Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

No. 91

John C. Harsanyi and Reinhard Selten

A Non-Cooperative Solution Theory with
Cooperative Applications

Chapter 1
Preliminary Discussion

March 1980

H. G. Bergenthal

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Universitätsstraße
4800 Bielefeld 1
Bundesrepublik Deutschland
Federal Republic of Germany

Chapter 1. Need for a new solution concept

1.1. Our solution concept. The purpose of this book is to propose a new solution concept, primarily defined for noncooperative games, but applicable also to cooperative games, because every cooperative game can be re-modelled as a bargaining game, having formally the structure of a non-cooperative game. For any noncooperative game, including non-cooperative bargaining games, our theory always selects one specific equilibrium point as the solution. By reducing co-operative games to noncooperative bargaining games, our approach unifies the theories of cooperative and noncooperative games into one general theory.

1.2. Cooperative and noncooperative games. In contrast, in classical game theory, cooperative and noncooperative games are treated quite differently, and the distinction between these two game classes plays a very fundamental role. Nash [1950a, 1951], who first introduced this distinction, defined cooperative games as games permitting both free communication and enforceable agreements among the players, in contrast to noncooperative games, permitting neither communication nor enforceable agreements.

Of course, a binary distinction based on two simultaneous criteria is logically unsatisfactory. We cannot define one category as a class of all objects possessing both properties A and B while defining the other category as a class of all objects possessing neither property. For if we do so then the question immediately arises, what about objects having property A but not B, and about objects having property B but not A?

Accordingly, it is preferable to use a one-criterion distinction, and to define cooperative games simply as those permitting enforceable agreements while defining noncooperative games as those not permitting them. How much communication is allowed among the players is, of course, also very important in many cases; but it turns out to be a less fundamental issue. To illustrate the problem consider the following prisoner's dilemma game. [For explanation of the term "prisoner's dilemma", see Luce and Raiffa, 1957, pp. 94-95]. In each cell of the payoff table, the number in the upper-left corner is player 1's payoff while that in the lower-right corner is player 2's. The rows of the table represent player 1's strategies C* and N*; whereas the columns represent player 2's strategies C** and N**.

	C**	N**
C*	10 / 10	-10 / 11
N*	11 / -10	1 / 1

<u>Fig. 1</u>

This game is completely symmetric between the two players, so that both of them have positions of equal strength. Therefore, it is natural to expect that they will agree on an outcome yielding them equal payoffs - - either by choosing the strategy pair C = (C*, C**), which would yield the payoffs (10,10), or by choosing the strategy pair N = (N*, N**), which would yield the payoffs (1,1). Accordingly, if the game is played as a <u>cooperative</u> game permitting enforceable agreements,then the players, assuming that they act rationally, will no doubt immediately agree to use the strategy pair C since C will give them much higher payoffs then N would. Thus, C = (C*, C**) may be called the <u>cooperative solution</u> of the game.

In contrast, if the game is played as a <u>noncooperative</u> game, i.e. if the players are unable to conclude enforceable agreements, then they cannot do any better than use the strategy pair N = (N*, N**), which, therefore, may be called the <u>noncooperative solution</u>.

To establish this point, we will first show that, if enforceable agreements are impracticable, then rational players cannot choose the strategy pair C = (C*, C**). This is so because even if they did agree to use their C-strategies, they could not rationally expect each other actually to keep to this agreement, so that any such agreement would be quite pointless. For suppose they would make such an agreement, and would in fact expect each other to keep it. Then, player 1 would immediately have an incentive to violate this agreement by using strategy N*, rather than C* because N*, and not C*, would be his best reply [1/] to player 2's expected strategy, viz. to C** . Likewise, player 2 would also have an incentive to violate the agreement by using strategy N**, rather than C** because N** , and not C**, would be his best reply to player 1's expected strategy, viz. to C*.

Thus, this strategy pair C cannot be chosen by rational players in a noncooperative game because it would be self-destabilizing: the very fact that the players would expect each other to abide by it would give them a clear incentive to deviate from it. Moreover, our analysis also shows the mathematical reason why C has this undesirable property. The reason is that the two players' C-strategies are not best replies to each other; rather, the best reply to C** is N*, and the best reply to C* is N**.

In contrast, the strategy pair N = (N*,N**) can be readily used by rational players in a noncooperative game because it is self-stabilizing: Since N* and N** are mutually the best replies to each other, if the two players for any reason expect each other to use their N-strategies, then both of them will have a clear incentive to make this expectation come true by in fact using their N-strategies.

Clearly, in playing this game, the decisive question is whether the players can make enforceable agreements or not; and it makes little difference whether they are or are not allowed to talk to each other. Even if they are free to talk and to negotiate any agreement, this fact will be of no real help if such an agreement is unenforceable and, therefore, has little chance of being kept. An ability to negotiate agreements is useful only if the rules of the game make such agreements fully binding and enforceable. (In real life, agreements may be enforced externally by courts of law and by other government agencies as well as by pressure from public opinion; or they may be enforced internally by the fact that the players are spontaneously unwilling to violate agreements, and also _know_ that this is the case, say, as a result of their personal moral attitudes.)

As Nash has already pointed out [1950a,1951], similar considerations apply to _all_ noncooperative games. Since in such games agreements are not enforceable, rational players will always choose a strategy combination that is _self-stabilizing_ in the sense that the players will have some incentive to abide by this strategy combination (or at least will have no incentive not to do so) - - if they expect all _other_ players to abide by it. Mathematically this means that they will always choose a strategy combination with the property that every player's strategy is a best reply to all other players' strategies.[A] strategy combination with this property is called an _equilibrium point_. (We will sometimes shorten this name simply to "equilibrium".)

In the two papers already quoted, Nash has also shown that every finite game [2/] has at least _one_ equilibrium point (in pure strategies or sometimes only in mixed strategies).

Before concluding this discussion, we must add that the definitions stated above are still in need of further clarification. As they stand, they may give the false impression that noncooperative games cannot be used at all for modelling game situations in which the players are able to make enforceable agreements (or to enter into other firm commitments [3/] such as irrevocable promises and threats). As we will see presently (in Section 1.3), this is not the case because it is perfectly possible to incorporate self-commitment moves explicitly into the extensive form of a noncooperative game.

To avoid this misleading impression, we propose to rephrase our definitions as follows. A _noncooperative_ game is a game modelled by making the assumption that the players are _unable_ to make enforceable agreements (as well as commitments of other sorts), except as far as the _extensive form of the_ game explicitly gives them an ability to do so. In contrast, a cooperative game is a game modelled by making the assumption that the players are able to make enforceable agreements (and possibly also commitments of other sorts) even if their ability to do so is not shown explicitly by the extensive form of the game.

1.3. Irrevocable commitments within a noncooperative game.

There are several alternative ways of incorporating self-commitment moves into the extensive form of a game. For instance, we can define the payoffs of the game in such a way that any violation of a commitment made by any player would carry heavy penalties. Or, we can add one or more extra players to the game whose task is to punish violators,etc..But the simplest method of doing it is this.

At a suitable point of the game tree, we give the relevant player a choice between two moves, say, α and β, where α is interpreted as a commitment to do or not to do something at some later stage(s) of the game, while β is interpreted as making no commitment. The commitment expressed by move α may be unconditional or it may become operative only conditionally, subject to occurrence of some future events. If the player chooses move β then from that point on the game will be governed by the remaining part of the original game tree, which we will call subtree T. But if he chooses move α, then from that point on the game will be governed by a modified version of subtree T, to be called T'. T' will differ from T by having all those branches removed that would correspond to moves violating the commitment that the player in question made when he chose mo e α. (In other words, moves violating his commitment will simply not be available to this player.)

Of course, it can happen that this removal of all commitment-violating moves will leave some of the player's information sets with one unique branch (i.e.,one unique move) originating from them, indicating that he has no real choice any more at any of these information sets.Such information sets (and these unique branches) can always be omitted, since information sets permitting no real choice are irrelevant.

This method can be easily generalized, of course, to cases where a given player can choose not only between making or not making a specific commitment, but rather can also choose among various alternative commitments, etc.

For example, the extensive form of the game discussed in Section 1.2 can be represented by the following game tree:

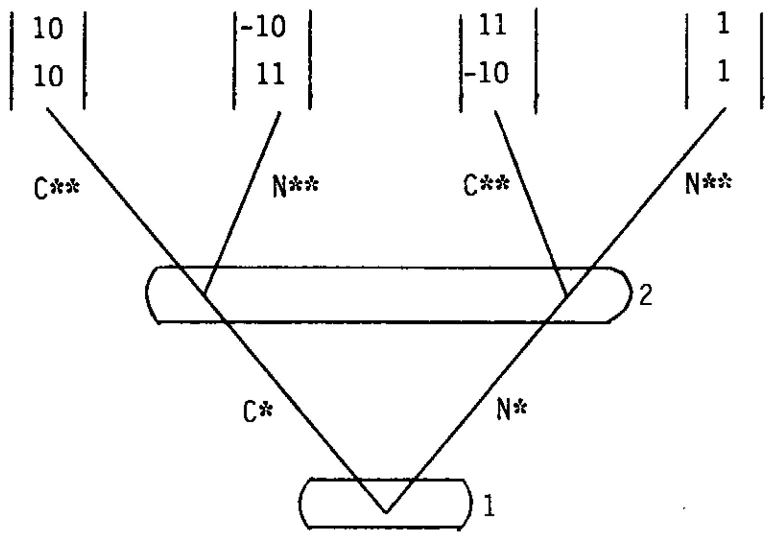

Fig. 2

The numbers 1 and 2 printed at the right of the two information-set symbols (i.e., the two ovals) indicate which player has a move at that particular information set.

Now, we can represent the players' ability to make an enforceable agreement about using their C-strategies as follows. At the beginning of the game, we give player 1 a choice between moves α^* and β^*, where α^* means, "I commit myself to using strategy C*, provided that player 2 will commit himself to using strategy C**," while move β^* means, "I make no commitment."

In case player 1 has actually chosen move α^*, we now give player 2 a choice between moves α^{**} and β^{**}, where α^{**} means, "Yes, I do commit myself to using strategy C** as player 1 has suggested," while move β^{**} means, "I make no commitment."

Now, we can distinguish three cases:
(1) If player 1 has chosen α* while player 2 has chosen α** then both players wil be <u>committed</u> to using their C-strategies. Consequently, the remaining part of the game will now be reduced to the subtree T_1, indicated below.

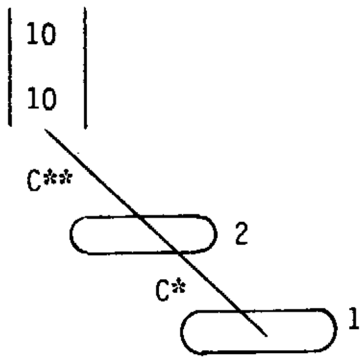

Fig. 3

But since each of the two information sets in T_1 has only <u>one</u> branch arising from it, we can now omit both of these information sets as well as these two branches (C* and C**) altogether, which amounts to replacing the entire subtree T_1 merely by the payoff vector $\begin{vmatrix}10\\10\end{vmatrix}$ generated by it.

(2) If player 1 has chosen α* while player 2 has chosen β** then the two players will be under no commitment restricting their freedom of action. Consequently, the remaining part of the game will be governed by a subtree T_2 which is simply a copy of the original game tree.

(3) If player 1 has chosen β* then, once more, the players will retain their freedom of action, and the remaining part of the game will be governed by a subtree T_3 which is again simply a copy of the original game tree.

Accordingly, the game tree of the enlarged game will be as follows:

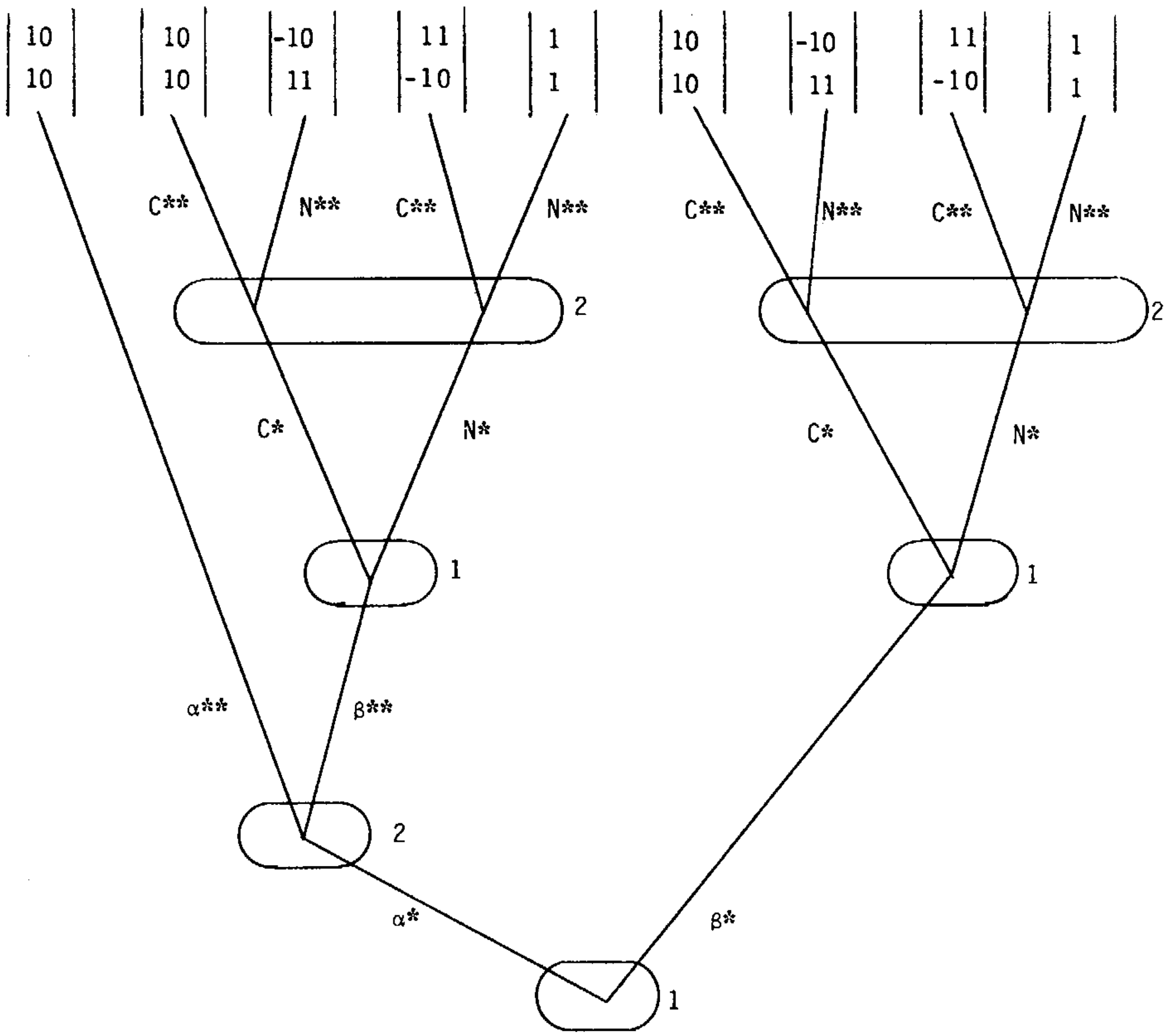

Fig. 4

In the normal form of the enlarged game, we can characterize each player's strategies by three symbols. For example, the first symbol (α* or β* for player 1, and α** or β** for player 2) may be used to indicate the player's choice between commitment and no commitment; the second symbol (C* or N* for 1, and C** or N** for 2) may indicate the strategy that he would follow in subtree T_2; and the third symbol (C* or N*, or alternatively C** or N**) may indicate the strategy he would follow in subtree T_3. Thus, one possible strategy of player 1 would be α*C*N*, etc. Obviously, either player will have $2^3 = 8$ different pure strategies.

It is easy to verify that the enlarged game has only one perfect [4/] equilibrium point in pure strategies, viz. $E_1 = (\alpha^* N^* N^*, \alpha^{**} N^{**} N^{**})$. In other words, if both players are able to commit themselves to their C-strategies then it will be clearly in their interest to do so in order to obtain the payoffs (10,10). At the same time, the definition of E_1 contains two N* and two N** symbols. These indicate the fact that each player would use his N-strategy if his opponent refused to commit himself to use his C-strategy. (Of course, this part of either player's strategy plan will not be implemented since in fact the opponent will make the required commitment.)

Intuitively, one can identify E_1 with the cooperative solution (C*, C**) of the original game. Thus, we can say that by incorporating the commitment moves α^* and α^{**} (as well as no-commitment moves β^* and β^{**}) into the extensive form of the game, we have essentially turned the cooperative solution (C*, C**) into an equilibrium point - - so as to make it an outcome achievable by rational players even if the game (or, rather, the enlarged version of the game) is played as a formally noncooperative game. Indeed, since E_1 is the only perfect equilibrium point of the enlarged game, we have turned E_1 into the only outcome consistent with rational behavior by both players. [5/] (As we will try to show in Sections 1.9 to 1.11, only perfect equilibrium points are compatible with rational behavior by all of the players in a noncooperative game.)

1.4. Limitations of the classical theory of cooperative games.

The classical theory of noncooperative games is essentially a theory of one basic solution concept, that of equilibrium points. In contrast, the classical theory of cooperative games offers a rich variety of alternative solution concepts, such as the von Neumann-Morgenstern stable sets [1944], the Nash solution for two-person bargaining games [1950b, 1953], the Shapley value [1953], the core [Gillies,1959], the Aumann-Maschler bargaining sets [1964], and a number of others.

Individually, each of these solution concepts is of great theoretical interest. But, taken as a group, they fail to provide a clear and coherent theory of cooperative games. Indeed, most of the different solution concepts have very little logical connection with each other, and even less can they be interpreted as special cases of one general theory.

One may think that this fact is merely a conceptual limitation of classical game theory, which may be of some importance to the logician, the methodologist, or the philosopher, but may be immaterial to the social scientist whose main interest lies in possible applications of game theory to economics, political science, and sociology. But, in actual fact, this conceptual limitation does create major problems also in empirical applications.

First of all, while classical game theory offers quite a number of alternative solution concepts for cooperative games, it fails to provide any clear criterion as to which particular solution concept is to be employed in analysing any given real-life social situation. Nor does it give a clear answer to the obvious question of why so many different solution concepts are needed in the first place.

Many solution concepts generate also some additional dimensions of indeterminacy. Even if the decision is made to analyze a given social situation in terms of one specific solution concept A, the latter will often fail to specify a well-defined unique outcome, but rather may tell us no more than that the actual outcome will be somehow chosen from some - - possibly very large - - set S of "acceptable" outcomes; indeed, it may say no more than that the outcome will be a point lying in one of several alternative sets S, S', S",..., each of them equally consistent with the axioms of the chosen solution theory A.

An even more serious shortcoming of classical game theory is its failure to provide any usable solution concepts at all for several theoretically and empirically very important classes of cooperative games (and of games closely related to cooperative games). These include:

(1) Games intermediate between fully cooperative games where all agreements are enforceable, and fully noncooperative games where none are (except in the cases mentioned in Section 1.3). Examples are games where some types of agreements are enforceable while others are not; games where some groups of players are able to make enforceable agreements but others are not; or games where enforceable agreements can be concluded at some stages of the game yet not at other stages, etc.

(2) Cooperative games with a sequential structure. (There is some overlaps between cases (1) and (2)).These are games involving two or more successive stages, and permitting agreements to be built up gradually in several consecutive steps. Unlike classical cooperative games, in which any agreement made is always final, such sequential games might allow renegotiation and modification of earlier agreements at later stages of the game under specified conditions.

(3) Cooperative games with incomplete information. (Since games with incomplete information, both cooperative and non-cooperative, raise some special problems, we will discuss them at some length in Section 1.5.)

All these difficulties of classical cooperative theory have their roots in one very basic limitation of that theory. In almost all cooperative games, bargaining negotiations among the players have an all-important role. Yet, classical cooperative game theory completely excludes the players' bargaining moves and countermoves from its formal analysis, by postulating that these moves - - often described as "pre-play negotiations" - - somehow occur before the game is actually played and, therefore, do not belong formally to the "game" itself. This amounts to voluntarily relinguishing any serious attempt to understand how the outcome of the game in fact depends on the specifics of the bargaining process among the players.

1.5. Games with incomplete information. One of the most serious deficiencies of classical game theory is its inability to deal with games involving incomplete information. We say that a game is one with complete information if all players know the nature of the game, in the sense of knowing the extensive form of the game (i.e., the game tree), or at least knowing the normal form of the game (i.e., the payoff matrix).

A game with complete information can be either a game with perfect information or one with imperfect information: it will be the former if the players not only know the nature of the game as such but also know all previous moves (made by other players or by chance) at any stage of the game; and it will be the latter if the players know the nature of the game but have less than full information about the earlier moves during the game.

On the other hand, a game involves incomplete information if the players have less than full information about each other's strategy possibilities and/or about each other's payoff functions. The latter problem may arise because the players have limited information about

(1) the physical consequences to be produced by alternative strategy combinations, or about

(2) the other players' preference ranking over these physical outcomes, or about

(3) the other players' attitudes toward risk-taking (or because of various combinations of these factors).

At the same time, the players may also be ignorant about the amount of information that the other players have about any given player's strategy possibilities and about his payoff function, ect.

Classical game theory cannot handle games with incomplete information at all (but does cover both games with perfect and with imperfect information as long as these have the nature of games with complete information). This is obviously a very serious limitation because virtually all real-life game situations involve incomplete information on the part of the players. (In particular, it very rarely happens that the participants of any real-life social sisutation have reasonably full information about each other's payoff functions. Uncertainty about the strategies actually available to the other players is also very common.)

It turns out, however, that we can bring a game with incomplete information within the scope of game-theoretical analysis by using a probabilistic model to represent the incomplete information that the players have about various parameters of the game [Harsanyi, 1967-68]. More specifically, let G be a game with incomplete information. Then, analysis of this game G can be reduced to analysis of a new game G* involving suitably chosen random moves. We will call G* a probabilistic model for G. In this new game G*, the fact that (some or all of) the players have limited information about certain basic parameters of the game is mathematically represented by the assumption that these players have limited information about the outcomes of these random moves.

Formally, this probabilistic model game G* will be a game with complete information. (But it will be a game with imperfect information because of the players' having less than full information about the outcomes of the random moves occurring in the game.) Thus, our approach essentially amounts to reducing the analysis of a game with incomplete information, G, to the analysis of a game with complete (yet imperfect) information, G*, which,being a game with complete information, is of course fully accessible to the usual analytical tools of game theory.

By constructing suitable probabilistic models of various types, we can produce games with any desired distribution of knowledge and of ignorance among the players, and can study how alternative informational assumptions will change the nature of the game. We can also study how any given player can infer some pieces of information originally denied to him, by observing the moves of those players who already possess this information. We can also investigate how each player can optimally convey information to some other players, or can optimally withhold this information from them, in accordance with his own strategic interests within the game.(On the problem of optimally conveying information, see our analysis of a two-person game with incomplete information on both sides, in Chapter of this book. On the problem of optimally withholding information, see Aumann and Maschler's discussion [1966, 1967, 1968] of infinitely repeated two-person zero-sum games under incomplete information; see also Stearns [1967].)

We must add, however, that use of such probabilistic models in general provides only a partial solution for the problem of how to analyze games with incomplete information. This is so because, as soon as a probabilistic-model game G* has been constructed as a more convenient mathematical representation for the originally given game with incomplete information, G, the problem immediately arises what solution concept to use for this newly constructed game G* itself.

In actual fact, if the game G we start with is a noncooperative game with incomplete information, then this question has an easy and natural answer. In this case, the probabilistic-model game G* derived from G will be itself also a noncooperative game (though of course one with complete information), and can be analized in terms of its equilibrium points: thus, the concept of equilibrium points can be extended to games with incomplete information without any difficulty [Harsanyi, 1967-68, pp. 320-329].

The situation, however, is very different if the game G we are trying to analyze is a cooperative game with incomplete information. In this case, we will find that typically the probabilistic-model game G* derived from G will not admit of analysis in terms of any of the cooperative solution concepts of conventional game theory.

For example, it turns out that the Nash solution for two-person bargaining games, which is such an attractive solution concept for such games in the case of complete information, cannot be used at all to define solutions for two-person bargaining games with incomplete information or for the probabilistic-model games derived from them: if we try to use the Nash solution for this purpose then we obtain completely nonsensical results [Harsanyi, 1967-68, pp. 329-334]. Other classical cooperative solution concepts give equally unsatisfactory results when applied to incomplete-information games. This lack of solution concepts applicable to games with incomplete information is another serious weakness of the classical theory of cooperative games.

1.6. Difficulties with the concept of equilibrium points. Compared with the classical theory of cooperative games, the classical theory of noncooperative games presents a much more satisfactory picture. First of all, it has a much higher degree of theoretical unity because it is wholly based on one specific basic solution concept, that of equilibrium points. It is also a more complete theory than the theory of cooperative games is because it tries to cover all aspects of any given game, and does not automatically exclude the players' bargaining moves from the scope of its analysis, in the way the theory of cooperative games does. Furthermore, as has already been mentioned, the concept of equilibrium points -- and, therefore, the classical theory of noncooperative games -- can be easily extended to games with incomplete information.

Finally, equilibrium points are one of the very few game-theoretical solution concepts that have direct application both to games in extensive form and to games in normal form, which enables the theory of noncooperative games to deal with both game forms in terms of a uniform theoretical framework. (This has many desirable consequences. One of them is the fact that the classical theory of noncooperative games, unlike the theory of cooperative games, can handle games possessing a sequential structure, without any difficulty.)

Yet, even though the concept of equilibrium points has many strong points, it also has at least three important weaknesses:

1. Almost every nontrivial game has many (sometimes even infinitely many) essentially different equilibrium points. Hence, a theory which could only predict that the outcome of a noncooperative game should be an equilibrium point,

without specifying which equilibrium point this actually were to be, would be an extremely weak and uninformative theory. This difficulty we will call the multiplicity problem.

2. Secondly, any mixed-strategy equilibrium point is, or at least appears to be, fundamentally unstable (see Section 1.8 below) and, therefore, to be unsuited to be the solution of a game. This gives rise to what we will call the instability problem: how are we to define a solution for a noncooperative game that has only mixed-strategy equilibrium points?
3. A third difficulty connected with equilibrium points was pointed out by Reinhard Selten [1965,1975]. He called attention to the fact that many equilibrium points require some or all of the players to use highly irrational strategies (see Sections 1.9 and 1.10). He proposed to call such equilibrium points imperfect equilibrium points, as distinguished from perfect equilibrium points, which involve no irrational strategies. The problem posed by the fact that many games contain imperfect equilibrium points we will call the imperfectness problem.

1.7. The multiplicity problem. For our purposes, among the three problems posed by the concept of equilibrium points, the multiplicity problem is of particular importance. To illustrate the nature of this problem, we will now consider a very simple two-person bargaining game, where two players have to agree on how to divide $ 100, the money being lost to them if they cannot agree. (We will assume that both players have linear utility functions for money.) This game can be represented by the following bargaining model. Each player has to name a real number, representing his payoff demand. The numbers named by players 1 and 2 will be called x_1 and x_2, respectively. If $x_1 + x_2 \leq 100$, i.e. if the two players' payoff demands are mutually compatible, then both will obtain their payoff demands, i.e. they will obtain the payoffs $u_1 = x_1$ and $u_2 = x_2$. In contrast, if $x_1 + x_2 > 100$, i.e. if their payoff demands are incompatible, then they will receive zero payoffs $u_1 = u_2 = 0$ (since this will be taken to mean that they could not reach an agreement).

If the players are free to divide the \$ 100 in all mathematically possible ways, then this game will have infinitely many essentially different equilibrium points in pure strategies since all possible pairs (x_1,x_2) satisfying $x_1 + x_2 = 100$, as well as $x_1 \geq 0$ and $x_2 \geq 0$, will be equilibrium points. But even if we restrict the players to payoff demands representing integer numbers of dollars, the game will still have 101 essentially different equilibrium points, from (0,100), (1,99),..., to (100,0). Clearly, a theory telling us no more than that the outcome can be *any one* of these equilibrium points will not give us much useful information. We need a theory selecting *one specific* equilibrium point as the solution of the game. In fact, the main purpose of our new solution concept is to provide a mathematical criterion that always selects one specific equilibrium point as the solution. In other words, its main purpose is to overcome the multiplicity problem. (But, as we will try to show, our theory also overcomes the two other problems posed by the concept of equilibrium points, viz. the instability problem and the imperfectness problem.)

1.8. The instability problem: a new justification for use of mixed-strategy equilibrium points. To illustrate the instability problem posed by games having only mixed-strategy equilibria, consider the following game:

	X	Y
A	45 30	0 90
B	30 75	60 45

Fig. 5

The only equilibrium in this game is in mixed strategies and has the form $E = (M,N)$, where $M = (1/3,\ 2/3)$ and $N = (4/5,1/5)$.

(In other words, player 1's equilibrium strategy M assigns the probabilities 1/3 and 2/3 to his two pure strategies A and B, respectively; while player 2's equilibrium strategy N assigns the probabilities 4/5 and 1/5 to his two pure strategies X and Y.) To facilitate analysis of this game, we will add a new row, corresponding to M, and a new column, corresponding to N, to the payoff matrix:

	X	Y	N
A	45 / 30	0 / 90	36 / 42
B	30 / 75	60 / 45	36 / 69
M	35 / 60	40 / 60	36 / 60

Fig. 6

As can be seen from this enlarged payoff matrix, if player 1 expects player 2 to use his equilibrium strategy N, then player 1 himself will have no real incentive to use his own equilibrium strategy M. This is so because he will obtain the same payoff $u_1 = 36$, regardless of whether he does use his mixed equilibrium strategy M, or uses either of his two pure strategies A and B,or uses any other mixed strategy whatever. Likewise, player 2 will have no real incentive to use his equilibrium strategy N even if he does expect player 1 to use his own equilibrium strategy M. This is so because player 2 will obtain the same payoff $u_2 = 60$, regardless of whether he uses his equilibrium strategy N,or uses either of his two pure strategies X and Y or uses any mixed strategy whatever.

This is what we mean by saying that the equilibrium point E = (M,N) is -- seemingly -- unstable: even if it does not provide any incentive for either player not to use his equilibrium strategy, it does not provide any incentive, either, that would make it positively attractive for him to _use_ his equilibrium strategy.

We now propose to argue that this instability of such mixed-strategy equilibrium points is only _apparent_. First of all, even if the players have as complete information about the payoff matrix of the game as they can possibly have, each player will always have some irreducible minimum of _uncertainty_ about the other player's actual payoffs. For example, even though the payoff matrix shows player 2's payoff associated with the strategy pair (A,X) to be $H_2(A,X) = 30$, player 1 will never be able to exclude the possibility that, at this very moment, this payoff may be in fact $30 - \varepsilon$ or $30 + \varepsilon$, where ε is a small positive number. This is so because every person's utility function is subject at least to some -- possibly very small -- unpredictable random fluctuation as a result of changes in his mood, or a possible sudden urge to use one of his pure strategies in preference to his other pure strategy, etc.

This means that a realistic model of any given game will not be one with _fixed_ payoffs but rather one with _randomly fluctuating_ payoffs, even though these fluctuations might be very small. Mathematical analysis shows that such a game will have no mixed-strategy equilibrium points at all. [6/] Rather, all its equilibrium points will be in pure strategies, in the sense that neither player will ever intentionally randomize between his two pure strategies: instead, he will always find that one of his two pure strategies will yield him a higher expected payoff, and this is the pure strategy that he will actually use.

At the same time, it can be shown that the random fluctuations in the two players' payoffs will interact in such a way that player 1 will find strategy A to be more profitable than strategy B almost exactly 1/3 of the time, and will find B more profitable than A almost exactly 2/3 of the time. As a result, even though he will make no attempt to randomize, he will in fact use his two pure strategies almost exactly with the probabilities actually prescribed by his equilibrium strategy M = (1/3, 2/3). By the same token, even though player 2 will make no attempt to randomize, he will in fact use his two pure strategies X and Y almost exactly with the probabilities prescribed by his equilibrium strategy N = (4/5, 1/5). [For detailed discussion and for mathematical proofs, see Harsanyi, 1973.]

To conclude, when a given game is interpreted as a game with fixed payoffs, then it will not provide any incentives for the players at a mixed-strategy equilibrium point to use their pure strategies with the probabilities prescribed by their equilibrium strategies. But if the game is, more realistically, reinterpreted as a game with randomly fluctuating payoffs, then it will actually provide the required incentives, so that the instability problem associated with mixed-strategy equilibrium points will disappear.

1.9. The imperfectness problem. In order to illustrate the imperfectness problem, we have to consider a game in extensive form, because in the normal form the distinction between perfect and imperfect equilibrium points often becomes unclear. We will use the following example:

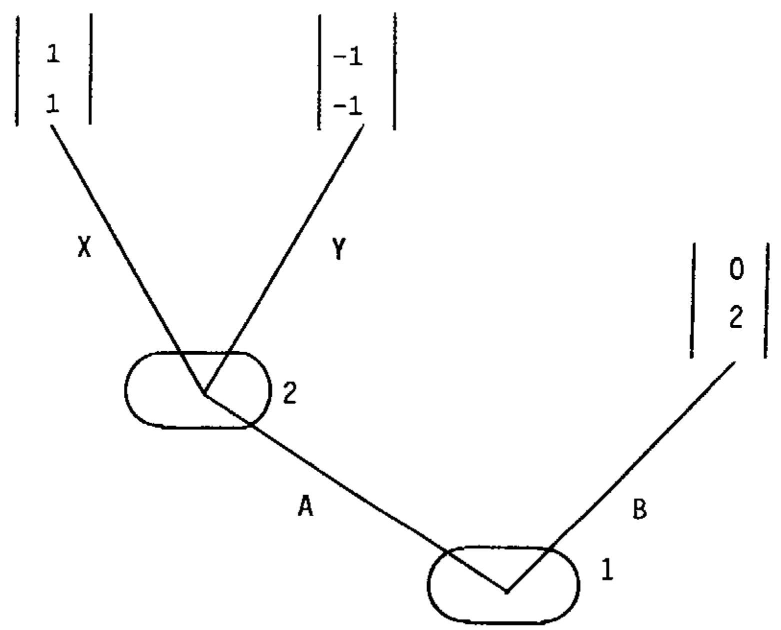

Fig. 7

In this game, player 1 has the first move: he can choose between move A and move B. If he chooses B then the game

will immediately end with the payoffs $u_1 = 0$ and $u_2 = 2$. But if he chooses A, then player 2 will also have a move: he will be able to choose between moves X and Y. If he chooses X then the payoffs will be $u_1 = u_2 = 1$ while if he chooses Y then the payoffs will be $u_1 = u_2 = -1$.
The normal form of this game is

	X	Y
A	1, 1	-1, -1
B	0, 2	0, 2

Fig. 8

Thus, player 1 has two pure strategies, viz. strategy A ("Choose move A") and strategy B ("Choose move B"). Player 2 also has two pure strategies, viz. strategy X ("Choose move X if player 1 has chosen A"), and strategy Y ("Choose move Y if player 1 has chosen A").

Obviously, the game has two pure-strategy equilibrium points, viz. the strategy pairs $E_1 = (A,X)$ and $E_2 = (B,Y)$. E_1 is a perfect equilibrium point. But we will now show that E_2 is an imperfect equilibrium point, involving irrational strategies.

In fact, strategy Y is surely irrational because it requires player 2 to choose move Y, which will yield him as well as player 1 the payoff $u_1 = u_2 = -1$, even though by choosing move X both he and player 1 could obtain the payoff $u_1 = u_2 = 1$. Strategy B is equally irrational: player 1 should know that if he chose move A then player 2 would surely choose move X, which would yield him the payoff $u_1 = 1$; therefore, player 1 should not choose move B which would yield him only $u_1 = 0$. [7/]

How is it possible that an equilibrium point should involve such irrational strategies? In particular, how can an equilibrium point require a player to choose a move like Y when choosing this move is inconsistent with maximizing his own payoff?

The answer is that a move like Y will in fact reduce a given player's expected payoff only if this move occurs with a positive probability. Yet, if the two players really act in accordance with equilibrium point E_2 = (B,Y), then player 2 will never come into a position of having to implement this irrational move Y, i.e., move Y will occur with zero probability and, therefore, will not actually reduce player 2's expected payoff.

More generally, an equilibrium strategy by difinition must maximize the relevant player's expected payoff if the other players' strategies are kept constant; and this means that no equilibrium strategy can possibly prescribe an irrational (i.e., a non-payoff-maximizing) move at any information set that will actually be reached with a positive probability if all players use their equilibrium strategies. But an equilibrium strategy may prescribe an irrational move for a given player at any information set that will be reached with zero probability. Imperfect equilibrium points are precisely those equilibtrium points that prescribe a move contrary to payoff maximization at some information set that will be reached with zero probability.

In terms of modern logic, the problem can be restated as follows. In our example, the assumption that player 2 will use strategy Y is equivalent to the following conditional statement S: "If player 1 were to make move A then player 2 would make move Y." If this conditional statement is interpreted as a Material Implication then it will automatically become vacuously true whenever the stated condition (viz. player 1's actually making move A) in fact does not arise. But if this statement S is interpreted as a Subjunctive Conditional - - and, grammatically, it is of course a Subjunctive Contional - - then this statement S will be simply false: If it really came to pass that player 1 made move A, then player 2 (assuming that he is a rational individual who tries to maximize his payoff) would most certainly not make move Y.

The strategy pair E_2 = (B,Y) _is_ formally an equilibrium point because for this to be the case, all we need is that statement S should be true when it is interpreted as a Material Implication. (It is, of course, a common practice in mathematics to regard any conditional statement as being true as long as it is true when interpreted as a Material Implication.) Nevertheless, our game-theoretical intuition judges E_2 to be an _irrational_ equilibrium point because this intuition would accept the truth of statement S only if it remained true even when interpreted as a Subjunctive Conditional - - which is obviously not the case.

Our distinction between perfect and imperfect equilibrium points is closely related to the question of whether the players can make any firm _commitment_ in a noncooperative game (see Sections 1.2 and 1.3 above). The game we have been discussing contains no self-commitment moves. Accordingly, our analysis has been based on the assumption that player 2 cannot commit himself in advance to choose move X rather than move Y at the time he actually reaches the information set where this choice has to be made. But it is easy to see that if he _could_ make such commitment then player 2 _would_ have a clear interest to do so (in order to frighten player 1 into making move A, rather than B). Yet, the proper way of enabling player 2 to make such commitment is to give him a _self-commitment move_ at the beginning of the game - - instead of misconstruing a game not containing anyself-commitment move as if it did contain one.

To add the desired self-commitment move, we can proceed as follows. At the beginning of the game, we permit player 2 a choice between moves α and β, where α can be interpreted as saying, "I commit myself to choose move Y if player 1 chooses move A," whereas β can be interpreted as saying, "I make no commitment." If player 2 actually chooses β then the future course of the game will be governed by subtree T_β which is an exact copy of the original game tree. In contrast, if he chooses α then the future course of the game will be governed by subtree T_α which differs from the original game tree in having move X omitted (since the latter is excluded by player 2's self-commitment move α).

Moreover, once branch X has been omitted, we can remove the entire information set from which branch X used to arise (since player 2 will not have any real choice any more at this point), together with the one remaining single branch Y, yet retaining the other components of T_α. This reduced version of subtree T_α will be called subtree T_α*. For the reader's convenience, T_α and T_α* are shown below:

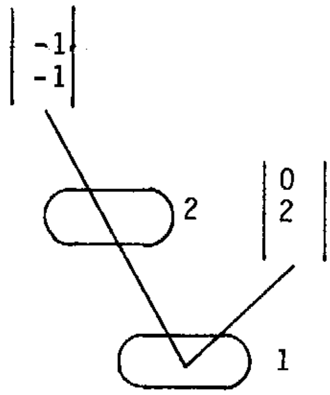

Fig. 9

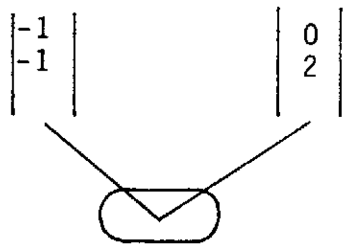

Fig. 10

Accordingly, the game tree of the new enlarged game will be as follows:

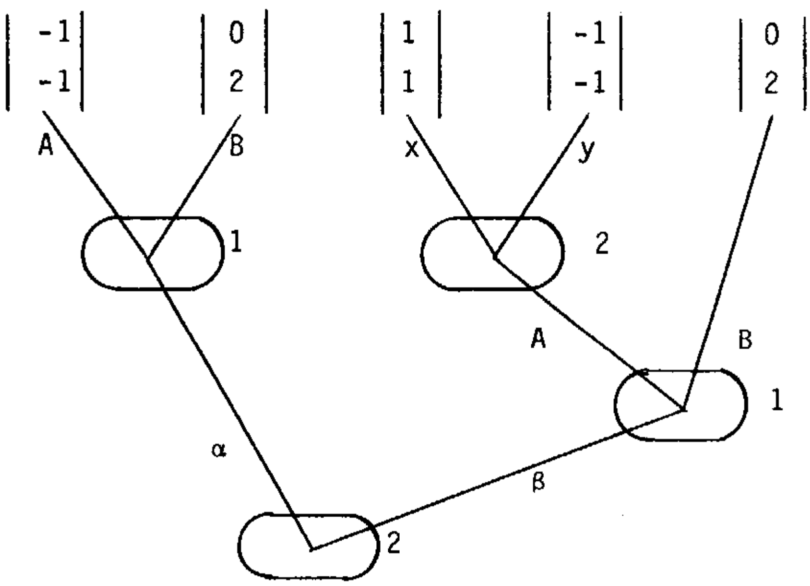

Fig. 11

In the normal form of the new game, each player will have four different pure strategies. Those of player 1 will be AA, AB, BA, and BB, where the first letter always indicates the move player 1 would make in subtree T_α* while the second letter indicates the move he would make in subtree T_β. Player 2's pure strategies are αX, αY, βX, and βY.

It is easy to verify that the new game has only one perfect equilibrium point in pure strategies, viz. $E_1^* = (BA,\alpha X)$. E_1^* has the following interpretation. At the beginning of the game player 2 will commit himself to make move Y, should player 1 make move A. This will deter player 1 from making move A. Instead, he will make move B, which will yield the payoffs (0,2), as desired by player 2. (Yet, if player 2 did not commit himself to move A, then the players would use the strategy pair (B,Y) instead.) Intuitively, we can identify E_1^* with the strategy pair $E_2 = (B,Y)$ of the original game - - except that in the original game this strategy pair was an imperfect equilibrium point whereas, in the new enlarged game, E_1^* is a perfect equilibrium point and, indeed, is the only perfect equilibrium point of the new game.

This is, of course, not at all surprising. Once player 2 is able to commit himself to use strategy Y, it becomes very rational for him to make this commitment and then, in view of this commitment made by him, it becomes likewise rational for player 1 to use strategy B, as desired by player 2. [8/]

1.10. Elimination of imperfect equilibtrium points from the game: the perturbed agent normal form. Up till now we have given only informal definitions for perfect and for imperfect equilibrium points: we have implicitly defined the former as equilibrium points assigning only rational moves (i.e. payoff-maximizing moves) to the players; and have implicitly defined the latter as equilibrium points assigning at least one irrational move (a move inconsistent with payoff maximization) to any player. Now we are going to describe a construction procedure which provides precise mathematical definitions for these two classes of equilibrium points and which at the same time also eliminates all imperfect equilibrium points from the game.

As a starting point of this construction procedure, we have to introduce a new game form, to be called the agent normal form, which in a sense is a game form intermediate between the extensive form and the ordinary normal form. The ordinary normal form itself could not be used for our purposes because it often obscures the difference between equilibrium points that do, and equilibrium points that do not, assign irrational moves to the player.This is so because two pure strategies of a given player which differ only in moves to be made by him at unreached information sets will be completely indistinguishable in the normal form, in the sense that both strategies will be associated with the very same payoff vectors if the other players' strategies are kept constant. They will be indistinguishable even if one strategy assigns very rational moves to the player at these unreached information sets whereas the other strategy assigns highly irrational moves to him.

To be sure, we could use the extensive form of the game as our starting point, but the latter is usually a much more complicated structure than the normal form is and is, therefore,less convenient to work with.

To put it differently, for our purposes the normal form contains too little information whereas the extensive form contains too much information - - including a good deal of information irrelevant from our point of view (such as the exact time order of the various moves, the detailed structure of the random moves, etc.).

The proposed intermediate game form can be constructed as follows. We replace every player by as many agents as the number of his information sets in the extensive form of the game. These agents will be considered to be the real players of the game. The pure strategies of each agent will be the alternative moves the main player used to have at the relevant information set, whereas his payoff function will be simply that of the main player. We define the agent normal form as the normal form of the new game played by these agents as the actual players.[9/]

In order to eliminate all equilibrium points assigning irrational moves to the players, we will use the following model. Suppose that player i (i.e. agent i, with $i = 1,\ldots,n$) has K_i different pure strategies, numbered as $1, 2,\ldots,K_i$. We will assume that, whenever player i tries to use a given pure strategy k, he will actually succeed in choosing the intended strategy only with probability $(1-\mu_{ik})$, where μ_{ik} is a very small positive number; and that with probability μ_{ik} he will always make a "mistake" by choosing one of his (K_i-1) unintended pure strategies $m \neq k$, the probability of his choosing any one particular unintended strategy m being η_{im}, where η_{im} is again a very small positive number.

Of course, the sum of these specific mistake probabilities η_{im} must be equal to the total probability μ_{ik} of making a mistake, so that

(1.10.1) $$\mu_{ik} = \sum_{m \neq k} \eta_{im} = \sum_{m=1}^{K_i} \eta_{im} - \eta_{ik}$$

for $i = 1,\ldots,n$ and for $k = 1,\ldots,K_i$. Moreover, by the positivity assumption,

(1.10.2) $$\mu_{ik} > 0 \text{ and } \eta_{ik} > 0$$

for $i = 1,\ldots,n$ and for $k = 1,\ldots,K_i$

A game G' we obtain, from any given game G in agent normal form, by imposing such mistake probabilities on the players will be called a perturbed agent normal form for this game G.

We have already pointed out that no equilibrium strategy q_i used by any player i at an equilibrium point $q = (q_1,\ldots,q_i,\ldots,q_n)$ can assign an irrational move to him at any information set that will be reached with a positive probability if all players use their equilibrium strategies $q_1,\ldots,q_i,\ldots,q_n$. Irrational moves can be assigned to him only at unreached information sets, i.e. at information sets reached with zero probability. Now, as is easy to verify, owing to assumption (1.10.2), in game G' all information sets will always be reached, regardless of the strategy combination chosen by the players. Consequently, G' will no longer contain any equilibrium points prescribing irrational moves for any player.

Next, we must make mathematically precise the assumption that all parameters μ_{ik} and η_{ik} are positive but are "very small". This can be done by letting all these parameters go to zero. More specifically, for any given game G in agent normal form, we will consider a sequence of perturbed agent normal forms $G^1, G^2, \ldots, G^j, \ldots$, satisfying

$$(1.10.3) \qquad \lim_{j\to\infty} \mu_{ik}^{\ j} = \lim_{j\to\infty} \eta_{ik}^{\ j} = 0$$

for $i = 1,\ldots,n$ and for $k = 1,\ldots,K_i$,

where $\mu_{ik}^{\ j}$ and $\eta_{ik}^{\ j}$ denote the parameters μ_{ik} and η_{ik} associated with game G^j ($j = 1,2,\ldots$). Such a sequence of perturbed games will be called a test sequence T for game G.

Consider a sequence of strategy combinations $q^1, q^2, \ldots, q^j, \ldots$, converging to a specific strategy combination q^o, with the property that each strategy combination $q^j (j=1,2,\ldots)$ is an equilibrium point of game G^j with the same superscript in the test sequence T. Then, q^o will be called a limit equilibrium point of this test sequence T. The set of all limit equilibrium points of T will be called L_T.

One can show that any such limit equilibrium point will be an actual equilibrium point of the original game G. Moreover, since every element q^j of the above sequence will be an equilibrium point making no use of irrational moves, the same will be true for their limit equilibrium point q^o. Accordingly, we can define the perfect equilibrium points of game G as those obtainable as limit equilibrium points of some test sequence for G; and we can define the imperfect equilibrium points of G as those not obtainable in this way. [For more detailed discussion and for the relevant mathematical proofs, see Selten, 1975].

1.11. The uniformly perturbed agent normal form. For many games, all possible test sequences T will generate the same set L_T of limit equilibrium points. But for other games, different test sequences will yield somewhat different sets L_T. In order that for every possible game we have a uniquely defined set L_T of limit equilibrium points from which to select a solution for the game, we will now make an additional assumption about the mistake probabilities η_{im}:

The uniformity assumption. The probability η_{im} that any given player i will choose any particular pure strategy m for his by mistake (when he is actually trying to use another pure strategy $k \neq m$) is the same for all players i and for all pure strategies m of any player i, so that

(1.11.1) $$\eta_{im} = \varepsilon \quad \text{for } i = 1,\dots,n \text{ and for } m = 1,\dots,K_i,$$

where ε is a very small positive number.

In view of (1.10.1),this implies that that the total probability μ_{ik} that any player i will make a mistake when he is trying to use any particular pure strategy k will be

(1.11.2) $$\mu_{ik} = (K_i-1)\varepsilon \quad \text{for } i = 1,\dots,n \text{ and for } k = 1,\dots,K_i.$$

Any perturbed game G_ε satisfying (1.11.1) and (1.11.2) for a specific value of ε will be called a uniformly perturbed agent normal form for game G.

A test sequence containing only uniformly perturbed agent normal forms G^1, G^2,... will be called a uniform test sequence. Any equilibrium point of game G that can be obtained as a limit equilibrium point of a uniform test sequence will be called a uniform perfect equilibrium point. Clearly, every uniform perfect equilibrium point will be a perfect equilibrium point but, in general, not every perfect equilibrium point will be uniform perfect. The set of these uniform perfect equilibrium points will be called L*. The uniformity assumption implies that we will have to choose the solution of any given game G from the set L* of its uniform perfect equilibrium points.

Operationally, the uniformity assumption means that, in analysing any given game G, our starting point must be the uniformly perturbed agent normal forms G_ε of G. More specifically, in order to define a solution for G, in the first instance we will always apply our solution theory, not to G itself, but rather to its uniformly perturbed agent normal forms G_ε. Suppose that, for any given game G_ε, our theory selects the equilibrium point q_ε of G_ε as a solution for G_ε. Then, the solution q* of the original game G will be defined as [10/]

$$q^* = \lim_{\varepsilon \to 0} q_\varepsilon . \quad (1.11.3)$$

<u>1.12. Analysis of cooperative games by means of noncooperative bargaining models</u>. The new solution concept we are proposing is formally a solution concept for noncooperative games. But actually it grew out of our research concerning cooperative games.

When it became clear to us that the Nash solution in its original form could not be used as a solution concept for two-person bargaining games with incomplete information [Harsanyi, 1967-68, pp. 349-334], we decided to follow Nash's [1951, p. 285] own suggestion that analysis of any *cooperative* game G should be based on a formal bargaining model B(G), involving bargaining moves and countermoves by the various players, and resulting in an agreement about the outcome of the game. Formally, this bargaining model B(G) would always be a *noncooperative* game in extensive form (or possibly in normal form), and the solution of the cooperative game G would be defined in terms of the equilibrium points of this *noncooperative* bargaining game B(G).

At the same time, we were fully aware of the fact that Nash's suggested approach could not possibly work unless we could find a way of overcoming at least the *multiplicity* problem (and, indeed, that it could not be fully successful unless we could resolve also the *instability* and the *imperfectness* problems).

Our first attempt to deal with the multiplicity problem was based on proposing an *ad hoc* modification of the Nash solution [1950b], specifically designed to overcome the multiplicity problem in that particular class of incomplete-information games we had been concerned with [Harsanyi and Selten, 1972]. But soon after this we came to the conclusion that thinking up new ad hoc solution concepts whenever a need for them arose was not really a satisfactory approach. Rather, a radically new theoretical departure was needed which would provide a general method of overcoming the multiplicity problem (as well as the instability and the imperfectness problems) for all possible noncooperative games.

Once these three problems can be overcome - - and our solution theory does overcome them - - an analysis of cooperative games by means of noncooperative bargaining models, as suggested by Nash, does provide a full remedy for the various problems posed by the classical theory of cooperative games. First of all, it yields a uniform approach to analyzing all classes of cooperative games. Even though different cooperative games may have to be analyzed in terms of very different bargaining models, the solution of each bargaining model, and therefore the solution of each cooperative game, will be defined in terms of the very same basic mathematical criteria, as specified by our solution theory.

Indeed, as we have already mentioned, the approach to be suggested will permit a unification, not only of cooperative game theory, but rather of game theory as a whole, including both the theories of cooperative and of noncooperative games. This is so because the problem of defining a solution for a cooperative game G will always be reduced to the problem of defining a solution for a noncooperative bargaining game B(G).

A further advantage of this approach is that it shows exactly how the solution (i.e. the theoretically predicted outcome) of any given cooperative game G will depend on the specific nature of the postulated bargaining process among the players, as indicated by the actual bargaining model B(G) used in analyzing this game G. For example, we can study how the outcome will depend on such factors as who can talk to whom, and who can talk to whom first, before anybody else can; what the rules are for concluding agreements, or for withdrawing from agreements already concluded, or again for making one's tentative agreements final and irrevocable; how easily coalitions can be formed, enlarged, dissolved, combined or recombined; what threats can be made by whom against whom, and to what extent such threats are irrevocable, etc.

In constructing bargaining models we can take advantage of the very great flexibility that bargaining games in extensive form provide for us - - a flexibility not available to the classical theory of cooperative games because of its insistence on using the much more restrictive normal form (or the even more rigid characteristic-function form). For example, we can easily construct bargaining models that represent cooperative games wholly inaccessible to classical theory, such as partially cooperative games or cooperative games possessing a sequential structure or cooperative games with incomplete information, etc.

Additional flexibility is provided by the fact that, by adding or not adding specific self-commitment moves, we can always give each player exactly the desired amount of self-commitment power in concluding enforceable agreements, and in making irrevocable promises and/or threats.

To sum up, we have considered the main difficulties of the classical theory of games, both in its cooperative and in its noncooperative branches. We have also indicated how our own solution theory proposes to resolve these difficulties.Of the three main difficulties arising in <u>noncooperative</u> game theory, the instability problem can be overcome by means of the theory of games with randomly fluctuating payoffs; the imperfectness problem can be overcome by means of the uniformly (or nonuniformly) perturbed agent normal form; while the multiplicity problem can be overcome by suitable mathematical criteria for selecting <u>one</u> specific equilibrium point as the solution for any given noncooperative game. Finally, the difficulties arising in the classical theory of <u>cooperative</u> games can be overcome by re-modelling any cooperative game as a noncooperative bargaining game.

Appendix 1.A1. The normal form of the enlarged game discussed in Section 1.3 is shown below. For easier readability, in the payoff matrix we have omitted the * symbols after the letters characterizing player 1's strategies, and have omitted the ** symbols after the letters characterizing player 2's strategies. Thus, for instance, player 1's strategy α*N*C* will be written simply as αNC, etc.

	αCC	αCN	αNC	αNN	βCC	βCN	βNC	βNN
αCC	10, 10	10, 10	10, 10	10, 10	10, 10	10, 10	-10, 11	-10, 11
αCN	10, 10	10, 10	10, 10	10, 10	10, 10	10, 10	-10, 11	-10, 11
αNC	10, 10	10, 10	10, 10	10, 10	11, -10	11, -10	1, 1	1, 1
αNN	10, 10	10, 10	10, 10	10, 10	11, -10	11, -10	1, 1	1, 1
βCC	10, 10	-10, 11	10, 10	-10, 11	10, 10	-10, 11	10, 10	-10, 11
βCN	11, -10	1, 1	11, -10	1, 1	11, -10	1, 1	11, -10	1, 1
βNC	10, 10	-10, 11	10, 10	-10, 11	10, 10	-10, 11	10, 10	-10, 11
βNN	11, -10	1, 1	11, -10	1, 1	11, -10	1, 1	11, -10	1, 1

Fig. 12

In the payoff matrix, cells corresponding to equilibrium points are marked by a heavy-set frame; and the cell corresponding to the unique perfect equilibrium point is indicated by a ♦ symbol.

As shown by the payoff matrix, apart from the perfect equilibrium pount E_1 = (α*N*N*, α**N**N**) the enlarged game also has five imperfect equilibrium points in pure strategies, viz. E_2 = (α*N*C*, α**C**N**), E_3 = (α*N*C*,α**N**N**),

E_4 = (α*N*N* α**C**N**), E_5 = (β*C*N*,β**N**N**), and

E_6 = (β*N*N* β**N**N**). Three of these, viz. E_2,E_3, and E_4, agree with the perfect equilibrium E_1 in involving commitments by both players to their C-stretagies, but differ from E_1 on the unreached subtrees T_2 and/or T_3.

In contrast, the other two imperfect equilibria, E_5 and E_6, involve refusals by both players to commit themselves to their C-strategies, and involve actual use of their N-strategies. In fact, E_6 can be intuitively identified with N = (N*, N**), the unique equilibrium of the original game which we called the noncooperative solution. E_5 is very similar to E_6 and differs from the latter only on the unreached subtree T_2.

Thus, we can say that, by introducing commitment moves into the game, we have made the noncooperative solution into an imperfect equilibrium and, therefore, into an outcome unavailable to rational players. (Cf. Sections 1.9 and 1.10 on the irrationality implied by imperfect equilibria). That is to say, in the enlarged game, rational players can always obtain the payoffs (10,10). Therefore, any strategy pair, such as E_5 and E_6,which yield only the payoffs (1.1),represents highly irrational behavior.

Appendix 1.A2. The normal form of the enlarged game discussed in Section 1.9 is shown below

	αX	αY	βX	βY
AA	-1, -1	-1, -1	1, 1	-1, -1
AB	-1, -1	-1, -1	0, 2	0, 2
BA	0, 2 ◆	0, 2	1, 1	-1, -1
BB	0, 2	0, 2	0, 2	0, 2

Fig. 13

As in Fig. 12, cells corresponding to equilibrium points are marked by heavy-set frame; and the cell corresponding to the unique perfect equilibrium point is indicated by a ◆ symbol.

As shown by the payoff matrix, the enlarged game has only one perfect equilibrium point, viz. $E_1^* = (BA,\alpha X)$. But it has five imperfect equilibrium points in pure strategies, viz. $E_2^* = (AA,\beta X)$, $E_3^* = (AB,\beta Y)$, $E_4^* = (BA,\alpha Y)$, $E_5^* = (BB,\alpha X)$, and $E_6^* = (BB,\alpha Y)$. Three of these imperfect equilibria, viz. E_4^*, E_5^*, and E_6^*, agree with the perfect equilibrium E_1^* in making player 2 commit himself to strategy Y, which, then, forces player 1 to use strategy B; but they differ from E_1^* on the unreached subtree T_β.

In contrast, E_2^* and E_3^* make player 2 refrain from any commitment (on the mistaken expectation that he could not induce player 1 to use strategy B even if he did commit himself to strategy Y - - this mistaken expectation is indicated by the letter A which is the first letter characterizing player 1's strategy). Nevertheless, E_3* still makes the two players use the strategy pair (B,Y) even though, as we have argued, this will represent irrational behavior when player 2 is not committed to strategy Y in advance.

On the other hand, E_2* makes the two players use the strategy pair (A,X) which is sensible enough once player 2 is not commited to strategy Y. Intuitively, E_2* can be identified with the strategy pair E_1 = (A,X) of the origional game - - except that, in the original game E_1 was a perfect equilibrium (in fact, the only one of the game) whereas, in the enlarged game, E_2* is an imperfect equilibrium because it is irrational for player 2 not to commit himself to strategy Y when he would be in a position to do so. This is so because, by making this commitment, he could obtain the payoff $u_2 = 2$ whereas, by not making this commitment, he reduces his payoff to $u_2 = 1$.

Footnotes

1/ A given strategy q_i of player i is a best reply to the other players' strategies $q_1,\ldots,q_{i-1}, q_{i+1},\ldots,q_n$ if this strategy q_i maximizes player i's payoff $H_i(q_1,\ldots,q_{i-1},q_i, q_i,q_{i+1},\ldots,q_n)$ when all other players' strategies are kept constant.

2/ A finite game is a game with a finite number of players, and with a finite number of pure strategies for every player.

3/ The great strategic importance of an ability or an inability to make firm commitments in playing a game was first pointed out by Schelling [1960].

4/ The distinction between perfect and imperfect equilibrium points will be discussed in Sections 1.9, 1.10, and 1.11 below.

5/ For a more detailed analysis of the enlarged game, see Appendix 1.A1 at the end of this chapter.

6/ Our theory assumes that the randon fluctuations in the payoffs are governed by an absolutely continuous joint probability distribution.

7/ It should be noted that E_2 = (B,Y) is an "undesirable" equilibrium point, not only because it is imperfect, but also because it uses a weakly dominated strategy (since strategy Y weakly dominates strategy X). In fact, it can be shown that, in any game containing only two information sets, an imperfect equilibrium point will always involve at least one weakly dominated strategy. But this theorem is not true for games containing three or more information sets. Therefore, the problem posed by imperfect equilibria cannot be reduced to the problem posed by dominated equilibria.

8/ For a more detailed analysis of the extended game, see Appendix 1.A2 at the end of this chapter.

9/ It can be shown that, for every game with perfect recall, its agent normal form will have exactly the same equilibrium points as the original game does in its extensive form (or, equivalently, in its normal form). [See Selten, 1975.] But in general this is not true for games with imperfect recall.

In view of this fact, we will always assume that any game we are dealing with has been modelled as a game with perfect recall. This is not a restrictive assumption because every game with imperfect recall can be easily transformed into one with perfect recall. This is so because any game with imperfect recall is always based on considering some team(s) (i.e.,some set(s) of players with identical interests) as single player(s); and it can always be transformed into a game with perfect recall by treating every member of any such team as a separate player. For instance, bridge is sometimes modelled as a two-person game with imperfect recall. But it can be just as readily modelled as a four-person game with perfect recall.

10/ We will follow the principle that, in the absence of specific reasons to the contrary, our analysis of any given game will always be based on the uniformity assumption and, therefore, on the uniformly perturbed agent normal form. In our opinion, the uniformity assumption is a very useful part of our theory: it is a rather natural assumption to make,and it greatly simplifies computation of the solution in many cases. But, even though it is very useful, it is not an indispensable assumption of our theory.

Should anybody feel that he had good reasons to think that in a given game G the players' mistakes would follow a nonuniform probability distribution, then all he had to do would be to select a specific family of nonuniform mistake-probability distributions Π_ε he felt was appropriate, and to use these distributions Π_ε for constructing the corresponding nonuniformly perturbed agent normal forms $\bar{G}_\varepsilon$ of game G. Then, he could apply our solution theory to these perturbed games $\bar{G}_\varepsilon$. All our theory would require is that these perturbed games should satisfy conditions (1.10.1) to (1.10.3).

On the other hand, if the analysis of any given game G were to be based on such a family of mistake-probability distributions Π_ε, then the latter would have to be included in the very definition of this game G, on a par with other defining characteristics, such as the players' strategy sets and payoff functions.

Thus, if two games had identical agent normal forms but were assumed to have different mistake-probability distributions Π_ε, then they would have to be regarded as being two different games and, therefore, our theory <u>might</u> very well define different solutions for them.

"WIRTSCHAFTSTHEORETISCHE ENTSCHEIDUNGSFORSCHUNG"

A series of books published by the Institute of Mathematical Economics, University of Bielefeld.

Wolfgang Rohde
Ein spieltheoretisches Modell eines Terminmarktes (A Game Theoretical Model of a Futures Market)
The model takes the form of a multistage game with imperfect information and strategic price formation by a specialist. The analysis throws light on theoretically difficult empirical phenomena.

Vol. 1 price: DM 24,80

Klaus Binder
Oligopolistische Preisbildung und Markteintritte (Oligopolistic Pricing and Market Entry)
The book investigates special subgame perfect equilibrium points of a three-stage game model of oligopoly with decisions on entry, on expenditures for market potential and on prices.

Vol. 2 price: DM 22,80

Karin Wagner
Ein Modell der Preisbildung in der Zementindustrie (A Model of Pricing in the Cement Industry)
A location theory model is applied in order to explain observed prices and quantities in the cement industry of the Federal Republic of Germany.

Vol. 3 price: DM 24,80

Rolf Stoecker
Experimentelle Untersuchung des Entscheidungsverhaltens im Bertrand-Oligopol(Experimental Investigation of Decision-Behavior in Bertrand-Oligopoly Games)
The book contains laboratory experiments on repeated supergames with two, three and five bargainers. Special emphasis is put on the end-effect behavior of experimental subjects and the influence of altruism on cooperation.

Vol. 4 price: DM 28,80

Angela Klopstech
Eingeschränkt rationale Marktprozesse (Marketprocesses with Bounded Rationality)
The book investigates two stochastic market models with bounded rationality, one model describes an evolutionary competitive market and the other an adaptive oligopoly market with Markovian interaction.

Vol. 5 price: DM 23,-- appr.

Orders should be sent to:

Pfeffersche Buchhandlung, Alter Markt 7, 4800 Bielefeld 1, West Germany.