma, frustrates desires for cooperation: If states could cooperate they could realize their highest values. But because of system-generated mistrust they cannot cooperate and are forced to seek their security independently, which

leads them into such "second-best" and conflictual outcomes as balances of power, alliances, arms races, tariff competition, etc.[11]

Arnold Wolfers poses the security dilemma as an alternative to the older notion of a "struggle for power" as the primary cause of conflict in international relations, and points out that it substitutes the motif of "tragedy" for the "evil" that was implied in power struggle theme. Wolfers stresses the self-defeating aspect of the theory, also reminiscent of the prisoners' dilemma:

> The insecurity of an anarchical system of multiple sovereignty places the actors under compulsion to seek maximum power even though this may run counter to their natural desires. By a tragic irony, then, all actors find themselves compelled to do for the sake of security what, in bringing about an all-around struggle for survival, leads to greater insecurity.[12]

The security dilemma has its most obvious manifestation in the arms race. But the theory has much broader implications. There are other sources of security besides arms--notably the control, or the independence or political allegiance, of territory and geographical positions beyond the boundaries of the state. Many of the "vital interests" of the state beyond its own frontiers are essentially generated by the anarchic structure of the system and the security dilemma. I.e., many interests are strategic rather than intrinsic in character; they are held as interests primarily because they represent potentially useful power to wage war in a world in which attack by other states is always possible. These are interests which would evaporate as interests if all states were interested only in security and could be sure that this was also the only motive of other states.[13] As in the arms race, measures to control or protect territory valued as strategic may appear as threatening to the opponent, causing him to take similar measures and perhaps to expand his strategic interests. Hence the security dilemma may take the form of a competition in the expansion and protection of strategic interests.

Strategic values are of course not the only values which determine interests; the United States has intrinsic interests in the protection of

Western Europe and other areas, and the Soviets presumably have an intrinsic interest in preserving and increasing the domain of Communism. Logically, between nuclear powers, strategic interests are decreasing in importance relative to intrinsic ones because of the declining value of sheer territory as a source of power, even though habit, tradition and intellectual lag may be retarding the recognition of this. Still, a good deal of the conflict in world politics stems from overlapping strategic interests which are directly attributable to system structure.

Of course a literal application of the Hobbesian notion of a "war of all against all" would be a gross exaggeration. States obviously do not fear all other states. In contrast to the mythical state of nature among individuals, the international state of nature is characterized by gross power inequalities. States tend to fear, and to identify as potential enemies, only those other states with the power to harm them, and security dilemmas arise only in such "relationships of tension." Such other factors as conflicts of intrinsic interest, ideological antagonisms and affinities, and geographical location may also contribute to the identification of enemies, but pure "power position" is a necessary and sometimes a sufficient condition. Enemy identification is more ambiguous and uncertain in a multipolar than in a bipolar system because power is roughly equally distributed among the main actors, because the target of each state's power accumulation is often uncertain, and because it may be equally as plausible to perceive other states as potential allies as potential enemies. Because of this ambiguity arising from power configuration alone, the "other factors" mentioned above will tend to be of relatively high importance in the identification of friend and foe. In a bipolar system, power distribution alone will tend to be more determinate. As Raymond Aron has put it, the United States and the Soviet Union are "enemies by position." That is, they are enemies not essentially because of ideological differences or because they have intrinsic interests in conflict, or because either has given specific

evidence of aggressive intent, but simply because of their "power position" as the two most powerful states in the world, each being the only other state with a potential capacity to harm the other militarily.[14] Of course, conflicts in ideology and intrinsic interests tend to exacerbate the underlying "conflict by position."

It follows that security dilemmas are likely to be less severe in multipolar systems, when the arms and security measures of particular other states do not necessarily appear as threatening to oneself, at least not until alliance line-ups are sharply drawn. The clear identification of adversaries in a bipolar system makes it much more likely that defensive power accumulation by one will be perceived as a threat by the other.[15]

It may be objected that two states cannot conceive of themselves as enemies until at least one of them takes some action which indicates possible aggressive intent. The security dilemma theorist would reply that given a latent "enmity-by-position", it takes very little in the way of apparently aggressive behavior to start the security dilemma operating and once started, it tends to feed on itself. As Butterfield puts it, "what seem to be little sins may have colossally disproportionate consequences."[16] A few "little sins" committed in the context of latent suspicion (and perhaps only for security reasons) will tend to confirm the suspicion and foster the growth of reciprocal images of hostility which are highly resistant to contradictory evidence. Occasional friendly gestures will be interpreted as "weakness", or as evidence that one's own strength is finally forcing the opponent to change his ways.[17]

The security dilemma theory was originated by so-called "realist" thinkers. A similar theory is held by some members of the "idealist" school, under the label of the "mirror image" hypothesis. Chiefly the property of psychologists, this theory is presented not as the inevitable result of system structure but as the consequence of certain psychological propensities. Briefly, the argument goes somewhat as follows: In a mental process characteristic of paranoia, each

side perceives the possible aggressive intentions of the other as probable or even certain. The adversary's capabilities are mistakenly equated with his intentions. Consequently, each side undertakes defensive measures which are misperceived by the other as aggressively motivated. Each party develops a "bogeyman" conception of the other which attributes a great deal more hostility to the other than is actually the case. Each side's image of the other becomes a "mirror image" of the other's image of itself; each sees the other as aggressive and hostile when neither is so in fact. The hostile image of the other is reinforced by a benevolent image of the self: the Americans (or the Soviets), perceiving themselves as entirely peaceful, are unable to understand why the Soviets (or Americans) are accumulating armaments, allies, bases, etc. unless it is because they are aggressive. And conversely the self-image depends on the enemy-image: the belief that one's own country is peaceful is preserved by the thought that one's own arms, etc. are only defensive reactions to the other's threat. Thus, as Ralph K. White puts it, "the two images, peaceful self and aggressive enemy, are mutually complementary and thoroughly interdependent."[18]

Although the dynamics of the security dilemma and mirror image theories are quite similar, it is important to understand the differences. The security dilemma theorists view the process as inherent in the structure of the system and therefore irremediable short of radical system transformation--i.e., short of world government (although none of them believe that world government is feasible). The mirror image people seem not to recognize any structural cause; it is all a question of misunderstanding, misperception, "wrong" or "outmoded ways of thinking," and certain semi-pathological psychological traits. Thus, the dilemma is remediable through the recognition and abandonment of these dysfunctional mental processes and the "learning" of more appropriate ones. An example is Charles Osgood's prescription of a program of "unilateral initiatives", in which the United States would undertake certain friendly moves to teach the Soviets that we are not really aggressive, and which hopefully they would

reciprocate, until the spiral of mutual suspicion is transformed into a spiral of increasing mutual trust--eventually leading to disarmament.[19] The mirror image theorists are thus "optimists"; the security dilemma people are "pessimists."

Secondly, the security dilemma thesis purports to be only a partial theory of international conflict; these theorists recognize that states often have aims other than security, including expansion, and consequently that policies of deterrence and protection of interests are necessary. The mirror image hypothesis, on the other hand, along with its companion psychological mechanisms, purports to be virtually a complete explanation of the Cold War, and its proponents tend to deny the long-run utility of deterrence and in fact stress its dangers and futility. For them, disarmament is the preferred alternative. They think far-reaching disarmament is both feasible and desirable; the security dilemma theorists, by and large, do not.

Finally, the mirror image theory asserts, or strongly implies, that the cause of the predicament is outright misperception or illusion concerning the adversary's intentions. Each side comes to <u>believe</u> the other is aggressive when it really is not. The security dilemma thesis, on the other hand, locates the trouble not in "illusion" but in uncertainty. Given uncertainty about the opponent's intentions, responsible decision-makers feel impelled to insure against "the worst." The core of the problem, as Butterfield says, is the difficulty of "seeing inside the other's mind," not wrongly interpreting the other's mind. The behavioral result may be the same under either interpretation but the difference in assigned cause is important for the chances of amelioration: "illusion", as the mirror image advocates claim, may be correctable by learning "better habits of thinking," but uncertainty interacting with a sense of responsibility is much less tractable.

The security dilemma and mirrage image theories are similar to the prisoners' dilemma in the basic sence that the parties wind up in a costly state

of conflict when cooperation would have yielded better payoffs. However, they are both truncated versions of the prisoners' dilemma since they tend to ignore or underplay the offensive incentives to defect from cooperation. They assert or imply that states are interested only in security and, consequently, that "defection" to a state of mutual conflict is solely a defensive reaction motivated by fear of or uncertainty about the opponent's intentions. In other words, they imply that there are no "real" incompatibilities of interest between the parties which either might attempt to resolve aggressively by force. This implication is stronger in the mirror image than the security dilemma; the latter retains the possibility of interest conflicts and aggressive behavior stemming from overlapping security goals. An extreme version of the mirror image hypothesis is shown in Figure 2. Since the mirror image phenomenon is said to follow from mutual misperception, two matrices are shown, one for each party's perceptions of the payoffs.

A's perceptions

A \ B	Disarm	Arm
Disarm	5, 5	-10, 10
Arm	-2, 0	-5, -5

B's perceptions

A \ B	Disarm	Arm
Disarm	5, 5	0, -2
Arm	10, -10	-5, -5

Fig. 2

In this depiction, the "offensive" incentive for defection (armament) is removed, but the "defensive" incentive remains. Thus, the "real" payoff of the "I arm, he disarms" outcome for each party is -2, representing simply the cost of the armaments, and the cost of this outcome for the other side is zero since no aggression is intended. However, neither can know that and suspects that the other may attack if it gains superiority. A, therefore tends to focus on the payoffs which he perceives in his upper-right hand box, and B on those in his lower left-hand box. These payoffs are surmised payoffs rather than real ones and they are of course misperceptions of the other side's values and

intent. They are nevertheless operative and produce a mutual incentive for "defensive doublecross." The parties thus end up in the lower right-hand box (mutual armament) because of mutual misperception of each other's intentions.

The mirror image model conflicts rather sharply with the theory of deterrence. The latter assumes that the threat from the opponent is no illusion: there are real incompatibilities of interest which the opponent may attempt to resolve by aggression unless deterred. Thus, deterrence theory tends to focus on the adversary's offensive incentive to defect. Most sophisticated deterrence theorists probably recognize the security dilemma or mirror image elements in the situation; that is, they realize that our deterrent measures may be misinterpreted as threatening by the adversary and thus generate a certain amount of "unnecessary" power competition. But they see this as the necessary price to be paid for protection against an opponent who would surely take advantage of our weakness. However, some of the less sophisticated proponents of deterrence may virtually ignore the adversary's defensive incentive to "defect" in the prisoners' dilemma. In extreme cases, they may hold assumptions like the following: (1) the opponent <u>is</u> aggressive, (2) he knows that we realize he is aggressive, and (3) he knows that we are not aggressive and consequently recognizes that our armaments are wholly defensive in purpose. This would also be a "truncated" version of the prisoners' dilemma, but one that is truncated in exactly the opposite direction from the mirror image version.

The central problem for policy-makers, in prisoners' dilemma terms, is to determine which incentive, the offensive or the defensive, is operating most strongly in the adversary's thinking. Prior to World War I, the defensive incentive was most prominent in all actors; the war apparently developed largely out of a spiral of suspicion and mutual fear. The proper policy in this case would have been measures to reduce these fears by cooperative, tension-reducing moves. In the 1930's, by contrast, one actor had strong offensive incentives which were met with the inappropriate policy of conciliation and cooperation when

strong deterrent postures were called for. The contemporary situation seems to be a mix: the Soviet Union undoubtedly would take advantage of opportunities for cheap aggression, yet it is also clear that she perceives U.S. deterrent measures as threatening and that the contemporary power competition is to some degree rooted in unfounded or exaggerated mutual suspicion. Robert Jervis suggests that the mirror image advocates (or "spiral theorists" as he calls them) tend to believe that the contemporary cold war is analogous to the situation before World War I, while the deterrence theorists tend to think we are in a situation like the 1930's. Jervis also argues that the essential difference between the two schools lies not so much in their values or their theoretical analyses as in their answer to a simple empirical question: What are the intentions of the Soviet Union?[20]

All three of the theories just duscussed do indeed capture a part of the truth. The security dilemma idea clarifies the determining effects of system structure, highlights the central importance of the security drive in international behavior and shows how the search for security can be tragically self-defeating. But it does not embrace objectives of power and expansion, or incompatibilities of interest, which are not fuelled by the security motive. Deterrence theory does comprehend the latter but may often underestimate the degree to which the competition is "illusory", i.e., based on false or exaggerated suspicions about the opponent's intentions and motives. The mirror image theory is strong on the latter point but tends to ignore "real" conflicts of interest which are unlikely to respond to changes in psychological attitudes or attempts to reverse "hostility spirals." The prisoners' dilemma is flexible enough to include all the underplayed or ignored aspects of these partial theories. It provides a more complete portrayal of the consequences of anarchic system structure because it allows both for the possibility of illusory conflict engendered by mutual suspicion or fear and for the possibility of actual incompatibilities of interest and aggressive intent not motivated by security

considerations. The "defensive" incentive to "defect" catches the essence of the mirror image hypothesis and most of the security dilemma, the "offensive" incentive captures "real" conflicts of interest and expansionary aims, whether these stem from security or non-security values, and deterrence, of course, is simply "defection" to forestall these latter aims in the adversary. It may fairly be said that the prisoners' dilemma poses more clearly than any other available model a central and eternal puzzle in international politics: how to deter a potentially aggressive adversary without setting off a self-defeating and mutually damaging spiral of armament, hostility and conflict. Or alternatively: how to achieve the highest rewards of mutual accommodation without unbearable risks to core national values.

The Game of "Chicken"

Another model which is sometimes advanced as analogous to certain aspects of international relations, chiefly crisis confrontations and military coercion, is the game of "chicken".[21] Although in matrix form the game looks rather similar to the prisoners' dilemma, the differences are extremely important. As will be shown, chicken situations in the real world of international politics are quite different than prisoners' dilemma situations. Figure 3 shows a game of chicken matrix.

		B: 1	B: 2
A	1	0, 0	-10, 10
	2	10, -10	-50, -50

Fig. 3

The game is typically played as follows: One (or both) player(s) threatens to play strategy 2, hoping thereby to persuade the other to play strategy 1 in order to avoid high mutual losses. The first player who firmly commits himself to strategy 2, and communicates this commitment, or the player who is able to demonstrate the highest degree of resolve to go through with the threat, forces

the other to "cooperate" and thereby wins, at either 1, 2 or 2, 1. If both choose strategy 1 (in the original game, both swerve aside) the outcome is no gain or loss for either. If both carry out strategy 2, the outcome is heavy mutual loss (collision).

Note that two sets of payoffs have been changed as compared to the prisoners' dilemma game. The payoffs for mutual cooperation are reduced to 0, 0 because in this game the common interest lies simply in avoiding the default cell 2, 2, rather than in creating new values to be mutually enjoyed as in the prisoners' dilemma. The other change, more significant, is that the payoffs in the lower right-hand cell have been increased sharply in the negative direction, so that this outcome is now more costly for both sides than the loss from being on the losing side in either the 1, 2 or 2, 1 combinations.

This latter change in the payoff structure creates a radically different game, with entirely different incentives and psychological properties than the prisoners' dilemma. In chicken there is no "tragedy" or "vicious circle". The parties are not in a predicament where they cannot cooperate "no matter how much they want or try to cooperate." They are not, as in the prisoners' dilemma, driven to the outcome of mutual punishment by the very logic or structure of the game. More precisely, they are not driven to choose a non-cooperative strategy by suspicion that the opponent will not cooperate--by fear that their own cooperation will be exploited by the adversary. The reason, of course, is that the penalty for mutual non-cooperation is worse than the penalty for exploited cooperation, while the reverse is the case in the prisoners' dilemma. In chicken, a rational player always chooses to cooperate ("surrender") when facing an opponent who is not expected to cooperate; he cannot "protect himself" by not cooperating himself, for mutual non-cooperation is the worst possible outcome.

Both chicken and the prisoners' dilemma are "mixed motive" games, that is, they involve a mixture of common and conflicting interests. The difference

is chiefly in the differing role of the common interest in the play of the game. In the prisoners' dilemma, the realization of the common interest may well be primary desire of both parties, but neither can trust the other to collaborate in realizing it; against the will of the parties, the situation degenerates into conflict. In "chicken", one party willfully creates a conflict by challenging the other and threatens to destroy an already enjoyed common interest if it does not get its way in the conflict; the defending party may reciprocate with a similar threat. Typically, the common interest in chicken is something that is manipulated as a means of coercion, not something that is mutually sought. The spirit or leading theme of the prisoners' dilemma is that of the frustration of a mutual desire to cooperate. The spirit of a chicken game is that of a contest in which each party is trying to prevail over the other. In both games, perceptions of the other party's intentions are crucial and the actors face a problem of establishing the credibility of their stated intentions. But in the prisoners' dilemma, establishing credibility means instilling trust, whereas in chicken it involves creating fear.[22]

In both of the two models there is the possibility of misperceiving the opponent's intent, but the nature of the misperception is different. One or both players in the prisoners' dilemma may perceive the opponent as non-cooperative or aggressive when he is not; as we have shown, such misperceptions are one of reasons why the game often produces the undesirable outcome of mutual conflict. In the game of chicken there is usually no uncertainty about the parties' basic intentions--each is trying to prevail over the other. There is room for misperception, however, of each other's degree of determination. Over-perception may induce unnecessary capitulation; under-perception could produce disaster. A further difference is that in the prisoners' dilemma the misperception tends to be generated or at least encouraged by the structure of the game itself (structure of the system in the international analogue) whereas in chicken the misperception is simply a mistaken prediction of the behavior of

another actor.

The chicken game need not eventuate either in victory for one contestant or in mutual disaster. There is, of course, another possible outcome, that of mutual compromise, i.e., outcome 1, 1. Karl Deutsch asserts that this is the "natural" outcome of chicken games. He reasons that strategy 1 is "unequivocally rational" for both players because the potential cost of playing strategy 2 is greater than the potential cost of "cooperating" by playing strategy 1, and also is clearly greater than the potential gain from playing 2 in the hope that the opponent will play 1.[23]

Although this reasoning has a certain plausibility, it does not truly catch the spirit of the chicken game either in its juvenile gang version or its political analogues. Compromise may be more likely in chicken than in the prisoners' dilemma, but this outcome is not "natural" in the same sense that mutual non-cooperation is natural in the prisoners' dilemma. In the latter, strategy 2 dominates strategy 1 for both players, so that the 2, 2 outcome is not just "natural" but logically necessary. By contrast, neither strategy is dominating in chicken. If one expects the adversary to play strategy 1, one's best policy is to play strategy 2, taking advantage of the other's cooperativeness. If one expects the other to play 2, then one plays strategy 1 to avoid maximum loss. Rationality in chicken is equivocal; what is rational depends on a player's expectations about the other's behavior, not primarily on the game's payoff structure. Furthermore, the Deutsch prescription ignores the reiteration of the game: to yield on one occasion in fear of the 2, 2 outcome creates an expectation that one will yield again on future occasions, which will encourage "toughness" in the adversary and put oneself at a disadvantage in future plays.

It is more in keeping with the spirit of the game to say that compromise occurs when both sides expect the other to be "tough"--to play strategy 2, at least with a probability too high to be risked. In the hotrod version, both cars then swerve. In a real-life analogue, compromise results from a process of bargaining. One party concedes something out of fear that if it does not

yield the other will precipitate a conflict. The other then makes a reciprocal concession because it believes the first party will fight rather than yield further. Thus the parties find their way to a compromise because of mutual perceptions of the other's "toughness".

This suggests an interesting comparison with the prisoners' dilemma. Mutual cooperation can occur in the prisoners' dilemma only when each side thinks the other is not committed to strategy 2 but will play strategy 1. Compromise occurs in chicken when both parties fear the other is or may be committed to strategy 2. A mutual perception of toughness in the opponent thus may lead to compromise in chicken, but similar perceptions in the prisoners' dilemma tend to produce conflict and mutual loss. Thus a clear difference between the models, at least in their formal or logical implications, lies in the kinds of perceptions which are conducive to mutual cooperation.

There may be some situations in international politics where the Deutsch formulation is closer to the mark. These are cases in which the parties find themselves at the brink of war by accident or the "course of events," and both sides are less interested in winning than in getting themselves out of their predicament. Then the desirability of a mutual backing off may loom much larger than the advantages of appearing tough or firm, and the reputational costs of accommodation are low because the situation has not been provoked by a deliberate challenge. Both parties then may choose strategy 1 out of mutual fears of "things getting out of hand" rather than because of perceptions of the other party's toughness or determination.[24]

The chicken analogy usually is applied to crisis confrontations.[25] It is thus a model of bilateral coercion or bilateral bargaining under the threat of violence. However, both the image of the juvenile contest and the simple matrix as developed thus far fail to encompass many of the complexities and nuances of real international crises. These include the variety of objects and values represented in the "stakes" (in addition to prestige and status), the diversity

of possible outcomes, non-absolute estimates of risk and various degrees of "resolve", and a wide range of communication and bargaining tactics which may be employed to manipulate the value of the stakes and perceptions of risk and resolve. What follows is an attempt to develop the model somewhat further to reflect more of these elements.

First, the idea of probabilities can be introduced. Probabilities enter into the play of chicken games in at least two respects: (1) each party's estimate of the probability that the opponent will play strategy 2 (the credibility of the opponent's threat), and (2) the degree of risk of the opponent's choosing strategy 2 which each party can "stand" without being induced to give way. Daniel Ellsberg has called the latter function a party's "critical risk."[26] A moment's reflection will make one appreciate that the juvenile hotrodder does not decide to swerve because he thinks the adversary will <u>certainly</u> drive straight ahead but because he thinks the risk of the other doing so is "too great." Similarly, he does not decide to drive straight ahead himself because he is certain the other will swerve but because he thinks the chances of this happening are "pretty good" or "high enough." A similar calculation may occur in an international crisis.

Implicitly, the calculation behind the judgments "too great", "pretty good" and "high enough" involves the relation between a party's critical risk and his appraisal of the credibility of the opponent's threat. If the latter is the higher of the two probabilities, a party must logically back down (play strategy 1). If the opponent's threat credibility is below the threshold of the party's critical risk, the party can rationally stand firm on his strategy 2. Figure 4 may clarify these relationships.

		B (defender)	
		Comply .50	Stand Firm .50
A (aggressor)	Comply .40	0, 0	-5, 5
	Stand Firm .60	10, -10	-20, -20

A's critical risk = .50
B's critical risk = .60

Fig. 4

This matrix portrays a crisis precipitated by an aggressor's demand that a defender yield something worth ten units, under threat of war. If the aggressor "stands firm" and the defender "complies", the payoffs are 10 and -10, respectively. If the defender is firm and the aggressor complies (fails to carry out his threat and lets the matter drop) the aggressor loses and the defender gains bargaining reputation, prestige, etc., worth five units. The consequence of both standing firm is war at a mutual cost of 20. The outcome comply-comply (analogous to "both swerve" in the game of chicken) we assume, for convenience, is a compromise with no net gain or loss to either party.

The critical risk threshold for either side is derived from a comparison of its payoff from complying with its payoffs for standing firm. B, for example, loses 10 by complying with the demand. If he stands firm he either gains 5 or loses 20 depending upon A's choice. If he estimates a .40 chance that A will comply and a .60 chance that A will be firm, B's "expected value" from standing firm is -10, just equal to the cost of compliance. In other words, when B estimates the credibility of A's threat at .60, B is indifferent between complying or standing firm. This is B's "critical risk"--i.e., the credibility of A's threat must be at least this high to force B to back down. A similar calculation will show that A's critical risk is .50--if A estimates the probability of B's firmness at higher than this, A must retreat (renege on his threat).[27]

If both parties feel that their critical risk is higher than the credibility of the opponent's threat, they will both commit themselves to fight if the other does not give way and the outcome is war. Conversely, if both perceive that their critical risk is lower than the opponent's threat credibility, the stage is set for mutual compromise.

International crises often involve a rich assortment of communicative moves and bargaining tactics designed to influence the adversary's perceptions and behavior. Many of these tactics can be related to our model. It is easy to see that the "bargaining problem" for each side is to arrange the other's utilities and perceptions so that the perceived credibility of its own threat of firmness is higher than the other's critical risk. Then the other must give way. Thus, there are two broad classes of coercive bargaining tactics for each side--those which attempt to increase its own threat credibility and those which seek to reduce the adversary's critical risk.[28]

Within the category of increasing threat credibility there are two subclasses: (1) changing the opponent's perception of one's own utilities and (2) increasing credibility without changing one's apparent utilities.

The opponent's estimate of the credibility of one's threat depends in part on how he perceives one's own valuation of the stakes at issue and one's own assessment of the costs of war. These perceptions are subject to manipulation. Either side, for example, can reduce its apparent costs of war by increasing its capabilities or by verbal statements indicating confidence of victory or indifference to war costs. Or the parties can increase their apparent costs of compliance (negative payoffs in the lower left and upper right boxes) by citing the legitimacy of their positions, tying them to principles and precedents, linking the issue involved to other issues, and deliberately engaging their prestige and future bargaining reputation.

The parties can also enhance the credibility of their threats without modifying their apparent utilities (In the model, manipulate the probabilities

but not the payoffs)--by such methods as "irrevocable commitment" which forecloses the alternative of compliance, appearing to lose control over bellicose subordinates, becoming committed to firmness by pledges to a constituency, or pretending recklessness or irrationality.

The adversary's critical risk is a function of his utilities, and these can also be manipulated. One can magnify the opponent's cost of war by increasing capabilities, stressing the danger of escalation or threatening to expand one's objectives after a war starts. Or the stakes can be devalued for the adversary, thus decreasing his costs of compliance, by tactics such as creating loopholes, offering a real or apparent quid pro quo, citing community values (e.g. "peace") which would be served by his compliance, or pointing to the illegitimacy or "abnormality" of his position. In short, the parties' perceptions of the "numbers in the boxes" and of the probabilities associated with each other's alternatives can be modified by various communication and bargaining tactics and most of the activity in international chicken games involves the use of such tactics.

Space forbids more than this very cursory discussion of bargaining processes.[29] Our only purpose here has been to show how the basic chicken model, in matrix form, can serve as a useful framework for classifying and analysing, in terms of function, a considerable range of these tactics.

Some Real-World Interactions and Comparisons

In the real-world analogues of our two models, the prisoners' dilemma appears as the more fundamental, pervasive and continuous since it is a function of the basic structure of the system. In the context of anarchy, any pair of states or blocs of roughly commensurate power, which identify themselves as adversaries, are likely to be caught in a prisoners' dilemma. Between such pairs, the typical state of affairs is the undesirable or less than optimum one of "double defection" (the 2, 2 outcome). The principal costs of being in

this predicament are arms burdens, the risks and constrictions of alliances, and the risk of war.

It is within the context of this shared fate that the game of chicken is played. That is to say, the prisoners' dilemma, in this primary form, is a kind of supergame whose characteristic outcome subsumes the potential for games of chicken. Thus, part of the "payoffs" of the 2, 2 outcome in the prisoners' dilemma are the expectations the parties hold about the outcomes of future games of chicken.

The distinction between preparedness policy and action policy also is useful in clarifying the interaction between the two games. The most prominent aspect of the 2, 2 outcome in the prisoners' dilemma is mutual armament--i.e., preparedness. Chicken games, on the other hand, are played at the level of action policy. Given that armaments exist, the threat to use them is the chief means of winning or avoiding defeat. And the cost of mutual non-cooperation in chicken is typically the cost of war (although it is possible to imagine chicken-type games in strictly political or economic contexts in which the 2, 2 outcome is non-violent). There is still a possibility, of course, for "secondary" prisoners' dilemmas--e.g., the "reciprocal fear of surprise attack" --to occur at the action level.

Thomas Schelling has remarked that it takes two not to play a game of chicken. On the drag strip, "if you are publicly invited to play chicken and say that you would rather not play, you have just played."[30] The international corollary is that a state which declines to rise to challenges must be willing to accept the settlements imposed upon it. An implication of this, however, is that two can decide not to play. If a challenge is not issued, the game is not played. By contrast, it is not possible for two (or any number) of the active participants in world politics to decide not to play the prisoners' dilemma because they are in this game not by choice but by fate.

As illustrated by the case of World War I, some crises may be prisoners' dilemmas of the secondary type. These are the most dangerous kind because of their tendency to trigger pre-emptive attack. Only slightly less dnagerous are crises which are basically chicken games but are subject to being transformed into prisoners' dilemmas at some point in their progress. This could happen if the costs of war are seen as relatively low compared to the value of the stakes or if the stakes are progressively magnified in the parties' minds by commitments and counter-commitments, to the point where it becomes less costly to fight than to concede. Figure 5 shows a situation which has undergone such transformation. Commitments of prestige and bargaining reputation to strategy 2 have made it more costly for each side to adopt an accommodating strategy (strategy 1) than to go through with the commitments and fight. Once both sides perceive the other's commitment, bargaining becomes superseded by efforts to prepare for, and perhaps to start, the "inevitable war." Some process such as this may be involved in the idea that there is some point in a crisis where "events get out of control," or a point where the parties become "locked in."[31] The central problem in "crisis management" is to stop a crisis before this point is reached.

		B: 1	B: 2
A	1	0, 0	-60, 10
	2	10, -60	-50, -50

Fig. 5

Our two models may be employed to distinguish between two kinds of escalation which may occur during the progress of a war. Morton Halperin has distinguished between escalation by "expansion" and escalation by "explosion."[32] Escalation by expansion is similar to a game of chicken. This is a deliberate, gradual, step-by-step intensification of nuclear violence, starting perhaps with

the restrained use of small tactical nuclear weapons on strictly military targets in the battle zone and moving upward through more and more powerful weapons, used with progressively less discrimination. Each step upward is deliberately calculated as a move likely to prompt the other side to capitulate out of fear of greater violence to come. In other words, each escalatory step is a threat-by-demonstration intended to convince the adversary of one's resolve to continue escalating if he does not come to terms. As in any chicken game, this process of competitive demonstration can spiral upward to the ultimate catastrophe.

Escalation by "explosion", however, involves a prisoners' dilemma. At some point in a war, probably after tactical nuclear weapons are introduced, one or both sides begin to fear that the other is about to launch a full-scale strategic first strike. Then one side decides to pre-empt to get the benefit of striking the first blow. It is easy to see, of course, that this kind of escalation is not undertaken to coerce the adversary by establishing certain expectations about one's probable future behavior--as in escalation by expansion. It is done to protect oneself against the worst possible outcome--a "doublecross" (first strike) by the enemy.

Similar reasoning can be employed to distinguish between "accidental" war and "inadvertent" war. The usual scenario for accidental nuclear war involves a prisoners' dilemma. Evidence is obtained or an event occurs--e.g., an accidental missile firing against one side's territory--which leads that side to believe that the other is about to undertake, or is already in the process of undertaking, an all-out attack. That side then decides to launch its own strike pre-emptively. Accidental war is therefore deliberate war which is started because of misperception or fear of the adversary's intentions. The accident triggers the misperception and the fear but it only indirectly causes the war.

"Inadvertent" war, on the other hand, is the outcome of miscalculations in a game of chicken. In a crisis confrontation, both sides make threats and commitments. One side underestimates the resolve of the other and commits itself

irrevocably to war, thinking the commitment will persuade the opponent to back away. But the other does not back away because, contrary to the first side's calculation, it has already become committed itself out of similar expectations. The result is war which "should not" have occurred, or would not have happened if the parties had not misperceived each others intentions.

Both chicken and the prisoners' dilemma are infinitely reiterated games in the real world. This means that the participants must always be aware of the effects of their behavior in a single instance on the development of others' expectations concerning their future behavior.

In the reiterated chicken game, this awareness takes the form of a need to protect one's position in the "balance of resolve". This is an important but often neglected component of the "balance of power", the other major component being the "balance of capability." The concept of balance of resolve captures the crucial importance of intentions, and the reciprocal perception of intentions, in the balance of power and transforms the notion of balance of power into more of a political and less a military concept. Often the reasons why the balance of power has been successful or unsuccessful in preventing war have turned on intent-perception rather than on perceptions of capability. Britain and France failed to deter Hitler in 1939 not so much because they were weak in material strength, but largely because Hitler discounted their will to fight after they had demonstrated weakness in crisis after crisis. In the nuclear age, intent-perception may be even more important, for the "balance of terror" depends not so much on relative capabilities as on reciprocally perceived intentions to invoke various possible levels of violence.

Each side's resolve and image of the other's resolve will fluctuate through time and tend to shift with events. For example, the outcome of the Cuban missile crisis apparently raised Russian perceptions of U.S. determination which had declined as a result of apparent U.S. weakness and vacillation on several earlier occasions. In any particular crisis, part of the payoff for firmness

is in the encouragement of expectations of one's future firmness. Logically, this future-oriented component should be discounted for both time and uncertainty since future values ordinarily are valued less than present ones and there will be some uncertainty about whether an impression of weakness can perhaps be repaired, or whether it will significantly affect the outcome of future confrontations. The magnitude of these discounts will vary among actors and situations. France and Britain apparently were induced to discount the future heavily in the 1930's by Hitler's assertions that "This is my last demand," implying that there would be no future occasions when the resolve of the Western powers would be tested. Conversely, President Eisenhower's "dominoe theory" and Secretary Rusk's repeated reference to "the interdependence of American commitments" in the context of the Vietnam debate implied low discounts for time and uncertainty.

The balance of resolve is logically unstable: if one side backs down once, the resolve of the other side increases, and the first side, realizing that it is expected to give way again on the next occasion, may find its own resolve weakening precisely as a result of this realization. With each successive capitulation, the resolve of the winning side increases and vice-versa, with the result that the preponderant side becomes even more preponderant and is eventually able to coerce the other at will. But another psychological mechanism may prevail over this one. One side, having given way one or two times, may say to itself "Never again!" and resolve to stand firm at the next challenge. If it is able to communicate this intent the balance may be stabilized. More dangerously, if it is unable to communicate its new determination, war may occur through miscalculation.

An example of the latter phenomenon concerns Russia before World War I. During the early 1900's, Russia had deferred several times to the demands of Austria and other powers in conflict situations in the Balkans. The Austrians may have concluded from this that Russia would not fight to defend Serbia after

Sarajevo. In fact, the Russians had vowed "Never again!" by the time of the Sarajevo crisis. The Austrians miscalculated because the Russians failed to communicate that they were now committed not to back down again.

The prisoners' dilemma game of the arms race is also infinitely reiterated since the antagonists repeatedly make new decisions about their armaments and sporadically negotiate about arms control and limitation. Although it is impossible to move completely to the 1, 1 solution (total disarmament), presumably the adversaries are both interested in reaching some limited variant of it, and the security value of each new quantitative or qualitative increase should be discounted by its effects on the probability of eventually getting there. The current debate about the ABM and MIRV largely concerns the proper size of this discount. And part of the payoff from any particular control or reduction agreement is the encouragement of trust in the opponent and a general atmosphere of cooperation which may make future agreements more likely. However, although laboratory experiments with the prisoners' dilemma indicate that it is possible to "teach" the opponent the virtues of trust over a series of reiterated plays,[33] in the real world this is more difficult because of the much more serious consequences of unreciprocated trust.

Playing prisoners' dilemma and chicken at the same time, the contestants may face difficult choices concerning the kinds of policies to adopt, and the kinds of images and attitudes they wish to encourage in the adversary. There is an obvious contradiction between instilling trust and instilling fear and respect in the opponent. To maximize one's chances of success in future crises, a general image of "toughness" is called for. But this may be counterproductive for the development of <u>detente</u> and cooperation, including arms control agreements, where images of benevolence and cooperativeness are likely to be most successful. Conversely, unilateral moves to reduce tension, or an attitude of flexibility and accommodation in arms control negotiations, may create an impression of weakening resolve in conflict situations. The line between

"firmness" and "aggressiveness", or between "flexibility" and "weakness" may be extremely clear in our own self-image but not clear at all in the image we project to the adversary.[34] Certain trends in contemporary history suggest that this apparent incompatibility may be considerably muted in practice, that the superpowers are able to discriminate between aspects of each other's behavior which are oriented to chicken-type competition and those which are aimed toward cooperation in the prisoners' dilemma. The Cuban missile confrontation did not prevent (may even have encouraged, it seems) subsequent moves toward collaboration and detente, and the adversary roles of the United States and the Soviet Union in the Vietnam conflict do not seem to have seriously weakened the detente. Still, in general, the co-existence of prisoners' dilemma and chicken games in international politics does give rise to a central problem of policy: how to protect one's interests in the dimension of conflict while at the same time maximizing the realization of common interests.

FOOTNOTES

[1]Conventionally, A's payoff is shown first, and B's second, in each cell,

[2]Anatol Rapoport and Albert Chammah (among others) have shown that some cooperation does take place in laboratory simulations, particularly when plays are reiterated many times. However, for purposes of this brief exposition of the logic of the model, we shall assume a single play. See Anatol Rapoport and Albert Chammah, Prisoner's Dilemma: A Study in Conflict and Cooperation (Ann Arbor: The University of Michigan Press, 1965).

[3]There is a close resemblance between the prisoners' dilemma and the theory of collective goods in welfare economics. Cf. Mancur Olson, The Logic of Collective Action (Cambridge: Harvard University Press, 1965). Olson points out that it is not rational for individual members of groups, such as labor unions, to contribute dues voluntarily to the group in order to obtain a good which is collectively enjoyed, because the individual's contribution has little or no effect on whether the good is obtained. If others are expected to "cooperate" by paying their dues, the defector enjoys the collective good without cost; if others also defect, the collective good is not obtained but the individual defector at least avoids being "suckered" into contributing to a futile cause. The only ways in which dues can be collected are by coercion or by providing "selective incentives"--i.e., rewards for dues-paying other than the collective good.

[4]The Franco-Russian counter-alliance was delayed for 15 years by Bismarck's brilliant diplomacy, ideological frictions between France and Russia, and French and Russian preoccupation with colonial competition with England, but in a long perspective it probably was inevitable.

[5]This statement assumes a multipolar international system with participants of roughly equal power, which has often been observed to be an essential

condition for the success of collective security.

[6]For examples of the "preclusive" motive for acquiring colonies, see William L. Langer, European Alliances and Alignments (New York: Random House, Vintage Books, 1964), pp. 284, 295. Of course we do not intend to argue that there were not other "autonomous" motives which contributed to imperialistic competition.

[7]Thomas C. Schelling, The Strategy of Conflict (Cambridge: Harvard University Press, 1960), pp. 207-230.

[8]Robert S. McNamara, Address to editors of United Press International, San Francisco, Sept. 18, 1967. New York Times, Sept. 19, 1967.

[9]The originator of the security dilemma theory in modern terms apparently was John Herz. See his Political Realism and Political Idealism (Chicago: University of Chicago Press, 1951), also his International Politics in the Atomic Age (New York: Columbia University Press, 1959), Chapter 10.

[10]Herbert Butterfield, History and Human Relations (London: Collins, 1951), p. 21.

[11]Kenneth N. Waltz, Man, the State and War (New York: Columbia University Press, 1959), Chapter VI.

[12]Arnold Wolfers, Discord and Collaboration (Baltimore: The Johns Hopkins Press, 1962), p. 84.

[13]Actually it is a little more complicated; all states would also have to be convinced that others perceived their own recognition that security was the only aim of others. Here one meets the familiar infinite regress of expectations, but in practice probably only the first one or two levels are important.

[14]Raymond Aron, Peace and War, translated from the French by Richard Howard and Annette Baker Fox (Garden City, N. Y.: Doubleday, 1966), p. xi.

[15]Certain other qualifications of the Hobbesian theme are worth making. When states become allies, their focus on a common enemy washes out, if only temporarily, any potential security dilemma between them. Some states pose no threat to each other because of geographical distance and others have developed bonds of community which rule out any possibility of war between them. Still, it would be a mistake, incidentally, to overlook the element of power configuration in certain cases where "community" seems to have displaced security concerns. For instance, if Canada were as powerful as the United States, it is quite conceivable that a "security dilemma", although perhaps only a moderate one, would arise between them. The community-like behavior among members of the European Common Market could hardly have developed without the common perception of the threat from the East. For an excellent discussion of various ways in which the consequences of international anarchy are mitigated by elements of "society," see Hedley Bull, "Society and Anarchy in International Relations," in Herbert Butterfield and Martin Wight (eds.), Diplomatic Investigations (London: George Allen and Unwin Ltd., 1966), pp. 35-51.

[16]Butterfield, op. cit., p. 23.

[17]On this point, see Ole Holsti, "The Belief System and National Images," Journal of Conflict Resolution, 6 (September, 1962), p. 242-252.

[18]Ralph K. White, "Images in the Context of International Conflict: Soviet Perceptions of the U.S. and the U.S.S.R.," in Herbert C. Kelman (ed.), International Behavior (New York: Holt, Rinehart and Winston, 1965), pp. 248-249. For other expositions of the mirror image phenomenon, and prescriptions for dealing with it (although not always under that label), see Urie Bronfenbrenner, "The Mirror Image in Soviet American Relations," Journal of Social Issues, 17, (1961), p. 52; Charles E. Osgood, An Alternative to War or Surrender (Urbana: University of Illinois Press, 1962); and Amitai Etzioni, The Hard Way to Peace

(New York: Collier Books, 1962).

[19]Charles E. Osgood, op. cit. This brief statement does not, of course, do justice to Osgood's formula.

[20]Robert Jervis, "The Security Dilemma, the Spiral Model, and Deterrence." Unpublished manuscript, Harvard University, Center for International Affairs, March 3, 1967. This paper has contributed a good deal to my analysis in this section.

[21]The original game of chicken is a game reportedly played for prestige and status in juvenile gangs. Two drivers straddle the white line on a road and drive toward each other. The first to turn aside is called "chicken" and suffers humiliation and loss of status.

[22]Anatol Rapoport, Strategy and Conscience (New York: Harper and Row, 1964), p. 116.

[23]Karl Deutsch, The Analysis of International Relations (Englewood Cliffs, N. J.: Prentice-Hall, 1968), pp. 119-120.

[24]A similar point is made by Thomas Schelling in "Uncertainty, Brinkmanship, and the Game of 'Chicken'," in Kathleen Archibald (ed.), Strategic Interaction and Conflict (Berkeley: University of California Institute of International Studies, 1966), p. 85.

[25]Of course, there are various types of crises, and the model fits some of them only imperfectly or not at all. The best "fit" is obtained for crises which begin with a demand or challenge backed by threats, which the victim then resists with counter-threats, but in which both parties hope to get their way without violence and would prefer capitulation to war. The Cuban missile crisis, for example, fits these specifications pretty well, as does the Moroccan crisis of 1911. Among other types of crises which are not appropriately modeled

by the chicken game are those which are merely preludes of tension prior to a war which one state deliberately provokes (e.g., the Franco-Prussian War), and those involving a challenge by a powerful state to a weak state, in which the latter by itself cannot threaten to inflict high costs (e.g., the Austrian Anschluss).

[26]Daniel Ellsberg, "The Theory and Practice of Blackmail," lecture delivered at Lowell Institute, Boston, March, 1959, in a series entitled "The Art of Coercion: A Study of Threats in Economic Conflict and War."

[27]The formula for critical risk is $\frac{uF_1 - uC_1}{uF_1 - uW}$ when uF_1 is the utility of successful firmness, uC_1 is the (negative) utility of compliance, and UW is the (negative) utility of war. The subscript "1" indicates that this is the situation after the first demand and counter-offer (In our example, B has, in effect, made a counter-offer of the status quo). Lesser demands and higher counter-offers will be made if both sides perceive the other's threat credibility to be higher than their own critical risk. If further bargaining ensues, the utilities and consequently the critical risks will change. When the parties move toward a compromise, logically, at each stage, the side with the greatest "gap" between its critical risk and its perceived risk that the other will carry out its threat will be the one to offer a concession, for this side can least "stand" the risks of a facedown. For a similar formula set in the context of labor-management negotiations, see Frederick Zeuthen, Problems of Monopoly and Economic Warfare (London: Routledge, 1930).

[28]There is a third class of bargaining tactics which we might call "bidding moves"--the making of new demands and offers and new settlement proposals--but here we are concerned only with "coercive moves."

[29]For fuller treatments, consult Thomas C. Schelling, The Strategy of Conflict (Cambridge: Harvard University Press, 1960), and his Arms and

Influence (New Haven: Yale University Press, 1966). Also see Herman Kahn, On Escalation: Metaphors and Scenarios (New York: Praeger, 1965).

[30]Thomas C. Schelling, "Uncertainty, Brinkmanship and the Game of 'Chicken'," in Kathleen Archibald, op. cit., p. 83.

[31]A similar notion is expressed in Bruce Russett's "point of no escape." See his "Cause, Surprise and No Escape," Journal of Politics, 24 (1962), pp. 3-22.

[32]Morton Halperin, Limited War in the Nuclear Age (New York: John Wiley & Sons, 1963), p. 3.

[33]Rapoport and Chammah, op. cit.

[34]A similar point is made in Robert Jervis, op. cit., pp. 114-116.