NON-COOPERATIVE GAME THEORY

Giacomo Bonanno

Department of Economics
University of California
Davis, CA 95616 – USA
e-mail: gfbonanno@ucdavis.edu

August, 2008

Abstract

This is the first draft of the entry "Game Theory" to appear in the Sage Handbook of the Philosophy of Social Science (edited by Ian Jarvie & Jesús Zamora Bonilla), Part III, Chapter 16.

1. Introduction

Game theory is a branch of mathematics that deals with interactive decision making, that is, with situations where two or more individuals (called *players*) make decisions that affect each other.[1] Since the final outcome depends on the actions taken by all the players, it becomes necessary for each player to try to predict the choices of his opponents, while realizing that they are simultaneously trying to put themselves in *his* shoes to figure out what *he* will do.

The birth of game theory is usually associated with the publication in 1944 of the book *Theory of Games and Economic Behavior* by the mathematician John von Neumann and the

[1] An example is a sealed-bid first-price auction, where each participant submits a bid for an object, in ignorance of the bids chosen by his opponents, and the object is assigned to the highest bidder who pays his own bid, while the others pay nothing and receive nothing.

economist Oskar Morgenstern, although important results had been obtained earlier.[2] Applications of game theory can be found in many fields, most notably biology, computer science, economics,[3] military science, political science and sociology.

Game theory has been traditionally divided into two branches: non-cooperative and cooperative. Cooperative game theory deals with situations where there are institutions that make agreements among the players binding. In such a setting the central question becomes one of agreeing on a best joint course of action, where 'best' could have different meanings, such as 'acceptable to all players and coalitions of players'[4] or 'satisfying some desirable properties'[5]. Non-cooperative game theory, on the other hand, deals with institutional settings where binding agreements are not possible, whether it is because communication is impossible or because agreements are illegal[6] or because there is no authority that can enforce compliance[7].

Because of space limitations we shall deal exclusively with non-cooperative games. Our focus will be on the philosophical and epistemological issues that arise in non-cooperative

---

[2] For a detailed historical account of the development of game theory see Aumann (1989).

[3] The Nobel prize in economics was awarded to game theorists three times: in 1994 to John C. Harsanyi, John F. Nash Jr. and Reinhard Selten; in 2005 to Robert J. Aumann and Thomas C. Schelling and in 2007 to Leonid Hurwicz, Eric S. Maskin and Roger B. Myerson.

[4] For example, the *core* of a co-operative game identifies a set of "best" agreements in this sense.

[5] For example, the *Shapley value* identifies the "best" agreement in this sense.

[6] For example, many countries have antitrust laws that forbid agreements among competing firms concerning prices or production.

[7] As is the case in the international arena.

games, in particular on the notion of rationality and mutual recognition of rationality. We shall begin with simultaneous or strategic-form games and then turn to games that have a sequential structure (extensive-form games).

2. Strategic-form games and common knowledge of rationality

A *game in strategic form with ordinal payoffs* consists of the following elements:

(1) the set $N = \{1,...,n\}$ of *players*;

(2) for every player $i \in N$, the set S_i of *strategies* (or choices) available to player *i*;

(3) the set of possible *outcomes* O;

(4) an *outcome function* $z: S \to O$ that associates, with every strategy profile (specifying a choice for each player) $(s_1,...,s_n) \in S = S_1 \times ... \times S_n$, the resulting outcome;

(5) for every player $i \in N$, a *weak total order* $\succsim_i$ on O representing player *i*'s ranking of the outcomes.[8,9]

---

[8] Thus $\succsim_i \subseteq O \times O$ is a binary relation on *O* which is connected (for all $o, o' \in O$, either $o \succsim_i o'$, or $o' \succsim_i o$, or both) and transitive (if $o_1 \succsim_i o_2$ and $o_2 \succsim_i o_3$ then $o_1 \succsim_i o_3$). The interpretation of $o \succsim_i o'$ is that player *i* considers outcome *o* to be *at least as good* as outcome o'. We denote that player *i prefers* outcome *o* to outcome o' by $o \succ_i o'$ and define it as $o \succsim_i o'$ and not $o' \succsim_i o$. Player *i* is *indifferent* between *o* and o', denoted by $o \sim_i o'$, if $o \succsim_i o'$ and $o' \succsim_i o$.

[9] For example a sealed-bid first price auction with two bidders, two legal bids ($10 and $15), a tie-breaking rule that declares player 1 the winner and where each player values the object more than $15 and has selfish preferences corresponds to the following strategic-form game: $N = \{1,2\}$, $S_1 = S_2 = \{10, 15\}$, $O = \{a, b, c\}$ (where *a* is the outcome "player 1 gets the object and pays $10", *b* is the outcome "player 2 gets the object and pays $15" and

In order to simplify the representation of a game the last three elements are usually collapsed into a *payoff function* $\pi_i : S \to \mathbb{R}$, for every player *i*, which is a numerical function ($\mathbb{R}$ denotes the set of real numbers) satisfying the following property: the strategy profile *s* is assigned a number greater than or equal to the number assigned to the strategy profile s' if and only if player *i* considers the outcome resulting from *s* to be at least as good as the outcome resulting from s'. Formally: for every $s, s' \in S$, $\pi_i(s) \geq \pi_i(s')$ if and only if $z(s) \succsim_i z(s')$.[10] Since $\succsim_i$ is a weak total order, if *O* is a finite set then such a payoff function always exists; furthermore, there is an infinite number of possible payoff functions that can be used to represent $\succsim_i$. It is important to note that the payoffs have no meaning beyond the ordinal ranking that they induce on the set of strategy profiles.

In the case of two players, a convenient way to represent a game is by means of a table where each row is labeled with a strategy of player 1 and each column with a strategy of player 2. Inside the cell that corresponds to the row labeled *x* and the column labeled *y* the pair of numbers $(\pi_1(x,y), \pi_2(x,y))$ is given, denoting the payoffs of player 1 and player 2, respectively. Table 16.1 represents a two-player game where $S_1 = \{a,b,c\}$ and $S_2 = \{d,e,f\}$. In this game, for example, player 1 is indifferent between $z(a,e)$ (the outcome associated with the strategy profile (a,e)) and $z(b,e)$; on the other hand, he prefers $z(a,e)$ to $z(c,e)$.

---

c is the outcome "player 1 gets the object and pays \$15"), *z*(10,10) = *a*, *z*(10,15) = *b*, *z*(15,10) = *z*(15,15) = *c*, $a \succ_1 c \succ_1 b$ and $b \succ_2 a \sim_2 c$.

		Player 2		
		d	*e*	*f*
Player 1	*a*	3 , 2	3 , 1	0 , 1
	b	2 , 3	3 , 2	0 , 1
	c	1 , 2	1 , 2	4 , 1

TABLE 16.1

The *epistemic foundation program* in game theory aims to identify, for every game, the strategies that might be chosen by rational and intelligent players who know the structure of the game and the preferences of their opponents and who recognize each other's rationality. The two central questions are thus: (1) under what circumstances is a player rational? and (2) what does 'mutual recognition of rationality' mean? The latter notion has been interpreted as *common knowledge* of rationality. Informally, something is common knowledge if everybody knows it, everybody knows that everybody knows it, … and so on, *ad infinitum*.[11] A defining characteristic of knowledge is truth: if a player knows E then E must be true. A more general notion is that of *belief*, which allows for the possibility of mistakes: belief of E is compatible with E being false. Thus a more appealing notion is that of common belief of rationality; however, in order to simplify the exposition, we shall restrict attention to knowledge and refer the reader to Battigalli and Bonanno (1999) for the analysis of common belief. The state of interactive knowledge among a set of players can be modeled by means of a set of *states* Ω and, for every player $i \in N$,

---

[10] In the example of the previous footnote the following are possible payoff functions: $\pi_1(10,10) = 3$, $\pi_1(15,10) = \pi_1(15,15) = 2$, $\pi_1(10,15) = 1$ and $\pi_2(10,15) = 2$, $\pi_2(15,10) = \pi_2(15,15) = 1$.

a partition $\mathcal{K}_i$ of Ω (thus $\mathcal{K}_i$ is a binary relation on Ω which is reflexive, transitive and symmetric, that is, an equivalence relation). Given a state $\omega \in \Omega$, we denote the cell of *i*'s partition that contains ω (that is, the equivalence class of ω) by $\mathcal{K}_i(\omega)$. The interpretation is that, at state ω, player *i* cannot distinguish between any two states in $\mathcal{K}_i(\omega)$, that is $-$ as far as she knows $-$ the true state could be any of the elements in $\mathcal{K}_i(\omega)$. The collection $\left\langle N, \Omega, \{\mathcal{K}_i\}_{i \in N} \right\rangle$ is called an *interactive knowledge structure*. A state $\omega \in \Omega$ is thought of as a complete description of the world and the subsets of Ω, which are called *events*, represent propositions about the world.

Knowledge pertains to propositions and a proposition is identified with the set of states where it is true. For every player *i*, we can define a *knowledge operator* $K_i : 2^{\Omega} \to 2^{\Omega}$ (where 2^{Ω} denotes the set of subsets of Ω) as follows: $K_i E = \{\omega \in \Omega : \mathcal{K}_i(\omega) \subseteq E\}$. Thus at state ω player *i* knows (the proposition represented by) event *E* if *E* is true at every state that player *i* considers possible: $\omega \in K_i E$ (*i* knows *E* at ω) if and only if $\omega' \in E$ (*E* is true at ω') for every $\omega' \in \mathcal{K}_i(\omega)$ (for every ω' that *i* considers possible at ω). This is illustrated in Figure 16.2 where $\Omega = \{\alpha, \beta, \gamma, \delta, \varepsilon\}$ and the cells of the partition of a player are denoted by rounded rectangles. Thus, for example, $\mathcal{K}_1(\beta) = \{\beta, \gamma\}$, that is, at state β player 1 is uncertain as to whether the true state is β or γ. Consider the event $E = \{\alpha, \beta, \delta\}$. Then $K_1 E = \{\alpha, \delta\}$ and $K_2 E = \{\alpha, \beta\}$, so that

[11] The notion of common knowledge was introduced independently by Lewis (1969) and Aumann (1976).

$K_1(K_2E) = \{\alpha\}$ while $K_2(K_1E) = \varnothing$ ($\varnothing$ denotes the empty set)[12]. Hence at state α both players know E ($\alpha \in K_1E$ and $\alpha \in K_2E$) and, while player 1 knows that player 2 knows E ($\alpha \in K_1K_2E$), it is not the case that player 2 knows that player 1 knows E ($\alpha \notin K_2K_1E$).

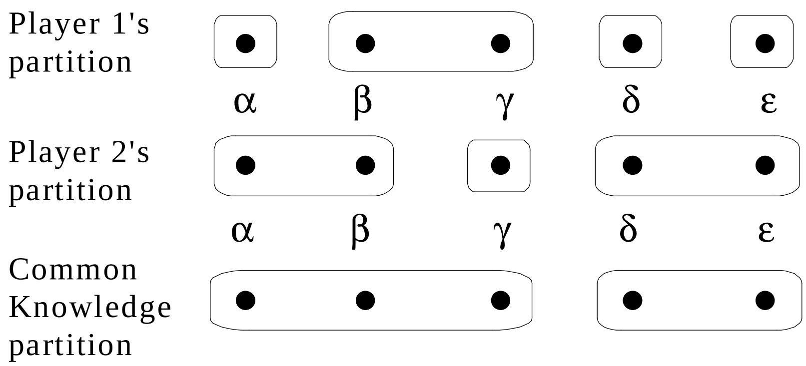


FIGURE 16.2

Given an interactive knowledge structure $\langle N, \Omega, \{\mathcal{K}_i\}_{i \in N} \rangle$, in order to determine whether an event E is common knowledge at some state ω we construct a new partition, called the *common knowledge partition*, as follows. Let $\omega, \omega' \in \Omega$. We say that ω' is *reachable* from ω if there is a sequence $\langle \omega_1, \omega_2, ..., \omega_n \rangle$ in Ω and a sequence of players $\langle j_1, j_2, ..., j_{n-1} \rangle$ in N such that (1) $\omega_1 = \omega$, (2) $\omega_n = \omega'$ and (3) for every $i = 1, ..., n-1$, $\omega_{i+1} \in \mathcal{K}_{j_i}(\omega_i)$. Let $\mathcal{K}_*(\omega)$ denote the set of states reachable from ω. For example, in Figure 16.2, $\mathcal{K}_*(\alpha) = \{\alpha, \beta, \gamma\}$. The common knowledge partition is obtained by enclosing two states in the same cell if and only if one is reachable from the other. Figure 16.2 shows the common knowledge partition constructed from

[12] From now on we shall write K_1K_2E instead of $K_1(K_2E)$.

the partitions of player 1 and player 2. We can now define a *common knowledge operator* $CK : 2^{\Omega} \to 2^{\Omega}$ as follows: $CKE = \{\omega \in \Omega : \mathcal{K}_*(\omega) \subseteq E\}$. In the example illustrated in Figure 16.2, $CK\{\alpha,\beta,\delta,\varepsilon\} = \{\delta,\varepsilon\}$ and $CK\{\alpha,\beta,\delta\} = \varnothing$. Thus, if $F = \{\alpha,\beta,\delta,\varepsilon\}$ then at state δ both players know *F* and both players know that both players know *F*, and so on; that is, at δ it is common knowledge that *F* has occurred. On the other hand, if $E = \{\alpha,\beta,\delta\}$ then, at state α both players know *E* but *E* is not common knowledge (indeed we saw above that at α it is not the case that player 2 knows that player 1 knows *E*).

Armed with a precise definition of common knowledge, we can now turn to the central question of what strategies can be chosen when there is common knowledge of rationality. In order to do this, we need to define what it means for a player to be rational. Intuitively, a player is rational if she chooses an action which is "best" given what she believes or knows. In order to make this more precise we need to introduce the notion of model of a game. Given a game *G* and an interactive knowledge structure $\langle N, \Omega, \{\mathcal{K}_i\}_{i\in N} \rangle$ we obtain a *model of G* by adding, for every player *i*, a function $\sigma_i : \Omega \to S_i$ that associates with every state a strategy of player *i*. The interpretation of $s_i = \sigma_i(\omega)$ is that, at state ω, player *i* plays (or chooses) strategy s_i. We impose the restriction that a player always knows what strategy he is choosing, that is, the function σ_i is constant on the cells of player *i*'s partition: if $\omega' \in \mathcal{K}_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$. The addition of the functions σ_i to an interactive knowledge structure yields an interpretation of events in terms of propositions about what actions the players take, thereby giving content to players' knowledge. Figure 16.3 reproduces the game of Table 16.1 and shows a model of, where

$\sigma_1(\alpha) = a,\ \sigma_1(\beta) = \sigma_1(\gamma) = c, \quad \sigma_2(\alpha) = \sigma_2(\beta) = e$ and $\sigma_2(\gamma) = f$.

		Player 2		
		d	*e*	*f*
Player 1	*a*	3 , 2	3 , 1	0 , 1
	b	2 , 3	3 , 2	0 , 1
	c	1 , 2	1 , 2	4 , 1

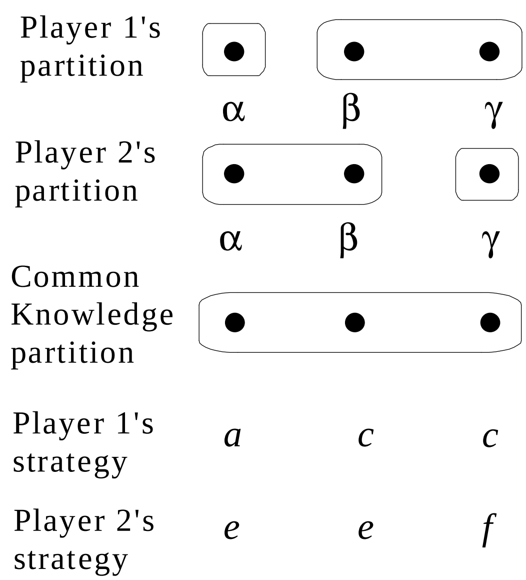


FIGURE 16.3

The following is a very weak definition of rationality: at a state a player is *rational* if it is not the case that he knows that his payoff would be greater if he had chosen a different strategy than the one he is choosing at that state. This definition can be stated formally as follows. First we label a player as *irrational* at state ω if there exists a strategy $s_i' \in S_i$ such that (1) $s_i' \neq s_i$, where $s_i = \sigma_i(\omega)$, and (2) for every $\omega' \in \mathcal{K}_i(\omega)$, $\pi_i(s_i', \sigma_{-i}(\omega')) > \pi_i(s_i, \sigma_{-i}(\omega'))$, where $\sigma_{-i}(\omega')$ denotes the strategy profile of players other than *i* at state ω': $\sigma_{-i}(\omega') = (\sigma_1(\omega'), ..., \sigma_{i-1}(\omega'), \sigma_{i+1}(\omega'), ..., \sigma_n(\omega'))$ [recall also that, by definition of model, $\sigma_i(\omega) = \sigma_i(\omega')$ for every $\omega' \in \mathcal{K}_i(\omega)$]. Secondly, we define a player to be *rational* at state ω if and only if he is not irrational at ω. For example, in the model illustrated in Figure 16.3 (viewed as a model of the game of Table 16.1), player 1 is rational at state β despite the fact that *a* would be a better choice than *c* there (since player 2 is choosing *e*), because he does not know that player 2 is choosing *e*: he is uncertain as to whether player 2 is choosing *e* or *f* and *c* is a best reply to *f*. On the other hand, at state γ, player 2 is *not* rational because she knows that player 1 is choosing *c* and she would get a higher payoff by playing *d*. Let R_i denote the event that player *i* is rational. For example, in the model illustrated in Figure 16.3, $R_1 = \{\alpha, \beta, \gamma\}$ and $R_2 = \{\alpha, \beta\}$. Let *R* be the event that every player is rational: $R = \bigcap_{i \in N} R_i$. In the model illustrated in Figure 16.3, $R = \{\alpha, \beta\}$ and there is no state where it is common knowledge that both players are rational: $CKR = \varnothing$. We are now in a position to express more precisely the question "what strategy profiles are compatible with common knowledge or rationality?" as follows. Suppose that $\omega \in CKR$ (that is, at ω it is common knowledge that all players are rational): what can we say about the strategy profile $\sigma(\omega) = (\sigma_1(\omega), ..., \sigma_n(\omega))$? The answer to this question we need to introduce the following definition. Fix a game and let $s_i, s_i' \in S_i$ be two strategies of player *i*. We say that s_i' *strictly dominates* s_i if $\pi_i(s_i', s_{-i}) > \pi_i(s_i, s_{-i})$ for every $s_{-i} \in S_{-i} = S_1 \times ... \times S_{i-1} \times S_{i+1} \times ... \times S_n$, that is, if s_i' is strictly better than s_i for player *i* against every possible profile of strategies of the opponents. For example, in the game of Table 16.1, strategy *d* of player 2 strictly dominates strategy *f*. A strategy of player *i* is *strictly dominated* if there is another strategy of player *i* that strictly dominates it. In the game of Table 16.1, with the exception of strategy *f* of player 2, there are no other strategies of either player that are strictly dominated.

A rational player would not play a strictly dominated strategy s_i, since he can obtain a higher payoff by switching to a strategy that dominates it. If the other players know that he is rational, they know that they are in fact playing the smaller game obtained by ruling out strategy s_i. In this smaller game there might be a player who has a strictly dominated strategy and thus, if rational, she will not play it. Hence this strategy can also be ruled out and the game can be reduced further. This procedure of elimination of strategies is called the *iterated deletion of strictly dominated pure strategies.*[13] For example, in the game of Table 16.2, deletion of the strictly dominated strategy f of player 2 leads to a smaller game where strategy c of player 1 becomes strictly dominated by a; deletion of c leads to a yet smaller game where strategy e of

[13] The precise definition of this procedure is as follows. Given a game $G = \left\langle N, \{S_i\}_{i \in N}, \{\pi_i\}_{i \in N} \right\rangle$ (where $\pi_i : S \to \mathbb{R}$ is the ordinal payoff function of player *i*), for every player *i* let $S_i^0, S_i^1, \ldots$ be the sequence of subsets of S_i defined recursively as follows: (1) let $S_i^0 = S_i$ and let $D_i^0 \subseteq S_i^0$ be the set of strategies of player *i* that are strictly dominated in *G*; (2) for $m \geq 1$ let $S_i^m = S_i^{m-1} \setminus D_i^{m-1}$, where $S_i^{m-1} \setminus D_i^{m-1}$ denotes the complement of D_i^{m-1} in S_i^{m-1} and D_i^{m-1} is the set of strategies of player *i* that are strictly dominated in the game whose strategy sets are given by $S_1^{m-1}, S_2^{m-1}, \ldots, S_n^{m-1}$. Define $S_i^\infty = \bigcap_{m \in \mathbb{N}} S_i^m$ (where $\mathbb{N}$ denotes the set of non-negative integers).

player 2 becomes strictly dominated (by *d*); after deleting *e*, strategy *b* of player 1 becomes strictly dominated by *a* and deletion of *b* leaves only the strategy profile (a, d).

One of the first and most important results in the epistemic foundations of game theory is the following: *common knowledge of the rationality of all the players implies the play of a strategy profile that survives the iterated deletion of strictly dominated pure strategies*.[14] For example, in the game of Table 16.1, if there is common knowledge of rationality then player 1 will play *a* and player 2 will play *d*, since (a, d) is the only strategy profile that survives the iterated deletion of strictly dominated strategies.

[14] Formally, in an arbitrary model of a game *G* if $\omega \in CKR$ then $\sigma(\omega) \in S_1^\infty \times ... \times S_n^\infty$ (where the sets S_i^∞ are as defined in the previous footnote). The converse is also true, in the sense that if $s \in S_1^\infty \times ... \times S_n^\infty$ then there is a model of *G* and a state ω such that $\omega \in CKR$ and $s = \sigma(\omega)$. For more details on the history and various formulations of this result see Bonanno (2008). In particular, if one allows for cardinal – rather than ordinal – payoffs (see Section 3) and/or the notion of rationality is strengthened, then it may be possible to eliminate more strategy profiles.

3. Nash equilibrium, cardinal payoffs and mixed strategies.

There are games where no strategy is strictly dominated and, therefore, common knowledge of rationality is compatible with *every* strategy profile. An example of such a game is given in Table 16.4.

		Player 2	
		c	*d*
Player 1	*a*	3 , 3	1 , 0
	b	4 , 1	1 , 2

TABLE 16.4

A weaker notion than the iterative deletion of strictly dominated strategies is that of *Nash equilibrium*. Given a game with ordinal payoffs $G = \left\langle N, \{S_i\}_{i\in N}, \{\pi_i\}_{i\in N} \right\rangle$, a strategy profile $s^* = \left(s_1^*,...,s_n^*\right)$ is called a Nash equilibrium if no player could obtain a higher payoff by unilaterally changing his choice, that is, if, for every $i \in N$, $\pi_i(s^*) \geq \pi_i(s_1^*,...,s_{i-1}^*, s_i, s_{i+1}^*,...,s_n^*)$ for every $s_i \in S_i$. For example, in the game illustrated in Table 16.4, (b,d) is a Nash equilibrium, while none of the other strategy profiles is.[15] A possible interpretation of Nash equilibrium is in terms of a *self-enforcing agreement*. Recall that in non-cooperative games it is assumed that players cannot reach enforceable agreements and thus an agreement is viable only

[15] For instance, (b,c) is not a Nash equilibrium because $\pi_2(b,c) = 1 < \pi_2(b,d) = 2$ and thus player 2 would be better off by unilaterally deviating from *c* to *d*.

if nobody has an individual incentive to deviate from it, assuming that the other players will follow it. A Nash equilibrium is precisely such an agreement.

The Nash equilibrium (b,d) of the game of Table 16.4 has the following feature: there is another strategy profile, namely (a,c), that gives rise to an outcome that both players strictly prefer to the one associated with (b,d). When this is the case, we say that the Nash equilibrium is *Pareto dominated* or *Pareto inefficient*. This is a generic phenomenon: in "almost all" games Nash equilibria are Pareto dominated (see Dubey, 1986). It is worth stressing that although in the game of Table 16.4 the strategy profile (a,c) yields a better outcome for both players than the Nash equilibrium, it is not a viable agreement: if player 1 expects player 2 to stick to the agreement by playing *c*, then he will gain by deviating from the agreement and playing *b*; realizing this, player 2 would want to play *d*, rather than the agreed-upon *c*. That is to say, (a,c) is not a Nash equilibrium.

There are games that have multiple Nash equilibria and games that have none. For example, if in the game of Table 16.4 one replaces the payoffs associated with (b,d) with (0,2) then the resulting game has no Nash equilibria. Nash (1950, 1951) proved that every game with finite strategy stets has an equilibrium if one allows for mixed strategies. A *mixed strategy* for player *i* is a probability distribution over his set S_i of "pure" strategies. The introduction of mixed strategies requires a theory of how players rank probabilistic outcomes or *lotteries*. For example, in the game of Table 16.4, suppose that player 2 uses the mixed strategy $\begin{pmatrix} c & d \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$.[16]

Then if player 1 chooses the pure strategy *a* he faces the lottery $\begin{pmatrix} o_1 & o_2 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$, where o_1 is the outcome associated with (a,c) and o_2 is the outcome associated with (a,d), while choosing the pure strategy *b* means facing the lottery $\begin{pmatrix} o_3 & o_4 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$, where o_3 is the outcome associated with (b,c) and o_4 is the outcome associated with (b,d). The *Theory of Expected Utility*, developed by von Neumann and Morgenstern (1944), provides a list of consistency-of-preferences-over-lotteries axioms which yield the following representation theorem: there exists a numerical function U defined over the set of basic outcomes $O = \{o_1, ..., o_m\}$ such that for any two lotteries $L = \begin{pmatrix} o_1 & o_2 & ... & o_m \\ p_1 & p_2 & ... & p_m \end{pmatrix}$ and $L' = \begin{pmatrix} o_1 & o_2 & ... & o_m \\ q_1 & q_2 & ... & q_m \end{pmatrix}$ the individual considers *L* at least as good as L' if and only if $EU(L) = p_1U(o_1) + ... + p_mU(o_m) \geq EU(L') = q_1U(o_1) + ... + q_mU(o_m)$. $EU(L)$ is called the *expected utility* of lottery *L*. Such a utility function is called a *von Neumann-Morgenstern utility function* or a cardinal utility function. In the mixed-strategy extension of a game the payoff of player *i* associated with a mixed-strategy profile is the expected utility of the corresponding lottery over basic outcomes.

[16] For example, he rolls a die and if the outcome is 1 or 2 then he chooses *c*, otherwise he chooses *d*.

It is worth noting that the transition from games with ordinal payoffs to games with cardinal payoffs (when mixed strategies are considered) is not an innocuous one. Given a game, it is implicitly assumed that the game itself (that is, the sets of players, strategies, outcomes and the players' rankings of the outcomes) is common knowledge among the players. Assuming common knowledge of ordinal rankings of the basic outcomes is far less demanding than assuming common knowledge of von Neumann-Morgenstern payoffs. A player might be fully aware of his own attitude to risk (that is, his own preferences over lotteries), but will typically lack information about the attitude to risk of his opponents.

4. Extensive-form games with perfect information

While strategic-form games represent situations where the players act simultaneously (or, equivalently, in ignorance of each other's choices), extensive-form games represent situations where choices are made sequentially. An *extensive-form game with perfect information* consists of a finite rooted tree, an assignment of a player to each non-terminal node and a ranking of the set of terminal nodes for each player. The terminal nodes correspond to the possible outcomes and, as usual, we represent the rankings of the outcomes by using an ordinal payoff function for each player. Nodes that are not terminal are called decision nodes. The arrows that emanate from a decision node represent the possible choices for the player assigned to that node. Figure 16.5 represents a perfect-information game with two players (ignore, for the moment, the fact that some arrows have a double edge).

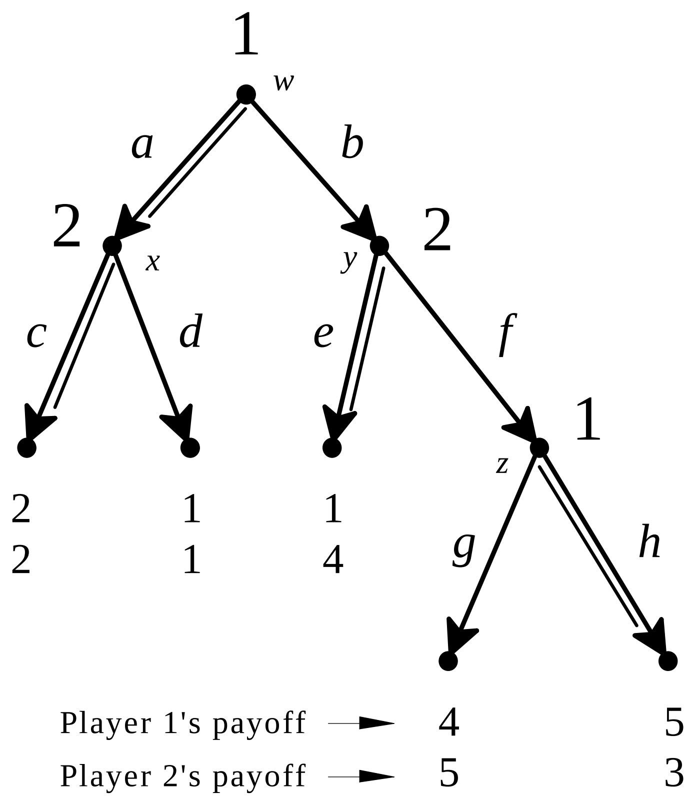
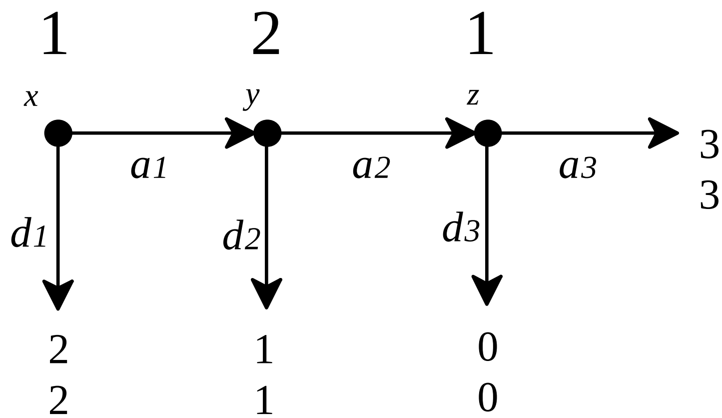


FIGURE 16.5

The solution concept most commonly used for this type of game is that of *backward induction*, which is the following algorithm. Start from a decision node whose immediate successors are all terminal nodes and select a choice at that node that maximizes the payoff of the player assigned to that node. Replace that decision node with the payoffs associated with the terminal node that follows the selected choice and repeat the procedure in the resulting smaller tree. When applied to the game illustrated in Figure 16.5, the backward induction procedure selects the choices highlighted by the double edges. The backward induction algorithm yields an actual play (in the game of Figure 16.5 the choice of *a* by player 1 followed by the choice of *c* by player 2) as well as hypothetical choices at nodes that are not reached by the actual play (for example, in the game of Figure 16.5, player 2's hypothetical choice of *e* at the unreached node *y*). A common interpretation of these hypothetical choices is in terms of *counterfactuals*. For example, in the game of Figure 16.5, player 2's choice of *e* is interpreted as the counterfactual statement "if player 2's node *y* were to be reached, player 2 would choose *e*". Furthermore, the backward induction solution is often presented as capturing the notion of common knowledge of rationality. For example, in the game of Figure 16.5, the reasoning would be as follows: "if node *z* is reached and player 1 is rational, then he will choose *h*; thus if node *y* is reached and player 2 knows that player 1 is rational, then player 2 knows that player 1 would follow with *h* if player 2 herself were to choose *f*; hence if player 2 is rational she will choose *e* at node *y*; etc." There is an ongoing debate in the game theory literature as to whether this reasoning is sound [see, for example, Aumann (1995, 1996), Binmore (1987, 1996), Bonanno (1991), Brandenburger (2007), Halpern (2001), Reny (1992), Samet (1996), Stalnaker (1998)]. The reason for doubting the validity of this interpretation of the backward induction solution can be illustrated in the game of Figure 16.5. If the backward induction solution is implied by common knowledge of rationality, then common knowledge of rationality implies that node *y* will *not* be reached. Hence the hypothesis "player 2 is rational and knows that player 1 is rational", which is used to conclude that player 2 would choose *e* at node *y*, will be false at node *y*. In particular, player 2 might conclude that player 1 is *not* rational and anticipate a choice of *g* by player 1 at node *z*, thus making *f* a better choice than *e* at *y*. In order to address these issues, once again one needs to have a precise definition of rationality as well as, possibly, a theory of counterfactuals.

A good starting point is the definition of rationality used in Section 2. That definition was formulated for strategic-form games and was based on the notion of model of a game, which associates with every state (in an interactive knowledge structure) a strategy profile. It is possible to associate with every perfect-information game a strategic-form game by using the following definition: a *strategy* for player *i* in a perfect-information game is a list of choices, one for each node assigned to player *i*. For example, a possible strategy for player 1 in the game of Figure 16.5 is (a,g) and a possible strategy for player 2 is (d,e). A strategy profile determines a unique path from the root of the tree to a terminal node and thus one can associate with that strategy profile the payoffs of the corresponding terminal node. Figure 16.6 illustrates a perfect-information game (whose backward induction play is $a_1a_2a_3$), its corresponding strategic form and a model of the strategic form.

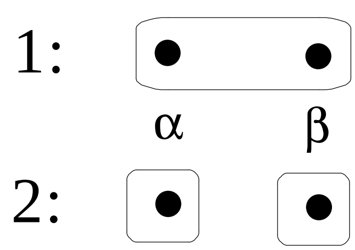
A perfect-information game:



The corresponding strategic form:

		Player 2	
		d2	*a2*
Player 1	*d1,a3*	2 , 2	2 , 2
	d1,d3	2 , 2	2 , 2
	a1,a3	1 , 1	3 , 3
	a1,d3	1 , 1	0 , 0

A model of the strategic form:



1's strategy	***d1,a3***	***d1,a3***
2's strategy	***d2***	***a2***

FIGURE 16.6

Using the definition of rationality introduced in Section 2, we have that both players are rational at every state (the only strictly dominated strategy is player 1's a_1d_3 and, after deleting it, there are no other strategies that are strictly dominated). It follows that at state α there is common knowledge of rationality, despite the fact that the associated strategy profile is (d_1a_3, d_2), which is different from the backward induction solution. The issue is whether at state α we can validly label player 2 as rational. At that state player 2 *knows* that her node *y* is not reached and therefore her payoff is not affected by her choice. Hence, she *is* rational in a weak sense. However, a stronger notion of rationality would require us to evaluate her choice of d_2 as a plan of *what she would actually do if her decision node were to be reached*. This is a counterfactual statement at state α, since her node *y* is not reached there. Aumann (1996) proposes a notion of rationality, which he calls *substantive rationality*, and shows that common knowledge of substantive rationality implies the backward induction *play* (but not necessarily the backward induction strategy profile). While accepting the correctness of this result within the framework adopted by Aumann, Stalnaker (1998, p.48) disputes its validity, arguing as follows:

> «Player 2 has the following initial belief: player 1 would choose a_3 on her second move *if* she had a second move. This is a causal 'if' – an 'if' used to express 2's opinion about 1's *disposition to act* in a situation that they both know will not arise. Player 2 knows that since player 1 is rational, if she somehow found herself at her second node, she would choose a_3 . But to ask what player 2 would believe about player 1 *if* he learned that he was wrong about 1's first choice is to ask a completely different question – this 'if' is epistemic; it concerns player 2's belief revision policies, and not player 1's disposition to be rational. No assumption about player 1's substantive rationality, or about player 2's knowledge of her substantive rationality, can imply that player 2 should be disposed to maintain his belief that player 1 will act rationally on her second move even were he to learn that she acted irrationally on her first.»

In order to be able to carry out a rigorous analysis of the implications of common knowledge of rationality in perfect-information games, we need to move away from the type of models that we have considered so far. The reason for this is that the association of a strategy profile with every state gives rise (implicitly) to two types of counterfactuals: (1) an objective statement about what the relevant player would do at a node that is not reached and (2) (with the help of the partitions) a subjective statement about what a player believes would happen if he were to take a different action from the one he is actually taking. The two can be disentangled by (1) associating, with every state, not a strategy profile but a play and (2) adding a set of relations that can be used to obtain a formal interpretation of counterfactual statements. We start from the latter. For every state $\omega \in \Omega$ let $\mathcal{P}_\omega$ be a "proximity-to-ω" binary relation on Ω and, for every $\omega' \in \Omega$ let $\mathcal{P}_\omega(\omega') = \{x \in \Omega : \omega' \mathcal{P}_\omega x\}$. The interpretation of $\beta \in \mathcal{P}_\omega(\alpha)$ or $\alpha \mathcal{P}_\omega \beta$ is that state α is closer to state ω than β is, so that $\mathcal{P}_\omega(\alpha)$ is the set of states that are not as close to ω as α is. We assume that the closest state to ω is ω itself and, for simplicity, that $\mathcal{P}_\omega$ is a strict ordering of Ω.[17] The truth of the counterfactual "if ϕ were the case then ψ would be the case" at state ω is then determined as follows: look for the closest state to ω at which ϕ is true, call it ω'; if ψ is true at ω' then the counterfactual is true at ω, otherwise it is false. Intuitively, closeness is interpreted as similarity: the closest state to ω where ϕ is true is interpreted as the most similar state to ω among the ones where ϕ is true. This theory of counterfactuals is due to Stalnaker (1968) and was later generalized by Lewis (1973).

---

[17] That is, $\mathcal{P}_\omega$ satisfies the following properties: (1) $\omega' \in \mathcal{P}_\omega(\omega)$, for all $\omega' \in \Omega \setminus \{\omega\}$ (centeredness), (2) for every $\alpha, \beta \in \Omega$ with $\alpha \neq \beta$, either $\alpha \in \mathcal{P}_\omega(\beta)$ or $\beta \in \mathcal{P}_\omega(\alpha)$ (connectedness), (3) for every $\alpha, \beta \in \Omega$,

We can use proximity orderings and counterfactuals to model strategies as well as hypothetical beliefs, by modifying our earlier definition of a model of a perfect information game as follows: (1) we replace the *n* function $\sigma_{\mathrm{i}} : \Omega \to S_i$ with a single function $d : \Omega \to P$, where *P* is the set of plays of the game, and (2) we add a set of proximity relations $\{\mathcal{P}_\omega\}_{\omega \in \Omega}$, one for each state. Thus, with every state, we associate a play rather than a strategy profile and, for each state, we give a proximity ranking of the states, with the state itself being the closest of all. Figure 16.7 illustrates a model of the perfect information game of Figure 16.6.

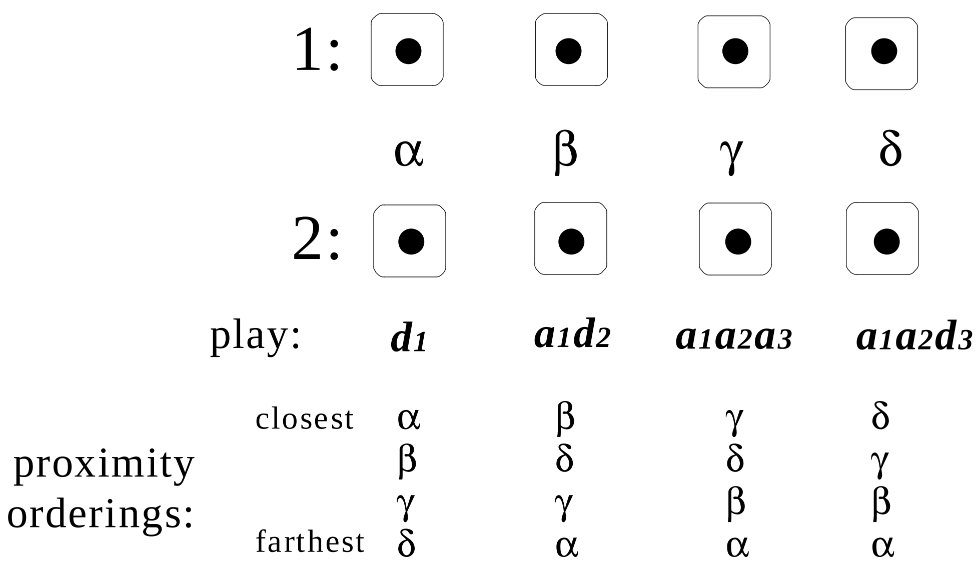


FIGURE 16.7

---

if $\beta \in \mathcal{P}_\omega(\alpha)$ then $\alpha \notin \mathcal{P}_\omega(\beta)$ (asymmetry) and (4) for every $\alpha, \beta \in \Omega$, if $\beta \in \mathcal{P}_\omega(\alpha)$ then $\mathcal{P}_\omega(\beta) \subseteq \mathcal{P}_\omega(\alpha)$ (transitivity).

In this model, at state α, node z of player 1 is not reached because his initial choice is d_1. Is it true, however, at state α, that if node z were to be reached player 1 would play d_3? In order to answer this question we use the proximity ranking at α to find the closes state to α at which node z is reached. That state is γ. Then we check whether at γ player 1 plays d_3. Since at state γ player 1 plays a_3 rather than d_3, the answer is negative: it is not true at state α that if node z were to be reached player 1 would play d_3. The strategy profile implicitly associated with state α is thus $(d_1 a_3, d_2)$. In order to determine what a player would know or believe if a node which is not reached were to be reached, we proceed the same way: we look for the closest state where the node is reached and determine (using the cell of the information partition of this player that contains that node) the player's state of knowledge at that node. For example, at state α player 2 knows that player 1 is playing d_1 and therefore knows that her node y is not reached. What would player 2 know if her node were to be reached? The closest state to α at which node y is reached is β and at β player 2 terminates the game by playing d_2 and collecting a payoff of 1. Is this a rational choice for player 2? The answer depends on what player 2 believes would happen if she played a_2 (if node y were to be reached). To determine this we look for the closest state to β at which node z is reached. It is state δ. At state δ player 1 plays d_3, giving a payoff of 0 to player 2. Thus player 2's choice of d_2 is indeed rational at state β. This conclusion makes player 1's choice of d_1 at state α rational. Furthermore, at state α player 2 is rational not only in a weak sense since she makes no choices at state α (because her node y is not reached) but also in the stronger sense that the choice that she would make if her node were to be reached (choice d_2 at state β) is rational, given her belief in that situation (that is, at state β). Since both players are rational at state α and the common knowledge partition coincides with the individual partitions, at state α there is common knowledge of rationality, despite the fact that the play at α is not the backward induction play. Thus, using this analysis based on a theory of counterfactuals one can conclude that common knowledge of rationality does not imply the backward induction play (let alone the backward induction solution, that is, the backward induction strategy profile).

In the model of Figure 16.7 at state α player 2 believes that if node *z* were reached then player 1 would choose a_3 (since the closest state to α where node *z* is reached is state γ and there player 1 chooses a_3); however, as we saw above, at state α it is also the case that player 2 would choose d_2 if her node *y* were to be reached (state β), based on the belief (at state β) that if she chose a_2 then player 1 would follow with d_3 (state δ). Hence what player 2 believes about player 1's behavior in the hypothetical world where node *z* is reached changes going from state α (where the game ends without node *y* being reached) to the closest state β where *y* is reached. Stalnaker (1998, p.48, quoted above) argues that there is nothing wrong with such a change. Halpern (2001) shows that if one imposes the constraint that such changes in beliefs are not allowed, then Aumann's result that common knowledge of substantive rationality implies the backward-induction play holds.

5. Extensive-form games with imperfect information

An extensive game is said to have *imperfect* information if at least one player is not fully informed about the choices made by other players in the past. To represent a player's uncertainty concerning past moves, we use *information sets* (which play the same role as the cells of the information partitions considered earlier). An information set of player *i* contains several nodes in the tree where player *i* has available the same choices and the interpretation is that the player cannot tell at which of these nodes her choice is being made. Figure 16.8 illustrates an extensive game with imperfect information. Player 2 has two information sets, one consisting of the two nodes *v* and *w* and the other consisting of the single node *x*.[18] The interpretation of information set $\{v, w\}$ of player 2 is that, when choosing between actions *D* and *E*, player 2 does not know whether player 1 chose *A* or *B*. Player 3 is in a similar situation at information set $\{y, z\}$ concerning the earlier choice of player 2 between *F* and *G*.

[18] When an information set is a singleton it is customary not to enclose it into a rectangle.

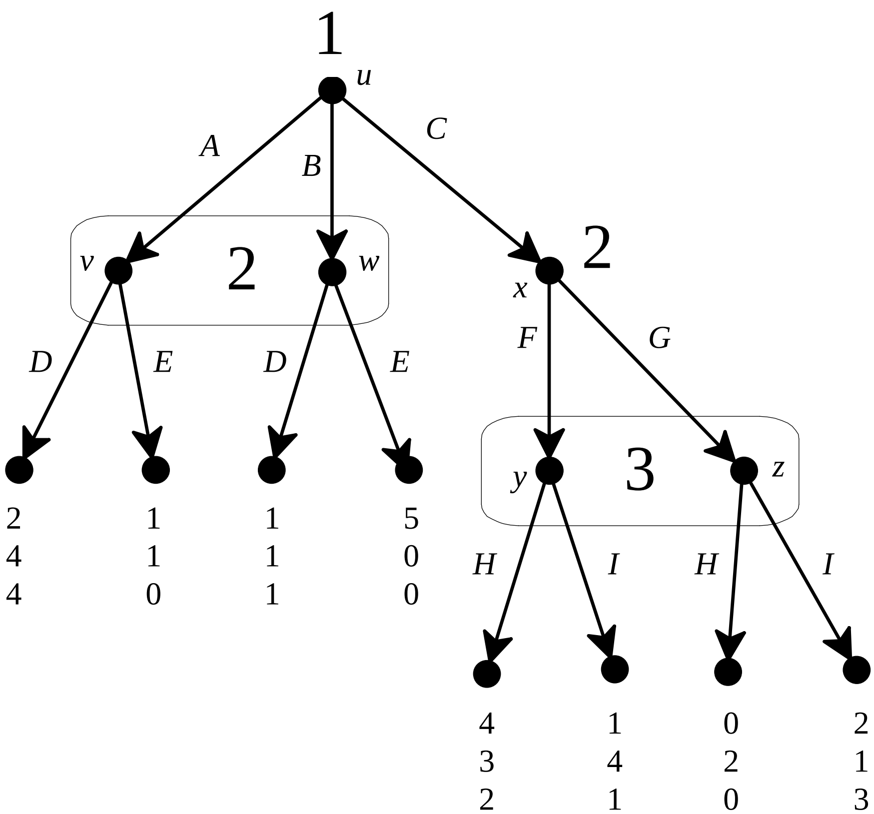


FIGURE 16.8

As in the case of perfect-information games, we can associate with every extensive-form game with imperfect information a strategic form using the following definition: a *strategy* for player *i* is a list of choices, one for every information set of player *i*.[19] For example, the set of strategies of player 2 in the game of Figure 16.8 is $S_2 = \{DF, DG, EF, EG\}$. The solution concept most used in extensive games with imperfect information is that of *subgame-perfect equilibrium*, which is a generalization of the notion of backward induction used in perfect-information games. A *subgame* is a portion of the entire game that (1) starts at a singleton information set {*x*} and includes *all* the successors of node *x* and (2) if *y* is a successor of *x* that belongs to information set *h* (of some player) then every node in *h* is a successor of *x*. For example, in the game of Figure 16.8, the portion of the tree that starts at node *x* of player 2 is a subgame; the only other subgame is the entire game. A *subgame-perfect equilibrium* is a Nash equilibrium of the entire game that satisfies the following property: for every subgame, the restriction of the strategy profile to that subgame is a Nash equilibrium of the subgame. For example, in the game of Figure 16.9, (C, DF, H) is the unique subgame-perfect equilibrium: it is a Nash equilibrium of the entire game and, furthermore, the restriction of (C, DF, H) to the subgame that starts at node *x*, namely (F, H), is a Nash equilibrium of that subgame. It was proved by Selten (1975) that if one allows for von Neumann-Morgenstern payoffs and mixed strategies (see Section 3) then every finite extensive-form game with perfect recall[20] has at least one subgame-perfect equilibrium in mixed strategies.

[19] This definition coincides with the earlier one in games with perfect information since in the case of perfect information the information sets of a player are all singletons.

[20] An extensive game has perfect recall if it satisfies the following property, for every player *i*: if *h* and *g* are two information sets of player *i* and there is a node in *g* which is a successor of a node in *h*, then every node in *g* comes after the same choice at *h*. Perfect recall implies that a player remembers his past choices as wells as what he knew in the past (see Bonanno, 2004).

6. Games with incomplete information

An implicit assumption in game theory is that the game is common knowledge among the players. The expression "incomplete information" refers to those situations where some of the elements of the game (e.g. the preferences of the players) are not common knowledge. In such situations the knowledge and beliefs of the players about the game need to be made an integral part of the model. Pioneering work in this direction was done by Harsanyi (1967,1968). Harsanyi suggested a method for converting a situation of incomplete information into an extensive game with imperfect information (this is the so-called *Harsanyi transformation*). The theory of games of incomplete information has been developed for the case of von Neumann-Morgenstern payoffs (see Section 3) and the solution concept proposed by Harsanyi is *Bayes-Nash equilibrium* which is simply a Nash equilibrium of the imperfect information game so constructed. Although the traditional definition of games of incomplete information is in terms of types of players and of probability distributions over types[21], we shall illustrate the Harsanyi transformation using the epistemic structures introduced in Section 2. States can be used to describe possible games and thus represent the uncertainty in a player's mind as to which game she is truly playing. Figure 16.9 illustrates a two-player situation of incomplete information using an interactive knowledge structure with the addition of a probability distribution for every cell of the information partition of each player.

[21] For an overview of the traditional approach see Battigalli and Bonanno (1999).

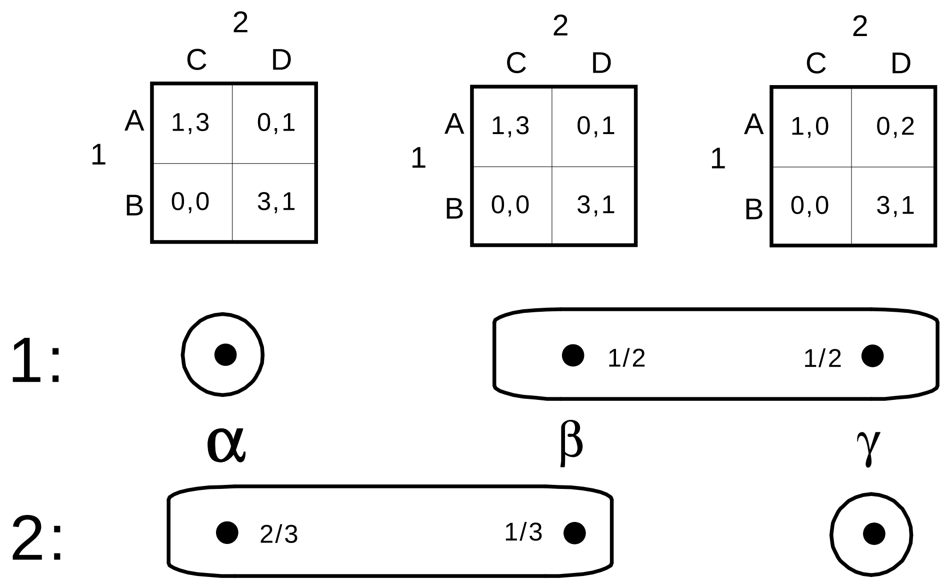


FIGURE 16.9

Associated with every state is a strategic-form game. Let G be the game associated with states α and β (it is the same game) and G' the game associated with state γ. Fix a state, say, state α. Then state α describes the following situation:

(1) both player 1 and player 2 know that the are playing game G,

(2) player 1 knows that player 2 knows that they are playing game G,

(3) player 2 is uncertain as to whether player 1 knows that they are playing G (which is the case if the actual state is α) or whether player 1 is uncertain (if the actual state is β) between the possibility that they are playing game G and the possibility that they are playing game G' and considers the two possibilities equally likely (that is, attaches probability $\frac{1}{2}$ to each); furthermore, player 2 attaches probability $\frac{2}{3}$ to the first case (where player 1 knows that they are playing game G) and probability $\frac{1}{3}$ to the second case (where player 1 is uncertain between game G and game G'),

(4) player 1 knows the state of uncertainty of player 2 concerning player 1 (as described sub (3) above),

(5) it is common knowledge that each player knows his own payoffs and that player 2 also knows player 1's payoffs.

Harsanyi's suggestion was to represent a situation of incomplete information such as the one illustrated in Figure 16.9 as a game with imperfect information where the initial move is assigned to a fictitious player, called Nature, whose role is to choose the state with predetermined probabilities. No payoffs are assigned to Nature and it makes no further choices. Information sets are then used to capture the uncertainty of the players concerning both the actual state and the choices made by the other players. In order for such a representation to be possible, it is necessary that the probabilistic beliefs of the players at the cells of their information partitions be consistent in the following sense: there is a probability distribution μ over the set of states, called a *common prior*, which yields those probabilistic beliefs upon conditioning on the information represented by a cell of an information partition. Conditional probabilities ought to be obtained from the common prior by using Bayes' rule. For example, in the situation illustrated in Figure 16.9 we want a function $\mu:\{\alpha,\beta,\gamma\}\to[0,1]$ such that

$$\mu(\beta\,|\,\{\beta,\gamma\})=\frac{\mu(\beta)}{\mu(\beta)+\mu(\gamma)}=\frac{1}{2}\ ,\ \mu(\alpha\,|\,\{\alpha,\beta\})=\frac{\mu(\alpha)}{\mu(\alpha)+\mu(\beta)}=\frac{2}{3}\ \text{and}\ \ \mu(\alpha)+\mu(\beta)+\mu(\gamma)=1.$$

In this case a common prior exists and is given by $\mu(\alpha)=\frac{2}{4}$ and $\mu(\beta)=\mu(\gamma)=\frac{1}{4}$. Using this common prior to assign probabilities to Nature's choices we obtain the imperfect-information game shown in Figure 16.10.

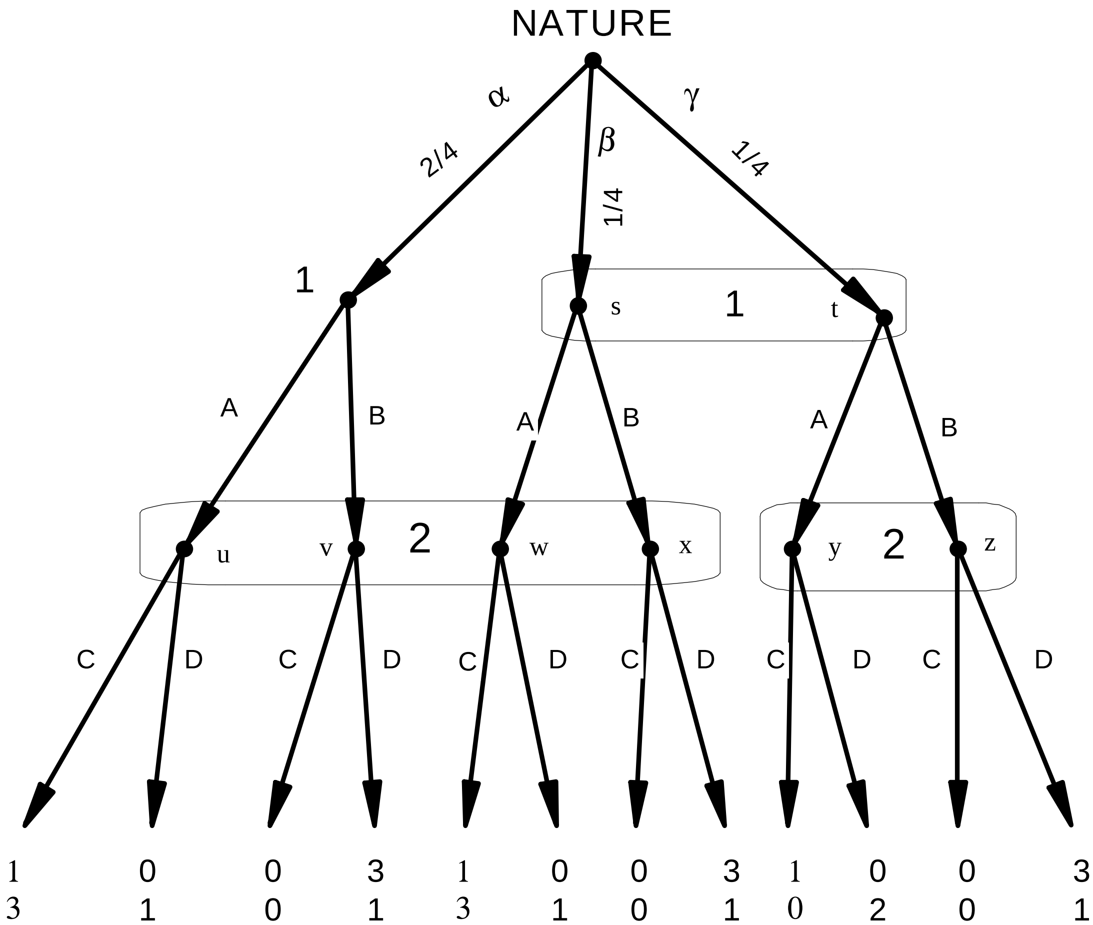


FIGURE 16.10

A Nash equilibrium of this imperfect-information game is called a *Bayes-Nash equilibrium* of the corresponding incomplete-information situation. The following pure-strategy profile is Nash equilibrium of the game of Figure 16.10: player 1's strategy is *AB* (that is, he plays *A* if informed that the state is α and plays *B* if informed that the state is either β or γ) and player 2's strategy is *CD* (that is, she plays *C* at her information set on the left and *D* at her information set on the right). To verify that this is a Nash equilibrium, we need to check that no player can increase his expected payoff by unilaterally changing his strategy. We begin with player 1: (1) at the singleton information set on the left, *A* gives player 1 a payoff of 1 (given player 2's choice of *C*) while *B* would give him a payoff of 0 (hence *A* is the optimal choice); (2) at the information set on the right, by Bayes' rule player 1 must assign probability $\frac{1}{2}$ to node *s* and probability $\frac{1}{2}$ to node *t*; thus (given player 2's strategy *CD*) choosing *A* would give him an expected payoff of $\frac{1}{2}1+\frac{1}{2}0=\frac{1}{2}$ while *B* gives him an expected payoff of $\frac{1}{2}0+\frac{1}{2}3=\frac{3}{2}$ (hence *B* is the optimal choice). Similarly for player 2: (1) at the information set on the left, by Bayes' rule (given that player 1's strategy is *AB*) player 2 must assign probability $\frac{2}{3}$ to node *u* and $\frac{1}{3}$ to node *x*; thus choosing *C* gives her an expected payoff of $\frac{2}{3}3+\frac{1}{3}0=2$ while *D* would give her an expected payoff of $\frac{2}{3}1+\frac{1}{3}1=1$ (hence *C* is the optimal choice); (2) at the information set on the right, by Bayes' rule (given that player 1's strategy is *AB*) player 2 must assign probability 1 to node *z*; thus choosing *C* would give her a payoff of 0 while *D* gives her a payoff of 1 (hence *D* is the optimal choice).

7. Conclusion.

There are several branches and applications of game theory that we were not able to discuss because of space limitations.[22] Besides co-operative game theory[23], which was mentioned in the introduction, we also left out evolutionary game theory[24], the theory of repeated games[25] and bargaining theory[26].

In the social sciences, game theory has become pervasive not only in economics but also in political science[27]. Of particular relevance to social science is also the game-theoretic approach to ethical matters and to fundamental questions of moral and political philosophy.[28] Finally it is also worth mentioning two new developments: experimental game theory and *neuroeconomics*. The former tries to test the predictions of game theory in controlled laboratory

---

[22] An excellent source of information and resources for educators and students of game theory is the web site www.gametheory.net.

[23] For a recent overview of co-operative game theory see Peleg and Sudhölter (2007).

[24] Evolutionary game theory was pioneered by the biologist John Maynard Smith (1982) and has been extensively applied not only in biology but also in the social sciences (see Samuelson, 1998, and Weibull, 1995).

[25] For a recent account of repeated games see Mailath and Samuelson (2006).

[26] See, for example, Osborne and Rubinstein (1990).

[27] Ordeshook (1986) is a classic reference. For a more recent account see McCarty and Meirowitz (2007).

[28] See Bicchieri (2006) and Binmore (1994, 1998).

settings, while the latter aims to systematically classify and map the brain activity that correlates with (individual and interactive) decision-making.[29]

References

Aumann, Robert (1976). Agreeing to disagree. *The Annals of Statistics*, 4: 1236-1239.

Aumann, Robert (1989). Game theory. In J. Eatwell, M. Milgate & P. Newman, editors, *The new Palgrave: a dictionary of economics*. W. W. Norton.

Aumann, Robert (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8: 6–19.

Aumann, Robert (1996). Reply to Binmore. *Games and Economic Behavior*, 17: 138–146.

Battigalli, Pierpaolo and Giacomo Bonanno (1999). Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53: 149-225.

Bicchieri, Cristina (2006). *The Grammar of Society: the Nature and Dynamics of Social Norms*. Cambridge University Press.

Binmore, Ken (1987). Modelling rational players I. *Economics and Philosophy*, 3: 179–214.

Binmore, Ken (1994). *Game Theory and the Social Contract*, Vol. 1: *Playing Fair*. MIT Press.

Binmore, Ken (1996). A note on backward induction. *Games and Economic Behavior*, 17: 135–137.

Binmore, Ken (1998). *Game Theory and the Social Contract*, Vol. 2: *Just Playing*. MIT Press.

Bonanno, Giacomo (1991). The logic of rational play in games of perfect information. *Economics and Philosophy*, 7: 37–65.

Bonanno, Giacomo (2004). Memory and perfect recall in extensive games. *Games and Economic Behavior*, 47: 237-256.

Bonanno, Giacomo (2008). A syntactic approach to rationality in games with ordinal payoffs. In: G. Bonanno, W. van der Hoek and M. Wooldridge, editors, *Logic and the Foundations*

---

[29] A special issue of *Economics and Philosophy* (Volume 24, No. 3, November 2008) is devoted to a discussion of the philosophical and methodological issues and challenges that have been raised by neuroeconomics.

of Game and Decision Theory. Texts in Logic and Games, Amsterdam University Press, 59-86.

Brandenburger, Adam (2007). The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory,* 35: 465–492.

Dubey, Pradeep (1986). Inefficiency of Nash equilibria. *Mathematics of Operations Research*, 11: 1-8.

Halpern Joseph (2001). Substantive rationality and backward induction. *Games and Economic Behavior,* 37: 425–435.

Harsanyi, John (1967,1968). Games with incomplete information played by "Bayesian players", Parts I-III. *Management Science*: 8, 159-182, 320-334, 486-502.

Lewis, David (1969). *Convention.* Harvard University Press.

Lewis, David (1973). *Counterfactuals.* Harvard University Press.

Mailath, George J. and Larry Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.

McCarty, Nolan and Adam Meirowitz (2007). *Political Game Theory: An Introduction*. Cambridge University Press.

Nash, John (1950). Equilibrium points in *n*-person games, *Proceedings of the National Academy of Sciences*, 36: 48-49.

Nash, John (1951). Non-cooperative games, *Annals of Mathematics,* 54: 286-295.

Ordeshook, Peter C. (1986). *Game Theory and Political Theory: An Introduction*. Cambridge University Press.

Osborne, Martin J. and Ariel Rubinstein (1990). *Bargaining and markets*. Academic Press.

Peleg, Bezalel and Peter Sudhölter (2007). *Introduction to the Theory of Cooperative Games*. Springer.

Reny, Phil (1992). Rationality in extensive form games. *Journal of Economic Perspectives*, 6: 103–11.

Samet Dov (1996). Hypothetical knowledge and games with perfect information. *Games and Economic Behavior*, 17: 230–251.

Samuelson, Larry (1998). *Evolutionary Games and Equilibrium Selection*. MIT Press.

Selten, Reinhard (1975), Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4: 25-55.

Smith, John Maynard (1982). *Evolution and the Theory of Games*. Cambridge University Press.

Stalnaker Robert (1968). A theory of conditionals. In N. Rescher, editor, *Studies in logical theory*. Blackwell, 98-112. Reprinted in E. Sosa, editor, *Causation and conditionals*, Oxford University Press, 1975, 165-179.

Stalnaker Robert (1998). Belief revision in games: forward and backward induction. *Mathematical Social Sciences*, 36: 31–56.

von Neumann, John and Oscar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Weibull, Jörgen W. (1995). *Evolutionary game theory*. MIT Press.