MANAGEMENT SCIENCE
Vol. 14, No. 3, November, 1967
Printed in U.S.A.

GAMES WITH INCOMPLETE INFORMATION PLAYED BY "BAYESIAN" PLAYERS, I-III

Part I. The Basic Model*†[1]

JOHN C. HARSANYI

University of California, Berkeley

The paper develops a new theory for the analysis of games with incomplete information where the players are uncertain about some important parameters of the game situation, such as the payoff functions, the strategies available to various players, the information other players have about the game, etc. However, each player has a subjective probability distribution over the alternative possibilities.

In most of the paper it is assumed that these probability distributions entertained by the different players are mutually "consistent", in the sense that they can be regarded as conditional probability distributions derived from a certain "basic probability distribution" over the parameters unknown to the various players. But later the theory is extended also to cases where the different players' subjective probability distributions fail to satisfy this consistency assumption.

In cases where the consistency assumption holds, the original game can be replaced by a game where nature first conducts a lottery in accordance with the basic probablity distribution, and the outcome of this lottery will decide which particular subgame will be played, i.e., what the actual values of the relevant parameters will be in the game. Yet, each player will receive only partial information about the outcome of the lottery, and about the values of these parameters. However, every player will know the "basic probability distribution" governing the lottery. Thus, technically, the resulting game will be a game with complete information. It is called the Bayes-equivalent of the original game. Part I of the paper describes the basic model and discusses various intuitive interpretations for the latter. Part II shows that the Nash equilibrium points of the Bayes-equivalent game yield "Bayesian equilibrium points" for the original game. Finally, Part III considers the main properties of the "basic probablity distribution".

---

* Received June 1965, revised June 1966, accepted August 1966, and revised June 1967.

† Parts II and III of "Games with Incomplete Information Played by 'Bayesian' Players" will appear in subsequent issues of *Management Science: Theory.*

[1] The original version of this paper was read at the Fifth Princeton Conference on Game Theory, in April, 1965. The present revised version has greatly benefitted from personal discussions with Professors Michael Maschler and Robert J. Aumann, of the Hebrew University, Jerusalem; with Dr. Reinhard Selten, of the Johann Wolfgang Goethe University, Frankfurt am Main; and with the other participants of the International Game Theory Workshop held at the Hebrew University in Jerusalem, in October–November 1965. I am indebted to Dr. Maschler also for very helpful detailed comments on my manuscript.

This research was supported by Grant No. GS-722 of the National Science Foundation as well as by a grant from the Ford Foundation to the Graduate School of Business Administration, University of California. Both of these grants were administered through the Center for Research in Management Science, University of California, Berkeley. Further support has been received from the Center for Advanced Study in the Behavioral Sciences, Stanford.

TABLE OF CONTENTS

Glossary of Mathematical Notation

I-game $\cdots$ A game with *incomplete* information.
C-game $\cdots$ A game with *complete* information.
G $\cdots$ The I-game originally given to us.
G^* $\cdots$ The Bayesian game equivalent to G. (G^* is a C-game.)
G^{**} $\cdots$ The Selten game equivalent to G and to G^*. (G^{**} is likewise a C-game.)
$\mathfrak{N}(G)$, $\mathfrak{N}(G^*)$, $\mathfrak{N}(G^{**})$ $\cdots$ The normal form of G, G^* and G^{**} respectively.
$\mathcal{S}(G)$, $\mathcal{S}(G^*)$ $\cdots$ The semi-normal form of G and G^* respectively.
s_i $\cdots$ Some strategy (pure or mixed) of player i, with $i = 1, \cdots, n$.
$S_i = \{s_i\}$ $\cdots$ Player i's strategy space.
c_i $\cdots$ Player i's attribute vector (or information vector).
$C_i = \{c_i\}$ $\cdots$ The range space of vector c_i.
$c = (c_1, \cdots, c_n)$ $\cdots$ The vector obtained by combining the n vectors $c_1, \cdots, c_n$ into one vector.

$C = \{c\}$ ··· The range space of vector c.

$c^i = (c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n)$ ··· The vector obtained from vector c by omitting subvector c_i.

$C^i = \{c^i\}$ ··· The range space of vector c^i.

x_i ··· Player i's payoff (expressed in utility units).

$x_i = U_i(s_1, \cdots, s_n) = V_i(s_1, \cdots, s_n; c_1, \cdots, c_n)$ ··· Player i's payoff function.

$P_i(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n) = P_i(c^i) = R_i(c^i \mid c_i)$ ··· The subjective probability distribution entertained by player i.

$R^* = R^*(c_1, \cdots, c_n) = R^*(c)$ ··· The basic probability distribution of the game.

$R_i^* = R^*(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n \mid c_i) = R^*(c^i \mid c_i)$ ··· The conditional probability distribution obtained from R^* for a given value of vector c_i.

k_i ··· The number of different values that player i's attribute vector c_i can take in the game (in cases where this number is finite).

$K = \sum_{i=1}^n k_i$ ··· The number of players in the Selten game G^{**} (when this number is finite).

s_i^* ··· A normalized strategy of player i. (It is a function from the range space C_i of player i's attribute vector c_i, to his strategy space S_i.)

$S_i^* = \{s_i^*\}$ ··· The set of all normalized strategies s_i^* available to player i.

$\mathcal{E}$ ··· The expected-value operator.

$\mathcal{E}(x_i) = W_i(s_1^*, \cdots, s_n^*)$ ··· Player i's normalized payoff function, stating his *unconditional* payoff expectation.

$\mathcal{E}(x_i \mid c_i) = Z_i(s_1^*, \cdots, s_n^* \mid c_i)$ ··· Player i's semi-normalized payoff function, stating his *conditional* payoff expectation for a given value of his attribute vector c_i.

D ··· A cylinder set, defined by the condition $D = D_1 \times \cdots \times D_n$, where $D_1 \subseteqq C_1, \cdots, D_n \subseteqq C_n$.

$G(D)$ ··· For a given decomposable game G or G^*, $G(D)$ denotes the component game played in all cases where the vector c lies in cylinder D. D is called the defining cylinder of the component game $G(D)$.

Special Notation in Certain Sections

In section 3 (Part I):

a_{0i} denotes a vector consisting of those parameters of player i's payoff function U_i which (in player j's opinion) are *unknown* to all n players.

a_{ki} denotes a vector consisting of those parameters of the function U_i which (in j's opinion) are unknown to some of the players but are *known* to player k.

$a_0 = (a_{01}, \cdots, a_{0n})$ is a vector summarizing all information that (in j's opinion) none of the players have about the functions $U_1, \cdots, U_n$.

$a_k = (a_{k1}, \cdots, a_{kn})$ is a vector summarizing all information that (in j's opinion) player k has about the functions $U_1, \cdots, U_n$, except for the information that (in j's opinion) *all* n players have about these functions.

b_i is a vector consisting of all those parameters of player i's subjective probability distribution P_i which (in player j's opinion) are unknown to some or all of the players $k \neq i$.

In terms of these notations, player i's information vector (or attribute vector) c_i can be defined as

$c_i = (a_i, b_i)$.

V_i^* denotes player i's payoff function *before* vector a_0 has been integrated out. *After* elimination of vector a_0 the symbol V_i is used to denote player i's payoff function.

In sections 9–10 (Part II):

a^1 and a^2 denote the two possible values of player 1's attribute vector c_1.

b^1 and b^2 denote the two possible values of player 2's attribute vector c_2.

$r_{km} = R^*(c_1 = a^k \text{ and } c_2 = b^m)$ denotes the probability mass function corresponding to the basic probability distribution R^*.

$p_{km} = r_{km}/(r_{k1} + r_{k2})$ and $q_{km} = r_{km}/(r_{1m} + r_{2m})$ denote the corresponding *conditional* probability mass functions.

y^1 and y^2 denote player 1's two pure strategies.

z^1 and z^2 denote player 2's two pure strategies.

$y^{nt} = (y^n, y^t)$ denotes a normalized pure strategy for player 1, requiring the use of strategy y^n if $c_1 = a^1$, and requiring the use of strategy y^t if $c_1 = a^2$.

$z^{uv} = (z^u, z^v)$ denotes a normalized pure strategy for player 2, requiring the use of strategy z^u if $c_2 = b^1$, and requiring the use of strategy z^v if $c_2 = b^2$.

In section 11 (Part II):

a^1 and a^2 denote the two possible values that *either* player's attribute vector c_i can take.

$r_{km} = R^*(c_1 = a^k \text{ and } c_2 = a^m)$.

p_{km} and q_{km} have the same meaning as in sections 9–10.

y_i^* denotes player i's payoff demand.

y_i denotes player i's *gross* payoff.

x_i denotes player i's *net* payoff.

x_i^* denotes player i's net payoff in the case $(c_1 = a^1, c_2 = a^2)$.

x_i^{**} denotes player i's net payoff in the case $(c_1 = a^2, c_2 = a^1)$.

In section 13 (Part III):

$\alpha, \beta, \gamma, \delta$ denote specific values of vector c.

$\alpha_i, \beta_i, \gamma_i, \delta_i$ denote specific values of vector c_i.

$\alpha^i, \beta^i, \gamma^i, \delta^i$ denote specific values of vector c^i, etc.

$r_i(\gamma^i \mid \gamma_i) = R_i(c^i = \gamma^i \mid c_i = \gamma_i)$ denotes the probability mass function corresponding to player i's subjective probability distribution R_i (when R_i is a discrete distribution).

$r^*(\gamma) = R^*(c = \gamma)$ denotes the probability mass function corresponding to the basic probability distribution R^* (when R^* is a discrete distribution).

$\mathcal{R} = \{r^*\}$ denotes the set of all admissible probability mass functions r^*.

E denotes a similarity class, i.e., a set of nonnull points $c = \alpha, c = \beta, \cdots$ similar to one another (in the sense defined in Section 13).

In section 16 (Part III):

$R^{(i)}$ denotes the basic probability distribution R^* as assessed by player $i(i = 1, \cdots, n)$.

$R^{*\prime}$ denotes a given player's (player j's) *revised* estimate of the basic probability distribution R^*.

$c'_i = (c_i, d_i)$ denotes player j's *revised* definition of player i's attribute vector c_i. (It is in general a larger vector than the vector c_i originally assumed by player j.)

R'_i denotes player j's *revised* estimate of player i's subjective probability distribution R_i.

1.

Following von Neumann and Morgenstern [7, p. 30], we distinguish between games with *complete information*, to be sometimes briefly called C-games in this paper, and games with *incomplete information*, to be called I-games. The latter differ from the former in the fact that some or all of the players lack full information about the "rules" of the game, or equivalently about its normal form (or about its extensive form). For example, they may lack full information about other players' or even their own payoff functions, about the physical facilities and strategies available to other players or even to themselves, about the amount of information the other players have about various aspects of the game situation, etc.

In our own view it has been a major analytical deficiency of existing game theory that it has been almost completely restricted to C-games, in spite of the fact that in many real-life economic, political, military, and other social situations the participants often lack full information about some important aspects of the "game" they are playing.[2]

It seems to me that the basic reason why the theory of games with incomplete information has made so little progress so far lies in the fact that these games give rise, or at least appear to give rise, to an infinite regress in reciprocal expectations on the part of the players, [3, pp. 30–32]. For example, let us consider any two-person game in which the players do not know each other's payoff functions. (To simplify our discussion I shall assume that each player knows his own payoff function. If we made the opposite assumption, then we would have to introduce even more complicated sequences of reciprocal expectations.)

In such a game player 1's strategy choice will depend on what he expects (or believes) to be player 2's payoff function U_2, as the nature of the latter will be an important determinant of player 2's behavior in the game. This expectation

[2] The distinction between games with *complete* and *incomplete* information (between C-games and I-games) must not be confused with that between games with *perfect* and *imperfect* information. By common terminological convention, the first distinction always refers to the amount of information the players have about the *rules* of the game, while the second refers to the amount of information they have about the other players' and their own previous *moves* (and about previous chance moves). Unlike games with *incomplete* information, those with *imperfect* information have been extensively discussed in the literature.

about U_2 may be called player 1's *first-order* expectation. But his strategy choice will also depend on what he expects to be player 2's first-order expectation about his own (player 1's) payoff function U_1. This may be called player 1's *second-order* expectation, as it is an expectation concerning a first-order expectation. Indeed, player 1's strategy choice will also depend on what he expects to be player 2's second-order expectation—that is, on what player 1 thinks that player 2 thinks that player 1 thinks about player 2's payoff function U_2. This we may call player 1's *third-order* expectation—and so on *ad infinitum.* Likewise, player 2's strategy choice will depend on an infinite sequence consisting of his first-order, second-order, third-order, etc., expectations concerning the payoff functions U_1 and U_2. We shall call any model of this kind a *sequential-expectations* model for games with incomplete information.

If we follow the Bayesian approach and represent the players' expectations or beliefs by subjective probablity distributions, then player 1's *first-order* expectation will have the nature of a subjective probablity distribution $P_1^1(U_2)$ over all alternative payoff functions U_2 that player 2 may possibly have. Likewise, player 2's first-order expectation will be a subjective probablity distribution $P_2^1(U_1)$ over all alternative payoff functions U_1 that player 1 may possibly have. On the other hand, player 1's *second-order* expectation will be a subjective probability distribution $P_1^2(P_2^1)$ over all alternative first-order subjective probability distributions P_2^1 that player 2 may possibly choose, etc. 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 alternative ($k - 1$)th-order subjective probability distributions P_j^{k-1} that the other player j may possibly entertain.[3]

In the case of n-person I-games the situation is, of course, even more complicated. Even if we take the simpler case in which the players know at least their own payoff functions, each player in general will have to form expectations about the payoff functions of the other $(n - 1)$ players, which means forming $(n - 1)$ different first-order expectations. He will also have to form expectations about the $(n - 1)$ first-order expectations entertained by each of the other $(n - 1)$ players, which means forming $(n - 1)^2$ second-order expectations, etc.

The purpose of this paper is to suggest an alternative approach to the analysis of games with incomplete information. This approach will be based on constructing, for any given I-game G, some C-game G^* (or possibly several different C-games G^*) game-theoretically *equivalent* to G. By this means we shall reduce the analysis of I-games to the analysis of certain C-games G^*; so that the problem of

[3] Probability distributions over some space of payoff functions or of probability distributions, and more generally probability distributions over function spaces, involve certain technical mathematical difficulties [5, pp. 355–357]. However, as Aumann has shown [1] and [2], these difficulties can be overcome. But even if we succeed in defining the relevant higher-order probability distributions in a mathematically admissible way, the fact remains that the resulting model—like *all* models based on the sequential-expectations approach—will be extremely complicated and cumbersome. The main purpose of this paper is to describe an alternative approach to the analysis of games with incomplete information, which completely avoids the difficulties associated with sequences of higher and higher-order reciprocal expectations.

such sequences of higher and higher-order reciprocal expectations will simply not arise.

As we have seen, if we use the Bayesian approach, then the sequential-expectations model for any given I-game G will have to be analyzed in terms of infinite sequences of higher and higher-order subjective probability distributions, i.e., subjective probability distributions over subjective probablity distributions. In contrast, under our own model, it will be possible to analyze any given I-game G in terms of *one* unique probability distribution R^* (as well as certain conditional probablity distributions derived from R^*).

For example, consider a two-person non-zero-sum game G representing price competition between two duopolist competitiors where neither player has precise information about the cost functions and the financial resources of the other player. This, of course, means that neither player i will know the true payoff function U_j of the other player j, because he will be unable to predict the profit (or the loss) that the other player will make with any given choice of strategies (i.e., price and output polices) s_1 and s_2 by the two players.

To make this example more realistic, we may also assume that each player has *some* information about the other player's cost functions and financial resources (which may be represented, e.g., by a subjective probability distribution over the relevant variables); but that each player i lacks exact information about *how much* the other player j actually knows about player i's cost structure and financial position.

Under these assumptions this game G will be obviously an I-game, and it is easy to visualize the complicated sequences of reciprocal expectations (or of subjective probablity distributions) we would have to postulate if we tried to analyze this game in terms of the sequential-expectations approach.

In contrast, the new approach we shall describe in this paper will enable us to reduce this I-game G to an equivalent C-game G^* involving four random events (i.e., chance moves) e_1, e_2, f_1, and f_2, assumed to occur *before* the two players choose their strategies s_1 and s_2. The random event $e_i (i = 1, 2)$ will determine player i's cost functions and the size of his financial resources; and so will completely determine his payoff function U_1 in the game. On the other hand, the random event f_i will determine the amount of information that player i will obtain about the cost functions and the financial resources of the other player $j (j = 1, 2$ and $\neq i)$, and will thereby determine the actual amount of information[4] that player i will have about player j's payoff function U_j.

Both players will be assumed to know the joint probability distribution $R^*(e_1, e_2, f_1, f_2)$ of these four random events.[5] But, *e.g.*, player 1 will know the actual outcomes of these random events only in the case of e_1 and f_1, whereas

[4] In terms of the terminology we shall later introduce, the variables determined by the random events e_i and f_i will constitute the random vector c_i $(i = 1, 2)$, which will be called player i's information vector or attribute vector, and which will be assumed to determine player i's "type" in the game (cf. the third paragraph below).

[5] For justification of this assumption, see sections 4 and 5 below, as well as Part III of this paper.

player 2 will know the actual outcomes only in the case of e_2 and f_2. (In our model this last assumption will represent the facts that each player will know only his own cost functions and financial resources but will not know those of his opponent; and that he will, of course, know how much information he himself has about the opponent but will not know exactly how much information the opponent will have about him.)

As in this new game G^* the players are assumed to know the probability distribution $R^*(e_1, e_2, f_1, f_2)$, this game G^* will be a C-game. To be sure, player 1 will have no information about the outcomes of the chance moves e_2 and f_2, whereas player 2 will have no information about the outcomes of the chance moves e_1 and f_1. But these facts will not make G^* a game with "incomplete" information but will make it only a game with "imperfect" information (cf. Footnote 2 above). Thus, our approach will basically amount to replacing a game G involving *incomplete* information, by a new game G^* which involves *complete* but *imperfect* information, yet which is, as we shall argue, essentially equivalent to G from a game-theoretical point of view (see section 5 below).

As we shall see, this C-game G^* which we shall use in the analysis of a given I-game G will also admit of an alternative intuitive interpretation. Instead of assuming that certain important *attributes* of the players are determined by some hypothetical random events at the beginning of the game, we may rather assume that the players *themselves* are drawn at random from certain hypothetical populations containing a mixture of individuals of different "types", characterized by different attribute vectors (i.e., by different combinations of the relevant attributes). For instance, in our duopoly example we may assume that each player $i(i = 1, 2)$ is drawn from some hypothetical population Π_i containing individuals of different "types," each possible "type" of player i being characterized by a different attribute vector c_i, i.e., by a different combination of production costs, financial resources, and states of information. Each player i will know his own type or attribute vector c_i but will be, in general, ignorant of his opponent's. On the other hand, both players will again be assumed to know the joint probability distribution $R^*(c_i, c_2)$ governing the selection of players 1 and 2 of different possible types c_1 and c_2 from the two hypothetical populations Π_1 and Π_2.

It may be noted, however, that in analyzing a given I-game G, construction of an equivalent C-game G^* is only a partial answer to our analytical problem, because we are still left with the task of defining a suitable solution concept for this C-game G^* itself, which may be a matter of some difficulty. This is so because in many cases the C-game G^* we shall obtain in this way will be a C-game of unfamiliar form, for which no solution concept has been suggested yet in the game-theoretical literature.[6] Yet, since G^* will always be a game with complete information, its analysis and the problem of defining a suitable solution concept for it, will be at least amenable to the standard methods of modern game theory. We shall show in some examples how one actually can define appropriate solution concepts for such C-games G^*.

[6] More particularly, this game G^* will have the nature of a game with *delayed commitment* (see section 11 in Part II of this paper).

2.

Our analysis of I-games will be based on the assumption that, in dealing with incomplete information, every player i will use the Bayesian approach. That is, he will assign a *subjective* joint probability distribution P_i to all variables unknown to him—or at least to all unknown *independent* variables, i.e., to all variables not depending on the players' own strategy choices. Once this has been done he will try to maximize the mathematical expectation of his own payoff x_i in terms of this probability distribution P_i.[7] This assumption will be called the *Bayesian hypothesis.*

If incomplete information is interpreted as lack of full information by the players about the *normal form* of the game, then such incomplete information can arise in three main ways.

1. The players may not know the *physical outcome function* Y of the game, which specifies the physical outcome $y = Y(s_1, \cdots, s_n)$ produced by each strategy n-tuple $s = (s_1, \cdots, s_n)$ available to the players.

2. The players may not know their own or some other players' *utility functions* X_i, which specify the utility payoff $x_i = X_i(y)$ that a given player i derives from every possible physical outcome y.[8]

3. The players may not know their own or some other players' *strategy spaces* S_i, *i.e.*, the set of all strategies s_i (both pure and mixed) available to various players i.

All other cases of incomplete information can be reduced to these three basic cases—indeed sometimes this can be done in two or more different (but essentially equivalent) ways. For example, incomplete information may arise by some players' ignorance about the amount or the quality of physical resources (equipment, raw materials, etc.) available to some other players (or to themselves). This situation can be equally interpreted *either* as ignorance about the physical outcome function of the game (case 1), *or* as ignorance about the strategies available to various players (case 3). Which of the two interpretations we have to use will depend on how we choose to define the "strategies" of the players in question. For instance, suppose that in a military engagement our own side does not know the number of fire arms of a given quality available to the other side.

[7] A *subjective* probability distribution P_i entertained by a given player i is defined in terms of his own choice behavior, cf. [6]. In contrast, an *objective* probability distribution P^* is defined in terms of the long-run frequencies of the relevant events (presumably as established by an independent observer, say, the umpire of the game). It is often convenient to regard the subjective probabilities used by a given player i as being his personal *estimates* of the corresponding objective probabilities or frequencies unknown to him.

[8] If the physical outcome y is simply a vector of money payoffs $y_1, \cdots, y_n$ to the n players then we can usually assume that any player i's utility payoff $x_i = X_i(y_i)$ is a (strictly increasing) function of his money payoff y_i and that all players will know this. However, the other players j may not know the specific mathematical form of player i's utility function for money, X_i. In other words, even though they may know player i's *ordinal* utility function, they may not know his *cardinal* utility function. That is to say, they may not know how much *risk* he would be willing to take in order to increase his money payoff y_i by given amounts.

This can be interpreted as inability on our part to predict the *physical outcome* (*i.e.*, the amount of destruction) resulting from alternative strategies of the opponent, where any given "strategy" of his is defined as firing a given *percentage* of his fire arms (case 1). But it can also be interpreted as inability to decide whether certain strategies are *available* to the opponent at all, where now any given "strategy" of his is defined as firing a specified *number* of fire arms (case 3).

Incomplete information can also take the form that a given player i does not know whether another player j does or does not have information about the occurrence or non-occurrence of some specified event e. Such a situation will always come under case 3. This is so because in a situation of this kind, from a game-theoretical point of view, the crucial fact is player i's inability to decide whether player j is in a position to use any strategy s_j^0 involving *one* course of action in case event e does occur, and *another* course of action in case event e does not occur. That is, the situation will essentially amount to ignorance by player i about the availability of certain strategies s_j^0 to player j.

Going back to the three main cases listed above, cases 1 and 2 are both special cases of ignorance by the players about their own or some other players' *payoff functions* $U_i = X_i(Y)$ specifying the utility payoff $x_i = U_i(s_1, \cdots, s_n)$ a given player i obtains if the n players use alternative strategy n-tuples $s = (s_1, \cdots, s_n)$.

Indeed, case 3 can also be reduced to this general case. This is so because the assumption that a given strategy $s_i = s_i^0$ is not *available* to player i is equivalent, from a game-theoretical point of view, to the assumption that player i will never actually *use* strategy s_i^0 (even though it would be physically available to him) because by using s_i^0 he would always obtain some extremely low (*i.e.*, highly negative) payoffs $x_i = U_i(s_1, \cdots, s_i^0, \cdots, s_n)$, whatever strategies $s_1, \cdots, s_{i-1}, s_{i+1}, \cdots, s_n$ the other players $1, \cdots, i-1, i+1, \cdots, n$ may be using.

Accordingly, let $S_i^{(j)}$ ($j = 1$ or $j \neq 1$) denote the *largest* set of strategies s_i which in player j's opinion may be *conceivably* included in player i's strategy space S_i. Let $S_i^{(0)}$ denote player i's "true" strategy space. Then, for the purposes of our analysis, we shall define player i's strategy space S_i as

$$S_i = \bigcup_{k=0}^{n} S_i^{(k)}. \tag{2.1}$$

We lose no generality by assuming that this set S_i as defined by (2.1) is *known* to all players because any lack of information on the part of some player j about this set S_i can be represented within our model as lack of information about the numerical values that player i's payoff function $x_i = U_i(s_1, \cdots, s_i, \cdots, s_n)$ takes for some specific choices of s_i, and in particular whether these values are so low as completely to discourage player i from using these strategies s_i.[9]

Accordingly, we define an I-game G as a game where every player j *knows* the strategy spaces S_i of all players $i = 1, \cdots, j, \cdots, n$ but where, in general, he does *not* know the payoff functions U_i of these players $i = 1, \cdots, j, \cdots, n$.

[9] Likewise, instead of assuming that player j assigns subjective probabilities to events of the form $E = \{s_i^0 \notin S_i\}$, we can always assume that he assigns these probabilities to events of the form $E = \{U_i(s_1, \cdots, s_i, \cdots, s_n) < x_i^0 \text{ whenever } s_i = s_i^0\}$, etc.

3.

In terms of this definition, let us consider a given I-game G from the point of view of a particular player j. He can write the payoff function U_i of each player i (including his own payoff function U_j for $i = j$) in a more explicit form as

$$(3.1)\quad x_i = U_i(s_1, \cdots, s_n) = V_i^*(s_1, \cdots, s_n; \quad a_{0i}, a_{1i}, \cdots, a_{ii}, \cdots, a_{ni}),$$

where V_i^*, unlike U_i, is a function whose mathematical form is (in player j's opinion) *known* to all n players; whereas a_{0i} is a vector consisting of those parameters of function U_i which (in j's opinion) are *unknown* to *all* players; and where each a_{ki} for $k = 1, \cdots, n$ is a vector consisting of those parameters of function U_i which (in j's opinion) are *unknown* to some of the players but are *known* to player k. If a given parameter α is known *both* to players k and m (without being known to all players), then this fact can be represented by introducing two variables α_{ki} and α_{mi} with $\alpha_{ki} = \alpha_{mi} = \alpha$, and then making α_{ki} a component of vector a_{ki} while making α_{mi} a component of vector a_{mi}.

For each vector a_{ki} $(k = 0, 1, \cdots, n)$, we shall assume that its *range space* $A_{ki} = \{a_{ki}\}$, *i.e.*, the set of all possible values it can take, is the whole Euclidian space of the required number of dimensions. Then V_i^* will be a function from the Cartesian product $S_1 \times \cdots \times S_n \times A_{0i} \times \cdots \times A_{ni}$ to player i's utility line Ξ_i, which is itself a copy of the real line R.

Let us define a_k as the vector combining the components of all n vectors $a_{k1}, \cdots, a_{kn}$. Thus we write

$$(3.2)\qquad a_k = (a_{k1}, \cdots, a_{kn}),$$

for $k = 0, 1, \cdots, i, \cdots, n$. Clearly, vector a_0 summarizes the information that (in player j's opinion) none of the players has about the n functions $U_1, \cdots, U_n$, whereas vector $a_k (k = 1, \cdots, n)$ summarizes the information that (in j's opinion) player k has about these functions, except for the information that (in j's opinion) *all* n players share about them. For each vector a_k, its range space will be the set $A_k = \{a_k\} = A_{k1} \times \cdots \times A_{kn}$.

In equation (3.1) we are free to replace each vector $a_{ki}(k = 0, \cdots, n)$ by the larger vector $a_k = (a_{k1}, \cdots, a_{ki}, \cdots, a_{kn})$, even though this will mean that in each case the $(n - 1)$ sub-vectors $a_{k1}, \cdots, a_{k(i-1)}, a_{k(i+1)}, \cdots, a_{kn}$ will occur vacuously in the resulting new equation. Thus, we can write

$$(3.3)\qquad x_i = V_i^*(s_1, \cdots, s_n; \quad a_0, a_1, \cdots, a_i, \cdots, a_n).$$

For any given player i the n vectors $a_0, a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n$ in general will represent *unknown* variables; and the same will be true for the $(n - 1)$ vectors $b_i, \cdots, b_{i-1}, b_{i+1}, \cdots, b_n$ to be defined below. Therefore, under the Bayesian hypothesis, player i will assign a subjective joint probability distribution

$$(3.4)\quad P_i = P_i(a_0, a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_n; \quad b_1, \cdots, b_{i-1}, b_{i+1}, \cdots, b_n)$$

to all these unknown vectors.

For convenience we introduce the shorter notations $a = (a_1, \cdots, a_n)$ and

$b = (b_1, \cdots, b_n)$. The vectors obtained from a and b by omitting the sub-vector a_i and b_i, respectively, will be denoted by a^i and b^i. The corresponding range spaces can be written as $A = A_1 \times \cdots \times A_n$; $B = B_1 \times \cdots \times B_n$; $A^i = A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_n$; $B^i = B_1 \times \cdots \times B_{i-1} \times B_{i+1} \times \cdots \times B_n$.

Now we can write equations (3.3) and (3.4) as

$$x_i = V_i^*(s_1, \cdots, s_n; a_0, a) \tag{3.5}$$

$$P_i = P_i(a_0, a^i; b^i) \tag{3.6}$$

where P_i is a probability distribution over the vector space $A_0 \times A^i \times B^i$.

The other $(n - 1)$ players in general will not know the subjective probability distribution P_i used by player i. But player j (from whose point of view we are analyzing the game) will be able to write P_i for each player i (both $i = j$ and $i \neq j$) in the form

$$P_i(a_0, a^i; b^i) = R_i(a_0, a^i; b^i \mid b_i), \tag{3.7}$$

where R_i, unlike P_i, is a function whose mathematical form is (in player j's opinion) *known* to all n players; whereas b_i is a vector consisting of those parameters of function P_i which (in j's opinion) are *unknown* to some or all of the players $k \neq i$. Of course, player j will realize that player i himself will *know* vector b_i since b_i consists of parameters of player i's own subjective probability distribution P_i.

The vectors $b_1, \cdots, b_{i-1}, b_{i+1}, \cdots, b_n$ occurring in equation (3.4), which so far have been left undefined, are the parameter vectors of the subjective probability distributions $P_1, \cdots, P_{i-1}, P_{i+1}, \cdots, P_n$, unknown to player i. The vector b^i occurring in equations (3.6) and (3.7) is a combination of all these vectors $b_1, \cdots, b_{i-1}, b_{i+1}, \cdots, b_n$, and summarizes the information that (in player j's opinion) player i lacks about the other $(n - 1)$ players' subjective probability distributions $P_1, \cdots, P_{i-1}, P_{i+1}, \cdots, P_n$.

Clearly, function R_i is a function yielding, for each specific value of vector b_i, a probability distribution over the vector space $A^i \times B^i$.

We now propose to eliminate the vector a_0, unknown to all players, from equations (3.5) and (3.7). In the case of equation (3.5) this can be done by taking *expected values* with respect to a_0 in terms of player i's own subjective probability distribution $P_i(a_0, a^i; b^i) = R_i(a_0, a^i; b^i \mid b_i)$. We define

$$\begin{aligned} V_i(s_1, \cdots, s_n; a \mid b_i) &= V_i(s_1, \cdots, s_n; a, b_i) \\ &= \int_{A_0} V_i^*(s_1 \cdots, s_n; a_0, a)\, d_{(a_0)} R_i(a_0, a^i; b^i \mid b_i). \end{aligned} \tag{3.8}$$

Then we write

$$x_i = V_i(s_1, \cdots, s_n; a, b_i), \tag{3.9}$$

where x_i now denotes the *expected value* of player i's payoff in terms of his own subjective probability distribution.

In the case of equation (3.7) we can eliminate a_0 by taking the appropriate *marginal* probability distributions. We define

$$P_i(a^i, b^i) = \int_{A_0} d_{(a_0)} P_i(a_0, a^i; b^i), \tag{3.10}$$

and

$$R_i(a^i, b^i \mid b_i) = \int_{A_0} d_{(a_0)} R_i(a_0, a^i; b^i \mid b_i). \tag{3.11}$$

Then we write

$$P_i(a^i, b^i) = R_i(a^i, b^i \mid b_i). \tag{3.12}$$

We now rewrite equation (3.9) as

$$x_i = V_i(s_1, \cdots, s_n; a, b_i, b^i) = V_i(s_1, \cdots, s_n; a, b), \tag{3.13}$$

where vector b^i occurs only vacuously. Likewise we rewrite equation (3.12) as

$$P_i(a^i, b^i) = R_i(a^i, b^i \mid a_i, b_i), \tag{3.14}$$

where on the right-hand side vector a_i occurs only vacuously.

Finally, we introduce the definitions $c_i = (a_i, b_i)$; $c = (a, b)$; and $c^i = (a^i, b^i)$. Moreover, we write $C_i = A_i \times B_i$; $C = A \times B$; and $C^i = A^i \times B^i$. Clearly, vector c_i represents the *total information* available to player i in the game (if we disregard the information available to all n players). Thus, we may call c_i player i's *information vector*.

From another point of view, we can regard vector c_i as representing certain physical, social, and psychological *attributes* of player i *himself*, in that it summarizes some crucial parameters of player i's own payoff function U_i as well as the main parameters of his beliefs about his social and physical environment. (The relevant parameters of player i's payoff function U_i again partly represent parameters of his subjective utility function X_i and partly represent parameters of his environment, *e.g.*, the amounts of various physical or human resources available to him, *etc.*) From this point of view, vector c_i may be called player i's *attribute vector*.

Thus, under this model, the players' incomplete information about the true nature of the game situation is represented by the assumption that in general the actual value of the attribute vector (or information vector) c_i of any given player i will be known only to player i himself, but will be *unknown* to the other $(n - 1)$ players. That is, as far as these other players are concerned, c_i could have any one of a number—possibly even of an infinite number—of alternative values (which together form the range space $C_i = \{c_i\}$ of vector c_i). We may also express this assumption by saying that in an I-game G, in general, the rules of the game as such allow any given player i to belong to any one of a number of possible "*types*", corresponding to the alternative values his attribute vector c_i could take (and so representing the alternative payoff functions U_i and the alternative subjective probability distributions P_i that player i

might have in the game). Each player is always assumed to know his own actual type but to be in general ignorant about the other players' actual types.

Equations (3.13) and (3.14) now can be written as

$$x_i = V_i(s_1, \cdots, s_n; c) = V_i(s_1, \cdots, s_n; c_1, \cdots, c_n) \tag{3.15}$$

$$P_i(c^i) = R_i(c^i | c_i) \tag{3.16}$$

or

$$P_i(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n) = R_i(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n | c_i). \tag{3.17}$$

We shall regard equations (3.15) and (3.17) [or (3.16)] as the standard forms of the equations defining an I-game G, considered from the point of view of some particular player j.

Formally we define the *standard form* of a given I-game G for some particular player j as an ordered set G such that

$$G = \{S_1, \cdots, S_n; C_1, \cdots, C_n; V_1, \cdots, V_n; R_1, \cdots, R_n\} \tag{3.18}$$

where for $i = 1, \cdots, n$ we write $S_i = \{s_i\}$; $C_i = \{c_i\}$; moreover, where V_i is a function from the set $S_1 \times \cdots \times S_n \times C_1 \times \cdots \times C_n$ to player i's utility line Ξ_i (which is itself a copy of the real line R); and where, for any specific value of the vector c_i, the function $R_i = R_i(c^i | c_i)$ is a probability distribution over the set $C^i = C_1 \times \cdots \times C_{i-1} \times C_{i+1} \times \cdots \times C_n$.

4.

Among C-games the natural analogue of this I-game G will be a C-game G^* with the same payoff functions V_i and the same strategy spaces S_i. However, in G^* the vectors c_i will have to be reinterpreted as being *random vectors* (chance moves) with an *objective* joint probability distribution

$$R^* = R^*(c_1, \cdots, c_n) = R^*(c) \tag{4.1}$$

known to all n players.[10] (If some players did not know R^*, then G^* would not be a C-game.) To make G^* as similar to G as possible, we shall assume that each vector c_i will take its values from the same range space C_i in either game. Moreover, we shall assume that in game G^*, just as in game G, when player i chooses his strategy s_i, he will know only the value of his *own* random vector c_i but will not know the random vectors $c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n$ of the other $(n - 1)$ players. Accordingly we may again call c_i the *information vector* of player i.

Alternatively, we may again interpret this random vector c_i as representing certain physical, social, and psychological attributes of player i himself. (But, of course, now we have to assume that for all n players these attributes are determined by some sort of random process, governed by the probability distribution R^*.) Under this interpretation we may again call c_i the *attribute vector* of player i.

[10] Assuming that a joint probability distribution R^* of the required mathematical form exists (see section 5 below, as well as Part III of this paper).

We shall say that a given C-game G^* is in *standard form* if

1. the payoff functions V_i of G^* have the form indicated by equation (3.15);
2. the vectors $c_1, \cdots, c_n$ occurring in equation (3.15) are random vectors with a joint probability distribution R^* [equation (4.1)] known to all players;
3. each player i is assumed to know only his own vector c_i, and does not know the vectors $c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n$ of the other players when he chooses his strategy s_i.

Sometimes we shall again express these assumptions by saying that the rules of the game allow each player i to belong to any one of a number of alternative *types* (corresponding to alternative specific values that the random vector c_i can take); and that each player will always know his own actual type, but in general will not know those of the other players.

Formally we define a C-game G^* in standard form as an ordered set G^* such that

$$G^* = \{S_1, \cdots, S_n; C_1, \cdots, C_n; V_1, \cdots, V_n; R^*\}. \tag{4.2}$$

Thus, the ordered set G^* differs from the ordered set G [defined by equation (3.18)] only in the fact that the n-tuple $R_1, \cdots, R_n$ occurring in G is replaced in G^* by the singleton R^*.

If we consider the normal form of a game as a special limiting case of a standard form (viz. as the case where the random vectors $c_1, \cdots, c_n$ are empty vectors without components), then, of course, every C-game has a standard form. But only a C-game G^* containing random variables (chance moves) will have a standard form non-trivially different from its normal form.

Indeed, if G^* contains more than one random variable, then it will have several different standard forms. This is so because we can always obtain new standard forms G^{**}—intermediate between the original standard form G^* and the normal form G^{***}—if we suppress *some* of the random variables occurring in G^*, without suppressing *all* of them (as we would do if we wanted to obtain the normal form G^{***} itself). This procedure can be called *partial* normalization as distinguished from the *full* normalization, which would yield the normal form G^{***}.[11]

5.

Suppose that G is an I-game (considered from player j's point of view) while G^* is a C-game, both games being given in standard form. To obtain complete similarity between the two games, it is not enough if the strategy spaces $S_1, \cdots, S_n$, the range spaces $C_1, \cdots, C_n$, and the payoff functions $V_1, \cdots, V_n$

[11] Partial normalization involves essentially the same operations as full normalization (see section 7 below). It involves taking the expected values of the payoff functions V_i with respect to the random variables to be suppressed, and redefining the players' strategies where necessary. However, in the case of partial normalization we also have to replace the probability distribution R^* of the original standard form G^*, by a *marginal* probability distribution not containing the random variables to be suppressed. (In the case of full normalization no such marginal distribution has to be computed because the normal form G^{***} will not contain random variables at all.)

of the two games are the same. It is necessary also that each player i in either game should always assign the same *numerical* proabability p to any given specific event E. Yet in game G player i will assess all probabilities in terms of his *subjective* probability distribution $R_i(c^i | c_i)$; whereas in game G^*—since vector c_i is known to him—he will assess all probabilities in terms of the objective *conditional* probability distribution $R^*(c^i | c_i)$ generated by the basic probability distribution $R^*(c)$ of the game G^*. Therefore, if the two games are to be equivalent, then numerically the distributions $R_i(c^i | c_i)$ and $R^*(c^i | c_i)$ must be identically equal.

This leads to the following definition. Let G be an I-game (as considered by player j), and let G^* be a C-game, both games being given in standard form. We shall say that G and G^* are *Bayes-equivalent* for player j if the following conditions are fulfilled:

1. The two games must have the same strategy spaces $S_1, \cdots, S_n$ and the same range spaces $C_1, \cdots, C_n$.

2. They must have the same payoff functions $V_1, \cdots, V_n$.

3. The subjective probability distribution R_i of each player i in G must satisfy the relationship

$$R_i(c^i | c_i) = R^*(c^i | c_i), \tag{5.1}$$

where $R^*(c) = R^*(c_i, c^i)$ is the basic probability distribution of game G^* and where

$$R^*(c^i | c_i) = R^*(c_i, c^i) \Big/ \int_{C^i} d_{(c^i)} R^*(c_i, c^i). \tag{5.2}$$

In view of equations (5.1) and (5.2) we can write

$$R^*(c) = R^*(c_i, c^i) = R_i(c^i | c_i) \cdot \int_{C^i} d_{(c^i)} R^*(c_i, c^i). \tag{5.3}$$

In contrast to equation (5.2), which ceases to have a clear mathematical meaning when the denominator on its right-hand side becomes zero, equation (5.3) always retains a clear mathematical meaning.

We propose the following postulate.

Postulate 1. Bayes-equivalence. Suppose that some I-game G and some C-game G^* are Bayes-equivalent for player j. Then the two games will be completely equivalent for player j from a game-theoretical standpoint; and, in particular, player j's strategy choice will be governed by the same decision rule (the same solution concept) in either game.

This postulate follows from the Bayesian hypothesis, which implies that every player will use his *subjective* probabilities exactly in the same way as he would use known *objective* probabilities numerically equal to the former. Game G (as assessed by player j) and game G^* agree in all defining characteristics, including the numerical probability distributions used by the players. The only difference is that in G the probabilities used by each player are subjective probabilities whereas in G^* these probabilities are objective (conditional) probabilities. But by the Bayesian hypothesis this difference is immaterial.

Of course, under the assumptions of the postulate, all we can say is that *for player j himself* the two games are completely equivalent for game-theoretical purposes. We cannot conclude on the basis of the information assumed that the two games are likewise equivalent also for some *other* players $k \neq j$. In order to reach this latter conclusion we would have to know that G and G^* would preserve their Bayes-equivalence even if G were analyzed in terms of the functions $V_1, \cdots, V_n$ and $R_1, \cdots, R_n$ postulated by these *other* players k, instead of being analyzed in terms of the functions $V_1, \cdots, V_n$ and $R_1, \cdots, R_n$ postulated by player j himself. But so long as we are interested only in the decision rules that player j himself will follow in game G, all we have to know are the functions $V_1, \cdots, V_n$ and $R_1, \cdots, R_n$ that player j will be using.

Postulate 1 naturally gives rise to the following questions. Given any I-game G, is it always possible to construct a C-game G^* Bayes-equivalent to G? And, in cases where this is possible, is this C-game G^* always unique? These questions are tantamount to asking whether for any arbitrarily chosen n-tuple of subjective probability distributions $R_1(c^1 | c_1), \cdots, R_n(c^n | c_n)$, there always *exists* a probability distribution $R^*(c_1, \cdots, c_n)$ satisfying the functional equation (5.3), and whether this distribution R^* is always *unique* in cases where it does exist. As these questions require an extended discussion, we shall answer them in Part III of this paper (see Theorem III and the subsequent heuristic discussion). We shall see that a given I-game G will have a C-game analogue G^* only if G itself satisfies certain consistency requirements. On the other hand, if such a C-game analogue G^* exists for G then it will be "essentially" unique (in the sense that, in cases where two different C-games G_1^*, and G_2^* are both Bayes-equivalent to a given I-game G, it will make no difference whether we use G_1^* or G_2^* for the analysis of G). In the rest of the present Part I of this paper, we shall restrict our analysis to I-games G for which a Bayes-equivalent C-game analogue G^* does exist.

As we shall make considerable use of Bayes-equivalence relationships between certain I-games G and certain C-games G^* given in standard form, it will be convenient to have a short designation for the latter. Therefore, we shall introduce the term *Bayesian games* as a shorter name for C-games G^* given in standard form. Depending on the nature of the I-game G we shall be dealing with in particular cases, we shall also speak of Bayesian two-person zero-sum games, Bayesian bargaining games, etc.

6.

In view of the important role that Bayesian games will play in our analysis, we shall now consider two alternative (but essentially equivalent) models for these games, which for some purposes will usefully supplement the model we have defined in Sections 4 and 5.

So far we have defined a Bayesian game G^* as a game where each player's payoff $x_i = V_i(s_1, \cdots, s_n; c_1, \cdots, c_n)$ depends, not only on the strategies $s_1, \cdots, s_n$ chosen by the n players, but also on some random vectors (information vectors or attribute vectors) $c_1, \cdots, c_n$. It has also been assumed that all players will know the joint probability distribution $R^*(c_1, \cdots, c_n)$ of these random vectors, but that in general the actual value of any given vector c_i will be known only to player i himself whose information vector (or attribute vector) it represents. This model will be called the *random-vector model* for Bayesian games.

An alternative model for Bayesian games can be described as follows. The actual individuals who will play the roles of players $1, \cdots, n$ in game G^* on any given occasion, will be selected *by lot* from certain populations $\Pi_1, \cdots, \Pi_n$ of potential players. Each population Π_i from which a given player i is to be selected will contain individuals with a variety of different attributes, so that *every* possible combination of attributes (i.e., every possible "type" of player i), corresponding to any specific value $c_i = c_i^0$ that the attribute vector c_i can take in the game, will be represented in this population Π_i. If in population Π_i a given individual's attribute vector c_i has the specific value $c_i = c_i^0$, then we shall say that he belongs to the *attribute class* c_i^0. Thus, each population Π_i will be partitioned into that many attribute classes as the number of different values that player i's attribute vector c_i can take in the game.

As to the random process selecting n players from the n populations $\Pi_1, \cdots, \Pi_n$, we shall assume that the probability of players $1, \cdots, n$ being selected from any specific n-tuple of attribute classes $c_1^0, \cdots, c_n^0$ will be governed[12] by the probability distribution $R^*(c_1, \cdots, c_n)$. We shall also retain the assumptions that this probability distribution R^* will be known to all n players, and that each player i will also know his own attribute class $c_i = c_i^0$ but, in general, will not know the other players' attribute classes $c_1 = c_1^0, \cdots, c_{i-1} = c_{i-1}^0, c_{i+1} = c_{i+1}^0, \cdots, c_n = c_n^0$. As in this model the lottery by which the players are selected occurs prior to any other move in the game, it will be called the *prior-lottery model* for Bayesian games.

Let G be a real-life game situation where the players have incomplete information, and let G^* be a Bayesian game Bayes-equivalent to G (as assessed by a given player j). Then this Bayesian game G^*, interpreted in terms of the prior-lottery model, can be regarded as a possible representation (of course a highly schematic representation) of the real-life random social process which has actually created this game situation G. More particularly, the prior-lottery model pictures this social process as it would be seen by an *outside observer* having information about some aspects of the situation but lacking information about some other aspects. He could not have enough information to predict the attribute vectors $c_1 = c_1^0, \cdots, c_n = c_n^0$ of the n individuals to be selected by this social process to play the roles of players $1, \cdots, n$ in game situation G. But he would have to have enough information to predict the joint probability distribution R^* of the attribute vectors $c_1, \cdots, c_n$ of these n individuals, and, of

[12] Under our assumptions in general the selection of players $1, \cdots, n$ from the respective populations $\Pi_1, \cdots, \Pi_n$ will not be statistically independent events because the probability distribution $R^*(c_1, \cdots, c_n)$ in general will not permit of factorization into n independent probability distributions $R_1^*(c_1), \cdots, R_n^*(c_n)$. Therefore, strictly speaking, our model postulates *simultaneous* random selection of a whole player n-tuple from a population Π of all possible player n-tuples, where Π is the Cartesian product $\Pi = \Pi_1 \times \cdots \times \Pi_n$.

course, also to predict the mathematical form of the payoff functions $V_1, \cdots, V_n$. (But he could not have enough information to predict the payoff functions $U_1, \cdots, U_n$ because this would require knowledge of the attribute vectors of all n players.)

In other words, the hypothetical observer must have exactly all the information *common* to the n players, but must not have access to any additional information private to any one player (or to any sectional group of players—and, of course, he must not have access to any information inaccessible to all of the n players). We shall call such an observer a *properly informed* observer. Thus, the prior-lottery model for Bayesian games can be regarded as a schematic representation of the relevant real-life social process as seen by a properly informed outside observer.

As an example, let us again consider the price-competition game G with incomplete information, and the corresponding Bayesian game G^*, discussed in Section 1 above. Here each player's attribute vector c_i will consist of the variables defining his cost functions, his financial resources, and his facilities to collect information about the other player.[13] Thus, the prior-lottery model of G^* will be a model where each player is chosen at random from some population of possible players with different cost functions, different financial resources, and different information-gathering facilities. We have argued that such a model can be regarded as a schematic representation of the real-life social process which has actually produced the assumed competitive situation, and has actually determined the cost functions, financial resources, and information-gathering facilities, of the two players.

Dr. Selten has suggested[14] a third model for Bayesian games, which we shall call the *Selten model* or the *posterior-lottery model.* Its basic difference from the prior-lottery model consists in the assumption that the lottery selecting the active participants of the game will take place only *after* each potential player has chosen the strategy he would use in case he were in fact selected for active participation in the game.

More particularly, suppose that the attribute vector c_i of player i ($i = 1, \cdots, n$) can take k_i different values in the game. (We shall assume that all k_i's are finite but the model can be easily extended also to the infinite case.) Then, instead of having one randomly selected player i in the game, we shall assume that the role of player i will be played at the same time by k_i different players, each of them representing a different value of the attribute vector c_i. The set of all k_i individuals playing the role of player i in the game will be called the *role class* i. Different individuals in the same role class i will be distinguished by subscripts as players $i_1, i_2, \cdots$. Under these assumptions, obviously the total number of players in the game will not be n but rather will be the larger (usually much larger) number

$$K = \sum_{i=1}^{n} k_i. \tag{6.1}$$

[13] Cf. Footnote 4 above.

[14] In private communication (cf. Footnote 1 above).

It will be assumed that each player i_m from a given role class i will choose some strategy s_i from player i's strategy space S_i. Different members of the same role class i may (but need not) choose different strategies s_i from this strategy space S_i.

After all K players have chosen their strategies, one player i_m from each role class i will be randomly selected as an *active player*. Suppose that the attribute vectors of the n active players so selected will be $c_1 = c_1^0, \cdots, c_n = c_n^0$, and that these players, prior to their selection, have chosen the strategies $s_1 = s_1^0, \cdots, s_n = s_n^0$. Then each active player i_m, selected from role class i, will obtain a payoff

$$x_i = V_i(s_1^0, \cdots, s_n^0; c_1^0, \cdots, c_n^0). \tag{6.2}$$

All other $(K - n)$ players not selected as active players will obtain *zero* payoffs.

It will be assumed that, when the n active players are randomly selected from the n role classes, the probability of selecting individuals with any specific combination of attribute vectors $c_1 = c_1^0, \cdots, c_n = c_n^0$ will be governed by the probability distribution $R^*(c_1, \cdots, c_n)$.[15]

It is easy to see that in all three models we have discussed—in the random-vector model, the prior-lottery model, and the posterior-lottery model—the players' payoff functions, the information available to them, and the probability of any specific event in the game, are all essentially the same.[16] Consequently,

[15] In actual fact, we could just as well assume that each player would choose his strategy only *after* the lottery, and *after* being informed whether this lottery has selected him as an active player or not. (Of course if we made this assumption then players not selected as active players could simply forget about choosing a strategy at all.) From a game-theoretical point of view this assumption would make no real difference so long as each active player would have to choose his strategy without being told the names of the *other* players selected as active players, and in particular without being told the attribute classes to which these other active players would belong.

Thus the fundamental theoretical difference between our second and third models is not so much in the actual *timing* of the postulated lottery as such. It is not so much in the fact that in one case the lottery precedes, and in the other case it follows, the players' strategy choices. The fundamental difference rather lies in the fact that our second model (like our first) conceives of the game as an *n-person* game, in which only the n active players are formally "players of the game"; whereas our third model conceives of the game as a *K-person* game, in which both the active and the inactive players are formally regarded as "players". Yet, to make it easier to avoid confusion between the two models, it is convenient to assume also a difference in the actual timing of the assumed lottery.

[16] Technically speaking, the players' effective payoff functions under the posterior-lottery model are not quite identical with their payoff functions under the other two models, but this difference is immaterial for our purposes. Under the posterior-lottery model, let $r = r_i(c_i^0)$ be the probability (marginal probability) that a given player i_m with attribute vector $c_i = c_i^0$ will be selected as the active player from role class i. Then player i_m will have the probability r of obtaining a payoff corresponding to the payoff function V_i and will have the probability $(1 - r)$ of obtaining a zero payoff whereas under the other two models each player i will *always* obtain a payoff corresponding to the payoff function V_i. Consequently, under the posterior-lottery model player i_m's expected payoff will be only r times $(0 < r \leqq 1)$ the expected payoff he could anticipate under the other two models. However, under most game-theoretical solution concepts (and in particular under all solution concepts we would ourselves choose for analyzing game situations), the solution of the game

all three models can be considered to be essentially equivalent. But, of course, formally they represent quite different game-theoretical models, as the random-vector model corresponds to an n-person game G^* with complete information, whereas the posterior-lottery model corresponds to a *K-person* game G^{**} with complete information. In what follows, unless the contrary is indicated, by the term "Bayesian game" we shall always mean the n-person game G^* corresponding to the random-vector model, whereas the K-person game G^{**} corresponding to the posterior-lottery model will be called the *Selten game.*

In contrast to the other two models, the prior-lottery model formally does not qualify as a true "game" at all because it assumes that the n players are selected by a chance move representing the first move of the game, whereas under the formal game-theoretical definition of a game the identity of the players must always be known from the very beginning, *before* any chance move or personal move has occurred in the game.

Thus, we may characterize the situation as follows. The real-life social process underlying the I-game G we are considering is best represented by the prior-lottery model. But the latter does not correspond to a true "game" in a game-theoretical sense. The other two models are two alternative ways of converting the prior-lottery model into a true "game". In both cases this conversion entails a price in the form of introducing some unrealistic assumptions. In the case of the posterior-lottery model corresponding to the Selten game G^{**}, the price consists in introducing $(K - n)$ fictitious players in addition to the n real players participating in the game.[17]

In the case of the random-vector model corresponding to the Bayesian game G^*, there are no fictitious players, but we have to pay the price of making the unrealistic assumption that the attribute vector c_i of each player i is determined by a chance move *after* the beginning of the game—which seems to imply that player i will be in existence for some period of time, however short, during which *he will not know yet* the specific value $c_i = c_i^0$ his attribute vector c_i will take. So long as the Bayesian game G^* corresponding to the random-vector model is being considered in its standard form, this unrealistic assumption makes very little difference. But, as we shall see, when we convert G^* into its normal form this unrealistic assumption implied by our model does cause certain technical difficulties, because it seems to commit us to the assumption that each player can choose his normalized strategy (i.e., his strategy for the normal-form version of G^*) *before* he learns the value of his own attribute vector c_i. An important advantage of the Selten game G^{**} lies in the fact that it does not require this particular unrealistic assumption: we are free to assume that every player i_m

will remain invariant if the players' payoff functions are multiplied by positive constants r (even if different constants r are used for different players).

In any case, the posterior-lottery model can be made completely equivalent to the other two models if we assume that each active player i_m will obtain a payoff corresponding to the payoff function $V_i/r_i(c_i^0)$, instead of obtaining a payoff corresponding to the payoff function V_i as such [as prescribed by equation (6.2)].

[17] This will be true even if we change the timing of the assumed lottery in Selten's model (see Footnote 15 above).

will know his own attribute vector c_i from the very beginning of the game, and will always choose his own strategy in light of this information.[18]

Thus, as analytical tools used in the analysis of a given I-game G, both the Bayesian game G^* and the Selten game G^{**} have their own advantages and disadvantages.[19]

7.

Let G be an I-game given in standard form, and let G^* be a Bayesian game Bayes-equivalent to G. Then we define the normal form $\mathfrak{N}(G)$ of this I-game G as being the normal form $\mathfrak{N}(G^*)$ of the Bayesian game G^*.

To obtain this normal form we first have to replace the strategies s_i of each player i by *normalized strategies* s_i^*. A normalized strategy s_i^* can be regarded as a conditional statement specifying the strategy $s_i = s_i^*(c_i)$ that player i would use if his information vector (or attribute vector) c_i took any given specific value. Mathematically, a normalized strategy s_i^* is a function from the range space $C_i = \{c_i\}$ of vector c_i to player i's strategy space $S_i = \{s_i\}$. The set of all possible such functions s_i^* is called player i's normalized-strategy space $S_i^* = \{s_i^*\}$. In contrast to these normalized strategies s_i^*, the strategies s_i available to player i in the standard form of the game will be called his *ordinary strategies*.

If in a given game the information vector c_i of a certain player i can take only k different values (with k finite) so that we can write

$$c_i = c_i^1, \cdots, c_i^k, \tag{7.1}$$

then any normalized strategy s_i^* of this player can be defined simply as a k-tuple of ordinary strategies

$$s_i^* = (s_i^1, \cdots, s_i^k), \tag{7.2}$$

where $s_i^m = s_i^*(c_i^m)$, with $m = 1, \cdots, k$, denotes the strategy that player i would use in the standard form of the game if his information vector c_i took the specific value $c_i = c_i^m$. In this case player i's normalized strategy space $S_i^* = \{s_i^*\}$ will be the set of all such k-tuples s_i^*, that is, it will be the k-times repeated Cartesian product of player i's ordinary strategy space S_i by itself. Thus we can write $S_i^* = S_i^1 \times \cdots \times S_i^k$ with $S_i^1 = \cdots = S_i^k = S_i$.

Under either of these definitions, the normalized strategies s_i^* will not have the nature of *mixed* strategies but rather that of *behavioral* strategies. Never-

[18] Moreover, as Selten has pointed out, his model also has the advantage that it can be extended to the case where the subjective probability distributions $R_1, \cdots, R_n$ of a given I-game G fail to satisfy the required consistency conditions, so that no probability distribution R^* satisfying equation (5.3) will exist, and therefore no Bayesian game G^* Bayes-equivalent to G can be constructed at all. In other words, for any I-game G we can always define an equivalent Selten game G^{**}, even in cases where we cannot define an equivalent Bayesian game G^*. (See Section 15, Part III.)

[19] We have given intuitive reasons why a Bayesian game G^* and the corresponding Selten game G^{**} are essentially equivalent. For a more detailed and more rigorous game-theoretical proof the reader is referred to a forthcoming paper by Reinhard Selten.

theless, these definitions are admissible because any game G^* in standard form is a game of perfect recall, and so it will make no difference whether the players are assumed to use behavioral strategies or mixed strategies [4].

Equation (3.15) can now be written as

$$(7.3)\quad x_i = V_i(s_1^*(c_1), \cdots, s_n^*(c_n); \qquad c_1, \cdots, c_n) = V_i(s_1^*, \cdots, s_n^*; c).$$

In order to obtain the normal form $\mathfrak{N}(G) = \mathfrak{N}(G^*)$, all we have to do now is to take expected values in equation (7.3) with respect to the whole random vector c, in terms of the basic probability distribution $R^*(c)$ of the game. We define

$$(7.4)\qquad \mathcal{E}(x_i) = W_i(s_1^*, \cdots, s_n^*) = \int_C V_i(s_1^*, \cdots, s_n^*; c)\, d_c R^*(c).$$

Since each player will treat his expected payoff as his effective payoff from the game, we can replace $\mathcal{E}(x_i)$ simply by x_i and write

$$(7.5)\qquad x_i = W_i(s_1^*, \cdots, s_n^*).$$

We can now define the normal form of games G and G^* as the ordered set

$$(7.6)\qquad \mathfrak{N}(G) = \mathfrak{N}(G^*) = \{S_1^*, \cdots, S_n^*; W_1, \cdots, W_n\}.$$

Compared with equations (3.18) and (4.2) defining the standard forms of these two games, in equation (7.6) the ordinary strategy spaces S_i have been replaced by the normalized strategy spaces S_i^*, and the ordinary payoff functions V_i have been replaced by the normalized payoff functions W_i. On the other hand, the range spaces C_i as well as the probability distributions R_i or R^* have been omitted because the normal form $\mathfrak{N}(G) = \mathfrak{N}(G^*)$ of games G and G^* does not any more involve the random vectors $c_1, \cdots, c_n$.

This normal form, however, has the disadvantage that it is defined in terms of the players' *unconditional* payoff expectations $\mathcal{E}(x_i) = W_i(s_1^*, \cdots, s_n^*)$, though in actual fact each player's strategy choice will be governed by his *conditional* payoff expectation $\mathcal{E}(x_i \mid c_i)$, because he will always know his own information vector c_i at the time of making his strategy choice. This conditional expectation can be defined as

$$(7.7)\quad \mathcal{E}(x_i \mid c_i) = Z_i(s_1^*, \cdots, s_n^* \mid c_i) = \int_{C^i} V_i(s_1^*, \cdots, s_n^*; c_i, c^i)\, d_{(c^i)} R^*(c^i \mid c_i)$$

To be sure, it can be shown (see Theorem I of Section 8, Part II) that if any given player i maximizes his *unconditional* payoff expectation W_i, then he will also be maximizing his *conditional* payoff expectation $Z_i(\cdot \mid c_i)$ for each specific value of c_i, with the possible exception of a small set of c_i values which can occur only with probability zero. In this respect our analysis bears out von Neumann and Morgenstern's Normalization Principle [7, pp. 79–84], according to which the players can safely restrict their attention to the normal form of the game when they are making their strategy choices.

However, owing to the special nature of Bayesian games, the Normalization

Principle has only restricted validity for them, and their normal form $\mathfrak{N}(G^*)$ must be used with special care, because solution concepts based on uncritical use of the normal form may give counterintuitive results (see Section 11 of Part II of this paper). In view of this fact, we shall introduce the concept of a semi-normal form. The *semi-normal form* $\mathcal{S}(G) = \mathcal{S}(G^*)$ of games G and G^* will be defined as a game where the players' strategies are the normalized strategies s_i^* described above, but where their payoff functions are the *conditional* payoff-expectation functions $Z_i(\cdot \mid c_i)$ defined by equation (7.7). Formally we define the semi-normal form of the games G and G^* as the ordered set

$$(7.8) \quad \mathcal{S}(G) = \mathcal{S}(G^*) = \{S_1^*, \cdots, S_n^*; C_1, \cdots, C_n; Z_1, \cdots, Z_n; R^*\}.$$

As the semi-normal form, unlike the normal form, does involve the random vectors $c_1, \cdots, c_n$, now the range spaces $C_1, \cdots, C_n$, and the probability distribution R^*, which have been omitted from equation (7.6), reappear in equation (7.8).

Instead of von Neumann and Morgenstern's Normalization Principle, we shall use only the weaker Semi-normalization Principle (Postulate 2 below), which is implied by the Normalization Principle but which does not itself imply the latter:

Postulate 2. Sufficiency of the Semi-normal Form. The solution of any Bayesian game G^*, and of the Bayes-equivalent I-game G, can be defined in terms of the semi-normal form $\mathcal{S}(G^*) = \mathcal{S}(G)$, without going back to the standard form of G^* or of G.

References

1. Robert J. Aumann, "On Choosing a Function at Random," in Fred B. Wright (editor), *Symposium on Ergodic Theory*, New Orleans: Academic Press, 1963, pp. 1–20.
2. ——, "Mixed and Behavior Strategies in Infinite Extensive Games," in M. Dresher, L. S. Shapley, and A. W. Tucker (editors), *Advances in Game Theory*, Princeton: Princeton University Press, 1964, pp. 627–650.
3. John C. Harsanyi, "Bargaining in Ignorance of the Opponent's Utility Function," *Journal of Conflict Resolution*, 6 (1962), pp. 29–38.
4. H. W. Kuhn, "Extensive Games and the Problem of Information," in H. W. Kuhn and A. W. Tucker (editors), *Contributions to the Theory of Games*, Vol. II, Princeton: Princeton University Press, 1953, pp. 193–216.
5. J. C. C. McKinsey, *Introduction to the Theory of Games*, New York: McGraw-Hill, 1952.
6. Leonard J. Savage, *The Foundations of Statistics*, New York: John Wiley and Sons, 1954.
7. John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1953.