Econometrica, Vol. 59, No. 4 (July, 1991), 909–924

COMMENTS ON THE INTERPRETATION OF GAME THEORY

BY ARIEL RUBINSTEIN[1]

The paper is a discussion of the interpretation of game theory. Game theory is viewed as an abstract inquiry into the concepts used in social reasoning when dealing with situations of conflict and not as an attempt to predict behavior. The first half of the paper deals with the notion of "strategy." Its principal claim is that the conventional interpretation of a "strategy" is not consistent with the manner in which it is applied, and that this inconsistency frequently results in confusion and misunderstanding. In order to prove this point, the term "strategy" is discussed in three contexts: extensive games in which players have to act more than once in some prespecified order, games normally analyzed using mixed strategies, and games with limited memory. The paper endorses the view that equilibrium strategy describes a player's plan of action, as well as those considerations which support the optimality of his plan rather than being merely a description of a "plan of action." Deviation from the perception of a strategy as a mere "plan of action" fits in well with the interpretation of the notion "game" which is discussed in the second half of this paper. It is argued that a good model in game theory has to be realistic in the sense that it provides a model for the *perception* of real life social phenomena. It should incorporate a description of the relevant factors involved, as perceived by the decision makers. These need not necessarily represent the physical rules of the world. It is not meant to be isomorphic with respect to "reality" but rather with respect to our perception of regular phenomena in reality.

KEYWORDS: Game theory, interpretation, strategy, game form, perception, relevance.

1. INTRODUCTION

I APPROACH THIS PAPER with the view that game theory is not simply a matter of abstract mathematics but concerns *the real world*. This does not mean that the object of game theory is to predict behavior in the same sense as the sciences do, or indeed, that it is capable of such a function. I view game theory as an analysis of the concepts used in social reasoning when dealing with situations of conflict. It is an abstract inquiry into the function and logic of social institutions and patterns of behavior. As game theory is at once abstract and concrete, we must build a bridge between the abstract formal concepts of the theory and reality. An interpretation is a mapping which links a formal theory with everyday language.

Despite the vital need to provide a plausible interpretation for the basic primitives of game theory, especially when game theory is applied to economics, I have found only a few discussions of this issue. Noteworthy exceptions are Aumann (1987a) and Binmore (1983, 1987b). I believe that discussion or application of game theory is utterly meaningless without a proper interpretation. This task cannot be left entirely to philosophers of science, for it consti-

[1] This paper is based on my Walras-Bowley Lecture delivered in the North American Econometric Society meeting at the University of Minnesota, Minneapolis, June 1988.

I am deeply grateful to my friends Dilip Abreu, Ken Binmore, John Moore, Martin Osborne, David Pearce, Avner Shaked, and Asher Wolinsky for their comments, suggestions, criticism, and most importantly their encouragement during the period in which this paper was written. Three referees provided exceptionally thorough comments on the first version of the paper.

tutes the very essence of the theory. In this paper I will bring together various comments on the interpretation of game theory.

How should we interpret game theory's two most basic primitives, "game form" and "strategy?" The standard interpretation of a game form is that it represents an exact and full description of the physical rules of a given of a situation. It includes a list of decision problems for each of the participants in the game. The information sets itemize the information available to the decision makers concerning random events and other players' previous moves. The game tree presents a chronology of potential events. The standard view is that this description is exhaustive. To quote Kohlberg and Mertens (1986, fn. 3): "We adhere to the classical point of view that the game under consideration fully describes the real situation—that any (pre)commitment possibilities, any repetitive aspect, any probabilities of error, or any possibility of jointly observing some random event, have already been modelled in the game tree." A strategy is usually interpreted as "a plan of action." M. Shubik refers to a strategy as "a complete description of how a player intends to play a game, from beginning to end." J. Friedman uses the phrase, "a set of instructions." These descriptions are linguistically consistent with the Oxford Dictionary, which defines the word "strategy" as "a general plan of action."

The first half of the paper deals with the notion of "strategy." Its principal claim is that the conventional interpretation of a "strategy" is not consistent with the manner in which it is applied, and that this inconsistency frequently results in confusion and misunderstanding. In order to prove this point, I discuss the use of the term "strategy" in three contexts: extensive games in which players have to act more than once in some prespecified order (Section 2), games normally analyzed using mixed strategies (Section 3), and games with limited memory (Section 4). I aim to endorse the view that equilibrium strategy describes a player's plan of action, as well as those considerations which support the optimality of his plan (i.e. preconceived ideas concerning the other players' plans) rather than being merely a description of a "plan of action."

Deviation from the perception of a strategy as a mere "plan of action" fits in well with the interpretation of the notion "game" which is discussed in the second half of this paper. It is argued that a good model in game theory has to be realistic in the sense that it provides a model for the *perception* of real life social phenomena (Section 5). It should incorporate a description of the relevant factors involved, as perceived by the decision makers. These need not necessarily represent the physical rules of the world (Section 6). It is not meant to be isomorphic with respect to "reality" but rather with respect to our perception of regular phenomena in reality (Section 7).

2. ON THE NOTION OF "STRATEGY" IN SEQUENTIAL GAMES

In an extensive game, a player's strategy is required to specify an action for each node in the game tree at which the player has to move. In other words, a player has to specify an action for every sequence of events which is consistent

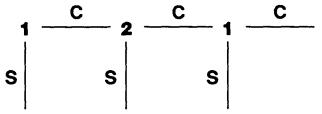


FIGURE 1

with the rules of the game. In games which require a player to make at least two consecutive moves (and most of the games which have been analyzed recently in economic theory fall into this category), a strategy must specify his actions even after histories which are inconsistent with the player's own strategy. For illustration, consider the two player game form shown in Figure 1. According to the natural definition of "strategy" as a complete "plan of action," player 1 is required to specify his behavior, "Continue" or "Stop," at the initial node, and if he plans to "Continue," to make provisional plans for his second decision node in the event that player 2 chooses *C*. However, the game theoretic definition of strategy requires player 1 to specify his action at the second decision node, even if he plans to "Stop" the game at the first node.

Why does the notion of strategy as used by game theorists differ from a "plan of action?"[2] Could we not narrow the formal definition of a "strategy" to specify an action only at the decision nodes which are not excluded from being reached by the strategy? If we were only investigating Nash equilibria of extensive games, then the game-theoretic definition would indeed be unnecessarily broad. The broad definition is, however, necessary for testing the rationality of a player's plan, both at the beginning of the game and at the point where he must consider the possibility of response to an opponent's potential deviation (the subgame perfect idea). Going back to the example of Figure 1, assume that each player plans to choose Stop at his first decision node. Testing the optimality of player 2's plan following player 1's deviation requires player 2 to specify his expectations regarding player 1's plan at his second decision node. The component of player 1's strategy after both players have chosen *C* provides these expectations. Player 1's strategy at his second decision node must be interpreted as what would be player 2's (as opposed to player 1's) *belief* regarding player 1's planned future play, should player 1 decide to deviate from what was believed to be his original plan of action. Thus, a strategy encompasses not only the player's plan but also his opponents' beliefs in the event that he does not follow that plan.[3]

[2] In some circumstances it is natural to require a player to plan his behavior after his own deviations. If the player is worried that he will make mistakes in the implementation of his decisions, then a comprehensive plan of action should include a move for each history including those which are the result of an error. This is the motivation behind Selten's trembling-hand perfect equilibrium concept. However, if the possibility of error is significant enough to enter into the player's considerations, then the technology of mistakes should be modeled explicitly as part of the game.

[3] This interpretation was mentioned in Rubinstien (1982) for the perfect equilibrium in a sequential bargaining model. Recently, Reny (1987) has used this interpretation to motivate his explicable equilibrium notion.

Interpreting a player's strategy after a deviation as the expectations of the other players about his future behavior makes it problematic to speak of "choice of a strategy." Player 1 does not choose player 2's belief. Game-theoretic models which use the notion of a subgame-perfect equilibrium (or any other solution concept which includes a sequential rationality requirement) therefore require re-examination. Consider, for example, the sequential bargaining literature in which the authors *assume* (rather than conclude) that strategies are stationary. That is to say, a player is confined by *hypothesis* to the use of offers and response patterns (response to offers made by the other player) that are independent of the history of the game. This literature presents this stationarity assumption as an assumption of simplicity of behavior. Consider, for example, player 1's strategy: "Demand 50% of the surplus and reject any offer which gives you less than 50%, independent of what has happened in the past." Simplicity of behavior implies that player 1 plans to make the same offer and make the same responses independently of how player 2 has reacted in the past. However, this strategy also implies that player 2 believes that player 1 would demand 50% of the surplus even if player 1 demanded 60% of the surplus in the first, let us say, 17 periods of bargaining. Thus, stationarity implies not only the simplicity of player 1's behavior but also the passivity of player 2's beliefs. This is unusual, especially if we assume simplicity of behavior. If player 2 believes that player 1 is constrained to choose a stationary plan of action, then 2 should believe (after 17 repetitions of the demand of 60%) that player 1 will continue to demand 60%. Thus, assuming passivity of beliefs eliminates a great deal of what sequential games are intended to model: namely, the changing pattern in players' behavior and beliefs, as they accumulate experience.

Note also, that in games with more than two players, the interpretation of a player's strategy following a deviation as the belief of the other players regarding his future behavior, implies that players 2 and 3 hold the same belief about player 1's future behavior, not only on the equilibrium path, but also after player 1 has deviated. This problem also arises when interpreting the sequential equilibrium notion. Assume that player 1 has to choose from (a, b, c) and that players 2 and 3 learn whether he chose within (a) or (b, c). If player 1 was supposed to choose a, but did not, then the beliefs of players 2 and 3 about player 1's actual move must (by the sequential equilibrium definition) coincide.

3. MIXED STRATEGIES

The second context in which I would like to examine the interpretation of the notion "strategy" is that of "mixed strategies." The concept of mixed strategy has often come under heavy fire. To quote Aumann (1987a):[4] "Mixed strategy equilibria have always been intuitively problematic...", and Radner and Rosenthal (1982): "One of the reasons why game-theoretic ideas have not found

[4] I will do injustice to Aumann if I do not add that in the same paper he also describes Harsanyi's justification of the mixed strategy concept as one which "rings so true."

more widespread application is that randomization, which plays a major role in game theory, seems to have limited appeal in many practical situations." The reason for the criticism is that the naive interpretation of a mixed strategy as an action which is conditional on the outcome of a lottery executed by the player before the game, goes against our intuition. We are reluctant to believe that our decisions are made at random. We prefer to be able to point to a reason for each action we take. Outside of Las Vegas we do not spin roulettes.

One must, however, be cautious. There are obviously cases in which players choose random actions. Consider, for example, an employer who has time to monitor only one of his two workers. The workers have to choose a level of effort, either High or Low. If a worker is caught being lazy, his salary is cut. A worker's tradeoff between effort and reduction in salary is such that if he faces a probability $1/2$ of being audited, he prefers to choose H rather than L. The employer is interested in maximizing the number of agents who choose H. Denote by p the probability with which the employer monitors worker A. His real choice is not between monitoring A and monitoring B, but among the possible numbers p. A suitable description of the situation is a leader-follower game in which the employer first commits himself to a number p after which the workers choose their level of effort. The policy $p = 1/2$ is the best policy for the employer and, using a random device, is strictly optimal. However, this choice is not a mixed strategy in which the two pure actions "monitoring A" and "monitoring B" are the pure strategies. The employer is not indifferent between $p = 1$ (auditing only worker A) and $p = 1/2$. He strictly prefers the latter. In contrast, in mixed strategy equilibrium, the players are indifferent among all lotteries which consist of the actions taken from the support of the mixed strategy.

It should also be mentioned that the use of mixed strategies is particularly problematic in any situation where their execution is costly in terms of devoting attention or time.[5] If implementing a mixed strategy is costly for the player, then he will strictly prefer to use any of the pure strategies which appear in the support of the mixed strategy rather than to waste the resources associated with implementing the lottery.

Having rejected the naive interpretation of mixed strategies let us discuss two other interpretations of the concept.

First, we will examine the large population case. One can think about a game as an interaction between large populations. Each occurrence of the game takes place after a random draw of the players from these populations.[6] Alternatively, a player stands for a large set of players, each of which chooses an action where the payoff depends linearly on the fraction of the population which selects each of the actions. In this context, a mixed strategy is viewed as the distribution of the pure choices in the population. However, this is a limited context which

[5] See the discussion in Abreu and Rubinstein (1988).
[6] See Rosenthal (1979).

does not justify the use of the notion in the case where a unique agent stands behind each of the players in the game.

The second interpretation is the purification idea. A player's mixed strategy is thought of as a plan of action which is dependent on private information which is not specified in the model. Although the player's behavior appears to be random, it is actually deterministic. If we add this information structure to the model, the mixed strategy becomes a pure strategy in which the action depends on the extraneous information. There are two major problems with interpreting a mixed strategy as a rule of behavior which depends on exogenous factors. First, it is hard to accept the assumption that behavior depends on factors which are payoff irrelevant. There are reasons behind the choices people make and their behavior is not motivated by what they consider to be irrelevant factors. If the additional factors are payoff relevant (as Harsanyi (1973) insists), then the model is incomplete in the sense that there are factors the players perceive as relevant (compare with Section 7) which are excluded from the model. Second, the predicted equilibrium behavior is highly fragile. If the manager's behavior is determined by the type of breakfast he eats, then factors outside the scope of the model, such as a change in diet or the price of eggs, may change the frequency with which players choose their actions, thus inducing changes in the beliefs of other players and causing instability.

Nevertheless, I agree that the purification idea provides a consistent interpretation of a mixed strategy as a plan of action. However, in applying this idea to a particular economic problem the economist needs to indicate which are the "real life" exogenous variables on which players base their behavior. Thus, if an industrial organization theoretician describes an equilibrium in price competition as an equilibrium with mixed strategies relying upon the purification concept, he must specify the unmodelled factors which actually serve as the basis for the firms' pricing policy. One has to show that the information structure is rich enough to span the set of all mixed strategy equilibria.[7] I am not familiar with a single case in which the use of mixed strategies in economics is justified by specifying a dependence of behavior on specific exogenous factors. In fact the mixed strategy equilibrium is widely used as a technique for calculating a solution in cases where the pure strategy equilibrium does not exist.

[7] Radner and Rosenthal (1982) investigate the following question: Given an information structure about some random elements (which may or may not be payoff relevant), can the mixed strategy equilibrium of the original equilibrium be attained as a pure Nash equilibrium of the perturbed game? Conditions are provided under which this is the case. These conditions are not mild and there are simple examples of informational structures by which a mixed strategy equilibrium could not be purified. Another important related reference is Harsanyi (1973) who shows that under wide conditions every mixed strategy equilibrium is a limit of *some* sequence of pure strategy equilibria in some sequence of games in which all payoffs are "slightly" randomly perturbed and each player gets exact information only about his own payoffs. From the perspective of looking for an interpretation of mixed strategy, this result is rather weak. At most we can conclude that there is some sequence of perturbations for which the mixed strategy equilibrium is an approximation.

Mixed strategy can alternatively be viewed[8] as the belief held by all *other* players concerning a player's actions. A mixed strategy equilibrium is then an n-tuple of common knowledge[9] expectations, which has the property that all the actions to which a strictly positive probability is assigned are optimal, given the beliefs. A player's behavior may be perceived by all the other players as the outcome of a random device even though this is not the case. Adopting this interpretation requires the reassessment of much of applied game theory. In particular, it implies that an equilibrium does not lead to a prediction (statistical or otherwise) of the players' behavior. Any player i's action which is a best response given his expectation about the other players' behavior (the other $n-1$ strategies) is consistent as a prediction for i's action (this might include actions which are outside the support of the mixed strategy). This renders meaningless any comparative statics or welfare analysis of the mixed strategy equilibrium and brings into question the enormous economic literature which utilizes mixed strategy equilibrium.

4. GAMES WITH LIMITED MEMORY

The extensive form game allows us to discuss situations in which it is common knowledge among the players that information concerning another player's move, a random event, or previously available information (perhaps even about his own actions) is to be withheld from a certain player.[10] Let us now consider situations in which a player knows that he might become doubtful about some relevant information which he possessed in the past. Such fear of memory confusion is clearly an important element in real life reasoning. In particular, it is often the motivation for preferring modes of behavior which keep things simple in order to avoid disasters resulting from this confusion.

Consider the following example which is illustrated in Figure 2. A driver is at A and wishes to reach C. He could drive to C by the long route, which would bring him to his destination directly without having to make any further decisions, or he could use a short but unmarked road, in which case he would have to make a turn at the second intersection. If he arrives at B or if he misses the second intersection and reaches D he will be stuck in traffic jams and hence waste several hours returning to C.

The driver knows that he is able to identify the turn to B but that when he arrives at the turn to C, he will become confused and believe that there is a 10% chance that he has not yet passed the first intersection. The driver believes that once he finds himself in this state of doubt, he will be alert and no longer be confused. We usually assume that a player makes all possible inferences from the information embedded in the game form; here, however, we assume that the

[8] This view has been promoted in the literature by Aumann; see, for example, Aumann (1987b).

[9] The important role of the common knowledge of expectations in this interpretation is emphasized in the discussion of rationalizability appearing in Bernheim (1984) and Pearce (1984).

[10] As in games of imperfect recall; see Alpern (1988).

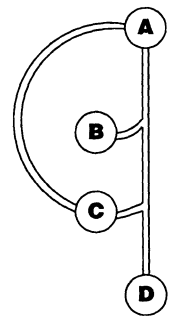


FIGURE 2

driver does not conclude from his state of doubt (which would occur only at the second intersection) that he is actually at the second intersection.

Any presentation of the situation as an extensive game must allow a path of length 4. It has to include a node, v_1, in which the choice between the short and long routes is made, a node, v_2, which corresponds to the decision problem at the turn to B, and an information set which corresponds to the state of doubt. The two nodes in this information set, v_3 and v_4, must be different from v_2, since at v_2 the decision maker does not have any doubts about his location. A chance player node which precedes this information set enables us to model the assumption that the decision maker believes with probability 0.9 that he is at the turn to C (and will not have any more decisions to make) and with probability 0.1 that he is at the turn to B (and if he continues he will have one more decision to make at the turn to C). When considering his action, at this information set, he realizes that he may pass through the intersection to C as well and thus the tree must include another successor node, v_5.

Can a strategy in this game be interpreted as a plan of action? No. Looking at the map, we observe that on any excursion the player could make at most 3 decisions (at A, at the turn to B, and at the turn to C). In the above extensive form game, there is a path (v_1, v_2, v_4, v_5) in which the driver has to make 4

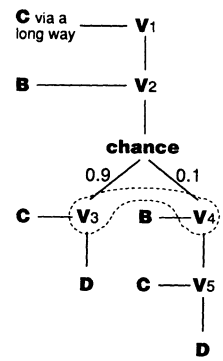


FIGURE 3

decisions. Thus, in practice the node v_5 is unreachable. The decision at v_5 is not a part of a plan of action made at A. It is added to the game form merely to allow us to discuss the player's reasoning in the state of doubt.

5. TOWARDS A PERCEPTIVE INTERPRETATION OF GAME THEORY

In the previous three sections I have tried to explain the problems of interpreting a strategy as a plan of action. In the case of sequential games, such an interpretation does not apply to the part of a player's strategy which supposedly describes a player's planned actions should he deviate from his original plan. A mixed strategy can rarely be interpreted as a set of instructions. In games with potential loss of memory, the game theoretic strategy has to be interpreted as including hypothetical plans which can never be realized. I have tried to argue that interpreting one's strategy as a description of the *other* players' beliefs about his behavior makes more sense in these contexts. In this section I argue that this interpretation fits neatly into a perceptive interpretation of the notion of a game.

Let me go back to the classical interpretation of the game form as a full description of the physical events in the modeled situation. Notice that it is rare that a situation involving a conflict of interests is described clearly and objectively by a set of rules. The exceptions I can think of are "games" in the colloquial sense. Unless the game instructions appear on the box bought at "toys 'r' us," I cannot see how we can avoid the interpretation of a game form as an abstract summary of the players' actual perceptions of the complicated situations they are in.[11] Players may be involved in a recurring situation but might ignore the repetitive aspect of their position; players may not be aware of some possibilities of error. They may consider only a few of the many available actions; they may combine several actions into one option; they may decide to ignore information which in principle they could take into account and so on.

As is often the case, we find an early reference to this point in John Nash's work. When describing the strategic approach to bargaining, he states (Nash (1953, p. 129)): "Of course one cannot represent all possible bargaining devices as moves in the noncooperative game. The negotiation process must be formalized and restricted, but in such a way that each participant is still able to utilize all the essential strengths of his position."

I like to think of a game as being a comprehensive description of the relevant factors involved in a specific situation as perceived by the players, rather than as a presentation of the physical rules of the game. Let me demonstrate this view while distinguishing between infinite and finite horizon games, particularly in infinitely and finitely repeated games. My thinking on repeated games has been influenced by the following quotation from Aumann (1959): "In the notion of a supergame that will be used in this paper, each superplay consists of an infinite number of plays of the original game G. On the face of it, this would seem to be

[11] See Selten (1978) for a discussion of "the level of imagination."

unrealistic, but actually it is more realistic than the notion in which each superplay consists of a fixed finite (large) number of plays of G.... Of course when looked at in the large, nobody really expects an infinite number of plays to take place; on the other hand, after each play we do expect that there will be more. A. W. Tucker has pointed out that this condition is mathematically equivalent to an infinite sequence of plays, so that is what our notion of supergames will consist of."

This paragraph[12] is an elegant statement of what in my opinion is the most basic principle in the art of formal modelling. By using infinite horizon games we do not assume that the real world is infinite. Models are not supposed to be isomorphic with reality. An infinitely repeated game is meant to assist in analyzing situations where players examine a long-term situation without assigning a specific status to the end of the world. In contrast, the finitely repeated game model corresponds to a situation in which the finite period enters explicitly into the players' considerations. Using the terminology of formal logic, we can say that finite horizon models are suitable only for modelling situations in which the last period appears as an "individual constant" (a specified element) in the players' model.

There is only a partial correspondence between the real length of a repeated game (assuming that the term "real length" is well defined) and the selection of a model to analyze it. Even short games may be better analyzed as infinite horizon games.[13] When subjects in a laboratory situation get instructions to play the prisoner's dilemma twenty times with payoffs of between one and four cents, it seems that the infinite horizon game captures their method of reasoning better then a finitely repeated game.

Many game theoreticians have expressed their unhappiness with the sharp discontinuities that exist between finite and infinite horizon game outcomes. I have no difficulty with these discontinuities. I view infinite and finite horizon models as representing two very different scenarios. I am equally unimpressed with the fact that the limit of a sequence of long finite horizon models is an infinite horizon game. For example, in the alternating bargaining model (see Rubinstein (1982)) the perfect equilibria for the finite horizon model converge to the infinite horizon unique perfect equilibrium (see Binmore (1987a)). However, the significance or insignificance of the infinite model does not depend on this result. The convergence result is, in my opinion, nothing more than a convenient tool for approximating the perfect equilibrium outcomes of finite horizon models when the horizon is very long.

I take a similar position on some other game theoretic modelling issues. For example, compare models with and without discounting. A model with discounting is appropriate when considering situations in which the players are con-

[12] Ignoring the problems associated with the term "mathematical equivalence" which I do not understand.

[13] This point of view weakens the significance of some game-theoretic results which are referred to in the game-theoretic jargon as "paradoxes" (e.g., the finitely repeated prisoner's dilemma and Selten's chain store paradox).

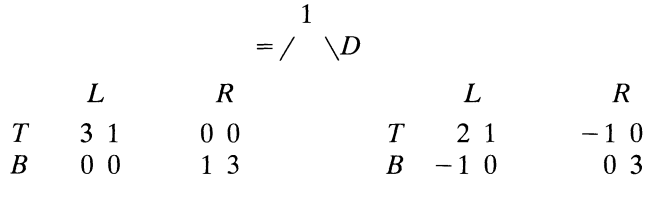
cerned with time and include time impatience in their strategic considerations. A model without discounting is better suited to situations in which the players ignore the timing component of an outcome. (This does not imply that *if asked*, a player in the no-discounting model would be indifferent between getting one million dollars today or getting the same amount in ten years time). Therefore, I do not see the need to approximate a world without discounting using a world in which the discount rate is close to 1. The similarity between the model with a discount rate equal to 1, and models with discount rates of "almost" 1 is purely mathematical. It is useful mainly as a technique to approximate the solutions for games with discount rates close to 1.

6. RELEVANCE

If we adopt the view that a game is not a rigid description of the physical rules of the world, then a game-theoretic model should include only those factors which are perceived *by the players* to be *relevant*. Modelling requires intuition, common sense, and empirical data in order to determine the relevant factors entering into the players' strategic considerations and should thus be included in the model. This requirement makes the application of game theory more an art[14] than a mechanical algorithm.

Adopting the view that a game should include only relevant factors entails implications for many game-theoretic issues and in particular for the recent discussion of "forward induction principles." This is demonstrated by one of the most intriguing examples that I have ever seen in game theory. The example is due to Eric van Damme (see van Damme (1989)[15] and also Ben-Porath and Dekel (1988) and Osborne (1990)) and is a development of ideas initiated in "Kohlberg's example" (see Kohlberg and Mertens (1986)).

EXAMPLE: Two players 1 and 2 participate in a battle of the sexes in which the nonzero outcomes are in dollar terms (3, 1) and (1, 3). Prior to their engaging in the battle, player 1 (and only player 1), is able to dispose of one dollar. Player 2 observes player 1 disposing or not disposing of the dollar. An extensive form game which describes the above scenario is:



[14] See Aumann (1987a).

[15] van Damme offers the example as a demonstration of the importance of sunk costs and he should not be indicted in any criticism of my position on the example.

In the normal form of this game each player has 4 strategies:

	LL	*LR*	*RL*	*RR*
$=T$	3 1	3 1	0 0	0 0
$=B$	0 0	0 0	1 3	1 3
DT	2 1	−1 0	2 1	−1 0
DB	−1 0	0 3	−1 0	0 3

van Damme invokes what is usually regarded as the rather weak solution concept of the successive elimination of weakly dominated strategies: It is easy to verify that successively, *DB* is dominated by $=T$, *RR* is dominated by *RL*, *LR* is dominated by *LL*, $=B$ is dominated by *DT*, *RL* is dominated by *LL*, *DT* is dominated by $=T$, and we are left with $(=T, LL)$. Thus, the mere fact that player 1 could dispose of one dollar skewed the game in his favor.

As economists we should ask ourselves whether we can conceive of a conflict of interests like this to be resolved by the identity of the player who has the ability to dispose of resources. I very much doubt this to be the case.

How is "relevancy" related to this discussion? If disposing of the dollar were a relevant consideration in the players' perception of the situation, then the result would (probably) make sense. However, I cannot believe that any reasonable person would consider a pre-game disposal of a dollar to be relevant in the analysis of the battle of the sexes. It is my opinion that a formal description of the situation should exclude the choice of disposal even in cases where a description of the game is given by a referee who specifies the possibility of disposing of the dollar (recall that there is rarely a referee).

We might think about "relevancy" as a matter of convention, which may be explained by game-theoretic arguments. I prefer to view the determination of relevancy as a stage in the game process which precedes the prediction of other players' behavior and the calculation of the best plan of action. I prefer to leave the explanation of the logic of perception to other branches of science (such as evolutionary biology). Still, we can attempt to organize our intuition of irrelevancy by appealing to general principles.[16] The following are three characterizations of the disposal of the dollar in van Damme's example which may be used to form a general sufficient condition for irrelevancy:

(a) The disposal decision does not affect the payoffs of the players in the battle of the sexes (when player 1 disposes of the dollar, the game is identical to the game in which player 1 does not dispose of the dollar).

(b) The disposal decision does not reveal any unknown information (in contrast to the role of spending money on advertising in the signalling literature). Even if disposal gives information about "the rationality of player 1," a sensible conclusion might be that a player who throws the dollar out of the window is just "crazy."

[16] See Harsanyi (1977, page 118).

(c) The disposal decision is not a part of a game which is identical or even similar to the choice problem which player 1 has to confront in the battle of the sexes. (In contrast, the problem which a player confronts in the first period of an infinitely repeated game is identical to his problem in the second period.) Thus, whether or not player 1 disposes of the dollar cannot provide information from which player 2 learns about player 1's behavior in the battle of the sexes.

The reader will probably feel that verbal statements made before a play of the battle of the sexes satisfy the above three conditions for irrelevancy and yet they are highly relevant. This point was made by Joe Farrell. Farrell (1985) suggests that if the players in the battle of the sexes have a language and if only player 1 could (cheaply) talk and state (in English) that he is going to play UP, then he has a serious advantage in determining the conflict's outcome. This sounds like a plausible argument and Farrell correctly modelled communication as a part of the solution concept rather then as a part of the game. However, this is not the argument which is used in van Damme's analysis. If instead of being able to throw a dollar out of the window, player 1 is allowed to throw a bill worth nothing out of the window (or nod his head), or even state that "I am going to play UP," then the process of successive elimination of weakly dominated strategies is not powerful and all equilibria of the battle of the sexes would survive. It is my impression that although language plays a crucial role in resolving conflicts, game theory has so far been unable to capture this role.

7. REGULARITY

There is one additional issue to be raised in connection with the interpretation of game theory. Game theory deals with regularities.[17] As Carnap (1966) writes: "The observations we make in everyday life as well as the more systematic observations of science reveal certain repetitions or regularities in the world. ..." The laws of science are statements expressing these regularities as precisely as possible. Events which are pure one-shot events, completely disconnected and uncorrelated with other events are outside the scope of any theory including game theory. Game theory should be employed as a descriptive science only after the social scientist observes or expects some regularities in a family of "similar" events. Accepting this view leads me to make the following three comments:

a. *Is a one-shot game an isolated event?* Game theoreticians often emphasize the one-shot character of the games analyzed. Careless interpretations of this emphasis have led to a widespread confusion in the applied game theory literature with respect to the distinction between one-shot and repeated games. It is the repetitive nature of a situation which makes it possible to talk about regularity. The distinction between a one-shot game and a repeated game is not based solely on the naive criterion of whether the players repeat the conflict

[17] See Lucas (1985) who expresses a similar view about economics.

again and again, or whether they are involved in an isolated incident. The proper criterion is concerned with whether the players take into account the effect of their choice today on similar future games in which they will participate. If they ignore the effect of their behavior on future games, then the framework of the game-theoretic *one-shot game* is appropriate. If the players calculate the effect of their behavior on future games, the *repeated game* framework is appropriate. If the players use their past experience to speculate about other players' future behavior without taking into account the effect of their own behavior, then we are dealing with *dynamics*.

b. *The meaning of nonexistence*. If what we are trying to model in game theory are situations in which we expect regular behavior, then it is not true that all descriptions of the world should have an equilibrium.[18] The mere fact that a game theoretician constructs a game does not mean that the game corresponds to a regular mode of behavior. The modeller should not, in the case of nonexistence, twist the analysis, but rather should check the adequacy of the model as a description of conceived regularity.

This brings me back to the mixed strategies issue. One of the reasons that mixed strategies are popular in both game and economic theory, in spite of being so unintuitive, is that many models do not have an equilibrium with pure strategies. However, the nonexistence of a solution concept in pure strategy does not necessarily mean that we should look for stochastic explanations. It means that the description of the game and the assumptions embedded in the solution concept are not consistent with regularity. Expanding the model or changing the basic assumptions are alternatives which the modeller should consider at least as favorably as mixed strategies.

c. *Regularity and language*. The observed regularity depends on the language employed. Behavior can be irregular in one language and regular in another. I became aware of this simple point after a game I played with my baby daughter when she was about one year old. Wanting to check her choice consistency and knowing that she recognized colors, I put Blue, Red, and Green cubes in front of her in an arbitrary order. I repeated this choice problem about a dozen times. Her choices revealed a clear inconsistency. However, before I became "overjoyed" with the idea that my cute baby daughter had refuted the basic postulate of choice theory, I realized that the baby was actually amazingly consistent. However, she was not consistent in her choice between Blue, Red, and Green. She was consistent in her choice between Left, Center, and Right. She always chose Left.... If my vocabulary did not include the words Left, Center, and Right, I would be unable to describe this regularity. Given that we can use two languages, we prefer to use the terminology in which we find the behavior to be more regular.

[18] My thinking on this matter was influenced by a statement made by my teacher, Menachem Yaari, in commenting on a famous paper which showed nonexistence of equilibrium in an economy with imperfect information. Yaari told the "shocked" economists that the problem lay neither in the solution concept nor in the real world, but rather... in the model.

8. CONCLUSION

The comments I have offered in this paper were intended to emphasize the inadequacy of the naive interpretation of game theory as a physical description of the world and to promote the view that in game theoretic modelling, we must regard as *given* the laws of perception, the bounds on rationality, and the processes of reasoning employed by the players.

There exists a widespread myth in game theory, that it is possible to achieve a miraculous prediction regarding the outcome of interaction among human beings using only data on the order of events, combined with a description of the players' preferences over the feasible outcomes of the situation. For forty years, game theory has searched for the grand solution which would accomplish this task. The mystical and vague word "rationality" is used to fuel our hopes of achieving this goal. I fail to see any possibility of this being accomplished. Overall, game theory accomplishes only two tasks: It builds models based on intuition and uses deductive arguments based on mathematical knowledge. Deductive arguments cannot by themselves be used to discover truths about the world. Missing are data describing the processes of reasoning adopted by the players when they analyze a game. Thus, if a game in the formal sense has any coherent interpretation, it has to be understood to include explicit data on the player's reasoning processes. Alternatively, we should add more detail to the description of these reasoning procedures. We are attracted to game theory because it deals with the mind. Incorporating psychological elements which distinguish our minds from machines will make game theory even more exciting and certainly more meaningful.

Department of Economics, Tel Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel

Manuscript received August, 1988; final revision received August, 1990.

REFERENCES

ABREU, D., AND A. RUBINSTEIN (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, 1259–1281.

ALPERN, S. (1988): "Games with Repeated Decisions," *SIAM Journal of Control and Optimization*, 26, 468–477.

AUMANN, R. (1959): "Acceptable Points in General Cooperative n-Person Games" in *Contributions to the Theory of Games IV*, Annals of Mathematics Study, 40. Princeton: Princeton University Press, pp. 287–324.

——— (1987a): "What Is Game Theory Trying to Accomplish?" in *Frontiers of Economics*, ed. by K. J. Arrow and S. Honkapohja. Oxford: Blackwell.

——— (1987b): "Correlated Equilibrium as an Expression of Bayesean Rationality," *Econometrica*, 55, 1–18.

BEN-PORATH, E., AND E. DEKEL (1988): "Coordination and the Potential for Self Sacrifice," Graduate School of Business, Stanford University, Research Paper 984.

BERNHEIM, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

BINMORE, K. (1983): "Aims and Scope of Game Theory," mimeo, LSE.

——— (1987a): "Perfect Equilibria in Bargaining Models," in *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Oxford: Blackwell.

——— (1987b): "Modeling Rational Players I," *Economics and Philosophy*, 3, 179–214.

CARNAP, R. P. (1966): *Introduction to Philosophy of Science*. New York: Basic Books, Inc.

FARRELL, J. (1985): "Credible Neologisms in Games of Communication," MIT Working Paper 386.

HARSANYI, J. (1973): "Games with Randomly Distributed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, 3, 211–225.

——— (1977): *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

KOHLBERG, E., AND J. F. MERTENS (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1038.

LUCAS, R. (1985): "Adaptive Behavior and Economic Theory," in *Rational Choice*, ed. by R. and M. Reder, pp. 217–242.

NASH, J. (1953): "Two Person Cooperative Games," *Econometrica*, 21, 128–140.

OSBORNE, M. J. (1990): "Signalling, Forward Induction, and Stability in Finitely Repeated Games," *Journal of Economic Theory*, 50, 22–36.

PEARCE, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

RADNER, R., AND R. ROSENTHAL (1982): "Private Information and Pure Strategy Equilibrium," *Mathematics of Operations Research*, 7, 401–409.

RENY, P. (1988): "Explicable Equilibria," mimeo, University of Western Ontario.

ROSENTHAL, R. W. (1979): "Sequences of Games with Varying Opponents," *Econometrica*, 47, 1353–1366.

RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–110 (and STICERD research report 80/13).

——— (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 39, 83–96.

SELTEN, R. (1978): "Chain Store Paradox," *Theory and Decision*, 9, 127–159.

VAN DAMME, E. (1989): "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48, 476–496.