5

Rationality and Rational Choice

Richard A. Wandling

Eastern Illinois University

The rationality concept has figured prominently in some of the most fascinating, heartfelt, and at times acrimonious scholarly exchanges among political scientists. This chapter focuses on five important intellectual developments in the study of rationality from a political science perspective: (1) the 1960s as an important era in scholarly exploration of the relationship between public policy making, decision making, and rationality; (2) Herbert Simon's seminal and hugely influential theorizing on decision making and the limits of individual rationality; (3) the legacy of bounded rationality, particularly in Graham Allison's models of decision making; (4) the seminal work of a group of economists and political scientists during the 1950s and 1960s who figured prominently in the emergence of modern rational choice theory; and (5) the modern scholarly debate over rational choice. A central theme of this survey is the tension between economic and political definitions of rationality and how these conceptions of rationality have shaped contemporary political science theory and research.

Policy Making, Decision Making, and Rationality

Charles Lindblom's "The Science of 'Muddling Through'" (1959) was an important milestone for a whole generation of theory and research on public policy making. Although an economist by training, Lindblom became a major figure in political science, particularly among scholars of public administration and public policy. While exploring the intersection of public policy making and administrative decision making, Lindblom compares two "methods" of policy analysis and choice, identified as "rational-comprehensive" and "successive limited comparisons" (p. 81). The first method is summarized as the "root" method and the latter, the "branch" method. Lindblom presents the rational-comprehensive method (or model) in a negative light, as not only empirically flawed social science but as normatively questionable as a guide for sound decision making and public policy making in a democracy.

The rational-comprehensive model assumes that policies are crafted through a process that involves advance specification of key values and goals, tightly configured means–ends analysis, extensive analysis that is at once comprehensive and characterized by high levels of information, and a prominent role for theory-driven analysis. Out of this analytically intensive and information-rich process emerges a policy choice that is the "best" relative to decisional elements such as values and goals, actual analysis, and means evaluation. The successive limited comparisons model, however, is the one embraced by Lindblom. With this model, also known as incrementalism, values and goals often are not distinct, analysis of relations between ends and means is limited and perhaps even inappropriate, the options considered are few in number and differ only marginally (or incrementally) from each other, and policy choices emerge out of a "succession of comparisons" (p. 81) among a limited set of options. If theory is important in the rational-comprehensive method, decision making in incrementalism is process oriented, with goodness of a decision defined as achieving agreement among analysts—that is, agreement rather than some objective evidence that the information, data, and analysis clearly point to the best option.

Lindblom's framework represents a broadside against application of the rational model to policy making and administrative decision making. This comprehensively and tightly specified version of rationality does not work as either description or explanation of public policy making. However, to Lindblom this does not mean that policy making lacks rationality or is characterized by irrationality. It comes down to how rationality is conceptualized. Lindblom does not portray a chaotic or random universe with irrationality run rampant; there is a science or logic to "muddling through." Decisions are made through a politicized process rather than based on compelling, objective logic of the facts, evidence, and information collected. In fact, to Lindblom the rationality of incrementalist-style policy making is preferable. Incrementalist-style rationality is very compatible with a pluralistic political system, particularly in producing options that rank high on political relevance and are grounded firmly in existing knowledge and information held by government officials.

Lindblom set the stage for further examination of rationality during the pivotal 1960s period of political science scholarship. Paul Diesing (1962) argued that rationality has multiple meanings and lamented the tendency to view rationality primarily as either technical or economic rationality concerned with organizational productivity and economic efficiency. Diesing develops a philosophy-oriented framework that argues for the study of three other forms of rationality—social, legal, and political. Aaron Wildavsky (1966), one of the 20th century's most influential political scientists, takes the cue and warns strongly against framing rationality in terms of decision-making strategies or techniques such as cost–benefit analysis, systems analysis, and program budgeting. For political science, the latter were flawed because they indicated an economics-oriented view of rationality. To Wildavsky, political rationality is important in its own right because government leaders must calculate political costs such as the resources needed to generate support for a policy, the implications of a policy decision for reelection, and the possibility of provoking hostility for decisions not well received.

Simon, March, and the Limits of Rationality

Herbert Simon greatly influenced theory and research in fields as disparate as organization theory, decision sciences, and bureaucratic policy making. His ideas also played a role in the development of rational choice theory—whether though his criticism or through efforts by some rational choice practitioners to incorporate Simon's rationality assumptions into their research.

Simon provides a synthesizing approach to rationality that incorporates both economic and psychological dimensions while exploring the limits or boundaries of individual and organizational rationality. A starting point is Simon's (1957) distinction between "objective" and "subjective" rationality. Objective rationality is evident if a decision or choice is the "correct behavior for maximizing given values in a given situation" (p. 76). With this version of rationality, a clear test is available to ascertain the correctness of a decision or choice. Subjective rationality incorporates psychological elements by considering the decision maker's actual knowledge—or knowledge limitations. In short, based on the information possessed by the decision maker, what might be concluded about the rationality of a decision? Simon's concern is that standards for achievement of objective rationality go well beyond the actual decision-making abilities of individuals, specifically individuals in complex organizations. The realities of psychology and human cognition mean that full knowledge of decision-related information is not possessed, and the full range of options also is not identified and evaluated in a comprehensive way.

Simon (1955) criticizes the rationality of classic economic theory and its model of "economic man" (p. 99), who is assumed to have extensive and intensive knowledge relative to the decision-making environment while possessing a well-organized and stable system of preferences, as well as a skill in computation that enables him to calculate the best alternative that reflects the highest point on his preference scale. The economic model of rationality is problematic for the development of a theory of the business firm or any type of organization, and this is the case whether the goal is empirically or normatively based theory. To Simon, real, empirical human rationality does not achieve the demanding standards of the classic economic model. Perhaps with a hint of things yet to come in the social sciences (including political science), Simon uses the term *rational choice* while inventorying key limits or constraints in "rational adaptation" behavior, particularly with respect to the range of alternatives considered, preferences, and decision maker knowledge of potential decision "payoffs" (p. 100). Simon also criticizes the "global rationality" assumptions that he sees embedded in game theory and castigates the economic rationality model as a "simplified model" that fails to capture the complex reality of a "choosing organization of limited knowledge and ability" (pp. 101, 114).

With this foundation, Simon fully develops his theory of *bounded rationality*—with important contributions from coauthor James March (March & Simon, 1958). The rationality of "administrative man" (p. 137) is compared and contrasted with the rationality requirements of classical economics—and statistical decision theory. In the latter versions of rationality, decision optimality is the standard in an environment with a full and clear specification of alternatives, knowledge of consequences of the alternatives, and a "utility ordering" (p. 138) in which key values at stake guide fully conscious assessment of the alternatives. March and Simon, however, argue that individuals in organizational settings are not guided by the quest for optimality (i.e., the best possible decision) but rather make decisions at the point that an alternative is deemed satisfactory. They assert that "most human decision-making, whether individual or organizational, is concerned with the discovery and selection of satisfactory alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives" (pp. 140–141). This point sets the stage for the much-referenced *satisficing* concept, which is a decision-making process in which the satisfactory standard is reached and the option selected is deemed as sufficient by the individual decision maker. In sum, the option is satisfactory—and it suffices. Satisficing is a major departure from the quest for the best possible choice as determined by extensively analyzing a wide range of alternatives and factoring in a full range of decision-related values or preferences. This model of decision making also parts company with the classic economic model in another way, through March and Simon's assertion that alternatives are evaluated sequentially rather than simultaneously. At some point, an alternative is considered to be acceptable, given organizational goals, values, and decision-maker knowledge; the decision process concludes at that point.

Satisficing, however, does not take place in a vacuum; it is embedded in an organizational context in which rationality is bounded by "repertoires of action programs" (March & Simon, 1958, p. 169) that circumscribe and also channel the decision-making process. March and Simon give particular emphasis to the role of organization structure as the setting for individual decision making. Organization structure comes to play an important role in establishing the "boundaries of rationality" (p. 171). In essence, when we speak of the rationality of individual decision makers, we also are considering the role that organizations play in funneling or channeling decision making and even compensating for the limits of human rationality.

Later, Simon (1985) shed additional light on this pathbreaking approach to rationality by noting that bounded rationality really is interchangeable with the term *proce dural rationality.* Rationality is rooted in an organizational process of identifying alternatives, collecting information, and considering important values. This is another way of saying that there is no such thing as a substantively or objectively optimal decision. Simon sees this distinction as parallel to the concepts of procedural and substantive due process, observing that "in the same way, we can judge a person to be rational who uses a reasonable process for choosing; or, alternatively, we can judge a person to be rational who arrives at a reasonable choice" (1985, p. 294). Bounded rationality is a way of focusing on the use of a reasonable process that helps to compensate for the limits of human rationality. And to avoid any misconceptions, Simon also contends that bounded rationality is not equivalent to irrationality. Objecting to the quality of choices or even the information that informed a decision is not the same as saying irrationality has prevailed. Individual decision makers do have goals and strive to make the best choices possible under the circumstances, which is another way of saying that they are "intendedly rational" (e.g., March & Simon, 1958, p. 170). Finally, Simon reminds us that bounded rationality has intellectual roots in psychological theory, specifically cognitive psychology. To Simon, cognitive psychology has a good appreciation of how individual choice making is limited in its computational abilities and involves a realistic understanding of individual problem-solving processes.

The Legacy of Bounded Rationality

The bounded rationality concept has figured prominently in political science, including influencing Lindblom's incrementalist theory of rationality. Bounded rationality is a robust concept that lends itself readily to multiple meanings and applications, and it continues to play a role in how political scientists frame rationality. To illustrate, Jones (2003) evaluates the contributions of bounded rationality in public administration and public policy scholarship and argues that the bounded rationality approach has yielded an enhanced understanding of how government organizations may produce unexpected or even unpredicted policy or program results. With public organizations not operating under full rationality conditions, administrators aspiring toward rationality may nonetheless find their goals undermined by a variety of forces, such as informational uncertainties and nonrational elements of organizational decision making.

Bounded rationality also plays an important role in Allison's (1971) three decision-making models for studying the Cuban missile crisis: rational policy, organizational process, and bureaucratic politics. The first and second models are most relevant to this chapter. Model 1 (rational policy) is Allison's version of the economic rationality model, with assumptions of advance specification of goals and objectives; identification and evaluation of a range of options; clear-headed knowledge of consequences of decision alternatives, particularly with respect to costs and benefits; and finally selection of the best option from the standpoint of value maximization. This model conceptualizes decision making by the U.S. government as a unified national actor coolly mapping out a set of different alternatives for careful, deliberate evaluation—major options such as doing nothing, diplomatic pressures, a surgical air strike, or a blockade. Model 2 (organizational process) focuses on organizational processes and outputs, seeing U.S. decision making as the result of complex bureaucratic properties. Simon's satisficing concept is evident in Allison's argument that decision making in Model 2 involves "sequential attention to goals" (p. 82). Bounded rationality also is evident in Allison's emphasis on "standard operating procedures" and "programs and repertoires" (p. 83) that coordinate the activities of individuals in government departments and agencies. These latter principles serve as the basis for Allison's much quoted examples of how organizational procedures and constraints may come to shape decision making at the highest levels of a presidential administration. Perhaps the most widely cited rationality example from Allison is Secretary of Defense Robert McNamara's argument for a political and internationally sensitive approach to blockade implementation, as opposed to admiral George Anderson's reluctance to deviate from the Navy's standard operating procedures for blockade placement.

Some scholars, however, have suggested that there may be problems with Allison's application of his decision-making models. To illustrate, Bendor and Hammond (1992) criticize Model 1 as unduly simplistic in its version of rational choice, and they contend that Allison has misinterpreted and misapplied bounded rationality theory. They argue that Allison's version of bounded rationality misinterprets Simon by viewing organizational structure, processes, and routines as a hindrance to quality decision making. Organizational properties such as standard operating procedures really are positive features in Simon's bounded rationality, by facilitating and assisting the decision-making process: In essence, complex challenges and difficult choices require that rationality be boosted through organizational processes, including processes as seemingly mundane as standard operating procedures. Organizations do not limit rationality; they facilitate rationality.

The Foundations of Rational Choice

The roots of modern rational choice theory generally are traced to the seminal contributions of a group of economists—primarily Arrow, Downs, Buchanan and Tullock, and Olson—and one path-breaking political scientist—Riker—through the 1950s to mid-1960s (e.g., see Almond, 1991; Ordeshook, 1990). Some scholars note the early formative role of social or economic philosophers such as Thomas Hobbes and Adam Smith (Monroe, 2001). Kenneth Arrow's (1963) social-choice approach to rationality is a good place to start. First developed in the early 1950s, it has contributed to decades of theory and research on the question of whether individual and collective rationality are inherently in conflict in democratic society. Individual rationality as indicated in expressed preferences might generate problematic collective social choices that lead to serious questioning of the possibility of coupling rationality with democracy—that is, without dictatorship to force choices on people. This puzzle is covered in rational choice investigations of what generally is identified as the *possibility theorem,* or alternatively the *impos sibility theorem.*

Anthony Downs's (1957) *Economic Theory of Democracy* is arguably the most important contribution from someone who is not a political scientist to rational choice in political science. While exploring the meanings of economic and political rationality, Downs presents a theory of rationality in which individuals in political and governmental arenas are guided by self-interest as they pursue choices with the highest levels of utility. The concept of utility figures prominently in economics and is a general way of summarizing the benefits choices bring to decision makers, and the utility concept makes regular appearances in the rational choice literature of political science. To Downs, benefits are not limited to a narrow monetary or financial nature; utility also may be derived from government services such as policing, water purification, and road repairs.

Downs is particularly well-known for his propositions on how self-interested voters assess the appeals of rationally oriented political parties in democratic political systems. These voters may also experience degrees of uncertainty and even information gaps, somewhat similar to what occurs in bounded rationality conditions. Kenneth Shepsle and Mark Bonchek (1997), coauthors of the standard text on rational choice, note the importance of Downs in spatial modeling to show how rational voters evaluate the merits of politicians and electoral candidates in ideological space. Governments themselves figure in Downs's analysis because government officials and political parties seek to maximize support from voters—for example, through spending on government programs or offering programs that appeal to voter self-interest. According to Downs (1957), governments are run by self-interested individuals whose primary concern is not an abstract ideal of social welfare maximization or the public interest; they are oriented toward developing government programs in relation to strategies to please voters.

James Buchanan and Gordon Tullock's (1962) *Calculus of Consent* presents a rationality model in which individuals choose according to the "more rather than less" principle (p. 18). The average individual seeks to maximize utility and secure more of what he or she values—rather than less of it—in the political arena as well as elsewhere. Buchanan and Tullock are particularly interested in the relationship between individual and collective rationality. Although they acknowledge that rationality in market-based decision making does not hold up as well in the governmental setting, they nonetheless argue for applying the logic of economic-based decision making to democratic political systems. Rational members of democratic society will decide in favor of political organizations and institutions that serve their respective individual interests, with competition among individuals also evident in this process. This competition becomes manifest as rational individuals in constitutional democracies pursue more rather than less for themselves in the political arena. Although there may be some slippage from the full rationality standard regarding information levels of individuals and even the extent to which self-interest may dominate, Buchanan and Tullock confidently assert that "each participant in the political process tries, single-mindedly, to further his own interest, at the expense of others if this is necessary" (p. 305). Furthermore, individual choice plays out in an existing constitutional system—for example, the institutions, processes, and rules of representative democracy. In this sense, Buchanan and Tullock embrace a version of bounded rationality in that constitutional democracy also sets the boundaries for political choice.

Mancur Olson's (1965) *Logic of Collective Action* represents a major challenge to traditional thinking on individual participation in groups in democratic society. Rational individuals may not have an incentive to join or participate in large voluntary associations, particularly those characterized as "latent" groups, if they can benefit from the collective or public goods provided by these groups without having to pay dues or incur other costs of membership (pp. 58–59). A key element of Olson's approach to rationality concerns the "objectives" pursued by individuals. Olson pointedly makes the following observation:

> The only requirement is that the behavior of individuals in large groups or organizations of the kind considered should generally be rational, in the sense that their objectives, whether selfish or unselfish, should be pursued by means that are efficient and effective for achieving these objectives. (p. 65)

Rational Choice Arrives in Political Science

William Riker's (1962) *Theory of Political Coalitions* is probably the most important scholarly work in the emergence of rational choice in political science. Riker takes the theories of economics and mathematics-based game theory and expressly applies them to political decision making, presenting an alternative to political science's long-standing focus on concepts such as power and authority. Riker sees rationality in terms of individuals who seek to win, rather than lose, in the context of various types of two-person games: "Politically rational man is the man who would rather win than lose, regardless of the particular stakes" (p. 22).

Whether considering topics such as voting choices or federal system design, Riker (1990) conceives of political rationality as involving actors who are "able to order their alternative goals, values, tastes, and strategies" and who "choose from available alternatives so as to maximize their satisfaction" (p. 172). In Riker we see the fusion of the rational actors of game theory and economics, transposed to the world of politics and government. Riker, however, sees his approach to rationality as transcending traditional arguments over pure economic and bounded rationality. The focus of rational choice theory should be on how individuals decide with information available to them, from knowledge of their own preferences or through the consequences of alternatives themselves. His definition of rationality "requires only that, within the limits of available information about circumstances and consequences, actors choose so as to maximize their satisfaction" (p. 173). Riker became one of the most controversial figures in modern political science, arguing for political science to openly embrace rational choice as its future, particularly because "in contrast to economists, political scientists frequently have been methodically unsophisticated" (p. 178).

Riker's approach to studying politics illustrates prominent features of modern rational choice. First, there is the common use of what may be called the "as if" assumption of rationality to guide empirical analysis and research (e.g., Moe, 1979). Individuals are assumed to act "as if" they decided according to principles such as utility maximization and the pursuit of self-interest (see Riker & Ordeshook, 1968), and then researchers go about the process of testing their propositions and hypotheses against empirical reality. The "as if" approach in rational choice theory has prompted great debate over rational choice's approach to knowledge in the social sciences, with one writer exploring tensions between "instrumentalist empiricism" and "scientific realism" in rational choice scholarship while asking whether the "as if" assumption approach represents a "useful fiction" (MacDonald, 2003).

A second feature is the tendency of rational choice practitioners to work out anomalies or counterevidence from within the rational choice tradition itself—that is, to focus on what some refer to as the maintenance of core elements of the rational choice theory as a way of explaining political reality—even in the face of potentially confounding data or developments (e.g., Shapiro, 2005). To illustrate, Riker and Ordeshook (1968) addressed the puzzle that voting itself might be an irrational act when considering individual costs and benefits; they find that there really is an underlying rational calculus to the decision to vote—or for that matter not to vote.

A third feature of rational choice is its ongoing evolution, as we would expect of any healthy scholarly approach. The rational choice of recent decades is not the same as that of the 1960s and 1970s. In Riker, this is seen in his devotion in the latter part of his career to a scholarly approach labeled *heresthetics,* which focuses on the strategic use of communications, such as sentences and languages, by political leaders and elites in important arenas such as agenda control and coalition formation (Shepsle, 2003).

Understanding Contemporary Rational Choice Theory

Rational choice theory draws from the general approach called *rational actor theory,* which Monroe (1991) identifies as emphasizing individuals who pursue goals and decide among competing alternatives while possessing extensive information, a coherent preference ordering, and a commitment to the principles of self-interest and utility maximization. Rational choice theorists, however, at times differ on how they incorporate these properties into their assumptions and empirical research. A major example is the distinction between *thin* and *thick* rationality. The thin version is the elemental approach to rationality that operates at a fairly broad level, not going much beyond general-purpose assumptions such as characterizing individuals as goal oriented, self-interested, and seeking utility maximization. A thickened version of rationality builds additional specifications into the rationality model—for example, actual belief systems, psychological needs, aspiration levels, cultural values, and even specific goals that may be important in the sociopolitical arena (e.g., see Ferejohn, 1991; Friedman, 1996). Rationality thus becomes richer or more substantive as it is thickened. The importance of understanding this distinction is underlined by Ostrom (2006), who criticizes the tendency in political science to "lump all scholars together who use a thin model of rationality together with those who are developing second- and third-generation behavioral theories" (p. 8).

A few examples from within rational choice scholarship illustrate efforts to broaden its framework and scholarly focus, particularly through the study of institutions. Shepsle and Barry Weingast (1994) assess the transition from the first generation of rational choice congressional research, which fused a behavioral orientation with a strong focus on majority cycles coupled with a relatively abstract notion of the legislature. The second and third generations of rational choice research on Congress, however, shifted toward incorporating institutional structure variables—such as committees, subcommittees, and their rules—along with parties and leadership in the postreform era. Terry Moe (2005) provides a critique from within rational choice that although supportive of the promise of rational choice for political science nonetheless calls for a much more substantial role for political power in rational choice and its study of institutions—in settings that range from the U.S. bureaucracy on through nation-to-nation interactions in international politics.

Richard Feiock (2007) develops a set of hypotheses on regional governance institutions based on what he identifies as a "second-generation model" that incorporates contextual factors that shape and underpin individuals as rational actors. A thin version of rationality is set aside, and contextual factors show how rationality may be bounded—and thus provide an example of integrating bounded rationality into modern rational choice. An excellent example of this synthesis is found in George Tsebilis (1990), who argues that rational choice has unique qualities in its ability to explain behavior of rational actors in the context of political and social institutions that establish the rules of the game in which individuals assess their options and seek utility maximization. Tsebilis's embrace of a rational choice that is bounded by institutional setting is particularly interesting in view of his application of it to comparative political analysis.

To this point, rational choice has been presented in a summative way to introduce the reader to its roots and key influences while providing some sense of its present concerns. It must be noted, however, that any survey of rational choice runs the risk of oversimplification, and the student may be wise to consider the statement by one well-known rational choice practitioner:

> I suspect the only thing all RC [rational choice] people would agree upon is that their explanations presume that individuals behave purposively. Beyond that, every manner of disagree ment theoretical, substantive, methodological can be found. RC is an approach, a general perspective, within which many different models can be located. (Fiorina, 1996, p. 87)

In addition, the undergraduate student with an interest in rationality will encounter multiple references to the public choice, social choice, and rational choice schools, and these terms often are used interchangeably—either accurately or inaccurately (e.g., Friedman, 1996; Monroe, 1991). Within political science, the term *public choice* certainly has definite connotations, primarily due to its association with a well-known political science couple, Elinor and Vincent Ostrom, whose unique and influential versions of rational choice theory and research have been identified by some as the *Bloomington school* (Mitchell, 1988). Illustrative of the sometimes tricky terrain, the term *public choice* may also represent a general ideological orientation to some political scientists who view public choice as having limited application to the discipline. These political scientists contend that public choice is too closely associated with a market-based model that ultimately sees politics and government as hindrances to individual and collective welfare. In sum, rational choice is a multifaceted subject with different schools of thought and even the potential for stirring some emotions.

Rational Choice Controversies

A full understanding of rational choice requires knowledge of controversies associated with this approach in the political science discipline. The decade of the 1990s represents a key turning point, with the emergence of open and occasionally heated debate over the value of rational choice to political science. This decade includes Donald Green and Ian Shapiro's *Pathologies of Rational Choice Theory* (1994) and subsequent scholarly exchanges such as those in *The Rational Choice Controversy: Economic Models of Politics Reconsidered* (Friedman, 1996). A survey of some representative criticisms from this era captures the intensity of this debate:

- Gabriel Almond (1991) asserts that the economic model of rational choice neglects scholarship in disciplines such as sociology, psychology, and anthropology, and its assumptions of human rationality, with their emphasis on utility-maximizing behavior, produce a conception of human rationality that has no "substantive content" and is akin to the Scrabble blank tile that "can take on the value of any letter" (p. 49).

- Green and Shapiro (1994) skewer rational choice as fundamentally flawed, both theoretically and methodologically. Although noting that it has constructed sophisticated formal mathematical models, they contend that the value of rational choice to political science is undermined by a set of deep-seated social scientific pathologies—for example, its theory-driven research with little interest in solving real political questions or problems and its research results that "do little more than restate existing knowledge in rational choice terminology" (p. 6).

- Stephen Walt (1999) criticizes rational choice's growing reliance on formal modeling, highly sophisticated mathematical analysis, and game theory applications, which he sees as not enhancing international security studies—with "rigor mortis" the more likely scholarly result than methodological "rigor."

The rational choice debate carried over into the first decade of the 21st century, though the intensity level of the debate certainly has waned in recent years. The Perestroika movement, which borrowed its name from the reform era of the Soviet Union, probably was the most significant development in the rational choice debate of the past decade. The year 2001 witnessed a multipronged effort by a coalition of disenchanted political scientists to reform the American Political Science Association and redirect political science scholarship in general.

The Perestroikan critics of the political science establishment grouped rational choice with formal modeling and quantitatively oriented research as they made their case against a style of political science perceived as actually diminishing genuine knowledge of government, politics, and policy. Perhaps the most colorful statement to represent the emergent criticism of rational choice is the following call to arms:

> William Riker was fond of saying that political science was a sinking ship, and rational choice theory was the only tugboat that might bring it to port. It is truer to say that Riker's disci ples have acted as pirates out to hijack political science to a rather barren island. Their piracy is doomed to fail. (Kasza, 2001, p. 599)

While the early fervor of the Perestroika heyday eventually dissipated, additional critiques of rational choice later emerged in an edited volume with the colorful title of *Perestroika! The Raucous Rebellion in Political Science* (Monroe, 2005). While rational choice was not by any means the sole object of attention of this volume, rational choice took its lumps from some high-profile political scientists such as Theodore Lowi and Samuel Beer.

Toward Reconciliation

It has not been all slings and arrows over the past decade. A case in point is the assessment provided by a scholar with a well-established record of questioning rational choice and who also has argued for an alternative framework rooted in political psychology—*perspective theory,* which focuses on identity at the individual, group, and societal levels. Kristen Monroe (2001) argues that the discipline "has wasted too much time debating the merits of rational choice theory" and that it is time to focus more fully on asking "what we have learned that may be utilized in the next stage of constructing more realistic theories of political life" (pp. 165–166). Ostrom's (2006) framing of the issue as "Rational Choice—An Evil Approach or a Theory Undergoing Change and Development?" (p. 8) also merits consideration. While embracing the value of rational choice as part of a diverse modern political science and certainly not seeing it as an evil approach, she nonetheless acknowledges, as a rational choice practitioner herself, that factionalism in today's political science may have multiple sources, including rigid adherence to a narrow definition of rationality: "Some of the factionalism does stem from the arrogance of those who consider the continued use of a narrow model of human rationality the essential qualification for doing good social science" (p. 8).

Conclusion and Disciplinary Directions

The past 50-plus years have shown great interest by political scientists in the meaning and applications of rationality. Lindblom's incrementalism ushered in an era of theory and research on the limits of rationality in crafting and choosing public policies, and Wildavsky expanded on incrementalist theory as he made the case for political rationality over economic rationality. Simon's seminal theorizing contributed greatly to knowledge of the realities and parameters of rationality by arguing that there are limits on decision-makers' abilities to acquire and process information and assess options. Rationality is circumscribed or limited, with bounded rationality the condition of individuals as they make important political, policy, and administrative choices. Starting with Riker, rational choice theory elevated the question of whether political actors—from voters on through institutional actors such as political parties, elected officials, government bureaucrats, or even nation-states—are motivated primarily by an economic-based sense of self-interest and utility maximization. Rational choice political scientists answered in the affirmative to this question as they drew from scholars such as Downs, Olson, and Buchanan and Tullock—all of whom cut their academic teeth in the economics discipline. With political scientists such as Riker and the Ostroms laying the foundations, rational choice would become an important force in the discipline.

Alternative conceptions of rationality have spurred debate among political scientists, including expressions of resistance to the notion that politics and government may be understood through the prism of an economics-oriented model of individual and organizational decision making and behavior. Scholars such as Lindblom, Wildavsky, and even Allison questioned the value of seeing policy making and government decision making as tightly structured processes of high-end rationality. Critics of rational choice argued against a political science that reduced the political arena to self-interested, utility maximizing political actors who could be studied through heavily assumption-laden theories and methodologies that make extensive use of formal modeling. Rational choice practitioners, however, have defended their scholarly approach while asserting that rational choice is not a monolithic enterprise, with scholars marching in lockstep. In response to criticisms of early versions of a stripped-down rationality, known alternatively as thin rationality, second- and third-generation versions of rational choice have emerged to incorporate more nuanced and developed understandings of rationality in politics and government—such as adopting bounded rationality assumptions and paying attention to the impact of institutional or cultural variables such as legislative rules and traditions.

Although the dialogue over rational choice has been animated and sometimes heated, it ultimately has been beneficial to modern political science. From the multi-pronged criticisms of rational choice theory, methodology and research voiced by Green and Shapiro in the 1990s on through the sometimes heated debates of the Perestroika movement at the dawn of the new century, political science certainly has indicated a willingness to address fundamental issues and questions. For example, what drives or motivates individuals or government officials to action? Are they fundamentally self-interested? Or are they capable of placing the public interest over personal, economic-oriented calculations of benefit or utility? What of the impact of social-psychological factors such as emotions, values, and identity? Is the political arena best understood as a venue explained by the basic concepts and tools of economics? Just how much information can political actors handle when making a decision—such as whether to vote for a candidate, align with a political party or ideology, express support for a public policy, or evaluate the performance of government officials? All these intriguing questions figure in the study of rationality in political science, and they no doubt will continue to shape future generations of theory development and empirical research.

References and Further Readings

Allison, G. T. (1971). *Essence of decision: Explaining the Cuban missile crisis.* Boston: Little, Brown.

Almond, G. A. (1991). Rational choice theory and the social sciences. In K. R. Monroe (Ed.), *The economic approach to politics: A critical reassessment of the theory of rational action.* New York: HarperCollins.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Bendor, J., & Hammond, T. H. (1992). Rethinking Allison's models. *American Political Science Review, 86,* 301 322.

Buchanan, J. M., & Tullock, G. (1962). *The calculus of consent: Logical foundations of constitutional democracy.* Ann Arbor: University of Michigan Press.

Diesing, P. (1962). *Reason in society: Five types of decisions and their social conditions.* Urbana: University of Illinois Press.

Downs, A. (1957). *An economic theory of democracy.* Boston: Addison Wesley.

Feiock, R. C. (2007). Rational choice and regional governance. *Journal of Urban Affairs, 29,* 47 63.

Ferejohn, J. (1991). Rationality and interpretation: Parliamentary elections in early Stuart England. In K. R. Monroe (Ed.), *The economic approach to politics: A critical reassessment of the theory of rational action* (pp. 279 305). New York: HarperCollins.

Fiorina, M. J. (1996). Rational choice, empirical contributions. In J. Friedman (Ed.), *The rational choice controversy: Economic models of politics reconsidered* (pp. 85 94). New Haven, CT: Yale University Press.

Friedman, J. (1996). Introduction: Economic approaches to politics. In J. Friedman (Ed.), *The rational choice controversy: Economic models of politics reconsidered* (pp. 1 4). New Haven, CT: Yale University Press.

Green, D. P., & Shapiro, I. (1994). *Pathologies of rational choice theory: A critique of applications in political science.* New Haven, CT: Yale University Press.

Jones, B. D. (2003). Bounded rationality and political science: Lessons from public administration and public policy. *Journal of Public Administration Research and Theory, 13,* 395 412.

Kasza, G. (2001). Perestroika: For an ecumenical science of politics. *PS: Political Science and Politics, 34,* 597 600.

Lindblom, C. E. (1959). The science of "muddling through." *Public Administration Review, 19,* 79 88.

MacDonald, P. K. (2003). Useful fiction or miracle maker: The competing epistemological foundations of rational choice theory. *American Political Science Review, 97,* 551 565.

March, J. G., & Simon, H. A. (1958). *Organizations.* New York: Wiley.

Mitchell, W. C. (1988). Virginia, Rochester and Bloomington: Twenty five years of public choice and political science. *Public Choice, 56,* 101 119.

Moe, T. M. (1979). On the scientific status of rational models. *American Political Science Review, 23,* 215 243.

Moe, T. M. (2005). Power and political institutions. *Perspectives on Politics, 3,* 215 233.

Monroe, K. R. (1991). The theory of rational action: Its origins and usefulness for political science. In K. R. Monroe (Ed.), *The economic approach to politics: A critical reassessment of the theory of rational action* (pp. 1 23). New York: HarperCollins.

Monroe, K. R. (2001). Paradigm shift: From rational choice to perspective. *International Political Science Review, 22,* 151 172.

Monroe, K. R. (2005). *Persestroika! The raucous rebellion in political science.* New Haven, CT: Yale University Press.

Olson, M. (1965). *The logic of collective action.* Cambridge, MA: Harvard University Press.

Ordeshook, P. C. (1990). The emerging discipline of political economy. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 9 30). Cambridge, UK: Cambridge University Press.

8 Game theory

Christine Chwaszcza

Introduction

Game theory is a branch of so-called Bayesian[1] rational choice theory (RCT). It has two distinct forms of application:

(i) explaining individuals' behaviour in social settings by their motives and reasons;

(ii) as an abstract model for the analysis of social structure, within the paradigm of methodological individualism (MI).

Game theory is explanatorily useful only to the extent that it models individuals' motives and reasons appropriately. Modelling, by contrast, aims not at replicating the world, but at artificially isolating features in order to study their potential or dynamics.[2] An explanatory approach fails if it cannot explain observable real-life behaviour. An abstract model, by contrast, can be a very fruitful analytical tool exactly when it fails if it is precise enough to tell us *why* it fails, and how the model can be enriched, changed or modified. Insights achieved from abstract modelling do not themselves explain phenomena but can be used in the development of explanatory hypotheses or even concept-formation; but these hypotheses then have to be tested independently.

The first section of this chapter clarifies the basic concepts and assumptions of RCT: rational choice, preference, expected utility and the structure of modern utility theory. The subsequent section turns to game theory proper and remarks on its relationship to the broader concept of RCT. For that purpose, we introduce two concepts of 'equilibrium' – the von Neumann–Morgenstern equilibrium and Nash's concept of equilibrium; and two of the best-studied types of game – the so-called prisoners' dilemma (PD), and a variety of co-ordination games. It is argued that game theory is best employed in the social sciences as an analytical tool. Turning to the more recent development of iterated and evolutionary games, the final section shows how the failure to model co-operation and co-ordination has contributed to a better understanding of those problems.

Bayesian framework of rational choice: basic concepts and assumptions

Game theory is a model for rational decision-making in situations of social interaction. Social interaction, here, is to be understood in Max Weber's sense: as action that involves two or more intentional actors, and that is guided by mutual expectations about how the other person(s) will behave. To the extent that intentional action is guided by reasons and/or rational deliberation, game theory provides a model for an ideal type of reasoning about what to do. In that sense it is not a model for action or behaviour proper, but for *reasoning*.

Originally, game theory was developed as one of three branches of the broader rational choice paradigm: decision theory, social choice theory and game theory.[3] The core idea is a refinement of the everyday concept of means–end reasoning (i.e. that the best means should be chosen to achieve a given end) into a calculus of decision-making that integrates *probabilistic* reasoning (Savage 1954). That refinement was made possible by the development of modern utility theory (MUT). Although game theory is not as closely tied to MUT as other branches of the rational choice paradigm, it was originally developed within that framework by von Neumann and Morgenstern (1944).

MUT was originally developed in applied mathematics for decisions in non-interactive situations characterized by risk. More simply, this means how a single individual would decide, faced with a range of choices whose consequences cannot be predicted with certainty because they depend on other events.

The intuitive idea that motivates modern utility theory is quite commonsensical. In order for a choice among alternative courses of action to be rational, it obviously ought not be guided by wishful thinking: choosing the course of action that yields your most preferred consequences, if everything goes well. Yet prudence – even in an ordinary sense – requires that we consider not only the desirability of each consequence, but also the likelihood of its occurrence, given the presence of external events. The basic idea of RCT says that one should choose the course of action that maximizes one's *expected utility*, that is, the overall sum of all positive and negative consequences of a course of action, weighed with the probability of their occurrence.

Given that probability estimates are commonly given in numerical terms, weighing the desirability of a consequence with the probability of its occurrence is informative only if desirability, too, can be expressed in numerical terms – or, more precisely, if 'desirabilities' can be measured along a cardinal scale that also provides information about how much one consequence is desired over another. These cardinal measures are usually called 'utilities'; modern utility theory defines the (formal) conditions under which it is possible to assign numerical values to desirabilities, thereby constructing utility measures.

The first step is to define the relevant properties of the problem. As an axiomatic theory, RCT is strictly defined by the terms and conditions specified in its axiomatic foundations. No concept or assumption not defined in the axioms, nor derivable from them, can be expressed within the theory. Given that decisions are only required where alternatives are open, a decision situation is defined by (i) the set of all feasible options, and (ii) the set of all possible events that might influence the consequences (outcomes) of one's action, where it is assumed that consequences can be specified for all possible combinations and evaluated by the deciding agent by means of pairwise comparisons.

These pairwise comparisons represent the preferences of an agent, that is, a relationship between two alternatives, A and B, such that one is ranked above the other. The concept is taken to be primitive and is not meant to represent some specific evaluative attitudes, such as egoistic, altruistic or hedonistic values, or a specific ideal of the good life. Most commonly, preferences of agents are considered to be empirically given, or to be given by the assumptions of the model. In economics this is often maximization of profits or monetary payoffs, but it need not be.

It is assumed that an agent can rank all possible consequences according to their desirability, that is, ordinally from best to worst. If that ordering fulfils certain requirements of consistency, it can be proved that there exists a mathematical function to rank preferences over consequences in a cardinal ordering. That function is commonly called a utility function. In modern utility theory, the definitional set is given by the ordinal ordering of preferences over consequences, while the set of values is the set of rational numbers. The two most important consistency requirements are completeness (that is, all pairs of alternatives can be ranked) and transitivity (that is, if I prefer A over B and B over C, then I must prefer A over C); further requirements concern mathematical properties and the applicability of rules of probability calculus.[4]

Given a cardinal ordering and the assignment of numerical measures, it is now possible to weigh the utility of each consequence with the probability of its occurrence, and to determine the expected utility for each course of action in a way that allows for a meaningful comparison of all alternatives open to an agent. We can now define the expected utility of each course of action as the sum of the utility of each of its possible consequences weighted by the probability of its occurrence. We can then select the one course of action with the highest expected utility.

Maximizing expected utility is the criterion recommended for rational choice in *decision theory* (we will qualify this for game theory below). The rational choice concept of *rationality* is primarily defined by the consistency requirements that must be met in order to construct a utility function. The criterion of maximizing expected utility is an extension of the common-sense concept of means–ends rationality for decision-making under risk. The contribution of decision theory for the clarification of means-end rationality consists in the specification of the conditions that must be fulfilled to reason or act in accordance with that criterion.

Accordingly, the model of reasoning in RCT must be characterized as a *logical* model of reasoning. It is definitely not a *psychological* account, but a formal account that specifies the ideal conditions under which a specific account of reasoning, maximization of expected utility, yields well-defined solutions.

It will not be necessary to go into the details of the axioms to recognize that conditions in RCT are highly technical and quite demanding; obviously, people's everyday practice of probability reasoning rarely involves mathematical probability calculus. But even completeness (all pairs of consequences can be compared) and transitivity are far from trivial requirements if one considers complex situations where evaluations include multiple perspectives and dimensions (Kahneman and Tversky 1981).

This causes no worries for mathematicians or economists. They seek a formal presentation of how to construct a utility function that suffices as a (mathematically) meaningful interpretation of such a function. They are interested neither in utilities – or preferences – *per se* nor in real-life decision-making.

Yet the technical nature of the conditions of consistency and the construction of a utility function required by the model do not necessarily meet the expectations and requirements of social scientists, who are interested in explaining the behaviour of persons in real-life situations. Average persons do not engage in probability estimates that would meet the standards of mathematical probability calculus (Allais 1953). Also the very idea that persons ought to aim at maximization of expected utility was criticized as too demanding by Simon (1982), who suggested a more modest model of imperfect instrumental rationality that aimed at a level of 'satisficing' rather than maximization. The first wave of critical objections to MUT was not that the concept of rationality employed was too narrow, but that it was too demanding.

The second point to emphasize is that the implicit account of evaluation employed in MUT is purely *consequentialist* – that is to say, outcome-oriented – and instrumental. Consequentialism seems to be an innocent assumption within the context of means-end reasoning, and when decisions are not considered to affect other persons. But it comes with two important implications:

(i) It implies that preferences are neutral as to moral or social descriptions of alternative courses of action – for example, whether an action conforms to social or moral norms or violates them;

(ii) Consequentialism is *strictly forward-looking*.[5] Notoriously, consequentialism cannot provide rational explanations for actions that are reactions to events in the past – such as actions of other persons or past commitments and promises – or are derived from norms, based on habits, and so on (Hollis and Sugden 1993; Nida-Ruemlin 1993; Zintl 2001).[6]

Non-consequentialist aspects are often decisive in the processes of reasoning and decision-making for real-life persons, but given the way in which the axiomatic theory is structured, these aspects cannot be integrated into the framework without major changes. Some theorists say, 'that's fine', because they do not consider means-end reasoning to be the only form of practical rationality, but simply one among others. Others are not concerned because they think these other aspects are irrational. But consequentialism then implies serious constraints on the general applicability of the model. It fits only specific types of choice, namely those where consequences are the unique – or at least the most important – aspects of evaluation.

These two points seem to be the most important shortcomings of rational choice theory in the social sciences. Whereas probabilistic reasoning plays a lesser role in game theory, the logic of consequentialism is the same.

Rationality in interaction: the search for equilibria

Game theory is connected to modern utility theory through the assumption that agents choose a course of action they expect will have the best consequences given the alternatives available. It recognizes, however, that straightforward maximization of expected utility is not a rational option in situations that are characterized by social interaction.

The criterion for rational choice in game theory is to aim at an *equilibrium point*. There exist different concepts of equilibrium points, not all of them identical to the maximization of expected utility. Yet all of them are strictly consequentialist. Game theory concerns rational decision-making in situations where the consequences of one's course of action are partly determined by one's own decision, and partly by the decisions of the persons with whom one interacts.

The challenge of social interaction arises because agents must base their choices on mutual expectations about how the other will decide. Since the second person's decision depends upon what she thinks the first person will choose, the first person has to base her choice on the expectation of how the second person will react to what she thinks will be the choice of the first person, and so on.

The mutual dependency of choices raises the threat that agents end up in an infinite regress or circular expectations about expectations. There is no way in which agents can make a choice that deserves to be called rational – as opposed to arbitrary – unless they can identify a rational stopping point at which the reflection about mutual expectations can end. The challenge for rationality here is not one of *maximization*, but of *stability*: to arrive at a choice to which one can stick even if the other person knows how one is going to decide. This is the problem which the concept of equilibrium answers.

Aiming at an equilibrium point can coincide with choosing an action that maximizes one's subjective preference satisfaction, but it need not. The so-called minimax theorem[7] proved by John von Neumann and Oskar Morgenstern, which originally started game theory, says that all two-person constant-sum games have an equilibrium point that guarantees the players a maximal minimum payoff and minimal maximum loss, respectively, if mixed strategies are accepted.

Constant-sum games are by definition characterized so that the gain of one person equals the loss of the other – the game is *strictly conflictive*. A mixed strategy is given by a probability distribution over all the (pure) strategies available to an agent. It selects the strategy to be acted upon by using, for example, a random device for deciding among the available courses of action. If, for example, an agent can do either X or Y and has the mixed strategy of choosing X with a probability of 2/3 and Y with a probability of 1/3, he might throw a die and perform X if 1, 2, 3 or 4 is obtained, and perform Y if 5 or 6 shows. In principle, each possible probability distribution over the set of available strategies is a mixed strategy. Rational actors are supposed to choose a mixed strategy that minimizes losses or maximizes gains. Unfortunately, the minimax theorem turns out to have a rather restricted scope.

The minimax theorem proves that for all two-person constant-sum games, there exists at least one combination of mixed strategies for the players such that if the same game were played a sufficiently high number of times, playing the mixed strategy would minimize the maximal loss and maximize the minimal gain of the players; and if there exists more than one such combination of mixed strategies, all resulting equilibria would be equivalent. The assumption, of course, is not that the game will in fact be repeated a high number of times, but that one should chose as if that would be the case, even though the game is played only once.

The concept of rationality employed in the minimax theorem is a variation of Laplace's principle of insufficient reason: if one does not have a good reason for thinking that one belief is more likely to be true than another, one should regard each as equally likely to be true. (See Neurath (1913) for a similar maxim of practical reasoning.)

Such reasoning is unlikely to be accepted as a rational method of deliberation outside academic classrooms. Even more mathematically minded theorists seem to have some doubts, if only because situations of strict conflict – as modelled by two-person constant-sum games – do not occur very frequently. Most situations of social interaction are so-called *mixed-motive games* – that is, situations where the gains of one player do not equal the losses of another, because, for instance, both can win or lose. Alternatives to the von Neumann–Morgenstern equilibrium of mixed strategies exist, and they are not only much easier to determine, but much less psychologically demanding. The concept of equilibrium that is most widely accepted in game theory is Nash's concept,[8] which says that one should choose the best counter-strategy to what one expects the other person(s)' choice will be. Note that the concept of Nash equilibrium is defined relative to the actual choice of one's co-player.

Nash's concept of an equilibrium point has the significant advantage of offering a rational criterion that can be applied even to games where only an ordinal ranking of preferences over outcomes is given. As the prisoners' dilemma shows, however, Nash equilibria do not necessarily select the course of action that maximizes preference satisfaction of the agents.

> *Game 1: Prisoners' dilemma (PD)*
> Two suspects are taken into custody and separated. The district attorney is certain that they are guilty of a specific crime, but he does not have adequate evidence to convict them at a trial. He points out to each prisoner that each has two alternatives: to confess to the crime the police are sure they have done, or not to confess. If they both do not confess, then the district attorney states he will book them on some very minor

Table 8.1. Game 1: Prisoners' dilemma (1)

Peter, Paul	Not confess (Co-operate (C))	Confess (Defect (D))
Not confess (Co-operate (C))	3, 3	1, 4
Confess (Defect (D))	4, 1	2, 2

Note: Here and in the following 4 > 3 > 2 > 1 always.

> trumped-up charge such as petty larceny and illegal possession of a weapon, and they will both receive minor punishments; if they both confess they will be prosecuted, but he will recommend less than the most severe sentence; but if one confesses and the other does not, the confessor will receive the lenient treatment for turning state's evidence whereas the latter will get 'the book' slapped at him (Hargreaves Heap, Hollis, Lyons *et al.* 1992: 99).

The payoffs obtained by each of the two prisoners, Peter and Paul, are shown in Table 8.1.

An alternative standard presentation (here showing the consequences for Peter) is displayed in Table 8.2.

As can easily be seen, each agent would be better off if both chose C rather than D because (C, C) > (D, D) for each of them. At the same time, each risks unilateral disadvantage if he or she commits him/herself to choose C, because the outcome (C, D) is worse than any other option. Since game theory – like modern utility theory – is strictly consequentialist, each agent must expect that the other's evaluation of the feasible courses of action is exclusively based on the consequences they will experience in the given situation. Neither of them, therefore, can expect that anybody would choose C if he expects the other to choose C, because (D, C) > (C, C) for each of them. Consequently, each knows that the choice of C is not rational for either of them under any circumstances, which makes D the dominant strategy[9] and (D, D) the unique equilibrium point of the game.

A common reaction to the dilemma is that it models a problem for egoists or persons tempted by self-interest. That reaction, however, rests on a misunderstanding, because the dilemma results from the structural properties of the game, not from any supposed theory of motivation. The structure of the game as given in the payoffs represents the preferences of the agents. It therefore does not make sense to ask whether altruists would 'prefer' C over D or (C, C) over (D, C), because *if* altruism versus egoism has any role to play in the evaluation, it is already reflected in the ranking of alternatives.

Table 8.2. Game 1: Prisoners' dilemma (2)

Peter, Paul	Co-operate (C)	Defect (D)
Co-operate (C)	R = reward	S = sucker
Defect (D)	T = temptation	P = punishment

A more sensible question to ask would be: can the prisoners' dilemma situation *occur* among non-selfish agents? That question, of course, is primarily an empirical one. To the extent that we consider real-life agents to be characterized by a mixed motivational structure that includes altruistic as well as selfish attitudes, the answer seems to be 'yes'. Such agents would resemble the average human being we know, and it seems that such agents find themselves in situations that structurally resemble the prisoners' dilemma. If not the two-person prisoners' dilemma, then at least the *N*-person prisoners' dilemma – also referred to as the 'Tragedy of the commons' (Hardin 1968) – seems to represent a rather common structural situation of social life.

> *Tragedy of the commons*
> The commons is a pasture open to all herdsmen of a village. Each herdsman can keep some of his cattle on the commons, the rest on his own land, and each herdsman can increase his herd by increasing the number of cattle sent to the commons. If each herdsman does so, the commons will be overgrazed.

This example has been applied to many real-life situations that require collective action or concern the provision and maintenance of public goods (see, for example, Olson 1971; Taylor 1987; Ostrom 1990)

Interestingly, real-life agents often do not end up at the Pareto-suboptimal equilibrium point, but actually co-operate – not only in daily life, but also in experimental settings (Rapoport and Chammah 1965).

Another assumption about what goes wrong in the model identifies consequentialism as the problem. An intuitive answer to why co-operation is successful in real-life environments is the existence of (coercive) institutions and (moral) norms or practices, such as promises or contracts that support and facilitate co-operation and overcome the constraints of rational individualism. This institutional solution, however, can only be integrated into the theoretical framework if its establishment and maintenance can be shown to be an equilibrium. (This question played an important role in the development of iterated and evolutionary games, which will be considered in a later section.)

In the simple one-shot game (Game 1), it can be easily shown that reference to attitudes of norm-obedience is unconvincing because of the consequentialist structure of the basic model. Assume that Paul *promises* Peter to choose C. Would that give Peter a 'reason' – compatible with the assumptions of modern utility theory – to choose C likewise? The answer of rational choice theorists is no. There are two reasons why not, a simple one and a more sophisticated one. The simple answer is that given Paul's promise, Peter would be tempted to exploit him – which, of course, can be foreseen by Paul and gives him an incentive to break his promise in the first place, which can be foreseen by Peter who consequently does not trust Paul's promise. Although both would be better off if they had the institution of promising, neither has a rational incentive to comply with it. The structure of the prisoners' dilemma repeats itself on the level of compliance (or enforcement) of institutions.

The more complicated answer points to the problem that consequentialism leaves no space for reasons or motives that derive from commitments (obligations) made in the past – such as a promise. Although such commitments are reciprocally advantageous, they cannot be introduced into the model because of the consequentialist structure of evaluation. An alternative path to take is to introduce more complex strategies such as 'co-operate with other co-operator', 'defect when meeting a defector'; but that changes the structure of the game: the PD becomes a *co-ordination game* (see Game 5 below).

The limits of consequentialism are most obvious in settings of social interaction, but can be equally observed in rational choice analysis of the political decisions of individuals. Consider, for example, Downs' (1957) economic theory of democracy. According to Zintl (2001) it provides an analytical test for assessing the limits and scope of conceptualizations of democracy as elite competition for votes – or, as one might say more generally, the *Homo economicus* model. Downs' ideal economic model of democracy analyses voting behaviour as utility-maximizing and party behaviour as competition for votes in order to maximize positions for party members. The assumption, famously, leads directly to the *voter's paradox* – the conclusion that voting is irrational. Given the minimal influence of each single vote, the costs of casting one's vote outweigh the potential gain to be received from it. Therefore, utility-maximizers should abstain. Although the ideal theory articulates only a foil against which Downs develops hypotheses about the role and significance of *prima facie* irrational attitudes (such as adherence to ideologies), neither the ideal nor the non-ideal model offers an escape from the voter's paradox. Although it is not obvious what follows, it definitely indicates the limits not

Table 8.3. Game 2: Traffic

Ann, Rosalind	Drive on the left-hand side	Drive on the right-hand side
Drive on the left-hand side	2, 2	0, 0
Drive on the right-hand side	0, 0	2, 2

Table 8.4. Game 3: Social trap

Jules, Jim	Meeting at the restaurant (A)	Meeting at the library (B)
Meeting at the restaurant (A)	2, 2	0, 0
Meeting at the library (B)	0, 0	1, 1

only of utility-maximization, but more generally of *consequentialist* reasoning within the explanation of socio-political behaviour.[10]

A second and different problem of identifying rational choice with the pursuit of equilibrium points is that in many types of game, more than one equilibrium exists; game theory does not indicate which one to choose. Such situations are commonly called *co-ordination problems* and are usually taken to model self-enforcing conventions (Lewis 1969). A standard co-ordination game is the following:

> *Game 2: Traffic*
> Two drivers, Ann and Rosalind, can drive either on the right-hand or on the left-hand side. Neither has a specific preference for one side over the other, but both prefer to drive on the same side of the road in order to avoid collisions.

The payoffs for this game are shown in Table 8.3. Game theory does not offer a well-defined solution for the problem, because neither Ann nor Rosalind has a basis for deciding independently on which side of the road to drive.

More intensely studied are co-ordination problems with several unequivalent equilibria, such as the following:

> *Game 3: Social trap*
> Two persons, Jules and Jim, plan to meet. Two meeting points are possible, the restaurant and the library, and both prefer to meet at the restaurant.

Game 3 (Table 8.4) has two Nash equilibria in (A, A) and (B, B) with (A, A) > (B, B) for each agent. As Hollis and Sugden (1993) show, however, neither agent has reason to choose A because that would be 'rational' only if

Table 8.5. Game 4: Battle of the sexes

Harry, Sally	Meet at the boxing match	Meet at the ballet
Meet at the boxing match	4, 3	2, 2
Meet at the ballet	1, 1	4, 3

he or she could expect the other also to choose A, and vice versa; but understood as the best counter-strategy to the other agent's choice, (B, B) is as rational a choice as (A, A). The concept of Nash equilibrium gives no reason to prefer one over the other. Intuitively, one would like to say that rational agents naturally choose the equilibrium that is better for all participants. But such a move is not part of the concept of Nash equilibrium, defined as the best counter-strategy to the other player's *actual* choice. In addition, such a move would be of limited help in co-ordination problems such as game 4 (see below and Table 8.5), where the two equilibria yield (4, 3) and (3, 4), favouring Harry in one case and Sally in the other.[11] It could therefore not replace the concept of Nash equilibrium, but would just define an additional concept and thereby repeat the co-ordination problem on a higher level, since it is only rational to adopt such a concept of rational choice if the other person does likewise.[12]

> *Game 4: Battle of the sexes*
> Harry and Sally have the overriding aim of spending the evening together, but Harry wants them to go to a boxing match, whereas Sally prefers that they see the ballet, each according to his or her personal preference for entertainment. They have no possibility to communicate their meeting point, but mutually know their preferences.

The intricacy of co-ordination problems has been extensively discussed by Schelling (1960) in *The Strategy of Conflict*, which included experimental settings with real persons. In the light of the empirical results, Schelling concluded that some equilibria somehow 'stand out' in the sense that they seem to 'have a special meaning' that made participants of the experiments select them. Schelling introduced the term *salience* to characterize the quality of standing out. But he also explicitly stated that salience cannot be adequately expressed within the theoretical framework of rational choice theory because it seems to presuppose a shared semantic practice. The point is far from trivial. Schelling implied that game theory is discontinuous with the Bayesian framework of rational choice theory (Schelling 1960; Spohn 1982).

Other theorists go even further, raising the question of whether the relevance of a shared semantic practice defies the project of methodological individualism (MI), one of the major assumptions of interest in game theory, because *salience* implies a form of holism – a 'common understanding' or 'meaning holism' (Hollis and Sugden 1993). Although meaning holism is a basic and fundamental prerequisite for any form of communication and reasonable interaction, it does not support any specific social ontology. It seems, therefore, insufficient to decide the debate between proponents of MI and holism; but it definitely increases the burden of arguments on the MI side.

'Too bad for the theory!' one might say. And so it may be if one expects game theory and rational choice to provide a straightforward explanatory approach for rational behaviour. The fact that game theory advises choosing the suboptimal equilibrium in situations of the prisoners' dilemma type has indeed been widely celebrated as a self-defeating result of the rational choice concept of rationality. The limits of rational choice detected in co-ordination problems, however, must be considered even more devastating in their implication that the notion of rational choice is ill-defined – that is to say, it does not provide a unique solution – for a rather significant number of games. As Hollis and Sugden (1993) remark, game theory, thus far, has failed to give us an adequate account of how two persons who meet each other in a narrow corridor should choose what to do.[13]

Taking stock

To return to the beginning: a judgement about the usefulness of game theory and rational choice theory in general depends not only on the *explanatory capacity* of the theory, but on the *use that is made of it.* Failures can be very instructive, if they allow for a precise diagnosis and theoretical improvement that goes beyond commonsensical objection or mere dogmatic opposition. They are most instructive when used for analytical purposes.

We can now come back to the possible uses and applications of game theory in the social sciences. The attractiveness of game theory for social theorists derives from a variety of motivations. The three most common seem to be the following:

(i) To the extent that rational choice theory was considered to provide an explanatory approach, one point of attraction seems to have been the expectation that it offers an alternative to behaviourism by opening up the 'black box' of the human mind (Monroe 2001). To the extent that causal explanations of agency are considered to require indications about (regular) psychological mechanisms (Hedstrom and Swedberg 1998), accounts of decision-making and reasoning are obviously attractive (see Héritier, ch. 4).

As a model of a specific account of reasoning, however, game theory competes with other approaches that also aim at explaining human behaviour by motives and reasons but endorse different accounts of 'practical rationality,' 'practical deliberation' or 'reasons for action'. The economic account of means-end rationality and the model of *Homo economicus* have sometimes been used unmodified as an ideal type for explanatory purposes[14] in both political science and sociology. More often, however, they are treated as ideal types and used as a device for the development of alternative and more realistic accounts for behavioural explanations.[15] Their results have also been transformed in explanatory accounts of institutional development and change as in Scharpf (1993), Aoki (2001), Congleton and Swedenborg (2006) and Héritier (2007). Evolutionary game theory, however, partly departs from the commitment to methodological individualism (MI).

(ii) On a more abstract level, game theory was welcomed as an agency-oriented approach by proponents of *methodological individualism* as an alternative to structuralist and functionalist approaches in social sciences. As Osborne and Rubinstein (1994: 1) remark, the models of game theory provide 'highly abstract representations of classes of real-life situations'. These models have been widely used for the analysis of the structure and the dynamic development of macro-phenomena such as institutions, norms and conventions in sociology (Coleman 1990), and in political theory in both its analytical and normative branches, especially social contract theory (Ullmann-Margalit 1977; Axelrod 1984; Taylor 1987; Bicchieri 2006). Since their attractiveness lies in their abstractness, these studies usually work with purely formal models.

(iii) Given its precise axiomatic foundations, the rational choice paradigm was appreciated as a path for the development of '*positive* (political) theory' – or rather *theorizing* – in the social sciences. Its success as a methodology obviously depends on the extent to which game theory allows us to derive explanatory models and hypotheses that are not only falsifiable, but also have the advantage of indicating rather precisely *where* and *why* they go wrong (Riker and Ordeshook 1973; Riker 1997). Although the precise axiomatic foundations of the rational choice paradigm do not entirely exclude controversial interpretations of the shortcomings of its models, it has indeed turned out to offer a fruitful method for the continuous development of research questions and – together with the development of statistics and computerized modelling – also improved models.[16]

In both political science and sociology, game theory has mainly been used as an analytical tool for theory-building, not as a straightforward account for explanation of individual behaviour or specific events.[17] As Zintl (2001) observes, there are two major areas of application in political science. The first is the analysis of institutional and social structures at a level where the motives or reasons of the individuals who constitute them are irrelevant – for instance, because the phenomena under consideration are macro-phenomena constituted by the actions of large numbers of persons with many different attitudes or reasons. Examples are phenomena such as general norms, social conventions or traditional practices – or in the analysis of social or institutional settings, where individual motives can be considered to be determined by structural aspects of the environment in which persons interact.

The second and most promising application of game theory, however, is on the level of conceptualization, the exposition of the problem or puzzles that one wants to study, and the construction of explanatory hypotheses. Zintl (2001) calls such applications 'sophisticated', contrasting them with straightforward endorsement of *Homo economicus* as a model for behavioural explanation, which he calls 'naïve'.

A classical example of the sophisticated application of game theory to explain political behaviour is probably Riker's *Theory of Political Coalitions* (1962). Starting from the assumption that the formation of minimal coalitions is the ideal rational choice for parties that try to optimize positions for their members, the frequency of non-minimal coalitions has set a research puzzle for more focused investigation of motives and incentives in coalition-building.

The major field of application for game theory, however, has been the analysis of institutions. Since no single article can give a satisfying picture of the scope of applications, and since studies in game theory are driven by problems, not by applications, the remainder of the chapter will focus on the most important analytical developments connected to prisoners' dilemma games.[18]

The final judgement on the usefulness of game theory, of course, will have to be made by the reader. But in order to provide some guidance, the final section will outline some of the more recent developments of game theory.

The use of game-theoretic models for analytical purposes

The prisoners' dilemma game is probably the most widely studied model in game theory, exactly because its game-theoretic solution is counterintuitive. Interestingly, although real-life situations seem to fit the structure of the prisoners' dilemma, co-operation is rather common in real life. One important reason seems to be that in real life, decision-making is facilitated by the existence of social and cognitive resources that support co-operation in PD cases.

The attempt to get a clearer picture of what those resources are has driven further analytical development.

The expectation is that to the extent that models can be modified, changed and revised, their study will reveal the conditions that must be satisfied for certain solutions to be possible or stable. The interest that drives the research is not so much the desire to make the model approximate reality – or to make reality compatible with the model – but rather the development of hypothetical scenarios that clarify the dynamics, structures and conditions of the stability or instability of certain forms of social structures. The more variations we get, the more information we receive. If, for example, the original model of single-shot game theory is developed into models of meta-games, iterated games and evolutionary games, the primary insight we can get from those variations concerns the conditions that facilitate or hinder the development of certain social structures, understood as patterns of individual behaviour.

In the remainder of this chapter, I will point to three results and further developments in game theory connected to the discussion of PD and co-ordination games:[19] the norm approach, which involves the transformation of the PD game into the so-called 'assurance' game (AG), also called the 'stag hunt' game; an interesting result from iterated PD games concerning group size; and some tools used in evolutionary game theory (EGT).

Norm-oriented reasons and the challenge of reciprocity

The first criticism of RCT has often taken the following form: its failure to offer an account of co-operative behaviour consistent with the basic assumptions of modern utility theory must be due to a bias in favour of egoism. Once we assume that personal preferences present not only egoistic concerns, but social – or moral – attitudes, the structure of interaction characterized by the PD does not occur. Instead, rational agents are confronted with a problem of reciprocity: the choice is not simply one between (a) to co-operate and (b) to defect, but between strategies or maxims for behaviour such as (a') co-operate with persons who are also willing to co-operate, and (b') defect if you encounter a person who is herself a defector. Such maxims can be called meta-strategies. A game that models the new interpretation is the so-called 'stag hunt' game (Table 8.6), named after a famous passage in Rousseau.

Game 5: The stag hunt

Two hunters can either jointly hunt a stag (an adult deer and a rather large meal) or individually hunt a rabbit (tasty, but substantially less

Table 8.6. Game 5: Stag hunt (assurance)

Peter, Paul	Stag hunt (C’)	Rabbit hunt (D’)
Stag hunt (C’)	3, 3	0, 2
Rabbit hunt (D’)	2, 0	1, 1

filling). Hunting stags is quite challenging and requires mutual co-operation. If either hunts a stag alone, the chance of success is minimal. Hunting stags is collectively most beneficial but requires a lot of trust among the hunters. It is a co-ordination game with two equilibria at (C’, C’) and (D’, D’), reciprocal co-operation (C’, C’) being Pareto-superior.

Co-ordination games are no less theoretically problematic than PD games. Rational actors have no incentive to co-operate with rabbit hunters, and given the fact that stag-hunting results in an equilibrium only if one stag-hunter meets another, the stag hunt game has no obvious solution in the terms of RCT. The problem of reciprocal co-operation as posed by the stag hunt game consists in (a) identifying co-operators and defectors, and (b) co-ordinating co-operators so that they interact with each other.

Unfortunately, neither problem can be solved with the theoretical resources offered by classical game theory. Nevertheless, the criticism moved the discussion a significant step forward. It made clear that the prisoners’ dilemma is less a problem of egoistic motivation that can be overcome by making persons more moral, than a *cognitive* one.

Iterated games – the challenge of free-riding

Another attempt to overcome the dilemma of the PD developed from the consideration that gains from repeated co-operation outweigh continuous mutual defection, and can even compensate for sporadic exploitation if reciprocal co-operation occurs frequently enough. This attempt remains within the consequentialist (outcome-oriented) structure of the broader rational choice paradigm, but it enriched the model by introducing a future orientation by through allowing for iteration and learning from experience. The latter development was made possible by (a) developing iterated games and (b) programming strategies that based decision-making on information about the outcomes of the previous round. Famously, Axelrod organized computer-based round-robin tournaments for PD games that were run with strategies sent in by the professional and non-professional publics. The tournaments modelled interaction between strategies for the iterated PD game, not between agents, and conducted iterated rounds of bilateral encounters. Some of these strategies were exploitative, some were co-operative, and the winning one – 'tit for tat' – played a strategy of reciprocity: 'tit for tat' always co-operates in the first move and then plays the strategy that was chosen by its partner in the previous round.

The interest in iterated games and evolutionary games concerns not so much questions of choice or strategic logic, but the conditions under which certain results or strategies can be achieved or expected to be stable. Accordingly, the attraction of the study of these games consists in identifying relevant parameters and modifying them in order to study their effects.

Axelrod (1984) summarizes a few general results. The tournament revealed that the success of co-operative strategies depends heavily upon their strategic environment; also, there is no single equilibrium, and several equilibria are possible. Although unconditional defection is always an equilibrium, co-operative equilibria can also occur under certain conditions, but only when co-operation is conditional on being reciprocated and when defection is punished. Unconditional co-operation encourages exploitative strategies. The strategy that received the highest average payoff, tit for tat, has been derived from empirical experiments conducted by Rapoport and Chammah (1965).

Although Axelrod summarized the results of his tournament optimistically as 'evolution of cooperation among egoists', his results are rather limited because the tournament consisted of repeated and aggregated bilateral encounters of each strategy with every other strategy over several rounds. The much more interesting case for the study of co-operative structures, and/or *general* social norms of co-operation, would have to be a genuinely *N*-person variation of the prisoners' dilemma game that is commonly used for modelling the provision of public goods (Hardin 1985; Taylor 1987).

The striking difference between the two-person case and the *N*-person case is that the payoffs are completely determined by the interaction between the two strategies in the first case, whereas in the second case they depend also on the degree of co-operation of those players with whom one does not interact.[20] This difference in the structure generates a serious free-rider problem in the iterated game and actually an incentive to boycott co-operation. Such games were construed and analysed by Taylor (1987), who found that the selection of a co-operative equilibrium in iterated *N*-person PD games is not excluded, but that the conditions under which it can occur are so strict that it is highly unlikely that they will ever be realized in practice.

As a side effect, Axelrod and Taylor's study of iterated PD games sheds light on an assumption that has been held by quite a number of sociological theorists, namely, that group size can make a difference and that duals, bilateral relations, follow a quite different dynamic from multilateral forms of interaction. Generally, the problem that game theory cannot isolate clear-cut equilibrium solutions for all games has resurfaced in the study of iterated games.

Evolutionary games – the instability of co-operation

Evolutionary game theory (EGT) studies the conditions under which pre-programmed strategies can become stable patterns of behaviour. EGT is primarily interested in the frequency of specific strategies within a population searching over time for dynamic equilibria. That allows one to analyse also the mutual influence or dependency that holds between individuals and the social environment. EGT has developed an impressive range of variations for both strategies and the construction of different social environments. Evolutionary simulations, for example, have used strategies that are capable of 'learning' or 'signalling'; others vary 'environmental' settings such as spatial locations of strategies, i.e. isolated or in clusters, and forms of encounter, which range from random combinations over the construction of 'neighbourhoods' to mechanisms for selecting partners.

Most interesting for the social sciences are two apparent motivations for the study of EGT: (i) the hope that it provides a better understanding of agency and the development of rationality in social (strategic) interaction; and (ii) the hope of arriving at a better understanding of the role of collective agencies (institutions) and the efficiency of spontaneous versus constructed orders.[21]

An important step towards evolutionary models was taken by the biologists Maynard Smith and Price (1973), who developed the concept of an *evolutionarily stable strategy* (ESS). Maynard Smith and Price were interested in the dynamics of selection of behavioural patterns within groups of individuals. The puzzle they addressed concerns the robustness of behavioural patterns against individuals or groups of invaders. A 'hawk–dove' game, which structurally resembles the chicken game (see note 11), is used in order to specify the conditions under which a population of doves can survive the invasion of hawks, and vice versa. For that purpose, an evolutionarily stable strategy is characterized by two properties that are familiar from the concept of a Nash equilibrium: (i) it is the best response to itself, and (ii) it is the best response to any other strategy in the environment.

The concept of ESS was also used by Axelrod (1984) for an evolutionary simulation of PD situations, which supported the result already achieved in iterated games: that (unconditional) co-operation is not an evolutionarily stable strategy. Although unconditional defection is always an ESS, conditional co-operation (following the logic of tit for tat) can also be stable in specific environments.

Another tool used in evolutionary modelling is so-called replicator dynamics. Replicator dynamics model strategy change in iterated games by changing the frequency of strategies within a given population in the following way: a strategy that does better than average increases in frequency at the expense of strategies that do worse than average.[22] The main interest in those studies again concerns the effect of the modifications of parameters, which is difficult to summarize. Two general results from the study of replicator dynamics in various games (chicken, hawk–dove, PD, stag hunt), however, seem to be as follows:

(i) Whereas equilibria for ESS are always also Nash-equilibria, there can exist equilibria in replicator dynamics that are not Nash-equilibria (Taylor and Jonker 1978).

(ii) Under certain conditions, replicator dynamics result in co-operative equilibria.

The latter point is especially strong in models that study reciprocal co-operation, such as the stag hunt game. A quite accessible presentation of the results of increasingly rich modulations of the stag hunt game is offered by Skyrms (2004), who also discusses their relevance for social science.

Paying tribute to the importance of contingencies in biological evolution, some models introduce random mutation (also called noise) in order to study the influence of contingent disturbances for dynamic equilibria. Equilibria that are resistant to small perturbations (noise) are often called asymptotically stable.

The results from evolutionary game theory show clearly that both the enrichment of cognitive resources (learning, signalling) and spatial closeness increase the likelihood of stable reciprocal co-operation. So far, it seems that the results do not indicate that institutional orders provide better mechanisms for equilibrium selection than do spontaneous orders, or vice versa.

A warning might, however, be appropriate. All strategies used in evolutionary game theory are algorithms that can model the behaviour of human beings as well as of bacteria or robots; that includes mechanisms of 'learning', which so far have been varieties of learning by reinforcement or imitation. Nevertheless, the EGT approach represents an agency-oriented approach, because social structures are perceived as being constituted by individuals' patterns of behaviour. Regarding the agent–structure debate and the MI paradigm, however, the results of EGT seem strongly to support the thesis that a mutual dependency between individual strategies and social environment exists, and that structures not only constrain individual behaviour, but also provide motives for agency (Hargreaves Heap and Varoufakis 2004: 264).

With EGT, in fact, we leave the paradigm of Bayesian RCT behind us. For evolutionary game theory does not model the choices of agents, but the success of different strategies for choice under varying circumstances by using algorithms. Algorithms, obviously, are quite different from agents, not only because of the lack of psychological properties, but also because they are in a sense deterministic. They are pre-programmed, even if they can learn. Thus, the later stages of dynamic evolutionary models are far removed from the original model of modern utility theory. Although it might be an open question whether algorithms that determine the behaviour of bacteria will provide us with insights into patterns of human behaviour, which can neither be affirmed nor excluded *a priori*, such algorithms provide an illustrative example of how theory development can proceed.

For further theoretical studies, however, one result seems especially crucial. Evolutionary games strongly indicate that the basic assumptions of the rational choice concept of rational agency have to be revised. If the social environment provides not only constraints, but also reasons for agency, basic assumptions of Bayesian rational choice theory have to be changed. As Hargreaves Heap and Varoufakis (2004: 264) conclude: 'The learning model, directed as it is instrumentally to payoffs, may be more realistic but it is not enough to lead unambiguously to some equilibrium outcome. Instead, if we are to explain actual outcomes, individuals must be socially and historically located in a way that they are not in the instrumental model. "Social" means quite simply that *individuals have to be studied within the context of social relations within which they live and which generate specific norms.*' (See Keating, ch. 6, and Steinmo, ch. 7.)

At the present stage, it is not easy to assess whether and what the social scientist can learn from EGT. But it certainly will reshape the scholarly debate, if not about human agency, then about Bayesian RCT.

NOTES

[1] This chapter will not consider non-Bayesian approaches.

[2] The distinction between the two applications is sometimes blurred because individual motives and reasons are often considered to be given by assumption, or to be irrelevant because the objects of study are large-*N*-person settings, or taken to be determined by the properties of the social setting under investigation.

[3] For a comprehensive selection of major contributions to all branches of RCT, see Allingham

(2006). Public choice theory can be considered to articulate a game-theoretic alternative to social choice theory (Buchanan and Tullock 1965; Mueller 1989).

4 For reasons of space, these conditions cannot be specified here. The most accessible presentation is still Luce and Raiffa (1957: ch. 2).

5 It is, however, neither necessarily amoral nor egoistic; utilitarianism is consequentialist too.

6 The view has been held that positive and negative evaluations of the course of action can be integrated if we consider 'psychological costs' that accompany the performance of a specific course of action, such as buying rather than stealing. The preference over owning a good if ownership is brought about by theft, and the preference over owning it if ownership is brought about by legal transfer from another person, need not be the same. Such a move is certainly possible in principle, but against the logical spirit of the model.

7 The minimax theorem is, as the name says, a theoretical proposition that can be proved. It should not be confused with the so-called minimax criterion for decision-making under uncertainty; for further clarification, see Luce and Raiffa (1957) or any other coursebook for decision and game theory.

8 An excellent and updated introduction to game theory is Hargreaves Heap, Hollis, Lyons *et al.* (1992).

9 A dominant strategy is a strategy that has better consequences than any other strategy available for all possible courses of events or strategies chosen by another agent.

10 For criticism and further development, see for example, Tsebelis (1990) and Brennan and Hamlin (2000).

11 The so-called 'chicken' game, which has been widely used as a model for threatening, is an even more intricate co-ordination problem: 'Two adolescents, Dean and Brando, decide to resolve a dispute by riding towards each other down the middle of a road. The first to turn away loses. If both continue straight ahead, they will crash and risk serious injury' (Hargreaves Heap, Hollis, Lyons *et al.* 1992: 106). The payoffs are shown in the following table.

Dean, Brando	Hold straight	Give way
Hold straight	0, 0	4, 1
Give way	1, 4	2, 2

12 For an exhaustive discussion, see Hollis and Sugden (1993).

13 For a solid and informed discussion of shortcomings of the rational choice concept of rationality, see Green and Shapiro (1994) and Friedman (1996).

14 For a general critique, see Sen (1977).

15 See Simon's account of 'bounded rationality' (Simon 1982) and Elster's studies on irrationality, preference change and the 'subversion of rationality' (Elster 1979, 1983, 2000). For applications of game theory in sociology, cf. Abel (1991), part III.

16 Some of the advanced models in evolutionary game theory even seem to come as close to experimental settings as non-natural sciences can be expected to come (Skyrms 2004).

17 An exception is bargaining theory, which seems to constitute a practice of interaction fit for the application of economic models if – or as long as – the questions at stake can be considered not to be exceptional. But obviously, bargaining is guided not only by logical strategies of choice, but also by psychological aspects; the more important the latter is considered

to be, the less reliable rational choice models become. For analysis and applications of game theory to problems of bargaining and negotiation, see for example, Brams (1990), Brams and Taylor (1996) and Raiffa and Richardson and Metcalfe (2002). For a criticism of the psychological shortfalls of rational choice theory, see Mercer (2005).

[18] It has to said, though, that this development was also supported by the improvement of computer technologies.

[19] A fourth development, psychological games, goes beyond the scope of the present chapter; interested readers are referred to Hargreaves Heap and Varoufakis (2004: ch. 7).

[20] One might think of the problem of building a dam to protect a small island against a flood. If the dam can be built in time by eighteen persons, and there are twenty-five persons living on the island, then seven of them can refrain from co-operating without defying the co-operative gains of the other eighteen.

[21] A more theory-immanent interest, of course, concerned the problem of selection of equilibrium points.

[22] The standard model was developed by P. Taylor and Jonker (1978). An easily accessible presentation is given in Hargreaves Heap and Varoufakis (2004: ch. 6); for a more formal presentation, see Weibull (1995).