GAMES WITH INCOMPLETE INFORMATION

Nobel Lecture, December 9, 1994

by

JOHN C. HARSANYI

Haas School of Business, University of California, Berkeley, USA

1. *Game theory and classical economics.*

Game theory is a theory of *strategic interaction.* That is to say, it is a theory of *rational behavior* in social situations in which each player has to choose his moves on the basis of what he thinks the other players' *countermoves* are likely to be.

After preliminary work by a number of other distinguished mathematicians and economists, game theory as a systematic theory started with von Neumann and Morgenstern's book, *Theory of Games and Economic Behavior,* published in 1944. One source of their theory was reflection on *games of strategy* such as chess and poker. But it was meant to help us in defining rational behavior also in *real-life* economic, political, and other social situations.

In principle, every social situation involves strategic interaction among the participants. Thus, one might argue that proper understanding of *any* social situation would require game-theoretic analysis. But in actual fact, classical economic theory did manage to *sidestep* the game-theoretic aspects of economic behavior by postulating *perfect competition*, i.e., by assuming that every buyer and every seller is *very small* as compared with the size of the relevant markets, so that nobody can significantly affect the existing market prices by his actions. Accordingly, for each economic agent, the prices at which he can buy his inputs (including labor) and at which he can sell his outputs are essentially *given* to him. This will make his choice of inputs and of outputs into a *one-person* simple maximization problem, which can be solved without game-theoretic analysis.

Yet, von Neumann and Morgenstern realized that, for most parts of the economic system, perfect competition would now be an *unrealistic* assumption. Most industries are now dominated by a *small number* of *large* firms, and labor is often organized in *large* labor unions. Moreover, the central government and many other government agencies are major players in many markets as buyers and sometimes also as sellers, as regulators, and as taxing and subsidizing agents. This means that game theory has now definitely become an important analytical tool in understanding the operation of our economic system.

2. *The problem of incomplete information.*

Following von Neumann and Morgenstern [1947, p. 30], we may distinguish between games with *complete information,* here often to be called C-games, and *games* with *incomplete information,* to be called *I-games.* The latter differ from the former in the fact that the players, or at least some of them, *lack* full information about the *basic mathematical structure* of the game as defined by its normal form (or by its extensive form).

Yet, even though von Neumann and Morgenstern did distinguish between what I am calling C-games and I-games, their own theory (and virtually all work in game theory until the late 1960s) was restricted to C-games.

Lack of information about the mathematical structure of a game may take many different forms. The players may lack full information about the other players' (or even their own) payoff functions, about the physical or the social resources or about the strategies available to other players (or even to them themselves), or about the amount of information the other players have about various aspects of the game, and so on.

Yet, by suitable modelling, *all* forms of incomplete information can be reduced to the case where the players have less than full information about each other's *payoff functions*[1] U_i, defining the *utility payoff* $u_i = U_i(s)$ of each player i for any possible strategy combination $s = (s_1,..,s_n)$ the n players may use.

TWO-PERSON I-GAMES

3. A model based on higher and higher-order expectations.

Consider a two-person I-game G in which the two players do not know *each other's* payoff functions. (But for the sake of simplicity I shall assume that they do know *their own* payoff functions.)

A very natural - yet as we shall see a rather *impractical* - model for analysis of this game would be as follows. Player 1 will realize that player 2's strategy s_2 in this game will depend on player 2's own payoff function U_2 Therefore, before choosing his own strategy s_1, player 1 will form some *expectation* e_1U_2 about the nature of U_2. By the same token, player 2 will form some *expectation* e_2U_1 about the nature of player 1's payoff function U_1, These two expectations e_1U_2 and e_2U_1 I shall call the two players' *first-order* expectations.

Then, player 1 will form some second-order expectation $e_1e_2U_1$ about player 2's first-order expectation e_2U_1 whereas player 2 will form some *second-order* expectation $e_2e_1U_2$ about player 1's first-order expectation e_1U_2 and so on.

Of course, if the two players want to follow the *Bayesian approach* then their expectations will take the form of *subjective probability distributions* over the relevant mathematical objects. Thus, player 1's first order expectation e_1U_2 will take the form of a subjective probability distribution $P_1{}^1(U_2)$ over all possible payoff functions U_2 that player 2 may possess. Likewise, player *2's first-order* expectation e_2U_1 will take the form of a subjective probability distribution $P_2{}^1(U_1)$ over all possible payoff functions U_1 that player 1 may possess.

On the other hand, player *1's second-order* expectation $e_1e_2U_1$ will take the form of a subjective probability distribution $P_1^2(P_2^1)$ over all possible first-order probability distributions P_2^1 that player 2 may entertain. More generally, the kth-order expectation *(k>1)* of either player i will be a subjective probability distribution $P_i^k(P_j^{k-1})$ over all the $(K-1)$-order subjective probability distributions P_j^{k-1} that the other player j $(j \neq i)$ may have chosen.[2]

Of course, any model based on higher and higher-order expectations would be even more complicated in the case of *n-person* I-games (with n > 2). Even if we retain the simplifying assumption that each player will know his own payoff function, even then each player will still have to form $(n - 1)$ different *first-order* expectations, as well as $(n - 1)^2$ different *second-order* expectations, and so on.

Yet, as we shall see, there is a *much simpler* and very much preferable approach to analyzing I-games, one involving only *one* basic probability distribution Pr (together with n different *conditional* probability distributions, all of them generated by this basic probability distribution Pr).

4. Arms control negotiations between the United States and the Soviet Union in the 1960s.

In the period 1965 - 69, the U.S. Arms Control and Disarmament Agency employed a group of about ten young game theorists as consultants. It was as a member of this group that I developed the simpler approach, already mentioned, to the analysis of I-games.

I realized that a major problem in arms control negotiations is the fact that each side is relatively *well informed* about *its own position* with respect to various variables relevant to arms control negotiations, such as its own policy objectives, its peaceful or bellicose attitudes toward the other side, its military strength, its own ability to introduce new military technologies, and so on - but may be *rather poorly informed* about the *other side's* position in terms of such variables.

I came to the conclusion that finding a suitable mathematical representation for this particular problem may very well be a *crucial key* to a better theory of arms control negotiations, and indeed to a better theory of all I-games.

Similar problems arise also in economic competition and in many other social activities. For example, business firms are almost always better informed about the economic variables associated with *their own* operations than they are about those associated with their *competitors'* operations.

Let me now go back to my discussion of arms control negotiations. I shall describe the *American* side as *player I,* and shall describe the *Soviet* side, which I shall often call the *Russian* side, as *player 2.*

To model the *uncertainty* of the Russian player about the true nature of the *American player; i.e.,* about that of *player I,* I shall assume that there are K *different* possible *types* of player 1, to be called types $t_1^1, t_1^2, \ldots, t_1^k, \ldots, t_1^K$. The Russian player, i.e., player 2, *will not know* which *particular type* of player 1 will actually be representing the American side in the game.

Yet, this fact will pose a serious problem for the Russian player because his own strategical possibilities in the game will obviously depend, often very strongly, on which *particular type* of the American player will confront him in the game. For each of the K possible types of this player might correspond to a *very different combination* of the possible characteristics of the American player - in terms of variables ranging from the true intentions of this American player to the availability or unavailability of powerful new military technologies to him, technologies sometimes very contrary to the Russian side's expectations. Moreover, *different* types of the American player might differ from each other also in entertaining different *expectations* about the true nature of the Russian player.

On the other hand, to model the *uncertainty* of the American player about the true nature of the *Russian* player i.e., about that of *player 2*, I shall assume that there are M *different* possible *types* of player 2, to be called types $t_2^1, t_2^2, \ldots, t_2^m, \ldots, T_2^M$. The American player, i.e., player 1, *will not know* which *particular type* of player 2 will actually represent the Russian side in the game.

Again, this fact will pose a serious problem for the American player because each of the M possible types of the Russian player might correspond to a very *different combination* of the possible characteristics of the Russian player. Moreover, *different* types of the Russian player might differ from each other also in entertaining different *expectations* about the true nature of the American player.[3]

5. A type-centered interpretation of I-games.

A C-game is of course always analyzed on the assumption that the *centers of activity* in the game are its players. But in the case of an I-game we have a choice between two alternative assumptions. One is that its centers of activity are its players, as would be the case in a C-game. The other is that its centers of activity are the various *types* of its players. The former approach I shall call a player-centered interpretation of this I-game, whereas the latter approach I shall call its *type-centered* interpretation.

When these two interpretations of any I-game are properly used, then they are always *equivalent* from a game-theoretic point of view. In my 1967 - 68 paper I used the *player-centered* interpretation of I-games. But in this paper I shall use their *type-centered* interpretation because now I think that it provides a more convenient *language* for the analysis of I-games.

Under this latter interpretation, when player *1* is of type t_1^k, then the strategy and the payoff of *player 1* will be described as the strategy and the payoff of this *type* t_1^k of player 1 rather than as those of player 1 as such. This language has the advantage that it enables us to make certain statements about type t_1^k *without* any need for further *qualifications*, instead of making similar statements about player *1* and *then* explaining that these statements apply to him only when he is of type t_1^k. This language is for us also a useful reminder of the fact that in any I-game the strategy that a given player will use and the payoff he will receive will often *strongly depend* on whether this player is of *one* type or is of *another* type.

On the other hand, one must keep in mind that any statement about a *given type* t_1^k can always be retranslated into *player-centered* language so as to make it into a statement about *player 1* when he *is* of type t_1^k.

A *type-centered* language about *player* 2 when he is of some *type* t_2^m can be defined in a similar way.

6. *The two active types and their payoff functions.*

Suppose that player 1 is of type t_1^k, whereas player 2 is of type t_2^m. Then we shall say that the two players are *represented* by their types t_1^k and t_2^m, and that these two types are the two *active types* in the game. In contrast, all types $t_1^{k'}$ with $k' \neq k$ and all types $t_2^{m'}$ with $m' \neq m$ will be called *inactive types.*

In a two-person C-game, the payoff of either player will depend only on the *strategies* used by the two players. In contrast, in a two-person I-game the payoffs v_1^k and v_2^m of the two active types t_1^k and t_2^m will depend not only on these two types' *strategies* s_1^k and s_2^m (pure or mixed) but also on their *types* as indicated by the *superscripts* k and m in the symbols t_1^k and t_2^m denoting them. Thus, we may define their payoffs v_1^k and v_2^m as

$$v_1^k = V_1^k(s_1^k, s_2^m; k, m) \tag{1}$$

and

$$v_2^m = V_2^m(s_1^k, s_2^m; k, m) \tag{2}$$

where V_1^k and V_2^m denote the payoff functions of t_1^k and of t_2^m.

Yet, I shall call V_1^k and V_2^m *conditional* payoff functions because the payoff of type t_1^k will be the quantity v_1^k defined by (1) *only* if t_1^k is an *active type* in the game and if the *other* active type in the game is t_2^m. Likewise, the payoff of type t_2^m will be the quantity v_2^m defined by (2) *only* if t_2^m is an *active type* and *if* the *other* active type is t_1^k.

More particularly, if either t_1^k or t_2^m is an *inactive type* then he will *not* be an actual participant of the game and, therefore, will *not* receive *any* payoff (or will receive only a *zero* payoff).

7. *Who will know what in the game.*

For convenience I shall assume that the *mathematical forms* of the two payoff functions v_1^k and v_2^m *will be known to all participants* of the game. That is to say, they will be known to *both players* and to *all types* of these two players.

On the other hand, I shall assume that player 1 *will know which particular type* t_1^k of his is representing him in the game. Likewise, player 2 *will know which particular type* t_2^m of his is representing him. In contrast, to model the *uncertainty* of each player about the true nature of the *other* player, I shall assume that *neither* player *will know which particular type* of the *other* player is representing the latter in the game.

In terms of *type-centered* language, these assumptions amount to saying that *all types* of both players *will know* that they are *active types* if they in fact *are*.

Moreover, they will know *their own identities*. (Thus, e.g., type t_1^3 will know that he is t_1^3, etc.) In contrast, *none* of the types of *player 1* will know the identity of *player* 2's active type t_2^m; and *none* of the types of *player 2* will know the identity of *player 1* 's active type t_1^k.

8. Two important distinctions.

As we have already seen, one important distinction in game theory is that between games with *complete* and with *incomplete* information, i.e., between C-*games* and *I-games*. It is based on the amount of information the players will have in various games about the *basic mathematical structure* of the game as defined by its normal form (or by its extensive form). That is to say, it is based on the amount of information the players will have about those characteristics of the game that must have been decided upon *before* the game can be played at all.

Thus, in C-games all players will have full information about the basic mathematical structure of the game as just defined. In contrast, in *I-games* the players, or at least some of them, will have only partial information about it.

Another, seemingly similar but actually quite different, distinction is between games with *perfect* and with *imperfect* information. Unlike the first distinction, this one is based on the amount of information the players will have in various games about the *moves* that occurred at *earlier stages* of the game, i.e., about some events that occurred *during* the time when the game was actually played, rather than about some things decided upon *before* that particular time.

Thus, in games with *perfect* information, all players will have full information at every stage of the game about *all moves* made at earlier stages, including both *personal moues* and *chance moves.* [4] In contrast, in games with *imperfect* information, at some stage(s) of the game the players, or at least some of them, will have only partial information or none at all about some move(s) made at earlier stages.

In terms of this distinction, chess and checkers are games with *perfect* information because they do permit both players to observe not only their own moves but also those of the other player.

In contrast, most card games are games with *imperfect* information because they *do not* permit the players to observe the cards the other players have received from the dealer, or to observe the cards discarded by other players with their faces down, etc.

Game theory as first established by von Neumann and Morgenstern, and even as it had been further developed up to the late 1960s was restricted to games with *complete* information. But from its very beginning, it has covered all games in that class, regardless of whether they were games with *perfect* or with *imperfect* information.

9. A probabilistic model for our two-person I-game G.

Up till now I have always considered the *actual types* of the two players, represented by the *active pair* (t_1^k, t_2^m) simply as *given*. But now I shall propose to *enrich* our model for this game by adding some suitable formal representation of the *causal factors* responsible for the fact that the American and the Russian player have characteristics corresponding to those of (say) types t_1^k and t_2^m in our model.

Obviously, these causal factors can only be *social forces* of various kinds, some of them located in the United States, others in the Soviet Union, and others again presumably in the rest of the world.

Yet, it is our common experience as human beings that the results of social forces seem to admit only of *probabilistic* predictions. This appears to be the case even in situations in which we are exceptionally *well informed* about the relevant social forces: Even in such situations the best we can do is to make *probabilistic* predictions about the results that these social forces may produce.

Accordingly, I shall use a random mechanism and, more particularly, a *lottery* as a formal representation of the *reluctant social forces*, i.e., of the social forces that have produced an American society of *one* particular type (corresponding to some type t_1^k of our model), and that have also produced a Russian society of *another* particular type (corresponding to some type t_2^m of our model).

More specifically, I shall assume that, *before any other moves are made* in game G, some lottery, to be called *lottery L*, will choose some type t_1^k as the type of the American player, as well as some type t_2^m as the type of the Russian player. I shall assume also that the *probability* that any *particular* pair (t_1^k, t_2^m) is chosen by this lottery L will be

$$Pr(t_1^k, t_2^m) = p_{km} \text{ for } k = 1,\ldots,K \text{ and for } m = 1,\ldots,M. \tag{3}$$

As player 1 has K different possible types whereas player 2 has M different possible types, lottery L will have a choice among H = KM different pairs of the form (t_1^k, t_2^m). Thus, to characterize its choice behavior we shall need H different probabilities p_{km}.

Of course, all these H probabilities will be *nonnegative* and will add up to *unity*. Moreover, they will form a K x M *probability matrix* $[p_{km}]$, such that, for all possible values of k and of m, its *kth row* will correspond to type t_1^k of player 1 whereas its *mth column* will correspond to type t_2^m of player 2.

I shall assume also that the two players will try to estimate these H probabilities on the basis of their information about the nature of the *relevant social forces*, using only information available to *both of them*. In fact, they will try to estimate these probabilities as an *outside observer* would do, one restricted to information *common* to both players (cf. Harsanyi, 1967 - 68, pp. 176 - 177). Moreover, I shall assume that, unless he has information to the contrary, each player will act on the assumption that the *other player* will estimate these probabilities p_{km} *much in the same* way as *he* does. This is often called the *common priors* assumption (see Fudenberg and Tirole, 1991, p. 210).

Alternatively, we may simply assume that both players will act on the assumption that *both of them know* the true numerical values of these probabilities p_{km} - so that the *common priors* assumption will follow as a *corollary*.

The mathematical model we obtain when we add a lottery L (as just described) to the two-person I-game described in sections 4 to 7 will be called a *probabilistic model* for this I-game G. As we shall see presently, this probabilistic model will actually convert this *I*-game G into a *C-game,* which we shall call the game G*.

10. *Converting our I-game G with* ***incomplete*** *information into* a *game G* with* ***complete*** *yet with* ***imperfect*** *information.*

In this section, I shall be using *player-centered* language because this is the language in which our traditional definitions have been stated for games with complete and with incomplete information as well as for games with perfect and with imperfect information.

Let us go back to the two-person game G we have used to model arms control negotiations between the United States and the Soviet Union. We are now in a better position to understand *why* it is that, under our original assumptions about G, it will be a game with *incomplete* information.

(i) First of all, under our original assumptions, player 1 is of type t_1^k, which I shall describe as *Fact I,* whereas player 2 is of type t_2^m, which I shall describe as *Fact II*. Moreover, both Facts I and II are established facts *from the very beginning* of the game, and they are not facts brought about by *some move(s)* made *during* the game. Consequently, these two facts must be considered to be parts of the basic *mathematical structure* of this game G.

(ii) On the other hand, according to the assumptions we made in section 7, player 1 *will know* Fact I but will *lack* any knowledge of Fact II. In contrast, player 2 will know Fact II but will *lack* any knowledge of Fact I.

Yet, as we have just concluded, *both* Facts I and II are parts of the basic mathematical structure of the game. Hence, *neither* player 1 *nor* player 2 will have full information about this structure. Therefore, under our original assumptions, G is in fact a game with *incomplete* information.

Let me now show that as soon as we reinterpret game G in accordance with our *probabilistic model,* i.e., as soon as we add *lottery* L to the game, our original game G will be *converted* into a new game G* with *complete* information. Of course, even after this reinterpretation, our statements under (ii) will *retain* their validity. But the status of Facts I and II as stated under (i) will undergo a *radical change*. For these two Facts will now become the results of a *chance move* made by lottery L *during* the game and, therefore, will *no longer* be parts of the basic mathematical structure of the game. Consequently, the fact that *neither* player will know *both* of these two Facts will no longer make the new game G* into one with *incomplete* information.

To the contrary, the new game G* will be one with *complete* information because its basic mathematical structure will be defined by our *probabilistic model* for the game, which will be *fully known* to *both* players.

On the other hand, as our statements under (ii) do retain their validity even in game G*, the latter will be a game with *imperfect* information because both players will have only *partial information* about the pair (t_1^k, t_2^m) chosen by the *chance move* of lottery L at the beginning of the game.

11. Some ***conditional*** *probabilities in game G*.*

Suppose that lottery L has chosen type t_1^k to represent player 1 in the game. Then, according to our assumptions in section 7, type t_1^k *will know* that he now has the status of an *active type* and *will know* that he is *type* t_1^k. But he *will not know* the identity of the *other active type* in the game.

How should t_1^k now assess the *probability* that the *other active type* is actually a *particular type* t_2^m of player 2? He must assess this probability by using the information he does have, viz. that *he*, type t_1^k, is one of the two *active* types. This means that he must assess this probability as being the *conditional probability*[5]

$$\pi_1^k(m) = Pr(t_2^m | t_1^k) = p_{km} / \sum_{k=1}^{K} p_{km}. \tag{4}$$

On the other hand, now suppose that lottery L has chosen type t_2^m to represent player 2 in the game. Then, how should t_2^m assess the *probability* that the *other active type* is a *particular type* t_1^k of player 1? By similar reasoning, he should assess this probability as being the *conditional probability*

$$\pi_2^m(k) = Pr(t_1^k | t_2^m) = p_{km} / \sum_{m=1}^{M} p_{km}. \tag{5}$$

*12. The **semi**-conditional payoff functions of the two active types.*

Suppose the *two active types* in the game are t_1^k and t_2^m. As we saw in section 6, under this assumption, the payoffs v_1^k and v_2^m of these two active types will be defined by equations (1) and (2).

Note, however, that this payoff v_1^k defined by (1) will *not* be the quantity that type t_1^k will try to maximize when he chooses his strategy s_1^k. For he *will not know* that his *actual* opponent in the game will be type t_2^m. Rather, all he will know is that his opponent in the game will be *one* of player 2's M types. Therefore, he will choose his strategy s_1^k so as to protect his interests not only against his unknown *actual* opponent t_2^m but rather against *all* M types of player 2 because, for all he knows, *any* of them *could* be now his opponent in the game.

Yet, type t_1^k will know that the *probability* that he will face any particular type t_2^m as opponent in the game will be equal to the *conditional probability* $\pi_1^k(m)$ defined by (4). Therefore, the quantity that t_1^k will try to maximize is the *expected value* u_1^k of the payoff v_1^k which can be defined as

$$u_1^k = U_1^k(s_1^k, s_2^*) = \sum_{m=1}^{M} \pi_1^k(m) V_1^k(s_1^k, s_2^m; k, m). \tag{6}$$

Here the symbol s_2^* stands for the strategy M-tuple[6]

$$s_2^* = (s_2^1, s_2^2, \ldots, s_2^m, \ldots, s_2^M). \tag{7}$$

I have inserted the symbol s_2^* as the second argument of the function U_1^k in order to indicate that the *expected payoff* u_1^k of type t_1^k will depend not only on the strategy s_2^m that his *actual* unknown opponent t_2^m *will use* but rather on the strategies $s_2^1,\ldots,s_2^M$ that anyone of his M *potential* opponents $t_2^1,\ldots,t_2^M$ *would use* in case he were chosen by lottery L as t_1^k's opponent in the game.

By similar reasoning, the quantity that type t_2^m will try to maximize when he chooses his strategy s_2^m will *not* be his payoff v_2^m defined by (2). Rather, it will be the *expected value* u_2^m of this payoff v_2^m, defined as

$$u_2^m = U_2^m(s_1^*, s_2^m) = \sum_{k=1}^{K} \pi_2^m(k) V_2^m(s_1^k, s_2^m; k, m). \tag{8}$$

Here the symbol s_1^* stands for the strategy K-tuple

$$s_1^* = (s_1^1, s_1^2, \ldots, s_1^k, \ldots, s_1^K). \tag{9}$$

Again, I have inserted the symbol s_1^* as the first argument of the function U_2^m in order to indicate that the *expected payoff* of type t_2^m will depend on *all* K strategies $s_1^1,\ldots,s_1^K$ that anyone of the K types of player 1 would use against him in case he were chosen by lottery L as t_2^m's opponent in the game.

As distinguished from the *conditional* payoff functions V_1^k and V_2^m used in (1) and (2), the payoff functions U_1^k and U_2^m used in (6) and in (8) I shall describe as *semi-conditional.* For V_1^k and V_2^m define the *payoff* v_1^k or v_2^m of the relevant type as being dependent on the *two conditions* that

(a) He himself must have the status of an active type and that

(b) The other active type in the game must be a specific type of the other player.

In contrast, U_1^k and U_2^m define the *expected* payoff u_1^k or u_2^m of the relevant type as being *independent* of condition (b) yet as being *dependent* on condition (a). (For it will still be true that neither type will receive any payoff at all if he is not given by lottery L the status of an *active type* in the game.)

As we saw in section 10, once we reinterpret our original I-game G in accordance with our *probabilistic model* for it, G will be converted into a C-game G*. Yet, under its *type-centered* interpretation, this C-game G* can be regarded as a *(K+M)-person* game whose real "players" are the K *types* of player 1 and the M *types* of player 2, with their basic payoff functions being the *semi-conditional* payoff functions U_1^k(k=1,. . .,K) and U_2^m(ml,. . .,M).

If we regard these (K+M) types as the real "players" of G* and regard these payoff functions U_1^k and U_2^m as their real payoff functions, then we can easily define the *Nash equilibria*7 of this C-game G*. Then, using a suitable theory of equilibrium selection, we can define one of these equilibria as the solution of this game.

N-PERSON I-GAMES

13. The types of the various players, the active set, and the appropriate sets in n-person I-game.5

Our analysis of two-person I-games can be easily extended to n-person I-games. But for lack of space I shall have to restrict myself to the basic essentials of the n-person theory.

Let N be the *set* of all n players. I shall assume that any player i (i=1,...,n) will have K_i different possible types, to be called $t_i^1,\ldots,t_i^k,\ldots,t_i^{K_i}$. Hence, the *total number* of different types in the game will be

$$Z = \sum_{i\in N} K_i. \tag{10}$$

Suppose that players $1,\ldots,i,\ldots,n$ are now represented by their types $t_1^{k_1},\ldots,t_i^{k_i},\ldots,t_n^{k_n}$ in the game. Then, the set of these n types will be called the *active set.* $\bar{a}$

Any set of n types containing exactly *one* type of *each* of the n players *could* in principle play the role of an active set. Any such set will be called an *appropriate set.* As any player i has K_i different types, the *number* of different appropriate sets in the game will be

$$H = \prod_{i\in N} K_i. \tag{11}$$

I shall assume that these H appropriate sets a will have been *numbered* as

$$a_1, a_2,\ldots,a_h,\ldots,a_H. \tag{12}$$

Let A_i^k be the *family* of all appropriate sets containing a particular type t_i^k of some player i as their *member* The *number* of different appropriate sets in A_i^k will be

$$a(i) = \prod_{\substack{j\in N\\ j\neq i}} K_j = H/K_j. \tag{13}$$

Let B_i^k be the set of all *subscripts h* such that a_h is in A_i^k. As there is a one-to-one correspondence between the members of A_i^k and the members of B_i^k, this set B_i^k, will likewise have $\alpha(i)$ different members.

14. *Some probabilities.*
I shall assume that, *before any other moues are made* in game G*, some lottery L will choose one particular *appropriate set* a to be the *active set* $\bar{a}$ of the game. The n types in this set $\bar{a}$ will be called *active types* whereas all types *not* in $\bar{a}$ will be called *inactive types.*

I shall assume that the *probability* that a *particular* appropriate set a_h will be chosen by lottery L to be the active set $\bar{a}$ of the game is

$$Pr(\bar{a} = a_h) = r_h \text{ for } h = 1,\ldots,H. \tag{14}$$

Of course, all these H probabilities r_h will be *nonnegative* and will add up to *unity*. Obviously, they will correspond to the H probabilities p_{km} [defined by (3)] we used in the two-person case.

Suppose that a particular type t_i^k of some player i has been chosen by lottery L to be an *active type* in the game. Then, under our assumptions, he *will know* that he is type t_i^k and *will know* also that he now has the status of an *active type*. In other words, t_i^k will know that

$$t_i^k \in \bar{a}. \tag{15}$$

Yet, the statement $t_i^k \in \bar{a}$ implies the statement

$$\bar{a} \in \mathrm{A}_i^k \tag{16}$$

and conversely, because A_i^k contains exactly those appropriate sets that have type t_i^k as their *member:* Thus, we can write

$$(t_i^k \in \bar{a}) \leftrightarrow (\bar{a} \in \mathrm{A}_i^k). \tag{17}$$

We have already concluded that if type t_i^k has the status of an *active type* then he will know (15). We can now add that in this case he will know also (16) and (17). On the other hand, he can also easily compute that the *probability* for lottery L to choose an active set $\bar{a}$ belonging to the family A_i^k is

$$Pr(\bar{a} \in \mathrm{A}_i^k) = \sum_{h \in B_i^k} r_h. \tag{18}$$

In view of statements (15) to (18), how should this type t_i^k assess the *probability* that the active set $\bar{a}$ chosen by lottery L is actually a *particular* appropriate set a,? Clearly, he should assess this probability as being the *conditional* probability

$$\pi_i^k(h) = Pr(\bar{a} = a_h | t_i k \in \bar{a} \tag{19}$$

Yet, in view of (17) and (18), we can write

$$\begin{aligned} Pr(\bar{a} = a_h | t_i^k \in \bar{a}) &= Pr(\bar{a} = a_h | \bar{a} \in A_i^k) = \\ = Pr(\bar{a} = a_h / Pr(\bar{a} \in A_i^k) &= r_h / \sum_{h \in B_i^k} r_h. \end{aligned} \tag{20}$$

Consequently, by (19) and (20) the required *conditional* probability is

$$\pi_i^k(h) = r_h / \sum_{h \in B_i^k} r_h. \tag{21}$$

15. Strategy profiles.
Suppose that the K_i types $t_i^1, \ldots, t_i^k, \ldots, t_i^{K_i}$ of player i *would* use the strategies $s_i^1, \ldots, s_i^k, \ldots, s_i^{K_i}$ (pure or mixed) in case they were chosen by lottery L to be *active types* in the game. (Under our assumptions, *inactive types* do not actively participate in the game and, therefore, do *not* choose any strategies.) Then I shall write

$$s_i{}^* = (s_i^1, \ldots, s_i^k, \ldots, s_i^{K_i}) \text{ for } i = 1, \ldots, n \tag{22}$$

to denote the *strategy profile*[8] of the K_i types of player *i.*

Let

$$s^* = (s_i^*, \ldots, s_n^*) \tag{23}$$

be the ordered set we obtain if we first list all K_1 strategies in s_1^*, then all K_2 strategies in s_2^*, . . . , then all K_i strategies in s_i^*,. . , and finally all K_n strategies in s_n^*. Obviously, s* will be a *strategy profile* of *all* types in the game. In view of (10), s* will contain Z different strategies.

Finally, let s*(*h*) denote the *strategy profile* of the *n* types belonging to a *particular* appropriate set a_h for $h = 1, \ldots, H$.

16. The conditional payoff functions.
Let a_h be an appropriate set defined as

$$a_h = (t_1^{k_1}, \ldots, t_i^{k_i}, \ldots, t_n^{k_n}). \tag{24}$$

The *characteristic vector c(h)* for a_h will be defined as the n-vector

$$c(h) = (k_1, \ldots, k_i, \ldots, k_n). \tag{25}$$

Suppose that this set a_h has been chosen by lottery L to be the *active set* $\bar{a}$ of the game, and that some particular type t_i^k of player *i* has been chosen by lottery L to be an *active type*. This of course means that t_i^k must be a *member* of this set a_h, which can be the case only if type t_i^k is identical to type $t_i^{k_i}$ listed in (24), which implies that we must have $k=k_i$

Yet, if all these requirements are met, then this set a_h and this type t_i^k together will satisfy all the statements (14) to (21).

As we saw in section 6, the payoff v_i^k of any active type t_i^k will depend both

1. On the *strategies* used by the *n active types* in the game, and
2. On the *identities* of these active types.

This means, however, that t_i^k's payoff v_i^k will depend on the *strategy profile s*(h)* defined in the last paragraph of section 15, and on the *characteristic vector c(h)* defined by (25).

Thus, we can write

$$v_i^k = V_i^k(s^*(h), c(h)) \text{ if } t_i^k \in \bar{a} = a_h. \tag{26}$$

The payoff functions V_i^k $(i=1,\ldots,n;k=1,\ldots K_i)$ I shall call *conditional* payoff functions. *Firstly*, any given type will obtain the payoff v_i^k defined by (26) *only* if he will be chosen by lottery L to be an *active type* in the game. (This is what the condition $t_i^k \in \bar{a}$ in (26) refers to.)

Secondly, even if t_i^k is chosen to be an active type, (26) makes his payoff v_i^k *dependent* on the set a_h, chosen by lottery L to be an active set $\bar{a}$ of the game.

17. Semi-conditional payoff functions.

By reasoning similar to that we used in section 12, one can show that the quantity any active type t_i^k will try to maximize will *not* be his *payoff* v_i^k defined by (26). Rather, it will be his *expected payoff;* i.e., the *expected value* u_i^k of his payoff v_i^k.

We can define u_i^k as

$$u_i^k = U_i^k(s^*) = \sum_{h=1}^{H} \pi_i^k(h)\, V_i^k(s^*(h), c(h)) \text{ if } t_i^k \in \bar{a}. \tag{27}$$

These payoff functions U_i^k $(i=1,\ldots,n;k=1,\ldots,K_i)$ I shall call *semi-conditional.* I shall do so because they are subject to the *first* condition to which the payoff functions V_i^k are subject but *not* to the *second.* That is to say, any given type t_i^k will obtain the *expected payoff* u_i^k defined by (27) *only if* he is an *active type* of the game. But, if he is, then his expected payoff u_i^k will *not* depend on which particular appropriate set a_h has been chosen by lottery L to be the active set $\bar{a}$ of the game.

It is true also in the *n*-person case that if an I-game is reinterpreted in accordance with our *probabilistic model* then it will be converted into a *C-game G**.

Moreover, this C-game G*, under its *type-centered* interpretation, can be regarded as a Z-person game whose "players" are the Z different types in the game. As the payoff function of each type t_i^k *we* can use his *semi-conditional* payoff function U_i^k.

Using these payoff functions U_i^k, it will be easy to define the *Nash equilibria* (Nash, 1951) of this Z-person game, and to choose one of them as its *solution* on the basis of a suitable theory of equilibrium selection.

[1]See Harsanyi, 1967-68 (pp. 167- 168).

[2]The subjective probability distributions of various orders discussed in this section all are probability distributions over function spaces, whose proper mathematical definition poses some well - known technical difficulties. Yet, as Aumann (1963 and 1964) has shown, these difficulties can be overcome. But even so, the above model of higher and higher-order subjective probability distributions remains a hopelessly cumbersome model for analysis of I-games.

[3]Let $\pi_1^k(m)$ for m =1,...,M be the probability that some type t_1^k of player 1 assigns to the assumption that the Russian side will be represented by type t_2^m in the game. According to Bayesian theory, the M probabilities $\pi_1^k(1), \pi_1^k(2), \ldots \pi_1^k(m), \ldots, \pi_1^k(M)$ will fully *characterize* the *expectations* that this type t_1^k entertains about the characteristics of player 2 in the game. On the other hand, as we shall see, the *probabilistic model we* shall propose for the game will imply that these *probabilities* $\pi_1^k(m)$ must be equal to certain *conditional probabilities* so that

$$\pi_1^k(m) = Pr\,(t_2^m | t_1^k) \text{ for } m = 1,\ldots,M.$$

A similar relationship will obtain between the *K [probabilities* $\pi_2^m(k)$ entertained by any given type t_2^m of player 2 and the *conditional probabilities* $Pr\,(t_1^k | t_2^m)$ for $k = 1,\ldots,K$.

[4]*Personal moves* are moves the various *players* have chosen to make. *Chance moves* are moves made by some *chance mechanism,* such as a roulette wheel. Yet, moves made by some players yet decided by chance, such as throwing a coin, or a shuffling of cards, can also count as chance moves.

[5]Cf. footnote 3 to section 4 above.

[6]Using player-centered language, in Harsanyi (1967-68, p. 180), I called the M-tuple s_2^* and the K-tuple s_1^* *(see* below), the *normalized strategies* of player 2 and player 1, respectively.

[7] As defined by John Nash in Nash (1951). But he actually called them equilibrium points.

[8]In Harsanyi, 1967-68, I called a strategy combination such as S_i^* the *normalized strategy* of player i (cf. Footnote 6 to section 12 above).

REFERENCES

Aumann, Robert J. 1963. "On Choosing a Function at Random." In Fred B. Wright (ed.), *Symposium* on *Ergodic Theory,* 1-20. New Orleans: Academic Press.

Aumann, Robert J. 1964. "Mixed and Behavior Strategies in Infinite Extensive Games." In M. Dresher, L.S. Shapley, and A.W. Tucker (eds.), *Advances in Game Theory,* 627-650. Princeton: Princeton University Press.

Fudenberg, Drew, and Tirole, Jean. 1991. *Game Theory.* Cambridge. MA: Cambridge University Press.

Harsanyi, John C. 1967-68. "Games with Incomplete Information Played by Bayesian Players." *Management Science* 14, 159-182, 320-334, and 486-502.

Nash, John F. 1951. "Noncooperative Games." *Annals of Mathematics* 54, 289-295.

von Neumann, John, and Morgenstern, Oskar. 1944, 1947. *Theory of Games and Economic Behavior* Princeton: Princeton University Press.

SELECTED LIST OF PUBLICATIONS

A. BOOKS:
Essays on Ethics, Social Behavior, and Scientific Explanation. With Foreword by Kenneth J. Arrow. Dordrecht, Holland: D. Reidel, 1976. xvi + 262 pp.
Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge, England: Cambridge University Press, 1977. x + 314 pp.
Papers in Game Theory. Dordrecht, Holland: D. Reidel, 1982. xii + 258 pp.
A General Theory of Equilibrium Selection in Games (joint work with Reinhard Selten). With Foreword by Robert Aumann. Cambridge, MA: MIT Press, 1988. xiv + 378.

B. SELECTED JOURNAL ARTICLES:
"Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," Journal of Political Economy, 61 (1953), 434-435.
"Welfare Economics of Variable Tastes," Review of Economic Studies, 21 (1953-54), 204 -213.
"Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," Journal of Political Economy, 63 (1955), 309-321.
"Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's Hicks's and Nash's Theories," Econometrica, 24 (1956), 144-157.
"Ethics in Terms of Hypothetical Imperatives," Mind, 67 (1958), 305-316.
"Measurement of Social Power, Opportunity Costs, and the Theory of Two-Person Bargaining Games," Bahavioral Science, 7 (1962), 67-80.
"Measurement of Social Power in n-Person Reciprocal Power Situations," Behavioral Science, 7 (1962), 81-91.
"A Simplified Bargaining Model for the n-Person Cooperative Game," International Economic Review, 4 (1963), 194-220.
"A Bargaining Model for Social Status in Informal Groups and Formal Organizations," Behavioral Science, 11 (1966), 357-369.
"Games with Incomplete Information Played by 'Bayesian' Players," Parts I to III, Management Science, 14 (1967-68), pp. 159-182, 320-334, and 486-502.
"Rational-Choice Models of Political Behavior vs. Functionalist and Conformist Theories," World Politics, 21 (1969), 513-538.
"Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," International Journal of Game Theory, 2 (1973), 1-23.
"Oddness of the Number of Equilibrium Points: A New Proof," International Journal of Game Theory, 2 (1973), 235-250.
"An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition," Management Science, 20 (1974), 1472-1495.
"Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory," American Political Science Review, 69 (1975), 594-606.
"Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" Theory and Decision, 6 (1975), 311-332.
"The Tracing Procedure: A Bayesian Approach to Defining a Solution for n-Person Noncooperative Games, International Journal of Game Theory, 4 (1975), 61-94.
"Rule Utilitarianism and Decision Theory," Erkenntnis, 11 (1977), 25-53.
"Advances in Understanding Rational Behavior." In R. E. Butts and J. Hintikka (eds.), Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, Part II. Dordrecht, Holland: D. Reidel, 1977, 315-343.
"Rationality, Reasons, Hypothetical Imperatives, and Morality," in Hal Berghel et al., Wittgenstein, the Vienna Circle, and Critical Rationalism. Vienna, Austria: Verlag Hoelder - Pichler - Tempsky, 1979, 463-475.
"Bayesian Decision Theory, Rule Utilitarianism, and Arrow's Impossibility Theorem," Theory and Decision, 11 (1979), 289-317.
'*Mathematics, the Empirical Facts, and Logical Necessity," Erkenntnis, 19 (1983), 167-192.

"Bayesian Decision Theory, Subjective and Objective Probabilities, and Acceptance of Empirical Hypotheses," Synthese, 57 (1983), 341-365.
"Acceptance of Empirical Statements: A Bayesian Theory Without Cognitive Utilities," Theory and Decision, 18 (1985), 1-30.
"Rule Utilitarianism, Equality, and Justice," Social Philosophy and Policy, 2 (1985), 115-127.
"Does Reason Tell Us What Moral Code to Follow, and Indeed, to Follow Any Moral Code at All?" Ethics, 96 (1985), 42-55.
"Utilitarian Morality in a World of Very Half-Hearted Altruists." In W. P. Heller et al. (eds.), Social Choice and Public Decision Making: Essays in Honor of K. J. Arrow, Vol. I Cambridge, England: Cambridge University Press, 1986, pp. 57-73.
"Von Neumann-Morgenstern Utilities, Risk Taking, and Welfare." In G. R. Feiwel (ed.), Arrow and the Ascent of Modern Economic Theory. New York: New York University Press, 1987, pp. 545-558.
"Assessing Other People's Utilities," in B. E. Munier (ed.), Risk, Decision, and Rationality. Dordrecht, Holland: D. Reidel, 1988, pp. 127-138.
"Problems with Act Utilitarianism and Malevolent Preferences." A Critique of Richard Hare's Theory. In D. Seanor and N. Fotion (eds.), Hare and Critics. Oxford, England: Clarendon Press, 1988, pp. 89-99.
"Equality, Responsibility, and Justice As Seen from a Utilitarian Perspective," Theory and Decision, 31 (1991), 141-158.
"Game and Decision Theoretic Models in Ethics." In R. J. Aumann and S. Hart (eds.), Handbook of Game Theory, Vol. 1. Amsterdam, The Netherlands: Elsevier (North-Holland), 1992; Chapter 19.