This is a section of doi:10.7551/mitpress/11252.001.0001

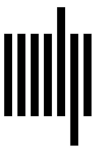
# The Handbook of Rationality

## Edited by: Markus Knauff, Wolfgang Spohn

## Citation:

*The Handbook of Rationality*
**Edited by:** Markus Knauff, Wolfgang Spohn
**DOI:** 10.7551/mitpress/11252.001.0001
**ISBN (electronic):** 9780262366175
**Publisher:** The MIT Press
**Published:** 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

# 9.1 Classical Game Theory

**Max Albert and Hartmut Kliemt**

**Summary**

Classical game theory extends rational choice theory to interactions among several actors. This chapter introduces the formal language of classical noncooperative game theory in a self-contained way. Emphasis on the extensive form of games and its presuppositions clarifies game theory's scope and limits as a modern *lingua franca* of the behavioral sciences. The discussion of rational behavior in games centers on the equilibrium concept and its most important refinement, subgame perfection. A final section relates this discussion to the interpretation of classical game theory in terms of "reasoning about knowledge."

## 1. Outline of the Chapter

"Classical game theory" extends rational choice theory to interactions among several actors. We use the term to distinguish our topic from related but distinct fields like epistemic, evolutionary, and behavioral game theory. We focus on noncooperative game theory, which is more fundamental than cooperative game theory (Harsanyi & Selten, 1988, chapter 1). Our presentation is mostly ahistorical; for historical accounts, see Weintraub (1992) and Leonard (2010). References were primarily selected to provide convenient entry points to the literature. For in-depth coverage of mathematical aspects, we refer readers to Fudenberg and Tirole (1991) and Maschler, Solan, and Zamir (2013).

Section 2 approaches games as mathematical structures. Section 3 focuses on applying game theory with the aim of explaining or predicting interactions among several actors. Section 4 explains, and comments on, the tension within applied game theory between the interpretation we present and an alternative interpretation in terms of "reasoning about knowledge."

## 2. Game-Theoretic Language

### 2.1 The Rules of a Game

Primitive terms of mathematical game theory are "player," "move," "payoff," and "probability." The use of these terms is restricted by certain formal requirements but otherwise unspecified. Depending on the application, a player can be an individual person, a group of persons (say, a firm), or even a bacterium. There are applications where a single person is modeled as a set of different players or where a player stands for logical operations. Probabilities satisfy the Kolmogorov axioms (see chapter 4.1 by Hájek & Staffel, this handbook), but their interpretation is left open. Payoffs are real numbers, but it remains unspecified what these numbers represent. For ease of exposition, we nevertheless speak of players "making," "selecting," or "choosing" moves and "receiving" or "earning" payoffs; moreover, we cast a female as our generic player and as players 1 and 3, as well as a male as player 2.

The definition of a game as a mathematical structure, often called "the rules of the game," specifies all moves that are possible in the game, implying that rule-breaking moves cannot exist. A "play of the game" is a sequence of moves, each made by a specific player. Each possible play of a game comes with a "payoff profile," that is, an assignment of one payoff to each player.

The rules of a game are summarized in the game's "extensive form," also called "game tree." Figure 9.1.1a, for instance, describes a game that begins with player 1 choosing between moves *A* and *B*. If she chooses *A*, player 2 must choose between *W* and *X*; if player 1 chooses *B*, player 2 must choose between moves *Y* and *Z*. After moves *X*, *Y*, or *Z* of player 2, the game ends. After move *W*, player 1 must choose between moves *C* and *D*, after which the game ends. A "path" is a succession of nodes and branches. Each complete path through the tree, for instance, *A*–*W*–*D* or *B*–*Y* (plus the relevant nodes), represents a possible play of the game. The associated payoff profile is displayed at the end of the path; thus, *A*–*W*–*D* is associated with the payoff profile (4, 6), that is, player 1 earns 4 and player 2 earns 6.

All game trees we consider begin with a unique initial node (indicated by a circle) and end with a finite number of final nodes represented by payoff profiles. Bullets represent intermediate nodes. Branches connect each node—except the initial node—with its immediate

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

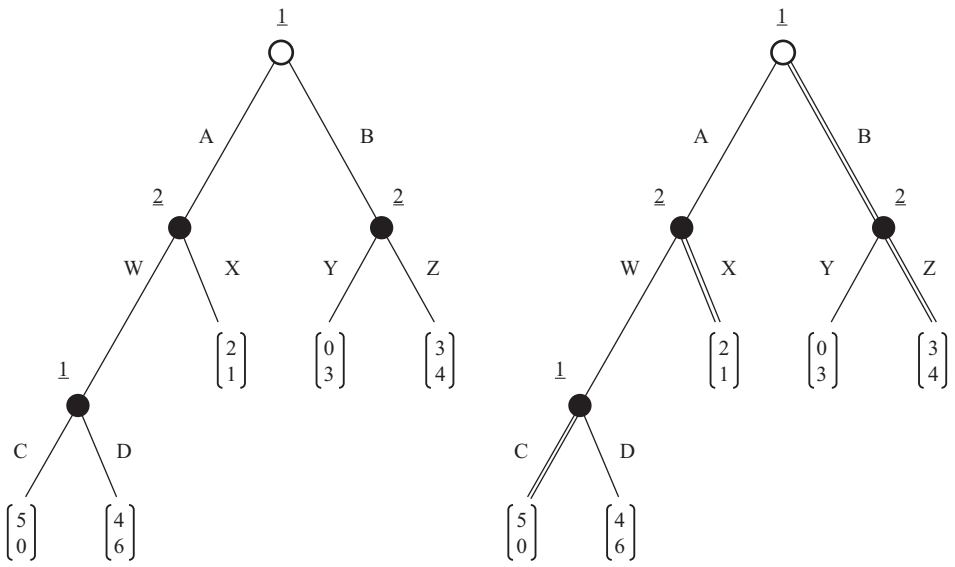



**Figure 9.1.1**
(a) A strictly sequential game. The pure strategies of player 1 are *AC*, *BC*, *AD*, and *BD*. Those of player 2 are *WY*, *XY*, *WZ*, and *XZ*. (b) The pure-strategy profile (*BC*, *XZ*), illustrated by the highlighted moves, implies the path *B*–*Z* with associated payoff profile (3, 4).

predecessor. Each node of a game tree is reached from the initial node through a unique path. The games we consider are finite in the sense that there are finitely many complete paths, each of finite length.

Each nonfinal node is assigned to one specific player. Names of players are numbers appearing (underscored) close to their nodes. There may, however, exist passive players without nodes who receive payoffs but cannot move. The branches emerging from a player's node correspond to different moves available to the player at the node. Names of players' possible moves are capital letters appearing at the respective branches.

The path reaching a node corresponds to the history of play up to the node; therefore, this path is also called the node's "history." A complete path ends with the payoff profile associated with the corresponding play of the game. Payoff profiles list payoffs in the order of the players' names, from 1 to *n*.

A situation in a game where a player must move is called "information set." In figure 9.1.1a, all information sets are singleton sets containing exactly one node; in this case, distinguishing between nodes and information sets is inessential. Each singleton information set, or node, of a player comes with its own set of moves. Thus, player 2 chooses among *W* and *X* at his left node and between *Y* and *Z* at his right node; his total number of different moves is 4.

This is different in figure 9.1.2a, where the two nodes of player 2 belong to the same information set (as indicated by the broken-line connection). The two moves possible at this two-element information set, *W* and *X*, are displayed at each of the two nodes; however, the two branches labeled *W* stand for a single move, as do the two branches labeled *X*. In total, player 2 has only two different moves (in contrast to figure 9.1.1a, where he has four). In applications, information sets are used to describe situations where a player is unable to observe (or remember, or notice in some other way) what had happened before she was called upon to move. The game, then, begins with player 1 choosing between *A* and *B*, followed by player 2 choosing between *W* and *X*. In case of *A* and *W*, player 1 chooses for a second time, between *C* and *D*.

Generally, each node of a player belongs to exactly one information set; any number of nodes of a player can belong to the same information set. At each of her information sets, a player has at least two possible moves, of which exactly one must be chosen if the play of the game reaches (a node of) the information set. The possible moves at an information set are displayed at every node of that information set, with their names indicating which moves at the different nodes are the same moves (so that exchanging names at one node but not at the others leads to a different game).

In a "simultaneous" game, each player has at most one information set, and each complete path runs through all information sets (as in figure 9.1.3). A nonsimultaneous game is often called "sequential," although it need not be "strictly sequential" in the sense that each information set is a singleton set (as in figure 9.1.1). Most games are neither simultaneous nor strictly sequential

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

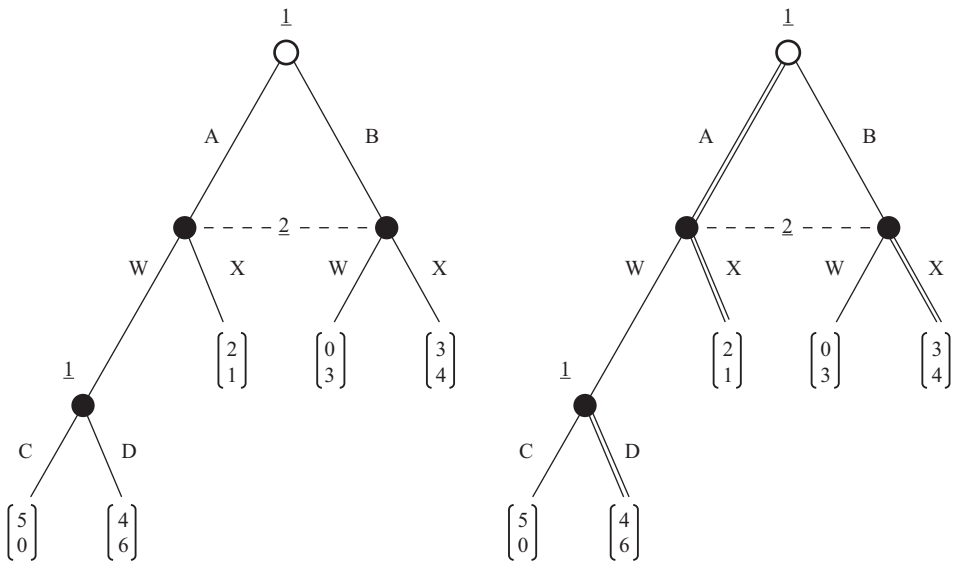



**Figure 9.1.2**
(a) A game that is neither strictly sequential nor simultaneous. Player 2 has only two pure strategies, *W* and *X*, since both his nodes belong to the same information set, meaning that he must choose the same move, *W* or *X*, at both nodes. (b) The pure-strategy profile (*AD*, *X*), illustrated by the highlighted moves, implies the path *A*–*X* with associated payoff profile (2, 1).

(see, e.g., figures 9.1.2, 9.1.4, and 9.1.5). Strictly sequential games are often (rather confusingly) called "games of perfect information."

### 2.2 Strategies and Equilibrium

A pure strategy of a player is a function assigning to each of her information sets exactly one of the possible moves. In figure 9.1.1, each player has two information sets with two possible moves each; therefore, each player has four ($2 \times 2$) pure strategies. We denote pure strategies by a sequence of letters spelling out the moves selected at each information set (in the order from left to right and top to bottom). The four pure strategies of player 1, then, are *AC*, *AD*, *BC*, and *BD*, while player 2's four strategies are *WY*, *WZ*, *XY*, and *XZ*.

A pure-strategy profile is a tuple of pure strategies, one for each active player, again in the order of players' names. In figure 9.1.1, a pure-strategy profile determines a unique complete path through the game tree, that is, a unique play of the game and, therefore, a unique payoff profile. For instance, the profile (*BC*, *XZ*) implies the path *B*–*X* with payoff profile (3, 4) (see figure 9.1.1b). Not each move selected by the strategies occurs in the play of the game generated by (*BC*, *XZ*): player 2 is not called upon to move at his left-hand node; nevertheless, his strategy selects *X*. Moreover, player 1's own choice of *B* prevents that she is called upon to move at her second node, where her strategy selects *C*. Yet, distinguishing between strategies like *BD* and *BC* turns out to be relevant.

In figure 9.1.2, player 2 has only two pure strategies, called, like the moves, "*W*" and "*X*" (accepting a minor ambiguity in the notation). Figure 9.1.2b illustrates the strategy profile (*AD*, *X*). In contrast to figure 9.1.1, player 2 has no pure strategy that discriminates, through its assignment of moves, between the different histories preceding his two nodes. Nonsingleton information sets, then, decrease the number of pure strategies.

The "strategic (normal) form" of a game is a list of all pure-strategy profiles together with the resulting payoff profiles. For two-player games, the strategic form is written in the form of a cross-tabulation. Table 9.1.1 displays the strategic forms of the games of figures 9.1.1 and 9.1.2. Strategies differing only at nodes that are out of reach due to the strategies themselves (like *BC* and *BD* in figure 9.1.1) generate identical entries in the strategic form. Different games can have the same strategic form.

Strategies can be stochastic. Let $\{E_1, \ldots, E_n\}$ be a finite set. A "mixture" of the set's elements, written as $p_1 \circ E_1 \oplus p_2 \circ E_2 \oplus \cdots \oplus p_n \circ E_n$, is a probability distribution assigning to each $E_k$ a probability $p_k \geq 0$ with $\sum_k p_k = 1$. Probability-0 elements can be dropped; the degenerate cases $1 \circ E_k$ are identified with $E_k$. Mixtures of elements from different sets are always stochastically independent.

A "mixed strategy" of a player is a mixture of her pure strategies. A "behavioral strategy" of a player is a function assigning to each of her information sets a mixture of the moves at that information set, called a "mixed move." Mixed and behavioral strategies include the pure

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

**Table 9.1.1**
The strategic forms of the games of figure 9.1.1 (left-hand side) and figure 9.1.2 (right-hand side)

| | | 2 | | | |
|---|---|---|---|---|---|
| | | *WY* | *WZ* | *XY* | *XZ* |
| 1 | *AC* | 5, 0 | 5, 0 | 2, 1 | 2, 1 |
| | *AD* | 4, 6 | 4, 6 | 2, 1 | 2, 1 |
| | *BC* | 0, 3 | 3, 4 | 0, 3 | 3, 4 |
| | *BD* | 0, 3 | 3, 4 | 0, 3 | 3, 4 |

| | | 2 | |
|---|---|---|---|
| | | *W* | *X* |
| 1 | *AC* | 5, 0 | 2, 1 |
| | *AD* | 4, 6 | 2, 1 |
| | *BC* | 0, 3 | 3, 4 |
| | *BD* | 0, 3 | 3, 4 |

*Note*: Underscored payoffs indicate best replies to the other player's pure strategies. A strategy profile leading to a cell where all payoffs are underscored—for example, (*AC*, *XY*) on the left—is an equilibrium.

strategies as degenerate cases. Further extensions of the strategy set, such as correlated (i.e., stochastically dependent) strategies (Maschler et al., 2013, chapter 8), go beyond classical game theory.

The payoff profile associated with a profile of mixed or behavioral strategies is, in general, a profile of expected (values of) payoffs (see also section 2.4 below). Subsequently, the term "payoff" covers "expected payoff"; we add "expected" only for special emphasis.

Exogenous stochastic events are represented by a player without payoffs (usually called "Nature"), playing, at each of its nodes, a specific mixed move. Although this extension has many important applications (see, e.g., Fudenberg & Tirole, 1991, parts III & IV, on so-called Bayesian games), we can disregard it for present purposes.

A "solution concept" selects, for each game in some class of games and given strategy sets (pure, mixed, or behavioral), a set of strategy profiles as solutions. The basic solution concept of classical game theory is the Nash equilibrium (subsequently, just "equilibrium").

An "equilibrium" is a strategy profile where the strategy of each player is a best reply to the "complementary strategy profile," that is, the tuple of strategies of the other players (Nash, 1950). "Best reply" of a player means: no other strategy of the player achieves a higher payoff—or expected payoff, if relevant—against the complementary strategy profile. In an equilibrium, then, each player's strategy maximizes the player's payoff given the other players' strategies; no player's payoff can be increased by changing only the player's own strategy while leaving the complementary strategy profile unchanged.

We focus on equilibria of two-player games since it usually does not matter whether the complementary strategy profile appearing in the definition of equilibrium involves one or several players.

### 2.3 Pure-Strategy Equilibrium

Figure 9.1.3 shows the general form and several examples of simultaneous games with two players where each player has two moves (simultaneous $2 \times 2$ games). We find the pure-strategy equilibria of these and all other finite games by underscoring, in the strategic form, those payoffs of each player that are associated with best replies to the other player's pure strategies. A pure-strategy equilibrium is associated with a cell where all payoffs are underscored (see table 9.1.1 and figures 9.1.3 and 9.1.4).

In most games, a player's equilibrium strategy cannot be determined independently of the equilibrium strategies of the other players. The Prisoner's Dilemma (figure 9.1.3a) is the most prominent exception: player 1's strategy $U$ is strictly dominated by her strategy $D$, that is, $D$ yields a higher payoff than $U$, no matter which (pure or mixed) strategy is played by player 2. Similarly, player 2's strategy $L$ is strictly dominated by his strategy $R$. Therefore, the only equilibrium is $(D, R)$.

An equilibrium is "strict" if and only if (iff) each player has exactly one best reply to the complementary strategy profile. With the exception of $(U, L)$ in figure 9.1.3d, the equilibria shown in figure 9.1.3 are strict.

A pure-strategy equilibrium in a game without moves of Nature selects a unique play of the game, that is, a unique complete path through the game tree, called an "equilibrium path." In an equilibrium with mixing (by Nature or by conventional players), the term "equilibrium path" denotes the set of complete paths with positive probability.

### 2.4 Mixed-Strategy Equilibrium

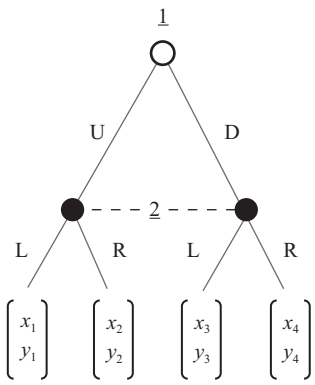
As shown by Matching Pennies (MP; figure 9.1.3b), there are games without pure-strategy equilibria. Yet, all finite games have at least one mixed-strategy equilibrium (Nash, 1950).

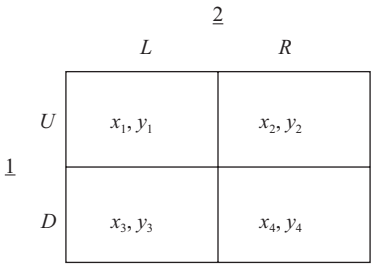
A profile of mixed strategies yields a probability distribution on the set of all payoff profiles and, thus, a profile of expected payoffs. For the general game of figure 9.1.3, we find that the mixed-strategy profile $(p \circ U \oplus 1 - p \circ D, q \circ L \oplus 1 - q \circ R)$ yields payoffs $pqx_1 + p(1-q)x_2 + (1-p)qx_3 + (1-p)(1-q)x_4$ for player 1 and $pqy_1 + p(1-q)y_2 + (1-p)qy_3 + (1-p)(1-q)y_4$ for player 2.

The definition of equilibrium remains unchanged. In MP, for instance, we find the unique mixed-strategy equilibrium $(\frac{1}{2} \circ U \oplus \frac{1}{2} \circ D, \frac{1}{2} \circ L \oplus \frac{1}{2} \circ R)$ with payoff profile $(0, 0)$.

Mixing maximizes a player's payoff iff all pure strategies played with positive probability are best replies to the complementary strategy and, therefore, yield the same payoff. This insight has three important consequences, which we spell out for simultaneous $2 \times 2$ games but which hold, *mutatis mutandis*, for other games

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024




| | |
|---|---|
| 3, 3 | 0, 4 |
| 4, 0 | 1, 1 |

(a) Prisoner's Dilemma: $x_3 > x_1 > x_4 > x_2$, $y_2 > y_1 > y_4 > y_3$.

| | |
|---|---|
| 1, –1 | –1, 1 |
| –1, 1 | 1, –1 |

(b) Matching Pennies (or Inspection Game): $x_1 = x_4 > x_2 = x_3$, $y_2 = y_3 > y_1 = y_4$.

| | |
|---|---|
| 4, 2 | 0, 0 |
| 0, 0 | 2, 4 |

(c) Battle of the Sexes: $x_1 > x_4 > x_3 = x_2$, $y_4 > y_1 > y_3 = y_2$.

| | |
|---|---|
| 4, 4 | 0, 3 |
| 4, 0 | 1, 1 |

(d) Nameless game.

**Figure 9.1.3**
Below the general form, extensive and strategic, of a $2 \times 2$ simultaneous game, three well-known named games (general structure indicated) and a nameless game are displayed.

where mixing occurs in equilibrium. First, no equilibrium where mixing occurs is strict. Second, if both players mix in equilibrium, one player's mixed strategy must equalize the expected payoffs of the *other* player's pure strategies (indicating how to compute the equilibrium). Third, small changes in one player's payoffs affect only the *other* player's equilibrium strategy. For instance, modifying MP by adding 1 to player 1's payoff in $(U,L)$ leads to the new equilibrium $(\frac{1}{2} \circ U \oplus \frac{1}{2} \circ D, \frac{2}{5} \circ L \oplus \frac{3}{5} \circ R)$.

Pure-strategy equilibria are special mixed-strategy equilibria. Accordingly, "Battle of the Sexes" (figure 9.1.3c) has three mixed-strategy equilibria: $(U,L)$, $(D,R)$, and $(\frac{2}{3} \circ U \oplus \frac{1}{3} \circ D, \frac{1}{3} \circ L \oplus \frac{2}{3} \circ R)$.

### 2.5 Subgame Perfection

Most games have several equilibria. So-called refinements of the equilibrium concept select subsets of these equilibria. We consider only the most important refinement, subgame perfection (Selten, 1965, 1975).

Subgame perfection is concerned with moves at information sets off the equilibrium path (i.e., information sets reached with probability 0 in equilibrium). Since these moves do not affect the players' equilibrium payoffs, they can fail to maximize payoffs *from the perspective of the information set where they are taken*. Examples are player 1's move $D$ and player 2's move $X$ in the equilibrium $(BD,XZ)$ in figure 9.1.1 (not highlighted; see also table 9.1.1). Subgame perfection aims at eliminating equilibria containing such non-payoff-maximizing moves.

Payoff maximization from the perspective of an information set requires that subsequent moves and, in the case of non-singleton sets, preceding moves are taken into account. Consider the equilibrium $(XU,R)$ in figure 9.1.4. All moves except the initial move $X$ of player 1

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

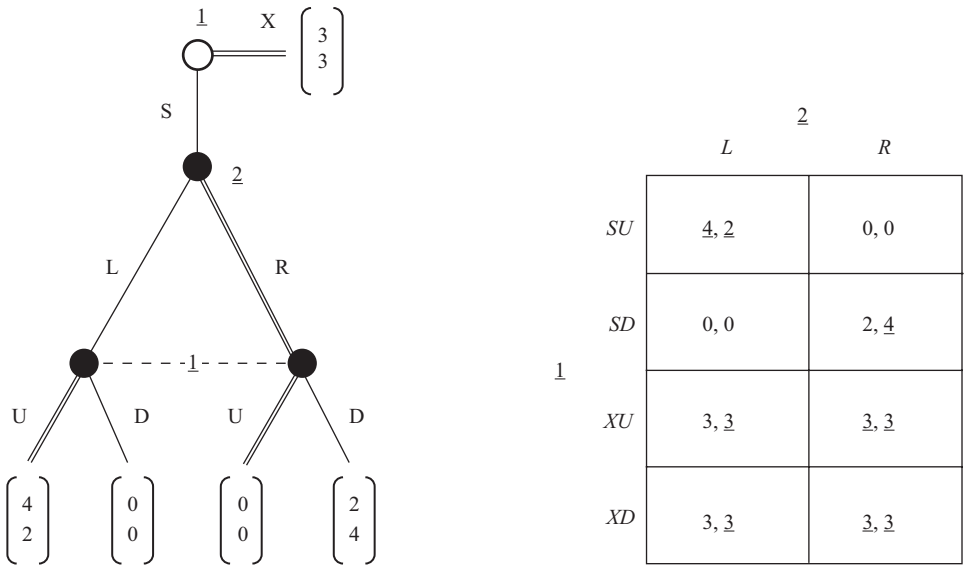



**Figure 9.1.4**

Player 1 decides whether to start (*S*) a subgame (type "Battle of the Sexes") or to exit (*E*) instead. The strategic form shows three pure-strategy equilibria; only (*XU*, *R*) (highlighted) is not subgame-perfect. An additional subgame-perfect equilibrium is (*XM*, *N*) with $M = \frac{2}{3} \circ U \oplus \frac{1}{3} \circ D$ and $N = \frac{1}{3} \circ L \oplus \frac{2}{3} \circ R$.

are off the equilibrium path. Given player 1's subsequent move *U*, player 2's move *R* does not maximize his payoff. Given player 2's preceding move *R*, player 1's move *U* does not maximize her payoff.

Subgame perfection addresses this interdependency problem by considering the section of the game beginning at player 2's node as a separate game, called a "subgame," requiring that the equilibrium of the complete game be consistent with some equilibrium of the subgame. The subgame has three equilibria, (*U*, *L*), (*D*, *R*), and ($\frac{2}{3} \circ U \oplus \frac{1}{3} \circ D$, $\frac{1}{3} \circ L \oplus \frac{2}{3} \circ R$), with payoff profiles (4, 2), (2, 4), and ($\frac{4}{3}, \frac{4}{3}$), respectively. In all three, both players' information sets are on the equilibrium path, implying payoff maximization by both.

In the complete game of figure 9.1.4, strategy *S* maximizes player 1's payoff iff her payoff in the subgame is at least 3. Accordingly, there are three subgame-perfect equilibria of the complete game: (*XD*, *R*), (*SU*, *L*), and a behavioral-strategy equilibrium (*XN*, *M*) with $N = \frac{2}{3} \circ U \oplus \frac{1}{3} \circ D$ and $M = \frac{1}{3} \circ L \oplus \frac{2}{3} \circ R$. The equilibrium (*XU*, *R*), in contrast, is not subgame perfect since (*U*, *R*) is not an equilibrium of the subgame.

Generally, a section of a game tree is a subgame iff it starts with a singleton information set, contains all successors to its starting node down to the final nodes with the payoffs, and contains no incomplete information sets (i.e., any non-singleton information set is either completely included or completely excluded). The complete game is an (improper) subgame of itself. A subgame considered in isolation, then, satisfies the definition of a game.

A strategy profile is a subgame-perfect equilibrium of a game iff, for all subgames, the profile's restriction to the subgame is an equilibrium of the subgame. Subgame-perfect equilibria are, in general, equilibria in behavioral strategies because mixed strategies cannot describe mixed moves at information sets that are, according to the strategy itself, reached with probability 0 (Fudenberg & Tirole, 1991, pp. 87–88, esp. figure 3.10). Every finite game has at least one subgame-perfect equilibrium in behavioral strategies (Selten, 1965).

In simultaneous and other games without proper subgames (figures 9.1.3 and 9.1.5), each equilibrium is subgame perfect by definition. In all other games, subgame-perfect equilibria are found by (a generalized version of) "backward induction." This solution method begins with finding one equilibrium for each of the "smallest" subgames (those that have no proper subgames themselves). It then goes on to more and more encompassing subgames, always preserving the equilibria of already solved subgames when solving the more encompassing ones (as in figure 9.1.4's game above).

In strictly sequential games, each nonfinal node is also the first node of a subgame. The smallest subgames begin at the last player nodes and have only one active player. In figure 9.1.1b, backward induction first selects *C* by player 1 and *Z* by player 2 at the two last nodes, then *X* at the left node of player 2, and, finally, *B* at the initial node. Thus,

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

the unique backward-induction solution is $(BC, XZ)$. This solution is an equilibrium (see table 9.1.1), and its restriction to any proper subgame is an equilibrium of the subgame—for instance, $(C, X)$ for the subgame beginning at player 2's left node. Therefore, the backward-induction solution is a subgame-perfect equilibrium.

In figure 9.1.2, the only proper subgame begins at the second node of player 1 and has a single equilibrium where player 1 chooses $C$. Of the two pure-strategy equilibria of this game (see table 9.1.1), then, only $(BC, X)$ is subgame perfect.

Yet, subgame perfection still allows for non-payoff-maximizing moves. Figure 9.1.5 illustrates two strategy profiles of Selten's (1975) "Horse Game." Both are subgame-perfect equilibria: no player can achieve a higher payoff by unilaterally changing her strategy, and there are no proper subgames. Yet, in figure 9.1.5a, the move of player 2 obviously fails to maximize his payoff from the perspective of his information set.

The problem is that subgame perfection addresses the issue of non-payoff-maximizing moves only indirectly, by considering subgames, which do not exist in the Horse Game. This suggests a more direct approach: require that each player at each information set maximize her payoffs. In the case of singleton information sets, this works. However, consider the equilibrium in figure 9.1.5b, where player 3's information set is not reached. Since this is a non-singleton information set, identification of a payoff-maximizing move would require an appeal to preceding moves reaching the information set—but there are none. Mathematically, payoff maximization at player 3's information set requires a probability distribution on the set of nodes, but any such distribution is arbitrary. In more complicated games, a similar problem occurs if several preceding moves reach a non-singleton information set off the equilibrium path. These problems provide the starting point for competing refinements beyond subgame perfection (e.g., Maschler et al., 2013, chapter 7).

## 3. Applying Game Theory

In applied game theory, game-theoretical models are used to explain and predict behavior. A model is a theory together with the description of a situation, real or hypothetical, to which the theory is applied in order to deduce predictions or explananda (Bunge, 1973, pp. 97–99).

We restrict attention to human behavior. In classical game theory, the relevant theory of human behavior is rational choice theory (RCT; cf. Sugden, 1991, and chapter 10.4 by Raub, this handbook), which is applied to situations described by a game tree. Players are individual persons. Payoffs are utilities representing players' preferences. Probabilities are either physical probabilities known to all players or players' subjective degrees of belief. In the latter case, the probabilities are, quite arbitrarily (Sugden, 1991, p. 768), assumed to be the same for all players (the "common prior assumption").

### 3.1 Rationality and Preferences

Applied game theory assumes that players are rational actors in the sense of RCT, know the game tree, and have true beliefs about the strategies played by the other players.

According to RCT, an actor is rational in the sense that she has, for any set $X$ of mutually exclusive and jointly exhaustive options, a complete preference ordering described by a complete and transitive relation $\succcurlyeq$ on $X$, called "weak preference" (see chapter 8.1 by Grüne-Yanoff, this handbook). For $a, b \in X$, $a \succcurlyeq b$ means either

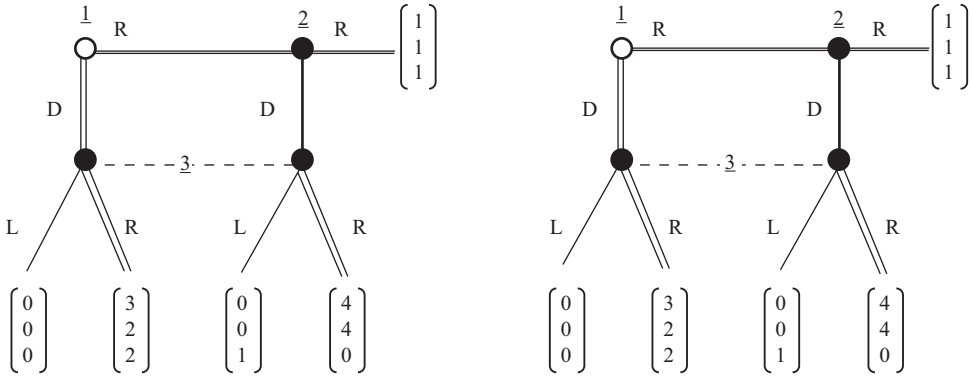



**Figure 9.1.5**

(a) Selten's Horse Game with equilibrium $(D, R, R)$ highlighted. Player 2 does not maximize his payoffs from the perspective of his information set. (b) Selten's Horse Game with equilibrium $(R, R, L)$ highlighted. It is unclear what it would mean for player 3 to maximize her payoffs from the perspective of her information set.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

that she strictly prefers $a$ to $b$ (symbolically, $a \succ b$) or that she is indifferent between the two (symbolically, $a \sim b$). A real-valued function $u$ defined on $X$ is said to represent the preference relation (and is called "utility function") iff, for all $a, b \in X$, $u(a) \geq u(b)$ implies $a \succcurlyeq b$ and vice versa. This ensures that $u(a) > u(b)$ iff $a \succ b$, and $u(a) = u(b)$ iff $a \sim b$.

Under the standard ("revealed-preference") interpretation, preferences only summarize behavioral dispositions: $a \succ b$ means that the actor would not choose $b$ if she had to choose from a subset $B$ of $X$ with $a, b \in B$. This subjunctive conditional exhausts the meaning of the strict preference relation. It implies that the actor would always maximize her utility: she would choose $c \in B$ only if $u(c) = \max_{x \in B} u(x)$. If $B$ contains several utility-maximizing options, the theory implies that the actor would choose one of them; apart from that, her behavior would be indeterminate.

This interpretation implies that RCT considers neither decision making nor deliberation or reasoning as mental processes; it is only concerned with the resulting choices. Neither utilities nor preferences are reasons for, or causes of, choices; they just summarize hypothetical choices, that is, behavioral dispositions.

Preferences themselves may result from all kinds of motivations—materialistic or nonmaterialistic, selfish or unselfish, and so on—but they are "all-inclusive": all factors influencing choices are already taken into account (Binmore, 1994, pp. 104–109). Although subjective well-being may motivate an actor's choices, her preferences may reflect other factors as well (e.g., moral considerations). Therefore, RCT is not wedded to the assumption that an actor strictly preferring $a$ to $b$ is subjectively better off if she gets $a$ than if she gets $b$ (Hausman, 2012, pp. 81–83).

In addition to some theory of motivation, predictions or explanations also require hypotheses (lacking in RCT) that specify the conditions under which players are rational and have true beliefs about the situation and other players' strategies—for instance, shared cultural background (Meyerson, 2009) or experience and sufficiently high stakes (Binmore, 2007, chapter 1). Given these possible specifications, RCT is not a single theory but a family of distinct theories relying on a common mathematical language. Applied game theory inherits this feature.

### 3.2 Preferences in Games

In game theory, actors are players who choose moves at information sets. This has two implications. First, the objects of choice are moves at information sets, not strategies. If a player has several information sets, she cannot choose a strategy in a single act of choice. Each information set, whether reached during play or not, requires its own (possibly hypothetical) choice act.

Second, a player's utilities attach to complete paths, which typically contain moves by other players. Therefore, the objects of preferences, namely, complete paths, differ from the objects of choice, namely, moves at information sets. As such, this is not problematic: given beliefs about other players' strategies, choosing moves at information sets amounts to choosing among complete paths (see section 3.5 below). However, it becomes a problem if the question of *who* chooses matters to the players: in these cases, preferences over complete paths defined in terms of hypothetical choices may not exist (Hausman, 2012, pp. 31–33).

Consider, for instance, a game where Eve must accept ($A$) or decline ($D$) Adam's marriage proposal. Assume that Adam wishes her to accept; *in this sense*, he prefers the $A$-path to the $D$-path. However, this is not the same as saying that, hypothetically, Adam would choose the $A$-path over the $D$-path *if it were his choice*. If, plausibly, the distinction between "the $X$-path chosen by Eve" and "the $X$-path chosen by Adam" matters to Adam, he cannot have preferences *in the revealed-preference sense* over paths chosen by Eve.

If, in contrast to this example, the objects of a player's desires are, physically or in the player's mind, separable from the paths (e.g., monetary gains), it does not matter who chooses. Preferences over complete paths derive from preferences defined in terms of hypothetical choices over these separable objects. This separability assumption is sometimes called "consequentialism" (Hausman, 2012, pp. 42–45).

### 3.3 Uncertainty and Beliefs

Extending RCT to "choice under uncertainty" (see also chapter 8.2 by Peterson, this handbook) requires a consequentialist distinction between actions, states, and consequences that are separable from actions and states (see table 9.1.2). The elements of each set (of actions, of states, and of consequences) are mutually exclusive and jointly exhaustive. An actor's set of actions describes the options among which she must choose. The set of states describes all aspects of the situation not under her control. States differ in those aspects about which she is uncertain. If an action is taken in some state, one of the consequences obtains.

Von Neumann and Morgenstern's (2004) theory of choice under risk assumes that a physical probability distribution on the set of states is given and known to the actor. Each action, then, implies a probability distribution on the set of consequences, called a "lottery." If

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

the actor has a complete preference ordering on the set of all conceivable lotteries, and if her preferences satisfy several further axioms, there exists a real-valued utility function on the set of consequences such that the lotteries' expected utilities represent the actor's preferences over actions.

Yet, if a player has nonconsequentialist motivations, her preferences may violate the axioms of the theory, and an expected utility representation may not exist (Bradley & Stefánsson, 2017). Examples are von Neumann and Morgenstern's (2004, pp. 629–630) "utility of gambling" and certain procedural-fairness motivations (Diamond, 1967; Verbeek, 2001).

In Savage's (1954) theory of choice under radical uncertainty, no physical probabilities are given; the probabilities are subjective. Both the utility function on the set of consequences and the subjective probabilities are derived from a preference ordering on the set of all conceivable actions (considered as functions assigning consequences to states). Again, the (subjective) expected utilities of the actions represent the actor's preferences.

In a strict revealed-preference interpretation of Savage's theory, the expected utilities of actions just summarize the actor's behavioral dispositions. Usually, however, the subjective probability of a state is interpreted as a personal degree of belief that the state obtains. In contrast to the utilities of the consequences, beliefs are mental states.

This interpretation of beliefs as mental states, on the one hand, and utilities of complete paths as expression of behavioral dispositions, on the other hand, is typical of classical game theory, whether or not beliefs are explicitly introduced as subjective probabilities.

### 3.4 Information, Causality, Time, and Memory

The game tree describes the information flow in plays of the game. Players called upon to move can observe at which information set they are, but, except for singleton information sets, not at which node. Since the transmission of information requires a causal link, the game tree describes at least some aspects of the causal relations relevant in the game. However, Newcomb's problem (recast as a game) shows that some conceivable causal structures cannot be described with the help of a game tree (Albert & Heiner, 2003; Binmore, 1994, pp. 242–256).

The flow of information also describes the flow of time: if a player observes a move, the move must have occurred earlier. Where no observations take place (as, e.g., in simultaneous games), the game tree is consistent with any timing.

Applied game theory mostly assumes "perfect recall": players forget neither their own moves nor anything they have known or observed before (Fudenberg & Tirole, 1991, p. 81). But models of imperfect recall are possible.

In figure 9.1.6a, player 1 forgets her initial move. It seems natural to assume that, at her second information set, she behaves as if she were an independent third player with player 1's preferences. The resulting equilibria—for instance, $(L, DX, L)$ (three-player notation) or $(LL, DX)$ (two-player notation)—assume that, as the third player, she has correct beliefs about her first-player strategy, although she forgot what she did in that role (which is not implausible with respect to routine behavior).

In figure 9.1.6b, player 1 does not know whether she has already moved ("absentmindedness"; Piccione & Rubinstein, 1997). Her equilibrium beliefs are the subject matter of philosophy's "sleeping beauty problem" (Cisewski, Kadane, Schervish, Seidenfeld, & Stern, 2016).

**Table 9.1.2**
Decision problem with four actions $A_i$, three states $S_j$, and five consequences $C_k$

| Actions \ States | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| $A_1$ | $C_1$ | $C_1$ | $C_1$ |
| $A_2$ | $C_1$ | $C_2$ | $C_3$ |
| $A_3$ | $C_5$ | $C_2$ | $C_1$ |
| $A_4$ | $C_4$ | $C_2$ | $C_1$ |

*Note*: Uncertainty about the states implies uncertainty about the consequences of actions.

### 3.5 Equilibrium and Beliefs about Other Players

A player's behavioral dispositions in a game are described by a behavioral strategy. A behavioral strategy is a complete contingent plan of action, that is, a list of subjunctive conditionals stating, for each information set of the player, which (possibly mixed) move the player would choose at this information set. According to RCT, a player would choose a move only if the move maximized her expected utility given her beliefs at the relevant information set.

The concept of equilibrium among rational players originates from economics; traditionally, it involves the assumption that everybody's beliefs are correct. If one assumes that *all* beliefs *everywhere* in the game are correct, RCT implies that all moves, those actually taken in the game as well as the hypothetical moves off the equilibrium path, maximize the relevant player's expected utility from the perspective of the information set where they are supposed to be made.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

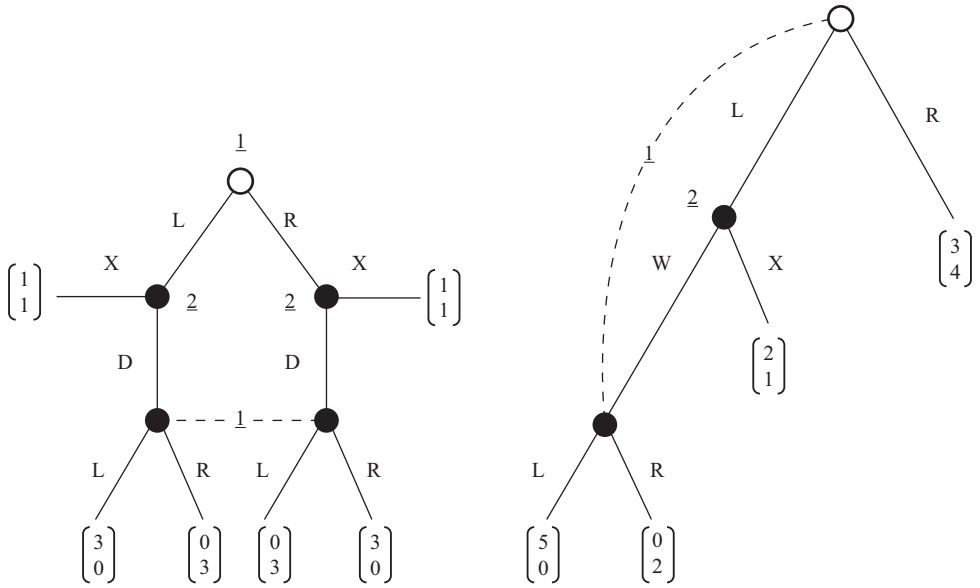



**Figure 9.1.6**
(a) At her second information set, player 1 has forgotten her initial move. (b) Player 1 does not know whether she moved before ("absentmindedness").

The game of Blackmail in figure 9.1.7 illustrates the basic argument. Eve (player 1) knows some damaging fact about Adam (player 2). She can keep silent (*S*) or, by threatening to reveal the fact, try to extort money from him (*X*). If Eve chooses *X*, Adam can give in and pay (*P*) or report Eve to the police (*R*), in which case she goes to jail but the damaging fact is revealed, to Adam's distress. In the strategic form, we find two pure-strategy equilibria, $(S,R)$ and $(X,P)$.

Players' utilities indicate how, given the opportunity, they would choose among complete paths. If Eve chose *X*, Adam could choose between complete paths *X*–*R* and *X*–*P*; preferring *X*–*P*, he would choose *P*. Since Eve has correct beliefs about what Adam would do and prefers the complete path *X*–*P* to the complete path *S*, she chooses *X*. Therefore, players play in accordance with the subgame-perfect equilibrium $(X,P)$.

More generally, a player's beliefs are conditional on information sets, that is, they take into account what the player knows at the information sets. Players anticipate their own beliefs and, therefore, their own moves. Even if a player believes that she would never reach a certain information set, her beliefs about how other players would move after this information set had been reached are the beliefs she would have if this—in the player's eyes, impossible—event happened. The requirement that beliefs are conditional on information sets and correct, even in the case of counterfactual considerations, implies that a player would not, upon encountering an unexpected move by another player, revise her beliefs about subsequent moves by this or any other player.

Consider, for instance, the unique subgame-perfect equilibrium $(BC,XZ)$ in figure 9.1.1b. The belief of player 2 at his left node that player 1 would play *C* at her last node already takes into account that player 1 would have to have played the unexpected move *A* in order for player 2 to find himself at his left node.

The whiff of inconsistency accompanying backward induction (e.g., Sugden, 1991, pp. 771–774) derives from the assumption that players derive their correct beliefs about other players' strategies from the belief that the other players are rational—a belief that might turn out to be false if an unexpected move occurred. In our presentation, however, we do not, and need not, refer to beliefs in the rationality of other players. It might be argued that such beliefs follow from the assumption that players know the extensive form of the game, which, after all, contains players' utilities, whose interpretation is tied to RCT. Yet, one need not assume that players know other players' utilities. It suffices to assume instead that each player has correct beliefs about the strategies played by the other players. This weaker assumption avoids any potential inconsistencies from the outset.

In many games, players' beliefs about the strategies played by the other players are sufficient to determine their utility-maximizing moves from the perspective of each information set. Not so in the equilibrium $(R,R,L)$ of figure 9.1.5b. Player 3's correct belief that players 1 and

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

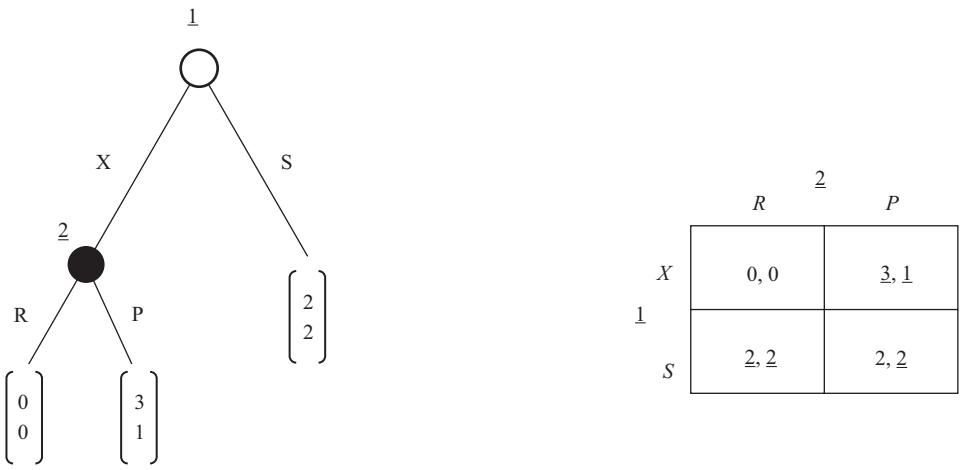



**Figure 9.1.7**
The game of Blackmail. The strategic form shows two equilibria, $(X, P)$ and $(S, R)$. Only $(X, P)$ is subgame-perfect.

2 would both play *R* implies radical uncertainty for the hypothetical situation where, unexpectedly, she finds herself called upon to move. In order to decide what she would do, player 3 must form beliefs about the alternative histories that, counterfactually, might have led to this situation.

Traditionally, RCT resolves radical uncertainty by assuming arbitrary beliefs in the form of a subjective probability distribution. This is a first step toward further refinements justifying and, often, restricting equilibrium beliefs.

One instructive example of such a refinement is "forward induction" (Maschler et al., 2013, p. 261). In figure 9.1.4, let player 2 know that player 1 would never play her—strictly dominated—strategies *SD* or *SM* (where $M = \frac{2}{3} \circ U \oplus \frac{1}{3} \circ D$). If player 2 found himself called upon to move, he would not believe, then, that player 1 would play either *D* or *M*: since player 1 must have had played *S* before, *D* or *M* would imply the strategies *SD* or *SM*, respectively, which are ruled out. Therefore, player 2 would not consider playing either *R* or $N = \frac{1}{3} \circ L \oplus \frac{2}{3} \circ R$ (his best replies to *SD* and *SM*, respectively). But then, only one subgame-perfect equilibrium remains: $(SU, L)$.

Equilibria where players are supposed to mix among strategies or moves lead to another problem with beliefs (cf., e.g., Sugden, 1991, pp. 765–768). In Matching Pennies (figure 9.1.3b), game theory predicts the mixed-strategy equilibrium $(\frac{1}{2} \circ U \oplus \frac{1}{2} \circ D,\ \frac{1}{2} \circ L \oplus \frac{1}{2} \circ R)$, while RCT says that the players' behaviors are indeterminate because they are indifferent between their pure strategies. This contradiction is resolved if one assumes that (a) each player's behavior is indeterminate so that the other player faces radical uncertainty, and (b) each player assigns subjective probabilities $\frac{1}{2}$ to the other player's choices. Thus, the probabilities are interpreted as subjective beliefs without physical correlates. Yet, while internally consistent, this interpretation clashes with empirical applications where the same probabilities appear as physical probabilities of actions.

The same problem occurs in nonstrict pure-strategy equilibria like $(U, L)$ in figure 9.1.3d, where player 1 is indifferent but player 2 nevertheless expects a specific move.

### 3.6 Credibility, Commitment, and the Explicitness Requirement

The strategic form represents a game as if players could choose strategies. However, as already explained, players can choose only moves. A player's strategy is a collection of subjunctive conditionals about her choices at her information sets.

The example of Blackmail (figure 9.1.7) illustrates an important consequence of this fact. Considered as such a subjunctive conditional, Adam's *strategy R* in Blackmail is an implicit threat to Eve: "I would play *R* if you played *X*." This threat, however, is not credible: Adam would not make the *move R* at this information set. The equilibrium $(S, R)$ in Blackmail assumes that Eve has false beliefs about what Adam would do.

One might wonder whether Adam, to deter Eve, could commit himself to report her if she blackmailed him. However, in marked contrast to cooperative game theory, noncooperative game theory requires games to be interpreted under the proviso that only those choices exist that are *explicitly* modeled as possible moves (Harsanyi & Selten, 1988, chapter 1). As it stands, the model of Blackmail contains no information set where Adam

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

could choose to commit himself. This leaves us with only two modeling options: either the model remains as is, expressing our belief that, in the situation we wish to model, a commitment option does not exist, or, if we believe otherwise, we must explicitly introduce a commitment option into the model.

Commitments can take several forms (Güth & Kliemt, 2007); for simplicity, we only consider removing unwanted options (proverbially, "burning one's bridges"). Consider an extended version of Blackmail (figure 9.1.8) where Adam can choose whether to commit himself in this sense to report Eve if she blackmails him and where Eve can observe whether he commits. In this game, the original game of Blackmail appears as a subgame reached only if Adam does not commit himself (move *N*). In the subgame reached after committing himself (move *C*), he has no choice between *R* and *P*; if Eve chose to blackmail him, she would automatically be reported to the police.

Noncooperative game theory is silent on the question of whether, or how, commitment is in fact possible. For instance, when revisiting Newcomb's problem, Spohn (2012) allows for (observable) commitment as a purely mental process, an exertion of willpower. In a game-theoretic model based on this hypothesis, however, such a commitment must explicitly appear as a move in the game tree (Binmore, 1994, pp. 165–167, 173–182). It is one of the merits of the explicitness requirement of noncooperative game theory that it forces us to represent all factual assumptions explicitly.

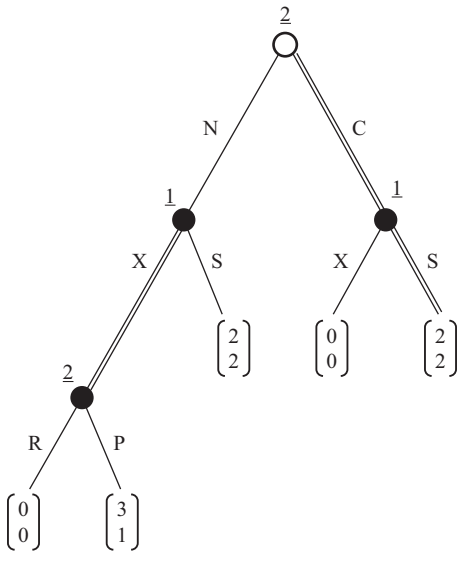



**Figure 9.1.8**
The extended game of Blackmail with initial commitment possibility for player 2. The only subgame-perfect equilibrium (highlighted) is (*CP*, *XS*).

The explicitness requirement and the assumption that preferences are all-inclusive play an important role in discussions concerning the Prisoner's Dilemma (figure 9.1.3a). Many contributors found it hard to swallow the conclusion that rational players play $(D, R)$ instead of $(U, L)$, which both players would prefer. However, in order to avoid this conclusion, one must change either the utilities or the structure of the game tree such that a different game results (Binmore, 1994, pp. 104–109).

## 4. Game Theory and "Reasoning about Knowledge"

The question of how players might come to have correct beliefs about other players' strategies becomes particularly relevant if one considers game theory not only as predicting what players will do but also as advising players what they should do in the light of the theory itself. The condition that all potential advisees believe in the predictions/prescriptions of a theory and can find no way to do better by deviating unilaterally implies equilibrium recommendations and equilibrium play (Dacey, 1976).

This insight instigated a tradition of game theory as "reasoning about knowledge." Von Neumann and Morgenstern (2004) already searched for a theory of rational play that would imply correct beliefs for rational players of unlimited intellectual capacities who know and accept the theory. Those who still work in this tradition replace knowledge of the theory by "common knowledge" of the game tree, that is, all players know the tree, all players know that all players know it, and so on, through all stages of meta-knowledge. Game theory becomes logical reasoning about common knowledge, including common knowledge of the theory of rational play (Fagin, Halpern, Moses, & Vardi, 1995).

This logico-philosophical rather than psychological "reasoning about knowledge" approach to interactive decision making appeals to game theorists (notably, economists) to the present day. Identifying equilibrium strategies by logical reasoning apparently keeps their discipline separate from psychology. However, their claim that game theory can, in this way, *explain* human interactions seems far-fetched and unlikely to survive the contact with reality. For this reason, Selten (1999, p. 303), while endorsing game theory for its philosophical interest, referred to it as "rationology," likening it to theology rather than science.

Yet, as our presentation emphasizes, classical game theory is not inevitably moored to the "reasoning about knowledge" approach and its most recent version, epistemic game theory (see chapter 9.2 by Perea, this handbook). Applications can often rely on simple psychological hypotheses about the circumstances under which players

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

may be rational and have correct beliefs about the situation and other players' strategies. Evolutionary game theory (see chapter 9.3 by Alexander, this handbook) shows simple mechanisms by which even players with low cognitive abilities learn to play sophisticated equilibria. Behavioral economics (see chapter 9.4 by Dhami & al-Nowaihi, this handbook) supplies not only new hypotheses about players' motivations but also alternatives to RCT as the basic theory of choice. It is a testimony of its intellectual achievement that many of these new developments still make use of, or at least take their inspiration from, classical game theory.

## Acknowledgments

For helpful comments and suggestions, we thank Volker Gadenne, Werner Güth, Sebastian Krügel, Fabian Meckl, Wolfgang Spohn, and an anonymous referee.

### References

Albert, M., & Heiner, R. A. (2003). An indirect-evolution approach to Newcomb's problem. *Homo Oeconomicus*, *20*, 161–194.

Binmore, K. G. (1994). *Playing fair: Game theory and the social contract* (Vol. I). Cambridge, MA: MIT Press.

Binmore, K. G. (2007). *Does game theory work?* Cambridge, MA: MIT Press.

Bradley, R., & Stefánsson, H. O. (2017). Counterfactual desirability. *British Journal for the Philosophy of Science*, *68*, 485–533.

Bunge, M. (1973). *Method, model and matter*. Dordrecht, Netherlands: Reidel.

Cisewski, J., Kadane, J. B., Schervish, M. J., Seidenfeld, T., & Stern, R. (2016). Sleeping beauty's credences. *Philosophy of Science*, *83*, 324–347.

Dacey, R. (1976). Theory absorption and the testability of economic theory. *Zeitschrift für Nationalökonomie*, *36*, 247–267.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, *75*, 765–766.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.

Güth, W., & Kliemt, H. (2007). The rationality of rational fools. In F. Peter & H. B. Schmidt (Eds.), *Rationality and commitment* (pp. 124–149). Oxford, England: Oxford University Press.

Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Hausman, D. M. (2012). *Preference, value, choice, and welfare*. Cambridge, England: Cambridge University Press.

Leonard, R. (2010). *Von Neumann, Morgenstern, and the creation of game theory: From chess to social science 1900–1960*. Cambridge, England: Cambridge University Press.

Maschler, M., Solan, E., & Zamir, S. (2013). *Game theory*. Cambridge, England: Cambridge University Press.

Meyerson, R. B. (2009). Learning from Schelling's *Strategy of conflict*. *Journal of Economic Literature*, *47*, 1109–1125.

Nash, J. F. (1950). Equilibrium points in *n*-person games. *PNAS*, *36*, 48–49.

Piccione, M., & Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, *20*, 3–24.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley.

Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit [Game-theoretic treatment of an oligopoly model with demand inertia]. *Zeitschrift für die gesamte Staatswissenschaft*, *121*, 301–324, 667–689.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*, 25–55.

Selten, R. (1999). Response to Shepsle and Laitin. In J. Alt, M. Levi, & E. Ostrom (Eds.), *Competition and cooperation: Conversations with Nobelists about economics and political science* (pp. 303–308). New York, NY: Russell Sage Foundation.

Spohn, W. (2012). Reversing 30 years of discussion: Why causal decision theorists should one-box. *Synthese*, *187*, 95–122.

Sugden, R. (1991). Rational choice: A survey of contributions from economics and philosophy. *Economic Journal*, *101*, 751–785.

Verbeek, B. (2001). Consequentialism, rationality and the relevant description of outcomes. *Economics & Philosophy*, *17*, 181–205.

von Neumann, J., & Morgenstern, O. (2004). *Theory of games and economic behavior* (60th anniversary ed.). Princeton, NJ: Princeton University Press.

Weintraub, E. R. (Ed.). (1992). *Toward a history of game theory*. Durham, NC: Duke University Press.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024

MIT Press Direct

© 2021 The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2161570/s041700_9780262366175.pdf by guest on 21 April 2024