Games with Incomplete Information†

By JOHN C. HARSANYI *

I was born in Budapest, Hungary, on May 29, 1920. The high school my parents chose for me was the Lutheran Gymnasium in Budapest, one of the best schools in Hungary, with such distinguished alumni as John von Neumann and Eugene Wigner. I was very happy in this school and received a superb education. In 1937, the year I graduated from it, I won the First Prize in Mathematics at the Hungary-wide annual competition for high-school students.

My parents owned a pharmacy in Budapest, which gave us a comfortable living. As I was their only child, they wanted me to become a pharmacist. But my own preference would have been to study philosophy and mathematics. Yet, in 1937 when I actually had to decide my field of study, I chose pharmacy in accordance with my parents' wishes. I did so because Adolf Hitler was in power in Germany, and his influence was steadily increasing also in Hungary. I knew that as a pharmacy student I would obtain military deferment. As I was of Jewish origin, this meant that I would not have to serve in a forced-labor unit of the Hungarian army.

As a result, I did have military deferment until the German army occupied Hungary in March 1944. Then I had to serve in a labor unit for a few months. In the last period of German occupation, from mid-November 1944 to mid-January 1945, the Jesuit fathers hid me in their monastery, which probably saved my life.

After the war, I reenrolled at the University of Budapest, this time to study philosophy and sociology. I obtained a Ph.D. in these subjects in June 1947. Then, for a year I was a junior faculty member at the University Institute of Sociology. It was there that I met a psychology student named Anne Klauber, who later became my wife. Ever since, her practical good sense and her unfailing emotional support have always been a great help to me. She has been always ready to discuss my ideas with me and to act as editor and proofreader of my work.

In June 1948, I had to resign from the Institute because of my commonly known opposition to Marxist ideology. It was Anne who convinced me at that point that we must leave communist Hungary if I ever wanted to resume an academic career.

In actual fact, we managed to leave Hungary only in April 1950. Then, after waiting for our Australian landing permits for a few months, we actually reached Sydney, Australia, only in December 1950.

As my English was not very good and as my Hungarian university degrees were not recognized in Australia, during most of our first three years there I had to do factory work. But in the evening I took economics courses at the University of Sydney. (I changed over from sociology to economics because I found the conceptual and mathematical elegance of economic theory very attractive.) I was given some credit for my Hungarian university courses so that I had to do only two years of further course work and had to write a thesis in economics in order to get an M.A. I received the degree late in 1953.

Early in 1954, I was appointed Lecturer in Economics at the University of Queensland in Brisbane. Then, in 1956, I was

†This article is the lecture John C. Harsanyi delivered in Stockholm, Sweden, December 9, 1994, when he received the Alfred Nobel Memorial Prize in Economic Sciences.

*Haas School of Business, University of California, Berkeley, CA 94720.

awarded a Rockefeller Fellowship, enabling me to spend two years at Stanford University, where I got a Ph.D. in economics, while Anne got an M.A. in psychology.

I had the good fortune of having Ken Arrow as an advisor and dissertation supervisor. I benefited very much from discussing many finer points of economic theory with him. But I also benefited substantially by following his advice to spend a sizable part of my Stanford time studying mathematics and statistics. These studies proved very useful in my later work in game theory.

In 1958, Anne and I returned to Australia, where I got a very attractive research position at the Australian National University in Canberra. But soon I felt very isolated, because at that time in Australia there was not much interest in game theory.

Then, in 1961, with Ken Arrow's and Jim Tobin's help, I was appointed Professor of Economics at Wayne State University in Detroit. In 1964 I became Visiting Professor, and then Professor at the Business School of the University of California in Berkeley. Later my appointment was extended also to the Department of Economics. Our only child, Tom, was born in Berkeley. I retired from the university in 1990.

In the 1950's I published papers on the use of von Neumann-Morgenstern utilities in welfare economics and in ethics, and on the welfare economics of variable tastes. My interest in game-theoretic problems in a narrower sense was first aroused by John Nash's four brilliant papers, published in the period 1950–1953 on cooperative and on noncooperative games, on two-person bargaining games, on mutually optimal threat strategies in such games, and on what we now call Nash equilibria.

In 1956, I showed the mathematical equivalence of Frederik Zeuthen's and of Nash's bargaining models and stated algebraic criteria for optimal threat strategies. In 1963, I extended the Shapely value to games without transferable utility and showed that my new solution concept was a direct generalization both of the Shapely value and of Nash's bargaining solution with variable threats. In a three-part paper (Harsanyi, 1967, 1968a,b), I showed how to convert a game with incomplete information into one with complete, yet imperfect, information. In 1973, I showed that "almost all" mixed-strategy Nash equilibria can be reinterpreted as pure-strategy equilibria of a suitably chosen game with randomly fluctuating payoff functions.

I have also published a number of papers on utilitarian ethics and have published four books. One of them, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (1977), was an attempt to unify game theory by extending the use of bargaining models from cooperative games to noncooperative games. Two books, *Essays on Ethics, Social Behavior, and Scientific Explanation* (1976) and *Papers in Game Theory* (1982) were collections of some of my journal articles. Finally, *A General Theory of Equilibrium Selection in Games* (1988) was a joint work with Reinhard Selten.

In 1993 and in 1994, I wrote two papers, proposing a *new theory* of equilibrium selection. My 1993 paper does so for games with *complete* information whereas my 1994 paper does so for games with *incomplete* information. My new theory is based on the theory in Harsanyi and Selten (1988) but is a simpler theory and is in my view an intuitively more attractive one. Both papers are soon to appear (and probably will have already appeared when these lines are being read) in *Games and Economic Behavior*.

I. 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 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.

II. The Problem of Incomplete Information

Following von Neumann and Morgenstern (1947 p. 30), one 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 1960's) 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, about the strategies available to other players (or even to themselves), about the amount of information the other players have about various aspects of the game, and so on.

Yet, by suitable modeling, *all* forms of incomplete information can be reduced to the case in which the players have less than full information about each other's *payoff functions* U_1^i, defining the *utility payoff* $u_i = U_i(s)$ of each player i for any possible strategy combination $s = (s_1, \ldots, s_n)$ the n players may use (see Harsanyi, 1967 pp. 167–68).

III. Two-Person I-Games

A. *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 will be seen 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

$e_2 e_1 U_2$ about player 1's first-order expectation $e_1 U_2$, 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 $e_1 U_2$ 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 $e_2 U_1$ 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 $e_1 e_2 U_1$ 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.[1]

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 one retains the simplifying assumption that each player will know his own payoff function, 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 will be seen, 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).

B. *Arms-Control Negotiations between the United States and the Soviet Union in the 1960's*

In the period from 1964 to 1970, 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 1* 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 1), 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.

[1]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 Robert J. Aumann (1963, 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.

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 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 $t_2^1, t_2^2, t_2^m, \dots, T_2^M$. 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.[2]

C. *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 1967–1968 papers 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 one 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 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

[2]Let $\pi_1^k(m)$ for $m = 1, \dots, M$ 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 $\pi_1^k(1), \pi_1^k(2), \dots, \pi_1^k(m), \dots, \pi_1^k(M)$ 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 will be seen, the *probabilistic model* I shall propose for the game will imply that these probabilities $\pi_1^k(m)$ must be equal to certain conditional probabilities so that

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

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 $k = 1, \dots, K$.

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.

D. *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 I 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_2^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, one may define their payoffs v_1^k and v_2^m as

$$(1) \qquad v_1^k = V_1^k(s_1^k, s_2^m; k, m)$$

and

$$(2) \qquad v_2^m = V_2^m(s_1^k, s_2^m; k, m)$$

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).

E. *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.

I shall also 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.

F. *Two Important Distinctions*

As I have already shown, 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 moves* and *chance moves*.[3] 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, and so on.

Game theory as first established by von Neumann and Morgenstern, and even as it had been further developed up to the late 1960's, 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.

[3]*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. 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.

G. *A Probabilistic Model for the 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 the 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 relevant social forces, that is, of the social forces that have produced an American society of one particular type (corresponding to some type t_1^k of the model) and that have also produced a Russian society of another particular type (corresponding to some type t_2^m of the 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

$$(3) \quad \Pr(t_1^k, t_2^m) = p_{km}$$

$$\text{for } k = 1, \dots, K \text{ and for } m = 1, \dots, M.$$

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 one needs 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 $K \times M$ *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 pp. 176–77). 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 Drew Fudenberg and Jean Tirole, 1991 p. 210). Alternatively, one 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 one obtains by adding a lottery L (as just described) to the two-person I-game described in Subsections B–E will be called a *probabilistic model* for this I-game G. As will be seen presently, this probabilistic model will actually convert this I-game G into a C-game, which I shall call game G^*.

H. *Converting the I-Game G with Incomplete Information into a Game G* with Complete Yet Imperfect Information*

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

Let us go back to the two-person game G 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 the original assumptions about G, it will be a game with incomplete information.

(i) First of all, under the 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 made in Subsection E, player 1 will know Fact 1 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 the original assumptions, G is in fact a game with incomplete information.

I will now show that as soon as one reinterprets game G in accordance with the probabilistic model (i.e., as soon as one adds lottery L to the game), the original game G will be converted into a new game G^* with *complete* information. Of course, even after this reinterpretation, the 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 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 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 the probabilistic model for the game, which will be fully known to both players.

On the other hand, as the 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.

I. *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 the assumptions in Subsection E, 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, namely, 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[4]

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

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

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

[4]See footnote 2.

J. *The Semiconditional Payoff Functions of the Two Active Types*

Suppose the two active types in the game are t_1^k and t_2^m. As was seen in Subsection D, 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 chose 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 now be 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

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

Here the symbol s_2^* stands for the strategy M-tuple[5]

$$(7)\qquad s_2^* = (s_2^2, s_2^2, \ldots, s_2^m, \ldots, s_2^M).$$

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 any one of his M *potential* opponents

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

$t_2^1, \dots, t_2^M$ 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

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

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

$$(9)\qquad s_1^* = (s_1^1, s_1^2, \dots, s_1^k, \dots, s_1^K).$$

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 $s_1^1, \dots, s_1^K$ that any one 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), I shall describe the payoff functions U_2^k and U_2^m used in (6) and in (8) as *semiconditional*. 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
(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 seen in Subsection H, once one reinterprets the original I-game G in accordance with the 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 *semiconditional* payoff functions U_1^k $(k = 1, \dots, K)$ and U_2^m $(m = 1, \dots, M)$.

If one regards these $K+M$ types as the real "players" of G^* and regards these payoff functions U_1^k and U_2^m as their real payoff functions, then one can easily define the *Nash equilibria*[6] of this C-game G^*. Then, using a suitable theory of equilibrium selection, one can define one of these equilibria as the *solution* of this game.

IV. *n*-Person I-Games

A. *The Types of the Various Players, the Active Set, and the Appropriate Sets in n-Person I-Games*

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

$$(10)\qquad Z = \sum_{i \in \mathcal{N}} K_i.$$

Suppose that players $1, \dots, i, \dots, n$ are now represented by their types $t_1^{k_i}, \dots, t_i^{k_i}, \dots, t_n^{k_n}$ 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

[6]As defined by Josh Nash (1951); but he actually called them *equilibrium points*.

will be

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

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

$$\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_h, \ldots, \mathrm{a}_H. \tag{12}$$

Let $\mathcal{A}_i^k$ 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 $\mathcal{A}_i^k$ will be

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

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

B. *Some Probabilities*

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

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

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)] 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 the 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

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

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

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

and conversely, because $\mathcal{A}_i^k$ contains exactly those appropriate sets that have type t_i^k as their member. Thus, one can write

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

I have already concluded that if type t_i^k has the status of an active type then he will know (15). I 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 $\bar{\mathrm{a}}$ belonging to the family $\mathcal{A}_i^k$ is

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

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

$$\pi_i^k(h) = \Pr\left(\bar{\mathrm{a}} = \mathrm{a}_h | t_i^k \in \bar{\mathrm{a}}\right). \tag{19}$$

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

$$\begin{aligned} \Pr\left(\bar{\mathrm{a}} = \mathrm{a}_h | t_i^k \in \bar{\mathrm{a}}\right) &= \Pr\left(\bar{\mathrm{a}} = \mathrm{a}_h | \bar{\mathrm{a}} \in \mathcal{A}_i^k\right) \\ &= \Pr(\bar{\mathrm{a}} = \mathrm{a}_h) / \Pr\left(\bar{\mathrm{a}} \in \mathcal{A}_i^k\right) \\ &= r_h \Big/ \sum_{h \in \mathcal{B}_i^k} r_h. \end{aligned} \tag{20}$$

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

$$\pi_i^k(h) = r_h \Big/ \sum_{h \in \mathcal{B}_i^k} r_h. \tag{21}$$

C. *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 the assumptions, inactive types do not actively participate in the game and, therefore, do not choose any strategies.) Then I shall write

$$\mathbf{s}_i^* = \left(s_i^1,\ldots,s_i^k,\ldots,s_i^{K_i}\right) \quad \text{for } i = 1,\ldots,n \tag{22}$$

to denote the *strategy profile*[7] of player-*i* types.

Let

$$\mathbf{s}^* = \left(s_1^1,\ldots,s_n^{K_n}\right) \tag{23}$$

be the ordered set obtained if one first lists all K_1 strategies in s_1^*, then all K_2 strategies in $s_2^*,\ldots,$ then all K_i strategies in $s_i^*,\ldots,$ and finally all K_n strategies in $\mathbf{s}_n^*$. Obviously, $\mathbf{s}^*$ will be a strategy profile of *all* types in the game. In view of (10), $\mathbf{s}^*$ will contain Z different strategies.

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

D. *The Conditional Payoff Functions*

Let a_h be an appropriate set defined as

$$\mathrm{a}_h = \left(t_1^{k_1},\ldots,t_i^{k_i},\ldots,t_n^{k_n}\right). \tag{24}$$

The *characteristic vector* $\mathbf{c}(h)$ for a_h will be defined as the n-vector

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

[7]In Harsanyi (1967, 1968a,b), I called a strategy combination such as s_i^* the *normalized strategy* of player i (see footnote 5).

Suppose that this set a_h has been chosen by lottery L to be the active set $\bar{\mathrm{a}}$ 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 $t_i^{k_i}$ listed in (24), which implies that $k = k_i$. 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)–(21).

As seen in Section III-D, the payoff v_i^k of any active type t_i^k will depend on both of the following:

(i) the *strategies* used by the n-active types in the game;
(ii) the *identities* of these active types.

This means, however, that t_i^k's payoff v_i^k will depend on the strategy profile $\mathbf{s}^*(h)$ defined in the previous subsection and on the characteristic vector $\mathbf{c}(h)$ defined by (25). Thus, one can write

$$v_i^k = V_i^k(\mathbf{x}^*(h), \mathbf{c}(h)) \quad \text{if } t_i^k \in \bar{\mathrm{a}} = \mathrm{a}_h. \tag{26}$$

The payoff functions V_i^k ($i = 1,\ldots,n$; $k = 1,\ldots,K_i$) I shall call *conditional payoff functions*. First, 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 $t_i^k \in \bar{\mathrm{a}}$ in (26) refers to.)

Second, 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 $\bar{\mathrm{a}}$ of the game.

E. *Semiconditional Payoff Functions*

By reasoning similar to that used in Section III-J, 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).

One can define u_i^k as

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

These payoff functions U_i^k ($i = 1,\dots,n$; $k = 1,\dots,K_i$) I shall call *semiconditional*. 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 condition. 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 $\bar{\mathrm{a}}$ of the game.

It is true also in the n-person case that if an I-game is reinterpreted in accordance with the 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 in the Z different types in the game. As the payoff function of each type t_i^k one can use his semiconditional 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.

REFERENCES

Aumann, Robert J. "On Choosing a Function at Random," in Fred B. Wright, ed., *Symposium on ergodic theory*. New Orleans: Academic Press, 1963, pp. 1–20.

______. "Mixed and Behavior Strategies in Infinite Extensive Games," in M. Dresher, L. S. Shapley, and A. W. Tucker, eds., *Advances in game theory*. Princeton, NJ: Princeton University Press, 1964, pp. 627–50.

Fudenberg, Drew and Tirole, Jean. *Game theory*. Cambridge, MA: MIT Press, 1991.

Harsanyi, John C. "Games with Incomplete Information Played by 'Bayesian' Players, Part I. The Basic Model." *Management Science*, November 1967, *14*(3), pp. 159–82.

______. "Games with Incomplete Information Played by 'Bayesian' Players, Part II. Bayesian Equilibrium Points." *Management Science*, January 1968a, *14*(5), pp. 320–34.

______. "Games with Incomplete Information Played by 'Bayesian' Players, Part III. The Basic Probability Distributrion of the Game." *Management Science*, March 1968b, *14*(7), pp. 486–502.

______. *Essays on ethics, social behavior, and scientific explanations*. Dordrecht: Reidl, 1976.

______. *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press, 1977.

______. *Papers in game theory*. Dordrecht: Reidel, 1982.

Harsanyi, John C. and Selten, Reinhard. *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press, 1988.

Nash, John F. "Non-cooperative Games." *Annals of Mathematics*, September 1951, *54*(2), pp. 286–95.

von Neumann, John and Morgenstern, Oskar. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press, 1944; revised edition, 1947.