PROTOCOL, PAYOFF, AND EQUILIBRIUM:

GAME THEORY AND SOCIAL MODELLING

Jack Hirshleifer

University of California, Los Angeles

UCLA Dept. of Economics
Working Paper #366
March 1985

PROTOCOL, PAYOFF, AND EQUILIBRIUM: GAME THEORY AND SOCIAL MODELLING

The purpose of this paper is mainly programmatic. My contention is that, in applications of game theory, it will be productive and clarifying always to maintain a sharp distinction between: (i) the payoff environment facing the participants, versus (ii) the protocol that specifies precisely how the players are permitted to interact. Consideration of these two elements, I will argue, is needed in order to understand the nature and validity of certain of the standard equilibrium concepts in game theory.

In applying game theory we are always attempting to model social situations. What I call the payoff environment is the set of distributed opportunities offered by Nature to the "society" comprised by the players of the game. It is typically displayed in the elementary normal-form matrices[1] known as Chicken, Prisoners' Dilemma, etc. The protocol represents a different aspect of the social situation, illustrated by the answers to such questions as: At each point in time who is allowed to make what sort of move? From what starting-point? What information is conveyed? How does the game end? In fact, the protocol is just what we call "the rules of the game" in common parlance. (Game theorists employ the game-tree or extensive form to picture at least some of these protocol elements, notably the sequence of moves and the state of information.) In short, the payoff environment consists of the available exogenously determined returns to the members of the social

[1] An elementary normal-form matrix has rows or columns only for what I term the "execution moves" of the players (see below).

grouping, while the protocol of play models certain aspects of the internal political and social institutions of the group.[2]

For concreteness, consider the social criterion of Pareto-efficiency. The question might be raised, for example, is a Pareto-efficient outcome more likely under social contexts corresponding to the game of Prisoners' Dilemma or under those corresponding to the game of Chicken?[3] Alternatively, we might be interested in the question of relative advantage: what social structures and corresponding games tend toward an equilibrium favoring one player over another? We cannot validly answer such questions without attending both to the payoff environment and to the protocol. As one instance, it is often said that in the single-play Prisoners' Dilemma the solution outcome is Pareto-inefficient. For many possible protocols that statement is indeed correct, but in others it is incorrect (as will be shown below). Or it might be that Pareto-efficiency is attained in one payoff environment only if the players move simultaneously, in another only if they move sequentially. And in addition, the protocol may be important in specifying whether the equilibrium

[2]Experimental economists have found it useful to distinguish between "environment" and "market institutions" in a way that parallels the payoff vs. protocol distinction employed here. Among the protocol aspects ("market institutions") explored in such studies have been, e.g., whether offers to buy or sell are to be public or secret, whether either side may bid or only buyers or only sellers, etc. (see, for example, Plott [1982] and Smith [1983]). (In contrast, those economists who emphasize the power of the Coase Theorem are essentially saying that the details of market protocols do not really matter for the final outcome.) The term "protocol" as used here also corresponds to the agenda in political choice, and of course it is well known that the details of the agenda may be crucial in determining the outcome achieved by majority voting or other political process. The role of various protocol elements has been emphasized by some game theorists, e.g., Schelling (1960) who calls them "institutional and structural characteristics of negotiation" (p. 28).

[3]For treatment of such questions see, e.g., Rapoport [1960, Ch. 11], Snyder and Diesing [1977, Ch. 2].

is unique or multiple, stable or cyclic, etc.

It is of course by no means a new idea that the final outcome for any game depends upon the specification of what I am calling a protocol. As already mentioned, the game-tree or extensive-form picture of the game does direct attention to the protocol, or at least to certain aspects thereof. Nevertheless, while game theorists surely have understood the significance of protocols, they have done surprisingly little in the way of systematic analysis of the implications of games conducted under one protocol rather than another. And in fact, even a taxonomy of the ways in which protocols might vary is lacking.

My purpose here is to make a start at these tasks. I will be showing systematically how various structural aspects of protocols — different "architectures of the game-tree", so to speak -- have predictable implications for the final outcomes of strategic interactions.

A. ELEMENTS OF ALTERNATIVE PROTOCOLS

In the interests of expository simplicity I will consider only 2-person, single-play games. In repeated-play games ("supergames," in the current jargon) the participants would earn multiple payoffs over time. In this analysis, in contrast, the ultimate outcome is a single payoff-pair from the elementary normal-form matrix. However, what is very important, the governing protocol might allow for one or more rounds of what I shall call "negotiation moves" -- offers, commitments, etc. -- before the players come to make their "execution moves." Only these final execution moves, however arrived at through the negotiation process, determine the payoff-pair received.

Communications — explanations, threats, promises, etc. — constitute one category of negotiation move that may or may not be permitted by the operative protocol. A commitment, by which a player can limit his own future freedom of choice among execution moves (in the hope of influencing the opponent's decision thereby) is another kind of negotiation move. To keep things still simple, however, I will be concerned here only with one very elementary type of negotiation move: what I will call a "try". A try is a tentative execution move, i.e., one that may or may not (depending upon the provisions of the protocol) become binding upon the player. The protocol might specify, for example, that Row has the first try (he makes a tentative selection among the execution moves available to him), then Column has a try, and finally Row has one last chance to change his mind. In this 1 1/2-round protocol Row's first move is a non-binding try, but the succeeding tries on each side lock into execution moves.

It is quite essential to maintain a strict distinction between "move" and "strategy". A player's <u>strategy</u> is a detailed and complete plan covering both negotiation and execution moves. The strategy could in principle be left in the form of instructions for an agent, telling him which (negotiation or execution) <u>move</u> to make on the player's behalf, in every context where the protocol allows a choice. A player could have a strategy in which his chosen negotiation or execution move at any point depends upon his opponent's previous visible moves. But, it is important to note, <u>one's chosen move at any point cannot in general be made to depend upon your opponent's strategy</u> -- since that will ordinarily not be fully visible in the negotiation phase. (Or, indeed, the opponent's strategy may not be fully visible even after the execution moves have completed the game.)

While in principle strategies can be "pure" or "mixed" (probabilistic), I will be limiting myself here to pure strategies. As still another simplification, I will be considering only protocols that strictly limit the players to the original payoff environment as displayed in the elementary normal-form matrix. Thus, no side-payments or transfers (between the players, or with third parties or with Nature) may take place. Also excluded are any agreements or threats or promises, except insofar as these may be implicitly conveyed by actual tries in the negotiation phase. In short, at any point the Row player can do no more than select a try from one row or another in the payoff matrix, and correspondingly for the Column player. Such protocols define what I shall call "transition games." Again for simplicity, I will always be assuming that both players have full knowledge of the payoff matrix, of the governing protocol rules, and of the past history of the negotiation or execution <u>moves</u> already made. (But the players are not, to repeat, generally in a position ever to know the opponent's underlying <u>strategy</u>.)

Among the different dimensions in which protocols may vary are: (1) the status quo position; (2) the sequence of turns; (3) ordering rules for allowable tries; and (4) the provision for termination.

1. The status quo position

Three different types of status quo or start-up positions can be described. First, play may commence with each player already "endowed with" an existing try. Thus, the negotiation phase begins with the parties initially located in one particular cell of the normal-form payoff matrix. Or, under the opposite assumption, the status quo could be an entirely clean slate, with no tries at all having yet been chosen. In this latter case there would have to be one round of "initializing" tries before transitions proper can begin. The third possibility is intermediate between the other two: one player is endowed with an existing try, which specifies the Row or the Column (as the case may be) of the start-up position, but the other party's slate is clean. Here a half-round of initialization would be required before transitions can commence.

2. The sequence of players' turns

The players may choose tries in alternation, or simultaneously. (Of course "simultaneity" must be understood as referring to the state of knowledge rather than to clock time: if your choice is later in time but must be made in ignorance of your opponent's current try, the moves are simultaneous from the decisional point of view.) Again, there are a number of intermediate cases possible. One such is partial information. For example, if the other player has options 1,2,3 you might know that on his current try he has not selected 3 but still be unable to determine whether his choice was 1 or 2. As another intermediate case, the protocol may provide for some more or less complicated switching between alternation and simultaneity in successive rounds of negotiation play. In the interests of still keeping things simple I will be dealing here only with protocols providing either for strict alternation or else for strict simultaneity of tries.

3. Ordering rules for negotiation moves

There may be some limitations upon the order in which negotiation moves may follow one another. In an ordinary commercial auction, for example, you are permitted to raise your previous bid but not to lower it.

4. The termination procedure

Here there are a great number of distinguishable and interesting possibilities. The following are some important examples:

(i) Fixed termination: The governing protocol may specify a fixed number of negotiation rounds. Thus, a 1 1/2-round game is dictated by a protocol like: "Row has a try first, then Column, then Row once more -- after which execution takes place."

(ii) "Natural" termination: The negotiation phase "naturally" terminates when all players are satisfied to stand fast (pass) rather than change their last tries. Thus, in an open public auction, the negotiation phase "naturally" terminates when no-one cares to raise his last bid.[4]

(iii) Player-biased termination: One player, but not the other, may have the option of declaring the negotiations ended (even if he has just switched).

(iv) Move-biased termination: A player's choice of one kind of try (of one particular row or column, as the case may be), but not of another, may bring about termination of the negotiation phase. As a practical example, some historians have argued that in the military situation before World War I the decision of any nation to mobilize -- given the strategic advantage of early mobilization and the absence of a stand-down procedure -- would have sufficed to put an end to further useful negotiations. As it happened Russia, with her great distances and inadequate transportation, felt it necessary to initiate mobilization first; that decision having been made, war became inevitable.

Combinations of termination rules are also conceivable. An example for the simultaneous-move game might be: "At the end of the first round of tries, Row but not Column has a termination option; should he not exercise it, the game ends anyway after two more rounds (unless it naturally terminates earlier)."

B. PRINCIPLES OF SOLUTION

I shall adopt the following principles (some already touched upon above) for arriving at a solution under any specified payoff matrix and protocol:

---

[4]Brams and Wittman [1981], in an article which has certain parallels with the present paper, limit their analysis to what I am calling "natural" termination protocols.

(1) Rational behavior: Each player maximizes his payoff, and is fully able to compute even very long and complex chains of possibilities on the order of "If I do this and then he does that...." Also, he assumes that his opponent is equally rational, and that this mutual rationality is "common knowledge".[5]

(2) Adherence to protocol: Every negotiation or execution move on the part of the players must be explicitly permitted in the protocol.

(3) Knowledge of the game: Each party is fully informed as to the operative payoff matrix and protocol.

(4) Memory yes, telepathy no: Each player can remember the history of all previous tries. However, the player does not know his opponent's future negotiation or execution moves (except insofar as he can infer them from the rationality principle). Thus, as has been mentioned, he cannot in general know his opponent's strategy.

These assumptions suggest that the way to find the solution is to apply the principle of dynamic programming (Bellman [1957]). Working backward through the game-tree the player, determining at every step the rational option for himself and his opponent, can ultimately calculate his optimal initial try.[6] Allowing for such behavior on both sides leads to the "Subgame-Perfect Equilibrium" (SGPE) concept (Selten [1975]). However, since I want to

---

[5]With common knowledge each player is rational himself, knows that his opponent is rational, knows that the other knows that he knows, etc. (see Aumann [1976]). If the mutual rationality were not common knowledge, I might correctly believe that you are rational but also believe that you mistakenly think I am not rational. Failure of common knowledge would in general affect the solution arrived at by two rational players.

[6]Dynamic programming may not lead to unique choices in a number of circumstances, but in the simple games to be considered here this difficulty will not arise.

highlight the sensitivity of the solution to the governing protocol, for purposes of emphasis I will employ the term "Protocol-Dependent Rational Equilibrium" (PDRE). Recall also that we are here considering only transition games. This structuring of the situation is what will allow us to explore the effects of the "protocol architecture" upon the equilibrium arrived at.

Figure 1 illustrates, for the payoff environment represented by Matrix 1, two alternative single-round protocols: in Protocol A Row has the first move, in Protocol B Column moves first. The ultimate payoff-pairs are indicated at the end of each decision path, at the far right. The payoff-pairs appearing at the earlier decision nodes represent the effect of allowing for optimal choices at all downstream points. Under Protocol A, should node y' be reached Row can predict that Column's later choice of move c_1 will lead to the payoff-pair (2,4).[7] Similarly, at node y'' Row can predict the ultimate payoffs (3,3). Since y'' is therefore superior to y' from Row's point of view, at his initial decision node x Row will select move r_2. Thus, the Protocol-Dependent Rational Equilibrium (PDRE) for this game is the path $\{r_2 \rightarrow c_2\}$ generating (3,3) as payoffs.[8] Similar reasoning under Protocol

---

[7]Following the standard notational convention, payoff vectors (shown in parentheses) will always be written with the Row payoff coming first.

[8]Where the temporal sequence of moves is shown, Column moving first will be indicated by a notation in braces such as $\{c_2 \rightarrow r_2\}$ — meaning here that Column has initially chosen move c_2 followed by Row's choice of r_2. Subscripted lower-case letters like r_1 or c_2 will be used to denote moves (tries) on the part of the Row and Column players, while similarly subscripted upper-case letters signify strategies. For the elementary normal-form matrices, which show the payoffs to execution moves only, it is the lower-case symbols that appear in the margins. But it will sometimes be useful to display in normal form some non-elementary matrices (e.g., Matrix 3A below) where the payoffs are shown as determined by pairs of more or less complicated player strategies — in such cases upper-case letters appear in the margins.

B (where Column has the first move) leads to a different PDRE — the path $\{c_1 \rightarrow r_1\}$ with payoffs (2,4).

[Figure 1 about here]

C. CRITIQUE OF CERTAIN SOLUTION CONCEPTS

When game theory is introduced in elementary textbooks, the story usually starts with the constant-sum game and two opponents who seem rather over-concerned with playing safe. The operative protocol is left vague, but the impression is conveyed that the players move simultaneously, i.e., in ignorance of the opponent's choice. The player on each side is supposed to be very preoccupied with his "security level" — the best he can achieve were his move to be discovered and optimally countered by the enemy. Thus the two contestants seem to behave like over-cautious military commanders, who always credit the enemy with the capacity of detecting and optimally responding to any chosen move. Such "maximin" reasoning leads, in the constant-sum Matrix 2 below, to respective choices of execution moves r_1 and c_1. So the maximin strategy-pair is $[R_1,C_1]$, with payoffs (2,3) to the associated execution moves.[9] Since each player here exactly achieves his security level, the solution is a "saddle point."[10]

---

fn. 8 cont. ...For some very simple protocols, such as single-round simultaneous play, the strategies on each side reduce down to choice among execution moves, and so we can equivalently use either upper-case or lower-case letters in the margins of the normal-form matrix.

[9]Here, by an obvious choice of notation made possible by the simplicity of the situation, Row's strategy R_i is identified with his move r_i and similarly Column's C_j is equivalent to his making move c_j.

[10]Not all constant-sum matrices possess saddle-points in pure strategies, but I will not be pursuing that point here.

Matrix 1

(see Figure 1)

	c_1	c_2
r_1	*2,4	4,1
r_2	1,2	3,3

Matrix 2

CONSTANT-SUM (with saddle-point in pure strategies)

	c_1	c_2
r_1	*2,3	3,2
r_2	1,4	4,1

Matrix 3

SILVER RULE

	c_1 (Good)	c_2 (Evil)
r_1(Good)	*4,4	1,3
r_2(Evil)	3,1	*2,2

Matrix 3A

SILVER-RULE EXPANDED (strategies for 1-round alternation protocol, Row moving first)†

	C_1	C_2	C_3	C_4
R_1	*4,4	1,3	*4,4	1,3
R_2	3,1	*2,2	2,2	3,1

†<u>Moves and strategies for Matrix 3A</u>

R_1 = Play r_1

R_2 = Play r_2

C_1 = Play c_1

C_2 = Play c_2

C_3 = Play c_1 if r_1
c_2 if r_2

C_4 = Play c_1 if r_2,
c_2 if r_1

While the maximin solution concept may possibly be defended on other grounds, the protocol implied by the introductory textbook tale is logically inconsistent. The generals cannot both be in the position of having their execution moves subject to optimal refutation by the enemy. In a 1-round alternating-move protocol, for example, only the first-mover needs to be concerned about the enemy's response to his move; in a simultaneous-move protocol neither player's move would be vulnerable to refutation. (Each general may believe that the other has such a capability, but if so at least one of them lacks adequate "knowledge of the game" — which would violate solution principle #3 above. The same holds for the more reasonable assumption that each commander believes only that there is a certain risk of his move being found out.)

Where the maximin principle implies irrationally over-cautious decisionmakers, just the opposite holds in the usual story presented to justify the equilibrium principle most commonly employed for non-constant-sum games -- the "Nash equilibrium" (NE) or "equilibrium point." An outcome is said to be an NE if, given the chosen strategy of the opponent, it does not pay either player to unilaterally modify his choice. Using Matrix 1 for illustration, this reasoning suggests that the (2,4) payoff vector corresponding to the $[R_1,C_1]$ strategy-pair would be the sole NE. (Here and elsewhere the NE's are indicated by asterisks.) Once again the specifics of the protocol are left vague, but it is argued that since at the NE Row's payoff is the highest in its column it does not pay him to switch, and similarly for Column since his payoff at the NE is the highest in its row. Whereas the maximin textbook story has each player thinking that his opponent has the last move, in the Nash textbook story each player in effect believes that he himself has the last move. Instead of over-cautious generals we have overconfident ones, each "myopically" sure that the enemy will not respond to his own chosen move.[11]

However, I must hasten to add, while social scientists who use game theory frequently do interpret the NE as just such a "myopic" solution, a skilled practitioner of the art would not be led astray. The Nash equilibrium should properly be defined in terms of strategies, which only in very simple protocols reduce down simply to execution moves on each side. An expert game theorist would therefore carefully examine the game-tree or extensive form (i.e., he would consider the protocol) operative in any actual situation, rather than draw conclusions simply from the elementary payoff matrix showing the returns to the execution moves alone.

When defined in terms of the full set of strategies permitted by the protocol, the Nash equilibrium solution concept cannot validly be criticized as "myopic."[12] It is possible to show the payoffs to the entire strategy set via a suitably expanded normal-form matrix (as will be illustrated below). The Nash equilibrium or equilibria in strategies can then be located by inspection.

However, some substantial difficulties with the NE still remain and are not so easily disposed of. First, while the true PDRE — the equilibrium under rational play, given the operative protocol -- will always be among the Nash equilibria turned up by this "expanded-strategy" process, there will

---

[11]For the constant-sum game it is a well-known result that the saddle-point and Nash equilibrium (in pure or mixed strategies) coincide. This is regarded by a number of analysts -- e.g., Shubik [1982, p. 221] and Zagare [1984, pp. 26-27] -- as strong evidence in favor of that solution. I cannot agree. Two invalid arguments, each based upon a logically inconsistent protocol, carry no more weight than one.

[12]The very valuable paper by Brams and Wittman [1981] is therefore somewhat unfair in attacking the Nash equilibrium as "myopic."

generally also be false solutions that have to be filtered out. Unfortunately, it is easy to be careless on this score and mistakenly believe that all NE's are of equal validity. Second, there may not be a valid informational basis for optimizing in response to the opponent's "given" strategy, which (depending upon the protocol) may or may not be visible to the opponent. These points will be illustrated by example.

Consider Matrix 3 which represents the payoff environment (to execution moves) that I will call SILVER RULE. The idea is that each player prefers to respond Good for Good, but under less pleasant circumstances would want to return Evil for Evil. Naive inspection of Matrix 3 suggests two seeming Nash equilibria: $[r_1,c_1]$ yielding both sides their ideal payoffs (4,4), but also the Pareto-inferior $[r_2,c_2]$ yielding them only (2,2). (There is also a third mixed-strategy NE which does not concern us.) Such a game matrix has been employed to picture, for example, one possible cause of war: each nation may behave aggressively because, once caught in the (2,2) NE, either side can only lose by changing unilaterally from Evil to Good.

A more careful analysis requires that the protocol be spelled out before any such conclusion is drawn. Specifically here, suppose the operative protocol provides for simple one-round alternation: starting from a clean-slate status quo, first Row moves, then Column, after which the game ends. (This protocol could of course be pictured as before in game-tree or extensive form, but doing so seems hardly necessary in such a simple case.) Evidently, with two rational (and common-knowledge) players the equilibrium path will be $\{r_1 \rightarrow c_1\}$ leading to the ideal payoffs (4,4). Row as first-mover can confidently choose Good, knowing that Column will then be motivated to respond in kind. Thus in Matrix 3 one of the seeming NE's has been determined to be the correct PDRE outcome under this protocol, while the other -- associated with the payoff-pair (2,2) -- has been shown to be incorrect.

Matrix 3A is the normal-form matrix in a suitably enlarged strategy space. Under the specified protocol Row's strategies remain the same as his execution moves, but Column now can adopt additional strategies which are contingent upon his opponent's prior move. More specifically, Row's R_1 is his simple execution move r_1 (Good) while his R_2 remains r_2 (Evil). But four distinct strategies are now available to Column, as follows:

C_1: Always play Good (execution move c_1)

C_2: Always play Evil (execution move c_2)

C_3: Answer Good with Good, and Evil with Evil

C_4: Answer Good with Evil, and Evil with Good

Of course only C_3 here is rational for Column. But in Matrix 3A we now see three pure-strategy NE's (once again indicated by the asterisks). The correct PDRE, which here appears as the strategy-pair $[R_1,C_3]$ with payoffs (4,4), is indeed among them. But two incorrect NE's have to be filtered out. Thus, the "sophistication" that takes the form of expanding the elementary matrix to allow for contingent strategies has not really helped. Filtering is still needed, and can only be done by bringing protocol considerations to bear.[13] Employing the PDRE in the first place would have precluded an irrelevant detour through the inadequate Nash equilibrium concept.

---

[13]The treatise by Van Damme [1983] is concerned to "refine" the Nash equilibrium concept, in order to filter out all but what he calls the "sensible" solutions. His underlying idea involves the same principle, of common-knowledge rational forethought, as postulated here. However, Van Damme's description of that principle seems somewhat flawed. His "sensible" NE's are strategy-pairs that would constitute "self-enforcing agreements" -- suggesting, what is not actually required, that the protocol has to allow negotiation or agreement-making moves prior to the execution phase.

The second and still more serious difficulty with the NE solution concept is that it implies -- inconsistent with assumption #4 above — that the players have a "telepathic" ability to read their opponents' strategy choices, when in general only the actual moves on each side are visible.

Consider once again Matrix 3A, the expanded version of SILVER RULE that applies when Row has the first move and Column the last move. Let us suppose that Row has chosen r_2 (Evil), followed by Column's choice of c_2 (Evil) as well. The game has ended, with payoffs (2,2). Now, we ask whether Row (if somehow given another chance to begin again and replay the game) would want to unilaterally revise his strategy. If he is already in an NE, he would not want to revise — that is the definition of a Nash equilibrium in strategies. But in Matrix 3A we can see that Row cannot, from having observed Column's move alone, know whether or not the outcome reached is an NE! For, from knowledge of Column's move alone, Row cannot tell which of C_2 or C_3 in Matrix 3A was Column's strategy. If Column's move c_2 was associated with strategy C_2, i.e., simply play Evil, then indeed the attained (2,2) payoff corresponds to an NE and Row should not shift. But if Column's strategy that led to move c_2 was really C_3, i.e., return Good for Good but Evil for Evil, then Row's shift would improve his payoff — i.e., the existing position is not an NE.

Thus, since a player with knowledge only of his opponent's moves cannot in general determine the opposing strategy, even after play is completed, he cannot know for sure whether or not it would pay him to revise his choice. But the criterion for Nash equilibrium is that neither player would want to unilaterally revise his own strategy choice. In sum, the so-called Nash equilibrium cannot be a valid equilibrium concept, since the players do not in general have an informational basis for knowing whether they are in "equilibrium" or not.

D. AN APPLICATION: FIRST MOVE VS. LAST MOVE

In this and the next section I shift to a more constructive mode in order to illustrate the utility of the payoff vs. protocol distinction. One question to which that distinction naturally leads is the following: Given some fixed protocol of play, how do the PDRE outcomes vary when we consider the range of possible payoff environments? Such a question will be considered in this section. In the section following, the opposite kind of issue will be addressed: Given a particular fixed payoff environment, how do different possible protocols affect the PDRE solution?

Many aspects of the solution arrived at might be of interest. Two main questions will be considered here: (1) Is the outcome Pareto-efficient? And, (2) Does one player or the other, Row or Column, have the advantage? For each question, of course, our main theme is that the answer in general will depend upon both payoff and protocol considerations.

Here a single-round alternating-move protocol will be assumed: starting from a clean slate first one player chooses, then the other, and the game ends. The range of payoff environments to be considered is the set of 2x2 matrices with ordinally ranked payoffs. For purposes of ordinal comparisons, and if ties are ruled out, the numbers appearing in the cells of the matrix for each player can be ranked 4,3,2,1 -- higher numbers corresponding to greater payoffs. There are 576 such matrices. However, Rapoport and Guyer (1966) show that many of these can be derived from others by inessential transformations, to wit: (i) re-labelling of rows, (ii) re-labelling of columns, or (iii) re-naming of players. (Or any combination of the above.) After elimination of inessentially different variants, only 78 matrices remain to represent all the possible qualitatively different payoff environments in

2x2 games, ties excluded. (The 78 such matrices are conveniently tabulated in the paper by Rapoport and Guyer.)

Let us start with the question of relative advantage -- who does better, first-mover or last-mover? In order to isolate this effect of priority, a protocol consideration, we need to set aside aspects of the payoff environment that may give one of the players the advantage whether or not he has the first move. One way of isolating the priority effect is to narrow the field of attention to the subset of the ordinally ranked 2x2 matrices that are symmetrical in terms of payoffs. The symmetrical matrices are those that "look the same" whether regarded from Row's or Column's point of view. I.e., an interchange of the Row and Column players would leave each with the same strategy choices and associated payoffs as before.

Of the 78 qualitatively distinct 2x2 matrices, exactly 12 are symmetrical in terms of payoffs. Of these, however, 6 are "uninteresting" in that they contain a (4,4) payoff element. Obviously, in a single-round alternating-move protocol the pair of execution moves leading to such a mutually preferred payoff-pair will always be achieved as the PDRE.[14] The remaining, more "interesting" payoff environments are tabulated below as Matrices D.1 through D.6.[15] (In each case the PDRE solution, for the case where Row has the first move, is indicated by the symbol #.)

Turning back for a moment to our first question, in the 6 "uninteresting" matrices not tabulated the PDRE represented by the (4,4) payoff-pair is of course Pareto-efficient. Inspection of the 6 tabulated matrices reveals that

---

[14]The (4,4) outcome will always be a Nash equilibrium as well, but as we have seen it may not be the sole NE.

[15]For a discussion of these matrices from another point of view see Rapoport (1967).

Matrix D.1

PRISONERS' DILEMMA

	c_1	c_2
r_1	3,3	1,4
r_2	4,1	2,2#

Matrix D.2

NON-DILEMMA

	c_1	c_2
r_1	3,3#	4,1
r_2	1,4	2,2

Matrix D.3

CHICKEN

	c_1	c_2
r_1	3,3	2,4
r_2	4,2#	1,1

Matrix D.4

ANTI-CHICKEN

	c_1	c_2
r_1	3,3#	4,2
r_2	2,4	1,1

Matrix D.5

DO IT MY WAY (I)

	c_1	c_2
r_1	2,2	4,3#
r_2	3,4	1,1

Matrix D.6

DO IT MY WAY (II)

	c_1	c_2
r_1	2,2	3,4
r_2	4,3#	1,1

Pareto-efficiency fails in only one case -- Matrix D.1 (PRISONERS' DILEMMA). Returning to the question of first-mover or last-mover advantage, the result is of course neutral when the (4,4) outcome is achieved in the 6 "uninteresting" matrices. As for the 6 "interesting" cases, exactly 3 are similarly neutral while the other 3 all give the advantage to the first-mover (Row as assumed here). Thus the possibly surprising result: In the symmetrical 2x2 game with strictly ordered payoffs, under the alternating-move protocol it is never advantageous to have the last move.

Recall that our ultimate interest lies in using game theory to model social outcomes and relationships. As an evident application, the question of first-move versus last-move suggests the traditional problem of who has the advantage in war (or conflict more generally) -- the offense or the defense? We would certainly be incorrect to assert that the defense (last-mover) never has the advantage in war! So the problem is to explain the seeming discrepancy between theory and observation.

First-move versus last-move advantage is connected with the distinction between the conflictual and the cooperative aspects of mixed-motive games. The two aspects can be visualized if the payoff-pairs for any game are plotted as in Figure 2 on P_R (payoff to Row) and P_C (payoff to Column) axes. Then, following a lead suggested by Snyder and Diesing (1977), the cooperative aspect is suggestively measured by the distance within the bounded region along or parallel to the main diagonal, and the conflictual element by the corresponding distance along or parallel to the secondary diagonal.[16] In Figure 2 the payoff environment represented by Prisoners' Dilemma (top

[16] We are shifting ground here from an ordinal to a ordinal interpretation of the tabulated payoffs.

diagram) is thus seen to be relatively more conflictual than the environment corresponding to Matrix D.5, termed here DO IT MY WAY (I)[17] (bottom diagram).

[Fig. 2 about here.]

Where the players' interests are more allied than in conflict, as in the two DO IT MY WAY matrices, the first-mover will often be able to capture more of the mutual gain. Having the first move, he can place the second-mover in a position where in trying to help himself the latter must automatically help his opponent even more.[18] Where the interests are sharply opposed, however, the last-mover in helping himself will tend to injure his opponent, hence the first-mover may be forced to go for a safer but less profitable strategy option. An obvious quantitative measure of the association of interests is the simple correlation coefficient. The argument preceding suggests that first-move advantage is more likely where the correlation of payoffs is high. Table 1 confirms this suggestion for the set of 6 "interesting" symmetrical 2x2 matrices considered here.

Table 1

Correlation of payoffs and first-move advantage

	Correlation Between Row and Column Payoffs	Instances of first-move advantage
Matrices D.1 and D.2	-.8	0 of 2
Matrices D.3 and D.4	+.2	1 of 2
Matrices D.5 and D.6	+.6	2 of 2

[17]The idea is, "Be reasonable -- we can both benefit if you do it my way." (Of course, "doing it my way" is relatively better for the speaker than for the person addressed.) Matrix D.6, DO IT MY WAY (II), represents a qualitatively similar though logically distinct game. Rapoport (1967) proposes the names Hero for Matrix D.5 and Leader for Matrix D.6, but descriptive titles do not seem very apt.

[18]On this see also Schelling [1960], p. 143.

But it is natural to ask why it is that, in the very similar matrices D.3 and D.4 (CHICKEN and what I have whimsically termed ANTI-CHICKEN)[19] which have the same correlation coefficient, the former is associated with a strong first-move advantage and the latter is not? The explanation is that, even though the payoff-pairs are identical, the available strategies link them in different ways. In Matrix D.4 (ANTI-CHICKEN) Row's execution move r_1 yields him (3,-) or (4,-) depending upon Column's move, while r_2 yields him (2,-) or (1,-). Thus, r_1 is completely dominant for Row, while analogously c_1 is completely dominant for Column.[20] This symmetrical pattern makes the $[r_1,c_1]$ strategy-pair the PDRE regardless of who moves first. In Matrix D.3 (CHICKEN), on the other hand, symmetry in this respect is lacking: there is no dominance. Here Row as first-mover can force (4,3) -- since second-mover's alternative strategy leads to the mutually undesired (1,1) -- and similarly Column as first-mover can force (3,4).

Although we see from the above that correlation of payoffs is not the sole factor at work, it is a major element in the determination of whether or not first-move advantage exists. Thus to find an instance of <u>second</u>-move advantage we should look for even stronger negative correlation of returns than in any of the matrices in Table 1 -- approaching the constant-sum condition in the limit. However, among the set of 12 symmetrical 2x2 matrices there are no constant-sum cases if payoff ties are excluded. In order to find symmetrical matrices that are constant-sum, we would have to allow for ties --

---

[19]Rapoport (1967) offers the name Exploiter for CHICKEN, but leaves ANTI-CHICKEN without a title.

[20]The concept of "complete dominance" employed here is stronger than ordinary dominance, because the player can see which strategy is better for him without even looking at his opponent's strategy options.

that is, for matrices with only three or even only two distinct payoff levels for each player.

A detailed investigation would lead into a number of issues that cannot be pursued here, but the following two-level constant-sum symmetrical matrix displays the possibility of a last-move advantage:

Matrix D.7

LAND OR SEA

	C_1	C_2
R_1	1,2#	2,1
R_2	2,1	1,2#

In the LAND OR SEA game, whether Row attacks by land or attacks by sea, Column as second-mover can defend in a corresponding way so as to refute his opponent's choice. The informational asymmetry in favor of the second-mover is of course one of the great advantages of having the defensive in warfare.

The information asymmetry that tends to favor last-mover can be made even stronger if we depart from our previously maintained protocol assumptions. In particular, we might very plausibly suppose (instead of our previous full-knowledge assumption) that each player know his own but not his opponent's payoff elements. Then, it will be evident, being able to see the enemy's chosen move before making one's own selection is a weighty advantage. Whether it is this informational factor, or the near-constant-sum aspect, that sometimes tells so heavily in favor of the defense in war is an interesting issue for further investigation.

E. SECOND APPLICATION: PRISONERS' DILEMMA UNDER DIFFERENT PROTOCOLS

The previous section examined aspects of the set of PDRE solutions attained under a single-round alternating-move protocol, over a range of alternative payoff environments. This example reverses the emphasis. Only a single payoff environment will be considered -- Prisoners' Dilemma (Matrix E.1). A range of alternative protocols will be examined, aimed at the question of whether the famous Pareto-inferior "trap" outcome, in which both sides play DEFECT with payoffs (2,2), can ever be escaped. (And, if so, how this corresponds to observable social institutions.) We will be dealing throughout as before only with the single-play game — which excludes, for example, possible escapes based upon "Tit for Tat"[21] or similar strategies which become possible in multiple-play games. We will however be considering protocols permitting multiple rounds of negotiation moves or "tries"; nevertheless, the negotiation-process must end with a single execution move on each side that determines the payoff-pair attained in the elementary Prisoners' Dilemma matrix.

Matrix E.1

PRISONERS' DILEMMA

	Loyal	Defect
Loyal	3,3	1,4
Defect	4,1	2,2

By an argument too familiar to require extensive justification here, any finite-termination protocol in Prisoners' Dilemms will inevitably lead into the (2,2) trap. If the ultimate execution moves are to be made in alternation, the party having the last execution move will surely play DEFECT. Knowing this, his opponent would rationally play DEFECT in his previous execution turn. Or, if the ultimate execution moves are to be made simultaneously,

[21] See Axelrod and Hamilton [1981].

DEFECT is dominant over LOYAL for each player and thus again would rationally be chosen by each. Thus, the only hope for escape from the trap in the single-play Prisoners' Dilemma is a termination rule that does not provide for a fixed number of negotiation rounds (tries). The most interesting possibility is the rule called "natural" termination above.

A negotiation process "naturally" terminates only when each player, though he has the option to change his previous try, chooses not to do so. It turns out that the PDRE solutions under natural termination differ importantly in the alternating-move versus the simultaneous-move cases.

Alternating moves

In an alternating-move game, once the negotiation process -- the haggling, we might say -- is well under way the game naturally terminates as soon as any player passes. The start-up condition also has to be specified, however. If the game starts with a clean-slate status quo, a pass in the first round is not meaningful -- to initialize play, each party has first to choose a row or column, as the case may be. If on the other hand the status quo is some particular cell of the matrix (an exogenously specified initial strategy-pair), the first-mover could pass. But, consistent with the previous discussion, the game will not be considered to have ended until the second-trier also has a chance to make a choice of move. Summing up: "natural" termination of the alternating-move game occurs when either party passes, once both players have had at least one try.

Figure 3 pictures the considerations bearing upon equilibrium under "natural" termination of the alternating-move Prisoners' Dilemma. The four cells of the underlying elementary Matrix E.1 are laid out here in diamond form. In the upper sketch Row's possible switches are indicated by the (dashed) lines in the NE <--> SW directions, while Column's switches are shown by the (solid) lines in the NW <--> SE directions.

[Figure 3 about here.]

Consider now Column's (solid) line labelled #1. It represents a transition from the payoff-pair (1,4) to the pair (3,3). Clearly, Column would prefer to pass-and-terminate rather than make this switch. This transition is therefore immediately ruled out, by a first-level analysis, and has accordingly been deleted from the lower diagram. And the same applies, by analogous reasoning, to Row's (dashed) #1 line.

Now consider Row's (dashed) #2 line, representing a transition from (2,2) to (1,4). Since as just argued Column would terminate if given the move at (1,4), Row knows that if he (Row) has the move at (2,2) he had better pass-and-terminate right off. Thus, a second-level analysis rules out this switch. Accordingly, Row's #2 line has also been deleted from the lower diagram, together with the corresponding (solid) #2 line for Column.

Finally, consider the #3 lines representing switches away from the efficient (3,3) outcome. By an evident process of third-level reasoning, once again either player having the move would do better to pass-and-terminate. Take Row. A switch by Row along his (solid) line #3 would lead to outcome (4,1) -- yielding a seeming profit. But this gain is illusory. Column in turn would surely then switch (solid un-numbered line) to (2,2). Since, as just seen, it does not pay either player to switch away from (2,2), failure to terminate at (3,3) will ultimately make the switcher (and the other party as well, of course) worse off. So the #3 lines have also been deleted in the lower diagram.

We end up with a lower diagram that can be interpreted as follows. Once the haggling is well underway, the (3,3) outcome is "retentive" for both parties but not "attractive".[22] The asymmetrical (4,1) is retentive only for Row, while (1,4) is retentive only for Column, and once again neither of these is attractive. Thus, so far everything still points to termination at (2,2) -- the Prisoners' Dilemma "trap".

However, we have not yet taken the start-up condition into account. Doing so, we find that <u>if (3,3) were the initial status quo</u> then neither player would have any incentive to switch away from it. For, as we have just seen, the efficient (3,3) outcome <u>is</u> retentive if the players ever find themselves there.[23]

What about the clean-slate status quo? Here the result is quite different. Each player will realize that, should either diverge on the start-up round from the strategy consistent with the efficient (3,3) outcome, the result will inevitably be the (2,2) trap. In particular, if the trap is to be avoided the <u>first</u>-mover (say, Row) must choose a try consistent with (3,3). Can the second-mover profitably exploit such "good behavior"? The answer is no. Suppose Row initially choose LOYAL and Column greedily responded with DEFECT, leading to the (1,4) position. Row can still profitably punish Column by switching into the trap. So Row <u>will</u> choose LOYAL, and Column will respond with LOYAL. Thus the efficient (3,3) solution is "attractive" after all in at least one case: starting up from a clean slate. (In the lower diagram, this is indicated by the arrow pointing toward (3,3) from above.) Summing up: under "natural" termination, in the alternating-move game there are two circumstances that permit escape from the trap in the Prisoners' Dilemma:

---

[22]The terminology is borrowed from Fiorina and Shepsle (1982).

[23]The efficient (3,3) outcome of the Prisoners' Dilemma is therefore a "reactive equilibrium" in the sense of Riley [1979].

(i) if the status quo condition has the players already at the efficient solution, or (ii) as a case with wider applicability, if the start-up is from a clean-slate status quo.

Simultaneous moves

If moves are simultaneous, under natural termination a non-pass try is always revocable. (There will always be another round of play unless both parties have passed.) This consideration makes it possible for a player to "take a chance" in trying for the efficient outcome of the Prisoners' Dilemma.

Suppose we start with the clean-slate status quo, so that passing in the first round is not allowed. Should Row try LOYAL but Column answer with DEFECT, Row can count on being able to switch on the next move. So there is no reason for Row not to try Loyal, and by symmetrical reasoning Column should do the same -- hence the efficient (3,3) outcome can indeed be achieved as the PDRE. (Row has no hope of ever achieving (4,1) since Column would surely switch on his next try, and similarly Column has no hope of ever achieving (1,4)).

Under this protocol, should the players ever find themselves away from the (3,3) position -- whether as an initial status quo, or as the consequence of mistaken play earlier on -- they can easily locate a path back to the efficient outcome. From the (2,2) "trap" position, for example, as just explained each should attempt to switch toward the efficient outcome -- and, if they each make the attempt, they will succeed. And of course either of the asymmetrical outcomes will always cause the aggrieved player to switch. So in general for the natural-termination, simultaneous-move game the Prisoners' Dilemma "trap" is ineffective; the efficient outcome will be achieved as the PDRE.

As an overall conclusion, then, consideration of possible alternative protocols reveals that the Prisoners' Dilemma trap is by no means such an irretrievable "black hole" outcome as is often painted. In some protocols the trap is indeed retentive, in others both retentive and attractive, but there are important cases in which not the trap but the efficient solution will be the PDRE.[24]

One instance that comes readily to mind is the problem of public goods. Self-interested agents who can mutually gain from joint contributions to the provision of public goods find themselves in a Prisoners' Dilemma — it pays each separately to contribute little or nothing, regardless of whether others contribute or not. Yet it is a notorious fact that individuals do far more in the way of voluntary private provision of public goods than standard economic theories allow for. Various more or less cogent explanations have been offered: for example, Margolis (1982) postulates a utility function containing an "altruistic" component, and Hirshleifer (1983) points to a class of public goods where each individual can plausibly regard himself as the "weakest link" in the chain of social provision (and therefore is inhibited from shirking since he himself will directly suffer from doing so). The discussion here suggests another possibility. To wit, that public goods ae often provided by a societal negotiation process that corresponds, in effect, to a protocol characterized by "natural" termination. As seen above, where one's "generous" try is not finally binding unless the other player or players respond appropriately, the individual's risk of being exploited largely or completely disappears.

The same argument can be put in a normative rather than positive way. If individual members of a society find themselves in a Prisoners' Dilemma trap,

---

[24]The conclusion here is entirely consistent with that in Brams and Wittman (1981).

they may often be enabled to escape it by agreeing upon a different protocol -- e.g., one providing for natural termination rather than a fixed number of rounds of negotiation moves.

Consider the arrangement called "silent trade" or "dumb barter" among primitive peoples:

> ...one party goes to the customary spot, lays down goods and retires....The other people then come, lay down what they consider to be articles of equivalent worth, and retreat in their turn. The first party then comes back and if satisfied with the bargain removes the newly-deposited goods; if not, these are allowed to remain until suitable additions are made. The people of the second party then take away the original wares and the transaction is concluded. (Firth [1947])

Assuming that there are effective sanctions against outright theft, this interaction can be regarded as a Prisoners'Dilemma environment where LOYAL = adequate reciprocation and DEFECT = inadequate. Clearly, a single round of tries would not escape the trap, whereas the alternating-move, natural-termination protocol described above evidently can do so. Furthermore, this protocol can be regarded as a paradigm of the near-universal process of bargaining. The point being that, since the transaction will not be executed until both parties are satisfied, the possibilities for mutual gain can be explored with little or no risk of being victimized by one's trading partner.

F. SUMMARY AND CONCLUDING COMMENTS

In applying game theory, the intention is to model patterns of social interaction. I argue here that it is programmatically fruitful, in attempting to achieve this end, to maintain a distinction between the exogenously given payoff environment that faces the players versus the protocol of play which corresponds to the endogenous institutional decision-making practices of the group. Thinking always in terms of both payoff and protocol leads, for example, in the single-play (non-repeated) game to an analytical distinction between the ultimate "execution moves" on the one hand, versus the preliminary "negotiation moves" whose precise form must also be dictated by the governing protocol.

The first element, the payoff environment, is best pictured by the elementary normal-form matrix defined in terms of the "execution moves" of the players; the second element, the protocol, is usually represented (in part at least) as the extensive form or game-tree. However, in current expositions of game theory there has been little or no analysis of how protocols may differ, so that the consequences of alternative "architectures" of the game-tree have not been systematically studied. It is shown here, in particular, that among the dimensions across which protocols may significantly vary are: (1) the status quo position, (2) the sequence of turns, (3) possible ordering rules for allowable moves, and (4) the procedure for termination.

The analysis reveals that, once we attempt to be rigorous about the implicit protocol of play, certain traditional equilibrium concepts much used in game theory become subject to serious question. In particular, "maximin" or "security level" reasoning as commonly employed for the constant-sum game is premised upon an internally inconsistent protocol -- in which, essentially, each player assumes that his opponent will have the last move. The "Nash equilibrium" (NE) concept widely used in the analysis of non-constant-sum games has been criticized for the opposite defect. For, in current usage what is described as the NE is often a "myopic" solution in which each player in effect acts as if he himself has the last move, i.e., as if his opponent will foolishly fail to respond to his move. However, this alleged "myopia" is not really a flaw of the underlying concept. Properly speaking, the Nash equilibrium ranges over all possible strategies permitted by the governing protocol, and in particular covers the complex or "expanded" strategies that do allow for the opponent's responses.

However, it remains a valid criticism of the Nash equilibrium concept that situations involving such complex strategies typically lead to non-unique NE's, of which the correct one or ones can only be located by attending explicitly to the operative protocol. Also, and even more importantly, since a player is normally only in a position to observe the opponent's moves (from which he cannot, in general, infer the underlying strategy), he will not ordinarily ever be in a position to know whether the current achieved outcome is an NE or not — and hence will not be able to decide whether a unilateral shift of strategy would or would not be profitable. So the very condition defining the Nash equilibrium, that neither player would want to revise his chosen strategy, is not in general within the power of the players to determine — once the informational aspects of the governing protocol are considered.

In the constructive portion of the paper the solution concept used is the so-called Subgame-Perfect Equilibrium (SGPE); in this particular context, in order to emphasize the dependence of the outcome attained upon the details of the protocol, the term Protocol-Dependent Rational Equilibrium (PDRE) was employed. This concept escapes the difficulty with the Nash equilibrium since, using the assumption of common-knowledge mutual rationality, the players can validly infer the strategy that the opponent is playing.

In the first applied example, the consequences of one particular protocol -- simple one-round sequential play -- were examined over a range of alternative payoff environments. The protocol employed was simple one-round sequential choice of move. The alternative payoff environments considered were the 12 possible symmetrical 2x2 ordinally ranked matrices. The questions examined concerned Pareto-efficiency and first-move vs. last-move advantage.

It was determined that 11 of the 12 PDRE outcomes are Pareto-efficient, the sole exception being Prisoners' Dilemma. For 9 of the 12 payoff environments the result is neutral as to first-move vs. last-move advantage, but for the other 3 the advantage is always to the first-mover. The priority effect favoring first-mover is somewhat counterbalanced by an informational effect favoring the last-mover, the information effect being relatively stronger the closer the payoffs approach the constant-sum condition. However, the overall advantage actually swings toward last-mover only for certain payoff environments (not among the 12 originally studied) in which outcome ties are permitted.

A second applied example reversed the emphasis, studying a single payoff environment -- Prisoners' Dilemma -- under a range of alternative protocols of play. It turns out that the famous Pareto-inefficient "trap" solution for Prisoners' Dilemma is by no means such an inescapable black hole as often assumed. The mutually preferred Pareto-efficient outcome would be achieved instead, it was shown, under a number of alternative protocols. In particular, under what is here called the "natural termination" procedure (where negotiations continue until each player is satisfied to execute his current try), the Pareto-efficient outcome is quite generally attained in the simultaneous-move game, and also in a number of cases in the alternating-move game. And indeed this is not surprising, since the natural-termination protocol can be regarded as a paradigm for the general process of bargaining through which individuals tentatively explore for mutually-preferred compromises.

REFERENCES

Aumann, Robert J., "Agreeing to Disagree," Annals of Statistics, v. 4 (1976), 1236-39.

Axelrod, Robert and William D. Hamilton, "The Evolution of Cooperation," Science, v. 211 (27 March 1981).

Bellman, Richard, Dynamic Programming (Princeton, NJ: Princeton U. Press, 1957).

Brams, Steven J. and Wittman, Donald, "Nonmyopic Equilibria in 2x2 Games," Conflict Management & Peace Science, v. 6 (Fall 1981).

Fiorina, Morris P. and Kenneth A. Shepsle, "Equilibrium, Disequilibrium, and the General Possibility of a Science of Politics," in Peter C. Ordeshook and Kenneth A. Shepsle, eds., Political Equilibrium (Martinus Nijhoff, 1982).

Firth, Raymond, "Trade, Primitive," in Encyclopedia Britannica (1947), v. 22, p. 345.

Hirshleifer, J., "From Weakest-link to Best-Shot: The Voluntary Provision of Public Goods," Public Choice, v. 41 (1983), pp. 371-86.

Margolis, Howard, Selfishness, Altruism, and Rationality (Cambridge: Cambridge U. Press, 1982).

Plott, Charles R., "Industrial Organization Theory and Experimental Economics," Journal of Economic Literature, v. 20 (Dec. 1982).

Rapoport, Anatol, Fights, Games, and Debates, (Ann Arbor, MI: U. of Michigan Press, 1960).

__________, "Exploiter, Leader, Hero, and Martyr: The Four Archetypes of the 2x2 Game," Behavioral Science, v. 12 (1967).

__________ and Melvin Guyer, "A Taxonomy of 2x2 Games," General Systems, v. 11 (1966).

Riley, John, "Informational Equilibrium," Econometrica, v. 47 (March 1979).

Schelling, Thomas C., The Strategy of Conflict (Oxford Univ. Press, 1960).

Selten, R., "Reexamination of the Prefectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, v. 4 (1975).

Shubik, Martin, Game Theory in the Social Sciences: Concepts and Solutions (Cambridge, MA: MIT Press, 1982).

Smith, Vernon L., "Microeconomic Systems as an Experimental Science," American Economic Review, v. 72 (Dec. 1982).

Snyder, Glenn H. and Paul Diesing, Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises, (Princeton, NJ: Princeton U. Press, 1977).

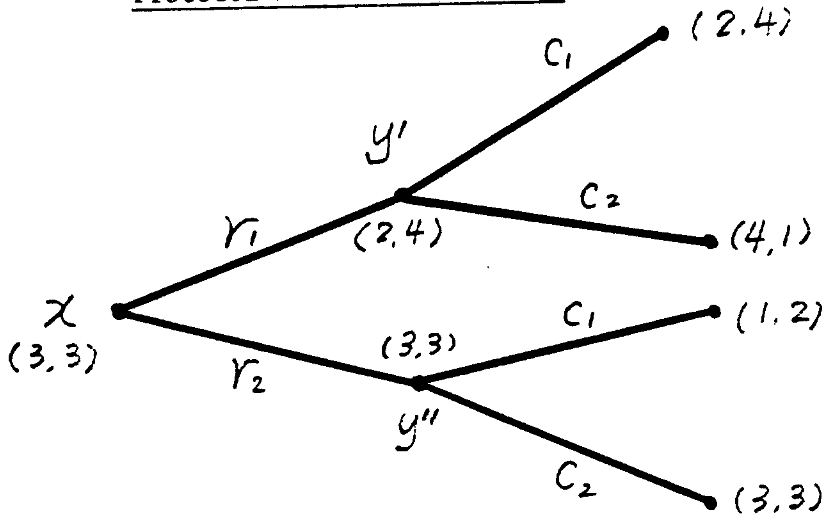
Van Damme, Eric E.C., Refinements of the Nash Equilibrium Concept, Ph.D. dissertation at Technische Hogeschool Eindhoven (1983).

Zagare, Frank C., Game Theory: Concepts and Applications (Beverly Hills, CA: Sage Publications, 1984).

Matrix 1

	c_1	c_2
r_1	*2,4	4,1
r_2	1,2	3,3

Protocol A: Row moves first



Protocol B: Column moves first

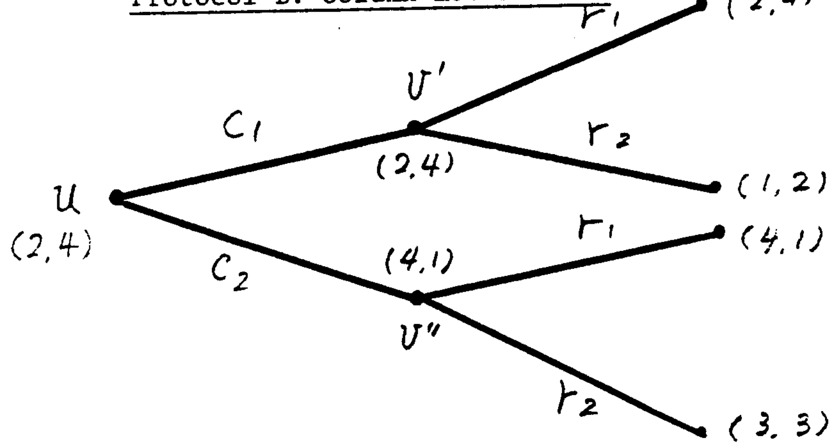
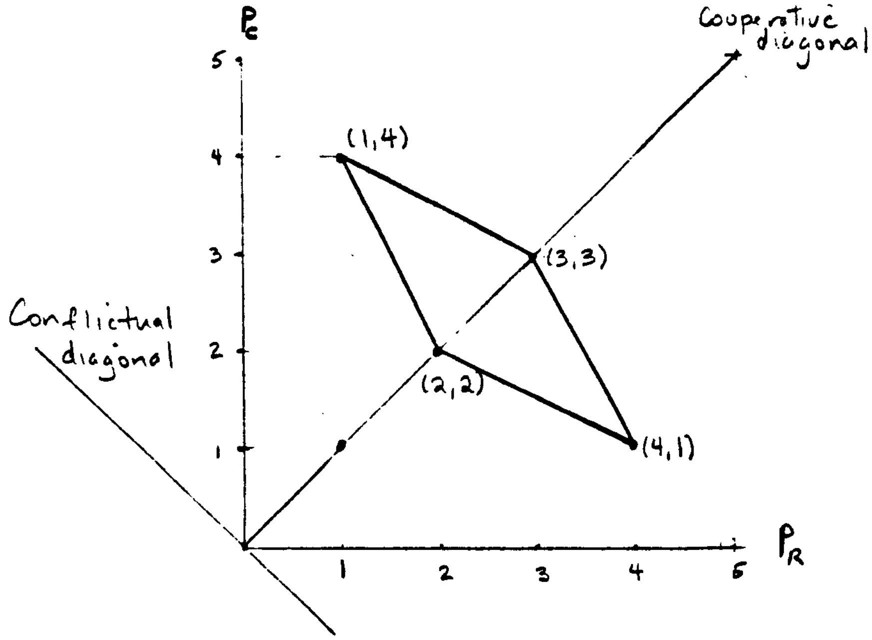


Figure 1: Alternative Single-round Protocols

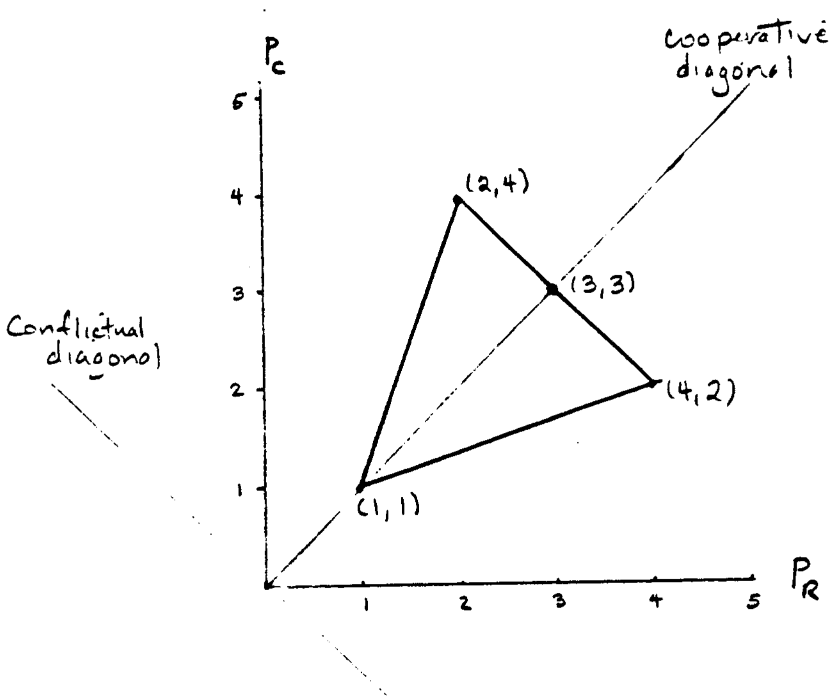
Figure 2: Conflictual versus cooperative aspects of mixed-motive payoffs.



Matrix D.1.

The Prisoners' Dilemma

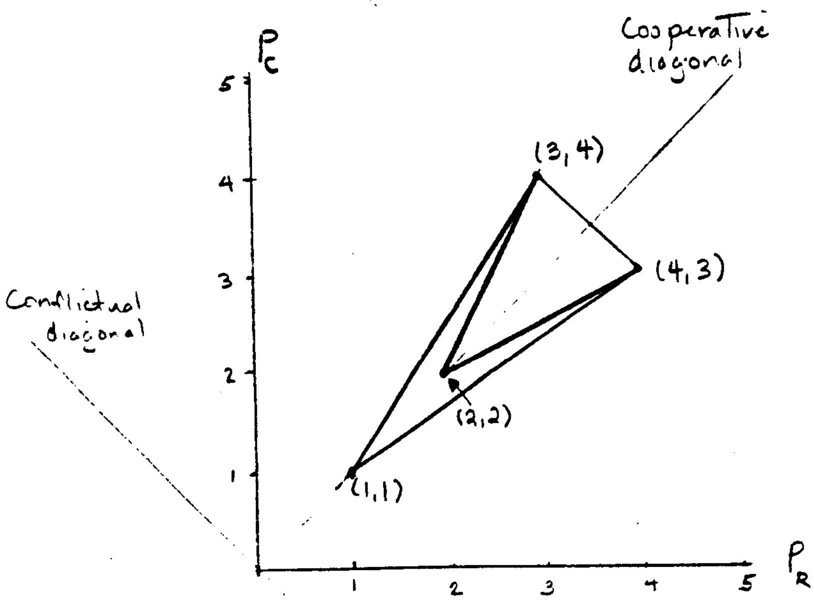
(3,3)	(1,4)
(4,1)	(2,2)



Matrix D.3.

Chicken

(3,3)	(2,4)
(4,2)	(1,1)



Matrix D.5.

Do It My Way (I)

(2,2)	(4,3)
(3,4)	(1,1)

PRISONERS' DILEMMA, ALTERNATING MOVES

(Natural termination, stability analysis)

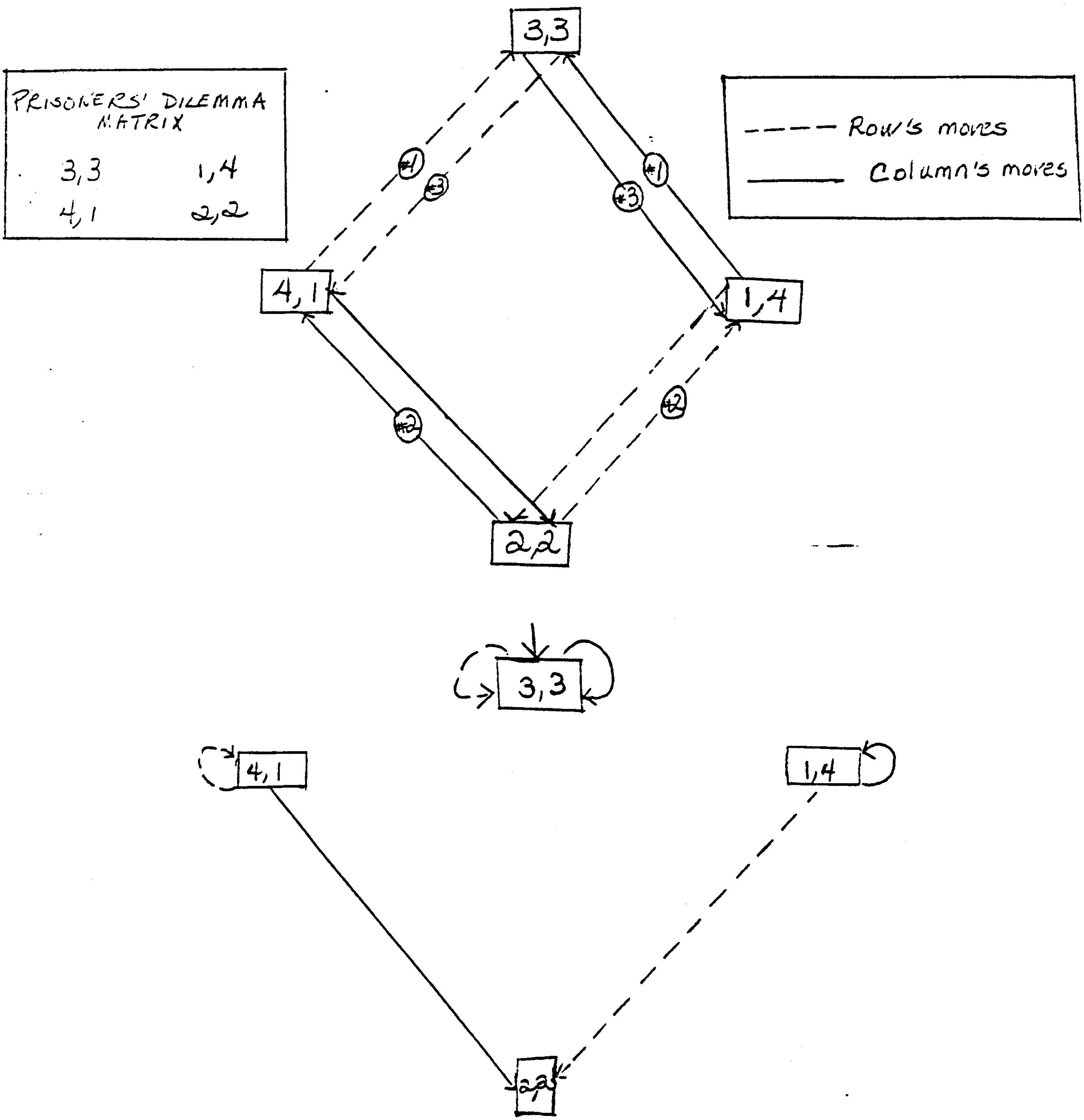


Figure 3: Stability analysis for Prisoners' Dilemma, alternating-move protocol with natural termination