The Rationalist Tradition and Political Science*

Jörgen Hermansson

The rational actor is a central tool for theorising in social science. Moreover, it is the distinctive feature of rational choice theory or, more generally, the rationalist approach, which today stands out as probably the most important tradition within positive social theory as well as moral and social philosophy. However, the rationalist tradition is in no way unchallenged, and apart from economics it has nowhere reached a position resembling hegemony. Political science and political philosophy, the disciplines of my own, may be said to represent the typical situation: rational choice theory executes its power by offering a common language suitable for many different topics. My ambition in this essay is to develop and elucidate this point.

Without going into technical details, I am trying to state my argument concerning the strength and weakness of rational choice theory as precisely as possible. Although I draw heavily on the extensive literature in many different disciplines,[1] my own pretensions are more restricted. I confine myself to political theory, which is the only discipline where I can claim to be something like a professional.

To begin with I will briefly introduce two of the leading and to some extent competing schools within the rationalist tradition. Next, I devote a section to a presentation of the central concepts of rational choice and to some remarks on the relevance of game theory for that approach. In the subsequent three sections I try to set out some of the limitations and shortcomings of rational choice theory. At first I will analyse some problems at the conceptual level. I then proceed to a discussion of rationality from an empirical point of view, i.e. how does rational choice theory manage as a positive political theory? And thirdly, I will offer an analysis of some normative aspects of rationality. The last section of the essay reflects my concerns for the future of rational choice theory.

*This essay summarises the main ideas of *Spelteorins nytta. Om rationalitet i vetenskap och politik* (The payoff of game theory. On rationality in science and politics, Acta Universitatis Upsaliensis. Skrifter utgivna av Statsvetenskapliga föreningen i Uppsala, 109). My work was pursued with the support from the Swedish Research Council for the Humanities and the Social Sciences (HSFR).

Towards hegemony for rational choice?

Rational choice theory or, more vaguely, the rationalist tradition has contributed a lot of new findings, which in the light of previous thinking appear as non-obvious and in some cases even counter-intuitive. The research on the track of Kenneth Arrow's impossibility theorem may be mentioned as an example.[2] To be sure, already in the eighteenth century Marquis de Condorcet clarified the paradox of voting and the preconditions for cyclical majorities, but not until Arrow these findings have emerged as a problem for the theorists of democracy.[3] It is due to Arrow and to modern game theory that we are able to discover the origins of rational choice theory in classics such as Hobbes, Hume and Rousseau. Moreover, as a result we are better equipped to scrutinise and disclose ambiguities and flaws in their argumentation.[4]

However, it is not necessarily the research results as such that should be emphasised. Fruitful scientific and philosophical ideas do not always give us new knowledge directly. Often a theoretical breakthrough consists of a new idea that helps us to put new questions and to clarify the old and well-known problems. That was the case with Mancur Olson's *The Logic of Collective Action* – a most powerful, although in some important respects faulty, analysis of the problem of voluntary co-operation, which has generated many interesting studies on the possibility of individual actors acting together as a collectivity. And in these studies the analytical tools of game theory have been indispensable.

When mentioning Mancur Olson it is common to assign him, together with James Buchanan and Anthony Downs, the role of a key representative for the public choice-school. The judgement and the characteristic of Olson's *The Logic of Collective Action* apply more or less to the entire school of thought. The same may be said about an even younger tradition, viz. rational choice Marxism,[5] the existence of which is due partly to the seminal work of Olson.

Rational choice Marxism and public choice resemble each other in several respects. Most important, they both apply rational choice theory with great methodological rigour to topics previously reserved for quite different approaches. However the dissimilarities are also obvious, though less important in this connection. They choose very different problems; indeed both schools almost follow a predictable ideological pattern. Perhaps there is also a dissimilarity between their ideological style. While the thesis of the public choice-school always have the expected ideological tendency, the lack of respect for the ideological heritage of its own tradition is characteristic of rational choice Marxism. The heresies of Jon Elster or John Roemer might be compared with something like James Buchanan arguing for the necessity of a radical redistribution every tenth year in order to realise the ideal of the market.

The very fact that modern rational choice theory is identified with James Buchanan as well as Jon Elster and John Roemer, who differ a lot with respect to their topical orientation and sources of inspiration (i.e. context of discovery as opposed to context of justification) may be seen as an illustration of both its strength and its limitation. On the one hand, it is a great advantage of rational choice theory that it is fairly neutral at least in the ideological terms of the West. It makes some demands on method and it forces the researcher to focus in certain directions, but that apart it is an apparatus that can be used by most social scientists and philosophers.[6] Because of this catholic position rational choice theory has enforced as well as paved the way for better communication between theorists of different schools. On the other hand, if almost every social scientist and theorist can approve the requirements imposed by rational choice theory, that also is a sign of its weakness or, at least, its limitation. There is a price to be paid for its catholic position. In substantive matters rational choice theory is marked by a lack of plainness and a poverty of content.

The emergency of the public choice-school may be seen as a reaction to this state. By postulating utility-maximizers but also egotists, public choice-theorists, as previously Hobbes, are able to reach more precise and, not to say provocable conclusions. However, at the same time, these theorists introduce an element into the theory, which others claim to be unnecessary and unfortunate, since it can be empirically refuted and since it appears to be controversial on normative grounds. The latter development imply a violation of an important rule for rational argumentation, viz. the parties ought to strive for consensus in the starting-points of the discussion. Of course, it is a good point if a debater is able to question ingrained opinions, but the controversial element of a theory should be a conclusion reached from arguments based on premises everyone has consented. Here a simple comparison with rational choice Marxism can be made. It deserves our interest, since it claims that Marxists should adopt rational choice theory. If its proposal were that advocates of rational choice should embrace Marxism, it would be considered a joke.

Not everyone will agree that lack of plainness and precision concerning empirical predictions and normative prescriptions should count as a weakness. However, as will be made clear later on, rational choice theory has other properties that definitely belong to the category of imperfections.

The notion of rational man and the significance of game theory

Rational man appears almost everywhere in the social sciences and philosophy and consequently is a well-known character for all of us. In spite of that there is some confusion about his/her distinguishing features. This confusion rests upon a dispersed and vague dissatisfaction with the substantial poverty of individual rationality. For example Amartya Sen points to the fact that it is possible to be "a rational fool".[7] Now and then, theorists have developed models of human agency that improve upon the capacity of rational man. However, throughout this essay I will stick to, and also to some extent argue for, a conventional and minimal concept of rationality, which basically calls for consistency and nothing more. Although, they deserve a lot more attention, I devote just a few remarks to the alternative conceptions.

Rational man is sometimes mistaken for two related but dissimilar notions: economic man and moral man. Roughly speaking, in both cases too many

qualities are ascribed to rational man.[8] The same may be said about a fourth variation – judicious man would be a fitting name – as launched by Jon Elster. Starting from the discontent with the empty formalism of rational man, Elster proposes a "broad theory of individual rationality" as a supplement to the conventional "thin theory". The broad theory requires more than mere consistency, although it does not demand either omniscience or goodness. Elster distinguishes judgement and autonomy as the characteristics of judicious man.[9]

However, rationality is essentially a matter of consistency. For an economist a rational actor is defined as an actor whose revealed preferences, i.e. choices, are transitive. This extremely minimal or descriptive conception of rationality is by no means restricted to human beings and it does not presuppose intentionality or agency.[10] It requires nothing but a consistent pattern. Hence it is an empirical problem whether the behaviour of animals or the resultant outcomes of collective action fall into such a pattern. For the purpose of model-building economists usually add completeness and continuity, viz. it is possible to represent preferences that are transitive, complete and continuous through a real-valued utility function. Moreover, if the preferences are assumed to be selfish, it may be said that we have all the characteristics of economic man.

A related and more interesting conception of rationality emphasises the prescriptive aspect of rational action, thereby accentuating the close ties between rationality and optimality. The general idea is that rational choice theory tells us what to do in order to attain our goals. In this context, it is common to give prominence to the idea that rationality only states a conditional or hypothetical imperative, i.e. it refers to the choice of means irrespective of the ends. To be rational is to choose the best possible means towards one's end.

In order to be truly useful for scientific purposes, it is important to stress the subjective nature of rationality. Among different perceived options, the actor chooses that alternative which he believes to be most adequate to fulfil his own desire. As an implication, the actor's reasons, i.e. his subjective rationality, may explain the chosen course of action. Such a rationalist explanation is an important sub-specie of intentional explanations.

To apply the notion of rationality may in some circumstances be straightforward. The most simple case will be to make a decision when all the parameters that can affect the outcome of different choices are known. However, situations where decision-makers operate under certainty are rare or of little interest. It is more likely that a social scientist focus on situations where the choices involve some risk and uncertainty. Unfortunately, the notion of rational choice will then be less simple to apply, although the general idea still holds.[11]

The main concern in this essay will be situations in which the actors by nature lack control over the outcome of their actions. I will concentrate on strategic situations, i.e. choice situations where the environment of the actor consists of other actors. To choose rationally in such a situation each actor has to anticipate the choices of other actors, i.e. he has to act strategically. And of course, each actor also has to take the strategic choices of the other into consideration. This idea promises to be quite complicated. It seems that we have to consider se-

		Y – column player		
		Stone	*Scissors*	*Paper*
X – raw player	*Stone*	1 , 1	2 , 0	0 , 2
	Scissors	0 , 2	1 , 1	2 , 0
	Paper	2 , 0	0 , 2	1 , 1

Figure 1: The Stone – Scissors – Paper Game.

quences of anticipations, but fortunately our situation is not that bad. Game theory makes it possible to analyse this kind of strategic interaction.[12]

Most children without any knowledge of the technicalities of game theory have grasped its essence or at least some of the important features of a game through a lot of practice. They know for example that playing games of pure conflict – or more generally, inessential games – soon may become rather boring. An extreme case is the well known "stone, scissors and paper" game (stone defeats scissors, which defeat paper, which defeats stone) where the players can not do better than picking randomly either of its three alternatives (figure 1).[13]

However, in a typical pure conflict game with two actors the notion of rational choice has a more clear-cut meaning. The best an actor can do is to play safe; i.e. when choosing a prudent strategy the actor selects that alternative which maximises the minimum payoff. Hence, this kind of strategy is known as maximin and the corresponding outcome is named a saddle-point.[14]

The analysis of games including some incentives for co-operation is far more interesting and sometimes also less straightforward. The extreme opposite of a pure conflict is a game of pure co-ordination (see figure 2a). If the actors in such a game lack the opportunity to communicate, there is no obvious final outcome, i.e. solution.[15] Both actors have a common interest. They want to co-ordinate their choices in order to reach one of the Pareto optimal outcomes (the outcomes with payoffs (2,2) in the upper-left and the lower-right corner of the matrix), but the best they can do is to arrange separate lotteries which single out strategy A or strategy B at random.

The matrix in Figure 2b illustrates a slightly different case. The game of generosity depicts a co-ordination problem, but it resembles the well-known Battle of the Sexes[16] in that the actors have separate favourite outcomes. Hence, the game of generosity exemplifies a choice situation in which the actors have mixed motives. In this specific case the actors have a common interest in reaching one of the two co-operative outcomes (A*A and B*B), but they are also in conflict as their opinions diverge on which of these is to be preferred.

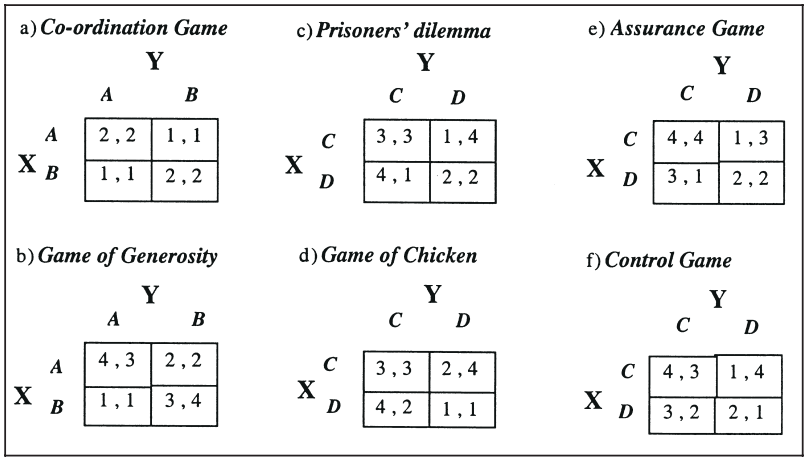

Figure 2. Six different two-person games.

The matrix in Figure 2c is an instance of the famous Prisoners' Dilemma.[17] Both actors have a dominant strategy, i.e. an alternative that is the best choice regardless of how others choose. Prisoners' Dilemma is peculiar in this respect: In choosing their respective dominant strategies – the course of action required for by individual rationality – the actors bring about an outcome (D*D) which according to the Pareto criterion is inferior to another feasible outcome (C*C). Since Pareto optimality may be seen as a specification – perhaps incomplete – of collective rationality, this case proves that there is no straight connection between individual and collective rationality. Hence, Prisoners' Dilemma illustrates, it is said, that there is a visible punch that is the counterpart of the invisible hand. Moreover, if the actors reach the co-operative outcome (C*C) – no matter how – this situation will not last long because they both have an incentive to defect, i.e. to switch from C to D.[18] As a matter of fact, the only stable outcome is D*D. Apart from being sub-optimal, this outcome is an equilibrium point, which means that the corresponding strategies are optimal against each other, i.e. the actors can not do better than sticking to the D-strategy as long as the other does.

In the remaining payoff matrices – the Game of Chicken, the Assurance Game and the Control Game[19] – both actors lack a dominant strategy. In Chicken (Figure 2d) there are two equilibria, viz. (4,2) and (2,4), but as in Prisoners' Dilemma the co-operative outcome is unstable. The Assurance Game (Figure 2e) also includes two equilibria and since one of them, the co-operative outcome (C*C), is preferred to any other outcome by both actors, it seems reasonable that the actors – at least if they have complete information about the choice situation – will reach the superior equilibrium point. Finally, the Control Game (Figure 2f) is characterised by entire instability. There is no equilibrium at all. No matter what outcome is taken as a starting-point, there is always someone who has an incentive to deviate.

Although these games may seem to be too simple in their structure, this presentation will still do as a survey of the interaction problems and perverse interdependencies analysed by game theorists. Some of the games are almost obligatory in every introductory textbook in game theory. That is not without reason: These games may serve well as models for some of the most fundamental problems in the history of political ideas. Just to mention a few examples: It is common to associate Thomas Hobbes' political theory with Prisoners' Dilemma, different games of co-ordination remind us of David Hume's ideas about a just society and the Assurance Game offers a telling picture of Rousseau's views on human co-operation.[20]

I will end this section by introducing a discussion on a minor conceptual problem, thereby laying the basis for the argument in the last section. The issue is whether the notion of collective rationality is at all meaningful except in a transferred sense referring to collective actors. As we all know, it is usual, at least in economics, to identify rationality with efficiency defined as Pareto optimality. By contrast, in a political context it almost seems suspect to use words like collective rationality. Hence, it is not at all obvious that there is a corresponding political notion of collective rationality.[21]

However, when raising empirical questions concerning the rationality of politics it is rather something about politics as a collective activity than the actions of individual politicians that interests us. In everyday speech we usually allude to objectivity and endeavour to find constructive solutions when stating that politics function rationally. Certainly, it is possible to link this to a means-end calculation associated with the conception of individual rationality. But primarily, this usage of language no doubt serves the purpose of transferring the positive value of rationality to a political style, which we appreciate exactly because of its objective and constructive approach. To be suitable as a scientific tool it is necessary to give the notion of political rationality a distinct and precise meaning, which is related to the original concept of rationality but still not coincident.

Here, political science can not offer any established terminological convention. However, an apparently reasonable and fruitful proposal has been launched by Jon Elster in *Sour Grapes*. As his point of departure Elster takes market-failures or collective irrationality in the economic sense, i.e. situations in which individually rational actions produce a sub-optimal outcome. The political notion of collective rationality then refers to the capability to bridge difficulties that obstruct voluntary co-operation.

> The economic notion of collective rationality implies that people, by individually rational actions, bring about an outcome that is good for all, or at least not bad for all. Failure of such collective rationality may occur in one of the three ways just described: by isolation, by perverse interaction structures, and by lack of information. Elsewhere I have referred to such failures as 'social contradictions'. The political notion of collective rationality implies that people by concerted action are able to overcome these contradictions.[22]

Thus, politics should – if it claims to be rational – help us to escape or avoid perverse interaction structures such as Prisoners' Dilemma or the Control Game.

I will return to this political conception of collective rationality, but before then some problems of rational action will be discussed.

The limits of rationality

Before asking if individuals (and possibly collectivities) do in fact act rationally and if this – true or not – always is desirable, rational choice theorists have to face a more fundamental question: is it always and in principle possible to tell what actions rationality prescribes? This is a theoretical or conceptual problem and it concerns the aspirations for generality of rational choice. Unfortunately, the literature is full of negative evidence on this issue. There are a lot of situations where rational choice theory fails to give us any answer to the very question it pretends to handle, viz. What is the best thing to do under these circumstances? Thus, rational choice theory is a failure in so far its ambition is to offer a general social theory.

I want to elaborate the conceptual aspects of this statement, while postponing further conclusions about its actual scope and capacity. As my point of departure I take the conceptualisation of rationality proposed by Felix Oppenheim in his *Political Concepts: A Reconstruction*.

Broadly speaking, an action x done by actor A in situation S will qualify as rational if, in the light of the information available to A in S, it is an optimal means to the attainment of some ultimate goal of his.[23]

I will pay attention to two different problems related to the application of this kind of definition to situations of strategic interaction.[24]

First, the shortcomings of rational choice theory manifest itself in non-cooperative game theory as the problem of how to develop an entirely satisfactory solution concept. As opposed to co-operative game theory, which may be said to study games from the point of view of the collectivity, we have to avoid the notion of a binding agreement. In relying only on individual rationality, the stability of a solution has then to be derived from some kind of self-enforcement. A couple of proposals will here be mentioned: solution in the sense of Nash and Luce and Raiffa's solution in the strict sense.[25]

According to John Nash, a game has a solution if and only if there is at least one equilibrium point and all equilibrium points (if there are many) are interchangeable.[26] Applying this solution concept to the six different games introduced in the previous section gives us only one game with a solution: In Prisoners' Dilemma there is only one equilibrium and by definition all its equilibrium points are interchangeable. By contrast, the Control Game has no solution, since there is no equilibrium point at all in that game. And in the Assurance Game, Chicken, the Generosity Game and the Game of pure co-ordination there are two equilibria, but in none of these games the two equilibria are interchangeable (though in the co-ordination game they are equivalent).

The proposal of Duncan Luce and Howard Raiffa was to single out Pareto optimal equilibrium points which are both interchangeable and equivalent.[27] At first this seems promising: the Assurance Game becomes a game with a solution in the strict sense. But in return, Prisoners' Dilemma is not solvable in the strict sense, since its equilibrium point (D*D) is sub-optimal.

The general idea behind these and other proposals is to link two elements to each other; viz. there is an equilibrium point (i.e. a stable outcome) which is the outcome of rationally chosen strategies. Elster makes the same point when he defines "the solution to a game as an equilibrium point towards which all the agents will tacitly converge".[28] Ignoring the debate on competing equilibrium concepts it may be said that the difficulty comes where we have to state what will count as the right signal to the actors on how to choose. It may then be said that a non-cooperative game has a solution if and only if the following stipulations are satisfied:

(1) There is at least one equilibrium point.

(2) There is some kind of focal point or strategic key, which tells the actors how to choose rationally.

(3) The resultant outcome of actions performed in accordance to (2) will be one of those equilibrium points referred to in (1).

Obviously, the second stipulation in this definition is rather vague and something more precise has to be said about what feature or characteristic of a game may function as a strategic key, i.e. something that tells the actor what is the best thing to do in the actual situation. However, the most I can offer is an extensive, but probably not exhaustive, list:

1. It is sometimes possible to find a univocal outcome by successive elimination of dominated strategies. This presupposes that each player is aware of everybody's preferences.

- A special case is that each actor has a dominant strategy (e.g. Prisoners' Dilemma). That is the only characteristic, which serves as a focal point in situations where the actors lack information about each other's preferences.

2. If the actors recognise that the situation is one of an inessential game, i.e. a game where there is nothing to gain by co-operation, then it is rational to adopt a prudent strategy (i.e. maximin).

- An important special case of this situation (but also of 1 above) is a two-person zero sum game with a saddle point.

3. In some games there are many equilibrium points, but one of them Pareto dominates the other (e.g. the Assurance Game).

4. In some games there are many equilibrium points, but they are all equivalent and interchangeable.

5. In games of co-ordination, i.e. where there are several Pareto optimal equilibrium points which are not interchangeable (they may but need not be equivalent), there are sometimes characteristics in the preference structure

which may serve as a focal point, i.e. a ”Schelling point”. However, if it is possible for the actors to communicate, they may correlate their strategy choices by constructing a common lottery, i.e. they construct a random signal that selects one of the equilibrium points. The signal to the actors may be understood as a non-binding suggestion to choose a certain strategy.[29] In this case it may be said that the actors themselves take the responsibility for the existence of a focal point.

None of these characteristics will help us to single out a solution in games such as Chicken or the Control Game. In a two-person Chicken game there are two Pareto optimal equilibrium points, but they are neither equivalent nor interchangeable. Consequently, the actors lack the strategic key that can help them to act rationally. The Control Game points to a different problem. It fails already on the first criterion, since there is no equilibria at all in that game.[30] Hence, the length of the list does not matter. If we stick to a reasonable solution concept, there will always be games without a solution.

The indeterminacy in some situations of strategic interaction may be seen as a failure of rational choice theory. It puts a definite limit on the applicability of the notion of rational action. However, this does not mean that the rationalistic approach has nothing to provide to the analysis of such situations. Of course, rational choice theory can not explain or predict the chosen courses of action or the final outcome in games without solution, but in spite of that it may be argued that the theory is able to shed light upon that which is of real interest in situations such as Chicken and the Control Game. In such situations it helps us understand why there is indeterminacy, i.e. it may explain and predict a random empirical pattern.

Apart from this conceptual problem it should be stressed that many applications of non-cooperative game theory have brought to light another weakness of its solution concept. It is in a way nothing but a practical problem for the researcher. However, it is nevertheless a disastrous problem, since it strongly reduces the applicability of game theory as an analytical tool. A typical application of non-cooperative game theory contains very few (as a rule only two) or very many (n-person games) actors. In n-person games the trick is to model the decision problem on the two person case, i.e. any actor plays against all the others. By contrast, stringent game theoretical analyses of strategic situations with five or six actors involved (most typical for a multi-party system) are rare. As Adam Przeworski has noted, this is not a coincidence. His diagnosis is worth quoting – though his implicit recommendation is somewhat ambiguous; he seems to suggest more empirical wariness as well as more mathematical sophistication.

> The technical apparatus of game theory is almost unusable for all situations that involve fewer than many but more than two actors.
>
> - - -
>
> ... and while the solution concepts remain the same for any number of players, technical difficulties become formidable when the number of strategic actors exceeds two. I am thus not particularly optimistic about the future of formal applications of game theory to the study of intergroup conflicts and, in particular, to the study of class alliances.
>
> What is apparent is that the formal game-theoretic analyses will remain unpersuasive unless the concepts of equilibria they employ are descriptive of particular historical conditions. I suspect, therefore, that methodological individualism will force us to be more cautious and more explicit in analysing strategic situations, especially class alliances and the relations among social groups and state apparatuses, but I do not see much future for formal deductive analyses of this kind, at least not before game theory crawls out of its mathematical infancy.[31]

The second conceptual problem concerns ”situation S” in Oppenheim’s definition. A simple and almost trivial observation may serve as our point of departure: most rational choice analyses use as an implicit assumption that the circumstances are given beforehand and that the situation is fixed and stable. For game theory analysts this notion corresponds to a methodological rule: When applying game theory it is important to depict a well-defined opportunity set, i.e. a set of alternative strategies. This is a prerequisite for a fruitful study. It seems that rational choice theory is not suitable for dynamic analyses, where the choice situation itself is the subject of analysis. As soon as we allow the actors to change the situation S, i.e. this possibility is an option within the opportunity set, and thus influences their possibility to act rationally in the future, problems arise which rational choice theory can not handle.

An example frequently used in this connection is the kind of difficulty which occurs if the actors are faced with the option that they may search for further information in order to get a more reliable basis for their choice. The relevant criticism of rationality in terms of optimality is then well known. It is impossible to answer questions about the rationality of searching for more information or about the optimal level of information. To find out an answer to these questions you have to know beforehand what information you can get by further research, and in that case the very reason for the original question disappears.[32] Once again rational choice theory, as in situations like Chicken and the Control Game, suffers to be at a loss for an answer.

More generally, it seems that game theory does not permit actors to reflect simultaneously upon alternatives within the frame of a fixed strategic situation and alternatives that amount to a change in the prerequisites for strategic action. To place these different types of alternatives on a par appears within game theory to be a kind of category mistake. It creates problems resembling those illustrated by the classical ’liar paradox’ and Bertrand Russell’s famous set of all elements which are not members of themselves.[33]

I take it that these two examples of conceptual problems are sufficient to show that the limitations of rational choice theory are indeed serious. Hence, it is not at all strange if one questions the prevalent occupation with the notion of rationality in contemporary social science and philosophy. But there is, on the other hand, a rather simple rejoinder to this reaction: The alternatives available to us are hardly superior. It is rather just the contrary. And as long as this prevails, most of us will stick to the rationalist approach, because ”one can’t beat something with nothing”.[34]

The empirical challenges to rational choice

To ask if individuals and collectives act rationally is, of course, to put an important empirical question. But it does not seem reasonable to settle this question once and for all. Evidently, we sometimes fail even though we attempt to be rational.[35] And it is also obvious that at other times we live up to our potential as rational creatures. The proper question is to ask if and to what extent the concept of rationality is fruitful in theorising about societal matters.

With respect to this latter issue, i.e. the scope of rational choice theory, two different but related views are frequently stated. First, it is common to point to the flaws of rational choice theory, which stem from the fact that it takes preferences as given. There are a few examples, often quoted, of theories including endogenous change of preferences, but these examples may be seen as exceptions that confirm the rule. As a consequence, rational choice theory is ill-suited to any social and political phenomenon that involves changing or shaping of people's desires and beliefs, e.g. an ordinary decision process in a parliamentary democracy. Second, it is often said that rational choice theory is essentially a static theory and it is therefore not capable of taking history into account. At most rational choice theory may help us to formulate some formal models of social change.

These criticisms, which I think are largely accurate, appear today as conventional wisdom.[36] Although both of these two critical points may be questioned as to the details,[37] I will instead turn to a slightly different discussion. Its point of departure is more modest than the ambition to develop a general social theory. I will restrict myself to a position where rational choice theory is conceived more or less as a heuristic device, and I will say something about its strength and weakness in relation to this position.

Rational choice theory is identified with more or less formal models. A typical rational choice theorist produces far-reaching idealisations of reality. The models of game theory make this poverty of substantial content explicit. As was announced in the first section, this may be a strength as well as a weakness. In this section I will consider some examples of theorising taken from political science. I begin with John D. Steinbruner's attack on rational choice in *The Cybernetic Theory of Decision*. By introducing Anthony Downs' *An Economic Theory of Democracy*, and Mancur Olson's *The Rise and Decline of Nations* I then want to reveal some of the problems, which the formalism produces.

The subject matter of my first example is the decision theory proposed by John D. Steinbruner as an alternative to the rational choice approach.[38] Steinbruner assisted Graham T. Allison in writing the famous *Essence of Decision: Explaining the Cuban Missile Crisis*. His own book *The Cybernetic Theory of Decision* was announced as a fourth model in addition to those presented by Allison. However, for several reasons Steinbruner fails in his attack on rational choice and his book does not by any means outshine Allison's.

It can be argued that Steinbruner misses the point when he chooses a different set of empirical data than Allison. Instead of the Cuban missile crisis, Steinbruner's case is the negotiations within NATO during the years 1956–1964 concerning the future organization of the alliance. More importantly, it appears on a closer look that Steinbruner has incorporated one of Allison's original models into his own, viz. the organizational process paradigm which takes the idea of bounded rationality as its starting-point. Moreover, but less important, the style of Steinbruner cannot match Allison's. Nevertheless, Steinbruner's book is an interesting challenge, since his purpose is to present a non-rationalistic decision theory, which claims to be superior to theories of the rationalistic approach.

The point of departure for Steinbruner is an analysis of political decision-making of today. His conclusion is that it has reached such complexity that the rational choice approach is doomed to fail because of its lack of realism. Instead, he proposes that we should adopt a cybernetic theory, which essentially is a combination of James March' and Herbert Simon's theory about administrative behaviour and some kind of psychological theory stressing the cognitive constraints on the individual.

With respect to substantial matters, there are two empirical problems that are blended in Steinbruner's book. To begin with, he wants to answer why the negotiations ended in the status quo, i.e. NATO survived without any organisational changes. This form of military co-ordination Steinbruner calls a multinational force (MNF) and is characterised by the fact that the different countries preserve their sovereignty. According to the MNF-doctrine there should be a co-ordination, but each individual country is in the end responsible for their own national forces. However, most of the interest is devoted to a related but still different question: How to explain that the United States took several different stands during the process of negotiation? Before the United States realised that it was impossible to change NATO and accordingly got stuck with the MNF-alternative, the American administration had launched two other proposals, both of which aimed at strengthening the position of the NATO-organisation against its members. According to one of the proposals the alliance should establish a common force with a single command. Power would be centralised in a supra-national organisation over which the members shared the responsibility: the so-called multilateral force proposal (MLF). Besides that the US Defence Secretary Robert McNamara introduced a proposal in May 1962 at a NATO-meeting in Athens. His idea was that the NATO-forces should be put under the command of the United States.

The main problem for Steinbruner is to explain the disorderly performance of the United States. His thesis is that this becomes intelligible only in the light of the theoretical perspective favoured by himself. The indecisive behaviour of the American administration is explained by the fact that no decision-maker in a central position surveyed the entire negotiation process and by the fact that different parts of the process were handled by different parts of the administration and according to different routines. It goes without saying that Steinbruner puts the finger on some important problems connected with political decision-making. It seems reasonable that limitations in organisational arrangements as

		US: *Swerve*	US: *Not swerve*
The US allies	*Swerve*	NATO preserved Joint forces Shared power	NATO preserved Joint forces US supremacy
	Not swerve	NATO preserved Co-ordination of forces Sovereignty	NATO preserved Joint forces US supremacy

Figure 3. Outcome matrix: The United States against its allies.

well as the mental capacity of humans sometimes have great impact on the outcome of decision-making.

In spite of that, Steinbruner's argument is still not convincing. As a decision theory his model has some obvious weak points. At most, it can explain why the performances of political actors, both individuals and collectives, are sometimes inconsistent. But according to this theory, what becomes the outcome of the decision process seems to be entirely arbitrary. Moreover, Steinbruner's book is filled with ad hoc explanations of details and common sense discussions without any connection to his own theory. All of this would be seen as challenges for future research and not only as decisive faults, if Steinbruner had convinced us that the shortcomings of the rationalistic approach were even greater. But that is definitely not the case.

As a matter of fact, the informal negotiation process between the allies in NATO may be analysed in terms of a game between the United States and its allies. The substantive matter of that game was the power of NATO, but also the power over NATO. The small nations, i.e. all but the United States, had to take a stand over whether they were ready to give up parts of their national sovereignty. Should they, in accordance with the wish of the United States, subordinate themselves to NATO in military affairs (*swerve*) or should they insist on their sovereignty (*not swerve*)? The United States had to decide if it was content with the present situation, i.e. to share the power over NATO with the others (*swerve*), or if it dared to claim that the NATO-forces should be put under American command (*not swerve*). This scenario may be depicted in the matrix in figure 3.

It is not difficult at all to deduce from Steinbruner's discussion how the United States was ranking these possible outcomes. Its first priority was that NATO should not become further impaired. It was bad enough that France went its own way. However, the ambition of the United States was to strengthen the military alliance. To achieve that all the forces should be under a joint command, preferably in the hands of the United States.

		US: *Swerve*	US: *Not swerve*
The US allies	*Swerve*	2 , 3	1 , 4
	Not swerve	4 , 2	3 , 1

Figure 4. The security policy game concerning the future of NATO.

From Steinbruner's text it also appears that the government of Great Britain was under no circumstances ready to give up its national sovereignty. Most of all, it wanted to preserve NATO in its present form. The alliance should co-ordinate the resources of all the members, but each nation was sovereign with respect to its own forces. Great Britain was ready to follow France: it would rather leave NATO than give up its sovereignty. Accordingly, the British government was not at all interested in an agreement on joint forces under joint command, and even less interested in the supremacy of the United States.

Since Great Britain, besides the United States, was the key actor within the alliance, the following pay-off matrix may illustrate the security policy game concerning the future of NATO (figure 4).

How, to use Steinbruner's question, should we explain the indecisive behaviour of the United States? It is obvious from the game matrix that the United States could do anything but swerve for Great Britain. The matrix expresses the fact that Great Britain (and all the other NATO-members) was in a veto position over the "constitution" of NATO. There was always the possibility to stay outside the co-operation. But why at all did the United States introduce proposals, which would imply that the other nations had to give up their national sovereignty?

It is hard to give a complete answer on the bases of Steinbruner's account, but it appears from what he writes that the American administration did not come to know the preferences of Great Britain until December 1962. President Kennedy realized that the British held a very stubborn attitude when he had a conversation with Prime Minister Macmillan during a meeting at Nassau. It seems as if that the United States believed up to then that they were playing the Game of Chicken (figure 5).

In this perspective the different proposals from the American administration become intelligible. The problem for the United States was to find the formula on which everyone within the alliance could reach an agreement.

The sometimes very tough attitude from the Kennedy administration can be explained as attempts to ascertain the will of the Europeans and, if possible, to

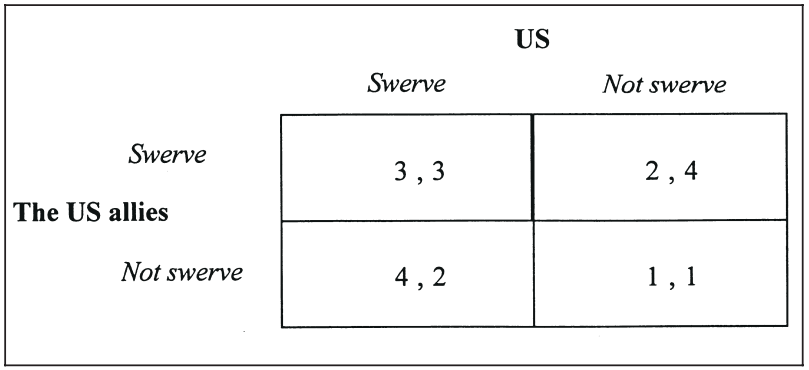

Figure 5. The US' picture of the security policy game before Nassau.

force them to swerve. After the meeting at Nassau the Americans never used this black mailing tactic anymore and gave in to the demands of the British. The established state of affairs was to prevail. It proved impossible to induce Great Britain to support further co-ordination of NATO.

The purpose of this somewhat lengthy exposition has been to indicate that the criticism of rational choice theory sometimes makes things too easy, and to illustrate (I hope convincingly) that the rationalistic approach still has a great potential at least as a framework for empirical analysis.[39] Now, performing something like a volte-face, I will turn to my two other examples. Since these examples are familiar to most political scientists, I will take a shorter path to my conclusion.

Downs' book on party politics in democratic states was published in 1957. It stands out as a pioneer-work not only for the public choice-school, but also for positive political theory as a whole. To put it very briefly, Downs develops an elaborated model of how politics works in a representative system by treating political parties as firms and voters as consumers. I do not intend to reduce the importance of Downs. However, Harold Hotelling had formulated the most central idea in *An Economic Theory of Democracy*, i.e. the median voter theorem, already in 1929. There are a lot of applications of this theorem, but among other things it tells us that in a two-party system public decisions always go in the way the median voter wants it.[40]

This theorem offers an intelligible and telling interpretation of American politics, especially when contrasted with the Nordic multi-party systems. In the same way as the previous models of game theory, the median voter theorem and Downs' economic theory of democracy provide powerful analytical tools.[41] Evidently, Downs' theory is an idealization of reality and it is easy to find empirical examples that refute its propositions, thereby throwing doubts on the soundness of its assumptions. As a reaction, Downs himself as well as his disciples have been engaged in modifications of some of the assumptions. They have, just to mention the most important, incorporated a more realistic view on the information held by the political actors. However, I am not convinced that this kind of theoretical development represents an improvement. Frankly, the more Downs reinforces his theory the less attractive it appears to me. I feel myself very much allied to Brian Barry's critical comment:

> Downs recognises that the assumptions of perfect information on both sides and of no abstentions are very strong, and suggests some results of weakening them. However, this increased realism is bought at a heavy price in loss of rigour. Part of the trouble is that Downs seems sometimes to introduce ad hoc premises in an attempt to be able to deduce from his model predictions that correspond to what is well known to be the case.[42]

Hence, it may be said that part of the problem is due to Downs himself.

However, I think there is a proper response from the rational choice theorist to this kind of criticism. In theories such as Downs', there seems to be a need for an even greater methodological consciousness. The median voter theorem and games such as Prisoners' Dilemma are best understood as formal models or Weberian ideal types, which epistemologically belong to the same category as analytical truths.[43] When dealing with the substantial political theory in An Economic Theory of Democracy, one has to distinguish between operational definitions of the variables in the formal models and hypotheses based on additional empirical premises. Only the latter can be refuted on empirical grounds.

I do think that this attitude is the accurate one, though as a reply to the critics of Downs I am not entirely happy about it. I am very much afraid that this kind of reasoning may be used in order to render theories immune against empirical tests. My last example will here serve as an illustration. In *The Rise and Decline of Nations* Mancur Olson uses the core idea from *The Logic of Collective Action* to shed light on everything from economic growth, stagflation and unemployment to the Indian caste system and other "social rigidities". It turns out that the "logic" of a collective action situation, i.e. essentially the same logic as in a Prisoners' Dilemma, provides the explanation of almost every social phenomena that can be cast under the headings "Rise" and "Decline".

The principal elements of Olson's theory are easy to grasp: The decline of nations and social rigidities can be explained by the successful pursuing of group interests at the expense of the general interest of society as a whole, i.e. efficiency defined in terms of Pareto optimality. But the fulfilment of a group interest depends on the success of a collective action, and this is according to the argument in *The Logic of Collective Action* hard, if not impossible, to achieve. In *The Rise and Decline of Nations* Olson qualifies this argument. He maintains that "'small' groups have disproportionate organizational power for collective action", but in addition he asserts that "this disproportion diminishes but does not disappear over time in stable societies".[44] Thus, countries which recently have experienced disruptions such as revolutions or wars (e.g. Germany and Japan) are more efficient than stable societies (e.g. Great Britain), since the old organizational structure of the former have been purged and new organizations have not yet been formed.

To evaluate this explanation I will again turn to Brian Barry, who in a critical comment has distinguished between three alternative positions with respect to

its pretensions to explanatory power.[45] A first possible claim would be that Olson offers a monocausal explanation, i.e. although there may be several explanatory factors, they will all be explained in terms of the monocausal explanation. Olson explicitly assures us that he does not make such a strong claim, but Barry correctly points out that he frequently writes as if he does.

A second, more modest claim would be that Olson's theory always provides the most important causal factor. This would imply that an exhaustive explanation of economic growth has to include factors, which can not be subsumed under his theory, although it still provides the principal explanation. It is not at all clear if Olson would agree to describe the status of his own theory in these terms, but that does not matter anyhow. Barry's objections make it clear that this claim would be far too strong.

The third claim would be "that the factor picked out (by Olson) is not always the most important but it is always present and will always emerge on top whenever other factors are not too strong". This seems more reasonable. It could be that organized interests of different kind, especially the unions, is part of the explanation of the slow rate of economic growth for Britain as compared to other OECD-countries during the post-war period. But there are a lot of other factors as well – Barry provides a list with seven different factors – and the problem is to evaluate the importance of just that factor picked out by Olson.

> So putting them all together we have no difficulty in explaining slow growth. The only trouble is that we have explained it so well that it is hard to say which factor is doing the most work. We can understand the question – if we took them out one at a time what difference would it make? – but we may not be able to answer it. Thus, Olson's explanation is fine, but I have no idea how important it is and I do not see how he can.[46]

This concludes my argument. It may be that the dilemma outlined can be resolved, but I prefer to leave it as a problem and as a memento.

The normative status of rational action

How should one act? Our subject matter in this essay seems to provide an obvious and reasonable answer. But is it always desirable that one acts rationally? This is the central issue in this section. To begin with we may notice that there is another answer – on the face of it almost as obvious and reasonable – to our original question, viz. one should act morally. Thus, we have at least two answers to take into consideration. We may then ask if this will cause us any problem. How are these two answers related to each other? What is the connection between rationality and morality?

In moral theory and in ethics, the notion of practical rationality is used when raising questions on how one ought to act. At the conceptual level it usually refers to something more qualified than rationality in the "thin" sense; to some variant of Kant's "pure rationality", where the distinguishing feature of rational action is its compatibility with the categorical imperative, which requires that one's actions are impressed by universality, i.e. that they do not

emanate from personal and subjective impulses and instincts.[47] Hence, following Kant, one may claim that it is never rational to behave immorally. The connection between rationality and morality is then a matter of definition; logic tells us that there is no rationality outside morality.[48] This may be a simple solution to our problem, but it is arrived at through a redefinition of rationality. Its core is no longer optimization or maximization, but rather universality.[49]

There are a few variants on this theme. Elster pleads for a "broad" theory of rationality including more than mere consistency, one that requires that desires are autonomously shaped and that beliefs are impressed by judgement. In the version of Richard Brandt the characteristic of rationality is its capacity to resist criticism. As a first approximation he refers to "actions, desires, or moral systems which survive maximal criticism and correction by facts and logic", The preferences of the actors are then taken as they are. However, for Brandt to accept an act to be fully rational, he also requires that the preferences will survive a similar test. He proposes "cognitive psychotherapy".[50]

Evidently, these attempts to define a more demanding concept of rationality have in common an urgent recognition of the need for a more attractive notion of rationality according to which "rational fools" are impossible by definition. They keep the close connection between rationality and consistency but resemble Kant in harmonizing rationality and morality (although they never reach it completely).

There are at least two other positions, which deny that, combining rationality and morality involves any difficulty, i.e. positions, which offer a "simple solution".[51] There is a second proposal founded on the reverse reductionism, which involves a similar dream about a straightforward conceptual connection between rationality and morality. According to this idea, recently stated by David Gauthier in his very sophisticated *Morals by Agreement*, it will always be morally justifiable to act rationally.[52] Then it would be possible to deduce morality solely from individual rationality. But that is to say that individual rationality is sufficient to avoid sub-optimalities in strategic situations such as Prisoners' Dilemma.[53] Hence, also this proposal is refuted.

What then about the third proposal to combine rationality and morality? By contrast it is not based on any reductionism. Instead its simplicity follows from the opposite idea; viz. rationality and morality have nothing at all in common. Rationality may be seen as a formal goal for the actor, while morality prescribes what substantial goal that should guide the action. Rationality demands consistency, but in addition to that no further restrictions is imposed on the actor's thoughts and behaviour.[54] This common sense proposal appears very attractive. The point would be that we as individuals in a community should perform in accordance with rationality as well as morality. Both are necessary, but neither of them is sufficient. The question is whether this solution will also be too simple. Can we claim that it is always desirable to act rationally, even though we admit that rationality is not enough?

In order to clarify this issue it is necessary that we are more precise about morality. Following Derek Parfit in his *Reasons and Persons* I confine myself to two different types of morality: act-consequentialism, and a common sense

morality with a deontological outlook. These two types of morality are distinguished from each other in two different ways. On the one hand, consequentialism, e.g. the traditional variant of utilitarianism in which the consequences are measured in terms of the total sum of utility,[55] is concerned with what will happen while common sense morality typically is concerned with what to do. On the other hand, consequentialism is characterized by its impartiality, i.e. it gives us agent-neutral aims, while the aims of common sense morality are agent-relative, e.g. it states that you shall take care of your own children as opposed to a morality which gives you the agent-neutral aim that parents shall take care of their own children.[56] The latter of these distinctions appears to be the most important, since it reflects the distinction between moralities that gives the actors common aims and moralities with different aims for different actors.[57]

Parfit demonstrates that the moral mathematics of act consequentialism include a lot of pitfalls. A minor problem is that it may be self-effacing, i.e. it may be the case that it tells us to believe in some other theory. This is not unique for consequentialism. It appears in most other moralities as well. Hence, it may be somewhat complicated to be moral, but this is not a failure of morality. It reflects the fact that life sometimes is complicated. However, the most important problem for a moral theory is that it can be self-defeating, i.e. it "fails even in its own terms, and thus condemns itself".[58] There are four different variants of this argument:

A theory T is

1. indirectly individually self-defeating "when it is true that, if someone tries to achieve his T-given aims, these aims will be ... worse achieved".

2. indirectly collectively self-defeating "when it is true that, if several people try to achieve their T-given aims, these aims will be worse achieved".

3. directly individually self-defeating "when it is certain that, if someone follows T, he will thereby cause his own T-given aims to be worse achieved than they would have been if he had not successfully followed T".

4. directly collectively self-defeating "when it is certain that, if we all successfully follows T, we will thereby cause the T-given aims of each to be worse achieved than they would have been if none of us had successfully followed T".[59]

If we suppose that a theory fails in its own terms then it follows that it is not compatible with rationality. Hence, this is the serious problem.

Obviously, an agent-neutral theory is not directly self-defeating, since it gives to all actors' common aims. However, it may be indirectly self-defeating: A theory such as consequentialism tells us all to be "pure do-gooders", but if each of us is a "pure do-gooder" the result may be worse for all (e.g. in a Prisoners' Dilemma situation) than if we were disposed to act in another way. However, in such a case the theory is still not failing on its own terms. Instead, the relevant conclusion is that the theory sometimes tells us not to be "pure do-gooders". The theory has to be revised. To save consequentialism Parfit argues that we ought to incorporate some elements of common sense morality.

The situation for an agent-relative theory such as common sense morality is hardly better. As for self-interest theory – the agent-relative theory, which tells you to strive for your own good – common sense morality, may be directly collectively self-defeating. Again the Prisoners' Dilemma provides us with our example. To avoid this fate it has to incorporate some agent-neutral elements.[60]

Thus, for Parfit the future for substantive moral theory lies in some kind of blend of agent-neutrality and agent-relativity. In this diagnosis and prospect for the future, he is not alone. Thomas Scanlon, Thomas Nagel and Amartya Sen – just to mention some prominent theorists – have all provided different versions of mixed moral theories.[61] Hence it seems to be possible to uphold the third position; namely that rationality and morality have nothing in common, although the price for this may appear to be too high. In order to give a consistent answer to our initial question about how one shall act, the actor has to rely on very sophisticated and complicated theorizing, e.g. it may be necessary to admit that there are situations where it is rational to act irrationally. As has already been noticed, this may not be regarded as a fault in the theory. It is the complexity of reality that forces us to develop a complicated theory.

Having reached the heights of current moral philosophy, we may feel some doubt. Our endeavour to find a convenient solution to the combination problem has carried us far off from theoretical simplicity. A more down to earth solution may after all appear as attractive. And even if we at the end return to Parfit's solution, the alternatives may be worth looking at.

There are at least two further positions that can be adopted: to give priority to rationality or to morality. Following Nietsche we may state that rationality always overrules morality or, more commonly, we may give priority to morality when it conflicts with rationality.[62] The second of these proposals is the most important. It is one of the main alternatives in the history of ideas. In its most typical form, it states that rights ought to function as some kind of constraint on individuals when they try to pursue their goals as rationally as possible. Today, it is the property rights theorist Robert Nozick, who is the best known proponent of such a deontological ethics. The main idea of Nozick is that an act is morally right if it does not infringe on anyone's rights. Obviously then, morality may be a restriction for the utility maximizing actor. Certain acts may simply be morally wrong, although they would be the best courses of action from a rational point of view. But this also means that this kind of moral theory makes the tension between morality and rationality weaker: as long as the individual acts within the restrictions of morality, he or she may choose whatever goals he or she wants.[63]

The idea of Ronald Dworkin about rights as some kind of trump may be interpreted in a similar way. Dworkin's criticism of utilitarianism focuses on the occurrence of external preferences, i.e. that our own satisfaction sometimes depends on other's well being (a more familiar notation is "other-regarding wants"). Rights then enter into our utility calculation in a very ruthless way: They simply purge those external preferences, e.g. racism, which are morally wrong.[64]

It would be premature to draw any far-reaching conclusions from this analysis. But if the diagnosis of the complexity of combining rationality and morality has something to say, it is that the problem for the individual may be interpreted as if he or she sometimes is put in choice situations where the rules of the game are badly shaped. The same insight is also brought about by Nozick's political philosophy and Dworkin's philosophy of law. Hence, it would be desirable if the rules of the games in politics and in society in general could be corrected so that value conflicts originating from individuals striving for rationality in strategic situations are avoided.

The future of rational choice theory

Two of the previous sections have ended in a similar idea: the task of politics is to create conditions – institutions – which make it possible for individuals to pursue their own goals without producing losses of efficiency or moral conflicts. Both Elster's idea about the political notion of collective rationality and Dworkin's theory about rights as trumps that regulate utility maximazation which otherwise would lead to morally doubtful conclusions, direct our attention to organizational problems. In a game theoretical analysis we ask how the rational actor should act, and this is followed by a question about what kind of outcome we expect as a result of the interaction of rational actors. However, the overall question may be another, viz. the one put at the foreground in social choice theory, i.e. how one should design rules in different contexts.[65]

Instead of asking how one should act, we then ask what kind of society the individual should favour if he or she is supposed to live his or her life in it. Both questions belong to practical philosophy. Moreover, if believing in Richard Brandt, they are the two most central questions of ethics:

> The constructive idea was the identification of two questions, different from the ones with which moral philosophers have traditionally been concerned, but which capture at least most of what is clear and important in the traditional questions. The first of these is: 'What would a person (perhaps all persons), if rational in the sense of having made optimal use of all available information, want and choose to do?' The second is: 'What kind of moral system, if any, would such a person support for a society in which he expected to live?'[66]

However, it is primarily the second of these questions which direct our interest to politics and consequently it is in connection with that question that the contributions from political science and political philosophy may be expected.

> An important, perhaps the most important task of political thought and action is to arrange the world so that everyone can live a good life without doing wrong, injuring others, benefiting unfairly from their misfortune, and so forth. Moral harmony and not only civil peace is the right aim of politics, and it would be desirable to achieve it without putting everyone through the type of deep personal conversion needed to make a clash between morality and the good life impossible.[67]

On closer inspection, it is precisely this kind of constitutional or administrative problem which a great deal of classical political theory deals with. This is obvious in the contract theories of Hobbes and Rousseau.

Rousseau's famous theory of the general will is usually identified with a romantic idea about people's unity and the common good. It has also been argued that Rousseau represents a totalitarian mode of thinking, where the general will may be derived independently from the wills or opinions expressed by individuals. And to be sure, the watchmaker-son from Geneva has inspired several different and sometimes contradictory systems of ideas. To some extent, this ambiguity is Rousseau's own fault. He was very fond of drastic formulations and he often made his argument too incisive. But never the less, it is definitely wrong to believe that Rousseau was not a realist with respect to the plurality of interests in society. His idea seems to have been that a society or any kind of association may be understood as a game in which the actors have common as well as diverging interests. As was mentioned before, it is possible to associate Rousseau with the Assurance Game. This is the most reasonable interpretation of the story about the stag hunt in *Discours sur l'origine et les fondements de l'inégalité*. However, when reading *Du contrat social* a situation such as the Prisoners' Dilemma comes naturally into one's mind.[68] The general will, which according to Rousseau is different from "the sum of all wills", may be interpreted as the co-operative outcome in that well-known game.[69]

In the Prisoners' Dilemma the actors have no difficulties on agreeing the urgency of co-operation. The real problem is compliance in the absence of external forces that would bind them to this agreement. Rousseau's problem has also to do with the guarantees for the individuals that ordinary legislation will be governed by the general will and that the co-operative outcome will be respected by the others. Rousseau's proposal for a solution consists of a careful specification of the constitutional rules of the society. The individuals themselves have to make up their own laws. This will also provide the solution to the problem of compliance and obedience; the only reason for the individuals to accept the role of a subject is that they as a collective perform the role as a sovereign.

When deciding on separate laws the general will is then to be realised through a procedure where the actors of the game appear as citizens and together participate in collective decision-making. One of the clauses of the social contract states that the general will is established in a referendum. Rousseau declares that such an arrangement will bring about a situation where the differences of the private wills may "cancel each other out" (The Social Contract, II, 3, 2). Russell Hardin has demonstrated that this is also the case in Prisoners' Dilemma if the rules are changed into an "agreeable" Prisoners' Dilemma. The co-operative outcome will then be a Condorcet winner, i.e. in a pairwise comparison with any other outcome there will always be a majority favouring the general will.[70]

According to Rousseau, majority rule may serve as a guarantee for the realisation of the general will: hence his importance for democratic theory. But to

make good laws is not sufficient even for Rousseau. Like Hobbes he is fully aware of the fundamental importance of the problem of obedience. Rousseau also needs a Leviathan, which forces the subjects to obey the laws. Therefore, the social contract includes a clause on individual compliance with decisions made by the collectivity and on the sovereign's right to use, if necessary, coercion to guarantee obedience to the laws. The difference is that the coercive power of Rousseau is the democratic state and that the social contract of Rousseau includes constitutional and procedural guarantees that the content of the laws will be in accordance with the interests the individuals have in common as citizens.

In this interpretation Rousseau's theory of the general will is far from a work by a romantic dreamer. Though he personally was an outsider, at odds with the dominant philosophy of his time, the Enlightenment, his constitutional ideas are a masterpiece in the spirit of constructivist rationalism and of enlightenment. By that I do not mean that his theory goes free from any objections. On the contrary, his social contract may be criticized because it presupposes too much fine-tuning. Rousseau's theory of the general will is, if anything, an ideal theory. It is a vigorous attempt to construct a perfect machinery. It then adds up to a quite complicated constitutional system, where all conceivable institutions are allotted their closely defined place and function, which they inevitably have to fulfil, if not the whole machinery shall break down and fail in its purpose, viz. to carry the general will into effect. In this sense Rousseau was a romantic, divorced from reality. Rousseau would have been more convincing if his ideal polity was marked by a greater strain of robustness.

This criticism of Rousseau is not directed against those streaks of rational choice theory, which at least can be gathered from his political writings. Instead, it is these traces of a game-theoretical approach, which still make it worthwhile to ponder on the strength and weakness of a political theory of Rousseauian kind. And from this we may derive the concluding viewpoint of this essay: The primary contribution of modern rational choice theory to political science and philosophy is that it makes the classical problems of political theory available for renewed and more precise analyses.

Notes

1. See especially the contributions from Jon Elster (e.g. 1983a, 1983b, 1984 and 1989), Amartya Sen (1982 and 1987), Hervé Moulin (1982) and Brian Barry (1978, 1982 and 1983).

2. Arrow 1963. See also Sen 1984 and the contributions in Barry & Hardin 1982.

3. See Elster & Hylland 1986, Nurmi 1987 and Riker 1982.

4. I am thinking of Jon Elster's monumental *Making Sense of Marx*, but there are a lot of other examples in recent literature. Just to mention two of the most interesting: In his *Morals by Agreement* David Gauthier starts from a Hobbesian position and makes use of Locke explicitly and Hume implicitly in taming Leviathan. Michael Taylor's *The Possibility of Cooperation* is another example of modern rational choice theory taking up the challenge of Hobbes. My own contribution to this discussion is an attempt to give a rationalistic interpretation of Rousseau's theory of the general will (Hermansson 1988 and 1992).

5. E.g. Jon Elster, John Roemer and Adam Przeworski. The label Analytical Marxists is sometimes used to name the same group of theorists. The important exception is G.A. Cohen's functionalist interpretation of Marx' theory of history, a typical example of analytical Marxism outside the rationalist tradition.

6. I am aware that I exaggerate this point somewhat.

7. Sen 1982:84-106.

8. Firstly, as will be further specified below, Economic man is one, but not the only, species of Rational man. Secondly, morality seems to presuppose rationality, while, at least according to most philosophers, it is possible to be rational and still fail to act morally. The treatment of this point will be postponed to a later section.

9. Elster 1983b:15-26.

10. This feature explains the curious fact that there are rational choice theorists who try to develop evolutionary social theories. See Axelrod 1984 and Sugden 1986.

11. See the surveys in Elster 1983b or 1986b, which I think represent the state of the art.

12. There are several variants of game theoretical analyses, but I confine myself to non-cooperative game theory and to games in normal form. For technical details, see von Neumann & Morgenstern 1953, Luce & Raiffa 1957, Moulin 1982, Shubik 1982 and Harsanyi 1986.

13. In this game a player may gain at most "two points". A draw gives both player "one point" and a loss nothing. Thus, they compete for a constant sum. Henceforth the payoffs represent the rank-order of the actors; i.e. the payoff 2 informs you that the actor values the outcome as better than an outcome with payoff 1.

14. I will disregard the notion of mixed strategies, although it plays an important role for the mathematical elaboration of game theory. Mixed strategies mostly force us into rather obscure interpretations when dealing with human agency (See Elster 1986b:18).

15. To begin with I use "solution" in a loose sense. The solution concept will be discussed in the following section.

16. See Luce & Raiffa 1957:90-94.

17. See Luce & Raiffa 1957:94-102 and Rapoport & Chammah 1965.

18. This is known as the 'free rider' problem.

19. These games are analysed in Brams 1975:39ff, Elster 1983a:78 and Sørensen 1984 respectively. See also Rapoport & Guyer 1966 for a general overview of all 2*2 games.

20. In my dissertation (Hermansson 1984; cf. Hermansson 1990) I use the Control Game to illustrate the orthodox Marxist-Leninist theory of proletarian revolution. The proletariat as row player (the controller) and the bourgeois as column player (the controlled) both choose between peaceful (C-strategy) and armed (D-strategy) class struggle. The outcomes are interpreted as follows: (4,3) – peaceful socialist revolution; (3,2) – almost peaceful socialist revolution; (2,1) – violent socialist revolution; (1,4) – violent capitalist counter-revolution.

21. Cf. Elster 1983b:26ff.

22. Elster 1983b:29. Elster here refers to Elster 1978, chapter 5.

23. Oppenheim 1981:125.

24. Cf. Elster 1986b:17-22. In Elster 1983b:2-26 my first problem appears as one of the "three cases where optimality breaks down", which according to Elster "provide the special argument for satisficing". My second problem resembles that kind of failure of rational choice, which in Elster 1983b provide "the general argument for satisficing". It may in fact be seen as a generalization of Elster's argument.

25. Cf. Luce & Raiffa 1957:106f.

26. I.e. suppose that both the strategy combinations x_1*y_1 and x_2*y_2 yield equilibria. These two equilibrium points are then interchangeable, if the strategy combinations x_1*y_2 and x_2*y_1 also yield equilibria.

27. Instead of Pareto optimality Luce & Raiffa (1957:107) use the notion of jointly admissible.

28. Elster 1983b:13.

29. Cf. Moulin 1982:200ff.

30. It will not even do to change the definition of equilibrium. In the Control Game there is also no non-myopic equilibrium (See Sørensen 1984).

31. Przeworski 1985b:399f.

32. This is the core of Herbert Simon's (1955) well-known critique of rational choice theory.

33. Most textbooks in mathematical logic or the history of philosophy include passages on the importance of self-referential statements in ordinary as well as formal language. However, Douglas Hofstadter's *Gödel, Escher, Bach: An Eternal Golden Braid* is by far the best and most cogent survey of this subject.

34. Elster 1986b:27.

35. See Quattrone & Tversky 1986 and Tversky & Kahneman 1986.

36. Rational choice theorists themselves admit that these features count as short-comings, although they may differ with regard to what conclusion that should be drawn. See e.g. Boudon 1986, Elster's contribution in Elster & Hylland 1986 and Przeworski 1985b.

37. The first critique may be right if one take preferences to mean desires, but it is wrong if one wants to widen the scope to include beliefs. As to the static nature of rational choice theory, one can argue that the recent development in game theoretical modelling of politics mostly make use of extensive games. Hence, it rather points to the dynamic character of politics. On both these points see Morrow 1994 and Gates & Humes 1997.

38. The term used by Steinbruner is the analytical paradigm.

39. This discussion can and should be taken further at least in two respects. Firstly, there is a need to specify the criterion to be used in comparing and evaluating different analytical models (see Bates et al. 1998 and Morton 1999). Secondly, it is rather obvious how the game theoretic model can be improved by introducing learning through some kind of signalling. The proper tools for this are extensive-form games and perfect Bayesian equilibria (see Morrow 1994).

40. Cf. Tullock 1976, Mueller 1979 and Ordeshook 1986.

41. See e.g. Brams 1978.

42. Barry 1978:108.

43. See especially Boudon 1986:200f.

44. Olson 1982:74.

45. Barry 1983, especially p. 19.

46. Barry 1983:23.

47. Lindley (1986) offers a lucid presentation of the notions of rationality and autonomy in the philosophies of Kant, Hume and Mill.

48. A somewhat weaker position is defended by Nagel 1986:200, viz. that it is never irrational to act morally, although the connection is then not only definitional. It may also be illuminating to think of Plato, where a theory of the right provide the answer on how to pursue one's own good.

49. Cf. Hare 1981:107-116 and Mackie 1977:83-102.

50. Elster 1983b:15-26 and Brandt 1979:10-16.

51. Cf. Williams 1985, chapter 1, and Nagel 1986, chapter X.

52. Historically this idea is usually associated with ethical theories which oppose that the notion of the good life may be reduced to something as simple as maximization of utility. However, both Bentham and Aristotle represent similar ideas, i.e. a teleological philosophy as opposed to a deontological one, in so far as that the morally right is defined in terms of pursuing the good (be that pleasure/utility or something more complex). As for Gauthier and his critics, see especially Kraus & Coleman 1987 and the articles in *Social Philosophy and Policy* 5 (2/1988).

53. In this context it is usual to refer to iterated Prisoners' Dilemmas (Axelrod 1984) and to the theory of supergames (Taylor 1976; 1987), and sometimes it is claimed that reiteration, i.e. the repetitive character of most strategic situations, provides a solution to the problems of cooperation in situations such as Prisoners' Dilemma. However, I do not agree. If at all relevant, I think that Axelrod's analysis primarily provides insight into the conflict of interest between present and future. In the case of Taylor and his supergames it can be argued that an iterated Prisoners' Dilemma actually becomes an Assurance Game. The possibility of cooperation rests on the possibility of co-ordinating behaviour in the long run, which crucially depends on actors' information about each other, which in turn is affected by the size of the community. Hence, the importance of smallness for collective action (Olson 1971 (1965) and Hardin 1982).

54. Cf. Parfit 1984:3.

55. See Smart & Williams 1973, Sen & Williams 1982 and Griffin 1982.

56. Apart from these cases Parfit also analyses the self-interest theory, which is a agent-relative theory, concerned with what happens to the agent himself.

57. See Sen's important "Rights and Agency", especially p. 214, which clarifies that this notion is the difference between viewer neutrality and viewer relativity. Sen distinguish between doer, viewer and self-evaluation neutrality, and correspondingly doer, viewer and self-evaluation relativity.

58. Parfit 1984:3.

59. The quotations are taken from Parfit 1984:5, 27 and 55.

60. Parfit 1984, especially pp. 24-28 and 40-43.

61. Scanlon 1988, Nagel 1986 and Sen 1988. See also Hare 1981.

62. To make my list exhaustive I may add another position, viz. one can argue that the eventual conflict between rationality and morality has to be solved differently in dissimilar situations. Michael Walzer (1983) and Bernard Williams (1981 and 1985) probably adhere to this (sixth) pluralist position.

63. Nozick 1974.

64. Dworkin 1977, especially pp. 231-238.

65. Cf. Riker & Ordeshook 1973:276-280 who distinguish between "questions of action" and "questions of regulation". With respect to social choice theory, see especially the works of Kenneth Shepsle and others, which analyse how complementary political institutions contribute to stability in collective decision making. One of the articles of Shepsle is titled "Structure-induced equilibrium and legislative choice".

66. Brandt 1979:v.

67. Nagel 1986:206.

68. I have developed this rationalistic interpretation of Rousseau's theory of the general will in Hermansson 1988 and 1992.

69. There are several passages in The Social Contract which provide suitable references, but the most important are book I, chapter 7, paragraph 7, and (using the same notation) II,1,1 and II,3,2. Cf. Runciman & Sen 1965 and Boudon & Bourricaud 1982.

70. Hardin 1982 (1971).

Literature

Abrams, Robert. 1980. *Foundations of Political Analysis.* New York: Columbia University Press.

Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis.* Boston: Little, Brown and Company.

Anckar, Dag, Nurmi, Hannu & Wiberg, Matti, eds. 1988. *Rationality and Legitimacy: Es-*

says on Political Theory. Jyväskylä: The Finnish Political Science Association.
Arrow, Kenneth J. 1963 (1951). *Social Choice and Individual Values.* 2nd edition. New York: Wiley.
Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.
Barry, Brian. 1978 (1970). *Sociologists, Economists and Democracy.* Chicago: The University of Chicago Press.
Barry, Brian. 1982. "Methodology Versus Ideology: The 'Economic' Approach Revisited", pp. 123-147 in Ostrom, Elinor, ed., *Strategies of Political Enquiry.* Beverly Hills: Sage.
Barry, Brian. 1983. "Some Questions about Explanation", *International Studies Quarterly* 27, 17-28.
Barry, Brian & Hardin, Russell, eds. 1982. *Rational Man and Irrational Society?* Beverly Hills: Sage.
Barry, Brian, ed. 1976. *Power and Political Theory. Some European Perspectives.* London: John Wiley & Sons.
Bates, Robert H. et al. 1998. *Analytical Narratives.* Princeton, New Jersey: Princeton University Press.
Binmore, Ken G. 1984. "Game Theory and the Social Contract, Part I", Mimeo, London School of Economics and Political Science, Department of Mathematics.
Binmore, Ken G. 1988. "Game Theory and the Social Contract, Mark II", Mimeo, London School of Economics and Political Science, Department of Mathematics.
Boudon, Raymond. 1986 (1984). *Theories of Social Change: A Critical Appraisal.* Cambridge: Polity Press.
Boudon, Raymond & Bourricaud, Francois. 1982. "Rousseau, Jean-Jacques", pp. 476-482 in *Dictionnaire Critique de la Sociologie.* Paris: Presses Universitaires de France.
Braithwaite, R.B. 1955. *Theory of Games as a Tool for the Moral Philosopher.* Cambridge: Cambridge University Press.
Brams, Steven J. 1975. *Game Theory and Politics.* New York: The Free Press.
Brams, Steven J. 1978. *The Presidential Election Game.* New Haven: Yale University Press.
Brams, Steven J. 1985. *Superpower Games. Applying Game Theory to Superpower Conflict.* New Haven/London: Yale University Press.
Brams, Steven J. & Hessel, Marek P. 1982. "Absorbing Outcomes in 2*2 Games", *Behavioral Science* 27, 393-401.
Brandt, Richard B. 1979. *A Theory of the Good and the Right.* Oxford: Clarendon Press.
Buchanan, James & Tullock, Gordon. 1962. *The Calculus of Consent.* Ann Arbor: University of Michigan Press.
Coleman, Jules L. & Ferejohn, John. 1986. "Democracy and Social Choice", *Ethics* 97, 6-25.
Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.
Dworkin, Ronald. 1977. *Taking Rights Seriously.* Cambridge, Mass.: Harvard University Press.
Dworkin, Ronald. 1978. "Liberalism", pp. 113-143 in Hampshire 1978.
Dworkin, Ronald. 1984. "Rights as Trumps", pp. 153-167 in Waldron, Jeremy, ed., *Theories of Rights.* Oxford readings in philosophy. Oxford: Oxford University Press.
Elster, Jon. 1978. *Logic and Society. Contradictions and Possible Worlds.* Chichester: John Wiley & Sons.
Elster, Jon. 1983a. *Explaining Technical Change. A Case Study in the Philosophy of Science. Studies in Rationality and Social Change.* Cambridge: Cambridge University Press.
Elster, Jon. 1983b. *Sour Grapes. Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press.
Elster, Jon. 1984 (1979). *Ulysses and the Sirens. Studies in Rationality and Irrationality.* Cambridge: Cambridge University Press.
Elster, Jon. 1985a. *Making Sense of Marx.* Cambridge: Cambridge University Press.
Elster, Jon. 1985b. "Rationality, Morality, and Collective Action", *Ethics* 96, 136-155.
Elster, Jon. 1987. "The Possibility of Rational Politics", *Archives Européennes de Sociologie,* 28, 67-103.
Elster, Jon & Hylland, Aanund, eds. 1986. *Foundations of Social Choice Theory. Studies in Rationality and Social Change.* Cambridge: Cambridge University Press.
Elster, Jon, ed. 1986a. *The Multiple Self.* Cambridge: Cambridge University Press.
Elster, Jon, ed. 1986b. *Rational Choice.* Oxford: Basil Blackwell.

Gates, Scott & Humes, Brian D. 1997. *Games, Information, and Politics. Applying Game Theoretic Models to Political Science.* Ann Arbor: The University of Michigan Press.
Gauthier, David. 1977. "The Social Contract as Ideology", *Philosophy and Public Affairs* 6, 130-164.
Gauthier, David. 1985. "Bargaining and Justice", *Social Philosophy and Policy* 2:2, 29-47.
Gauthier, David. 1986. *Morals by Agreement.* Oxford: Oxford University Press.
Gauthier, David. 1987. "Taming Leviathan", *Philosophy and Public Affairs* 16, 280-298.
Gauthier, David. 1988. "Morality, Rational Choice, and Semantic Representation: A Reply to My Critics", *Social Philosophy and Policy* 5:2, 173-221.
Griffin, James. 1982. "Modern Utilitarianism", *Revue Internationale de Philosophie* 141, 331-375.
Hamburger, Henry. 1979. *Games as Models of Social Phenomena.* San Francisco: Freeman and Company.
Hardin, Russell. 1982 (1971). "Collective Action as an Agreeable n-Prisoners' Dilemma", pp. 123-135 in Barry & Hardin 1982.
Hardin, Russell. 1982. *Collective Action.* Baltimore: Resources for the Future, The Johns Hopkins University Press.
Hare, R.M. 1981. *Moral Thinking. Its Levels, Method and Point.* Oxford: Clarendon Press.
Harsanyi, John C. 1986 (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge: Cambridge University Press.
Hermansson, Jörgen. 1984. *Kommunism på svenska? SKP/VPK:s idéutveckling efter Komintern.* Stockholm: Almqvist & Wiksell International.
Hermansson, Jörgen. 1988. "The Majority as a Guarantee for the General Will: A Note on the Constitutional Theory of Rousseau", pp. 96-126 in Anckar m.fl. 1988.
Hermansson, Jörgen. 1990. *Spelteorins nytta. Om rationalitet i vetenskap och politik.* Stockholm: Almqvist & Wiksell International.
Hermansson, Jörgen. 1992. "Rousseau on Justice.
Hofstadter, Douglas R. 1980 (1979). *Gödel, Escher, Bach: The Eternal Golden Braid.* New York: Penguin Books.
Kraus, Jody S. & Coleman, Jules L. 1987. "Morality and the Theory of Rational Choice", *Ethics* 97, 715-749.
Lindley, Richard. 1986. *Autonomy.* London: Macmillan.
Luce, R. Duncan & Howard Raiffa. 1957. *Games and Decisions.* New York: John Wiley & Sons.
Mackie, John L. 1977. *Ethics. Inventing Right and Wrong.* London: Penguin Books.
May, K.O. 1982 (1952). "A Set of Independent, Necessary and Sufficient Conditions for Simple Majority Decision", pp. 299-303 in Barry & Hardin 1982.
Morrow, James D. 1994. *Game Theory for Political Scientists.* Princeton, New Jersey: Princeton University Press.
Morton, Rebecca B. 1999. *Methods and Models. A Guide to the Empirical Analysis of Formal Models in Political Science.* Cambridge: Cambridge University Press.
Moulin, Hervé. 1982. *Game Theory for the Social Sciences. Studies in Game Theory and Mathematical Economics.* New York/London: New York University Press.
Mueller, Dennis C. 1979. *Public Choice.* Cambridge: Cambridge University Press.
Nagel, Thomas. 1986. *The View from Nowhere.* Oxford: Oxford University Press.
Nash, John. 1951. "Non-cooperative Games", *Annals of Mathematics* 54:286-95.
Nelson, Alan. 1988. "Economic Rationality and Morality", *Philosophy and Public Affairs* 17, 149-166.
Neumann, John von & Morgenstern, Oskar. 1953 (1944). *The Theory of Games and Economic Behavior.* Princeton: Princeton University Press.
Niskanen, William A. 1971. *Bureaucracy and Representative Government.* Chicago, New York: Aldine-Atherton.
Nozick, Robert. 1974. *Anarchy, State, and Utopia.* Oxford: Basil Blackwell.
Nurmi, Hannu. 1987. *Comparing Voting Systems.* Dordrecht: Reidel.
Olson, Mancur. 1971 (1965). *The Logic of Collective Action.* Cambridge, Mass.: Harvard University Press.
Olson, Mancur. 1982. *The Rise and Decline of Nations. Economic Growth, Stagflation, and Social Rigidities.* New Haven: Yale University Press.

Oppenheim, Felix. 1981. *Political Concepts: A Reconstruction.* Oxford: Basil Blackwell.
Ordeshook, Peter C. 1986. *Game Theory and Political Theory. An Introduction.* Cambridge: Cambridge University Press.
Parfit, Derek. 1984. *Reasons and Persons.* Oxford: Clarendon Press.
Popper, Karl R. 1963. *Conjectures and Refutations. The Growth of Scientific Knowledge.* London: Routledge and Kegan Paul.
Przeworski, Adam. 1985a. *Capitalism and Social Democracy.* Cambridge: Cambridge University Press.
Przeworski, Adam. 1985b. "Marxism and Rational Choice", *Politics & Society* 14, 379-409.
Quattrone, George A. & Tversky, Amos. 1986. "Self-Deception and the Voter's Illusion", pp. 35-58 in Elster 1986a.
Rae, Douglas W. 1969. "Decision-Rules and Individual Values in Constitutional Choice", *American Political Science Review* 63, 40-56. (also in Barry & Hardin 1982).
Rapoport, Anatol. 1960. *Fights, Games, and Debates.* Ann Arbor: The University of Michigan Press.
Rapoport, Anatol. 1966. *Two-Person Game Theory. The Essential Ideas.* Ann Arbour: The University of Michigan Press.
Rapoport, Anatol & Chammah, Albert M. 1965. *Prisoner's Dilemma. A Study in Conflict and Cooperation.* Ann Arbor: The University of Michigan Press.
Rapoport, Anatol & Guyer, M. 1966. "A Taxonomy of 2*2 Games*", General Systems: Yearbook of the Society for General Systems Research* 11: 203-214.
Rawls, John. 1971. *A Theory of Justice.* Oxford: Oxford University Press.
Riker, William H. 1982. *Liberalism Against Populism. A Confrontation Between the Theory of Democracy and the Theory of Social Choice.* San Francisco: Freeman and Company.
Riker, William H. & Ordeshook, Peter C. 1973. *An Introduction to Positive Political Theory.* Englewood Cliffs, N.J.: Prentice-Hall.
Roemer, John E. 1982. *A General Theory of Exploitation and Class.* Cambridge, Mass.: Harvard University Press.
Roemer, John E. 1985a. "Should Marxists Be Interested in Exploitation?", *Philosophy and Public Affairs* 14, 30-65 (also in Roemer 1986).
Roemer, John E. 1985b. "Equality of Talent", *Economics and Philosophy* 1, 151-188.
Roemer, John E. 1986a. "An Historical Materialist Alternative to Welfarism", pp. 133-164 in Elster & Hylland 1986.
Roemer, John E. 1986b. "The Mismarriage of Bargaining Theory and Distributive Justice", *Ethics* 97, 88-110.
Roemer, John E. 1986c. "New Directions in the Marxian Theory of Exploitation and Class", pp. 81-113 in Roemer 1986 (originally published in *Politics and Society,* vol. II, nr 3, 1982).
Roemer, John E. 1989. "What Is Exploitation? Reply to Jeffrey Reiman*", Philosophy and Public Affairs* 18, 90-97.
Roemer, John E., ed. 1986. *Analytical Marxism.* Cambridge: Cambridge University Press.
Roth, Alvin E. 1979. *Axiomatic Models of Bargaining.* Berlin: Springer Verlag.
Runciman, W.G. & Sen, Amartya K. 1965. "Games, Justice and the General Will", *Mind* 74, 554-562.
Satterthwaite, M.A. 1975. "Strategy-Proofness and Arrow's Conditions", *Journal of Economic Theory* 10, 187-217.
Scanlon, T.M. 1982. "Contractualism and Utilitarianism", pp. 103-128 in Sen & Williams 1982.
Scanlon, T.M. 1988 (1978). "Rights, Goals, and Fairness", pp. 74-92 in Scheffler 1988.
Scheffler, Samuel, ed. 1988. *Consequentialism and Its Critics.* Oxford: Oxford University Press.
Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: Norton.
Schelling, Thomas C. 1980 (1960). *The Strategy of Conflict.* Cambridge, Mass.: Harvard University Press.
Schelling, Thomas C. 1984. *Choice and Consequence. Perspectives of an Errant Economists.* Cambridge, Mass.: Harvard University Press.
Sen, Amartya. 1982. *Choice, Welfare and Measurement.* Oxford: Basil Blackwell.
Sen, Amartya. 1984 (1970). *Collective Choice and Social Welfare.* Amsterdam: North-Holland.
Sen, Amartya. 1987. *On Ethics and Economics.* Oxford: Basil Blackwell.

Sen, Amartya. 1988 (1982). "Rights and Agency", pp. 187-223 in Scheffler 1988.
Sen, Amartya & Williams, Bernard, eds. 1982. *Utilitarianism and Beyond.* Cambridge: Cambridge University Press.
Shepsle, Kenneth A. & Weingast, Barry R. 1981. "Structure-Induced Equilibrium and Legislative Choice", *Public Choice* 37, 503-519.
Shubik, Martin. 1982. *Game Theory in the Social Sciences.* Cambridge, Mass.: MIT Press.
Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics* 69, 99–118.
Smart, J.J.C. & Williams, Bernard. 1973. *Utilitarianism: For and Against.* Cambridge: Cambridge University Press.
Snyder, Glenn H. & Diesing, Paul. 1977. *Conflict among Nations.* Princeton: Princeton University Press.
Sørensen, Rune J. 1984. "The Application of Game-Theory in the Study of Hierarchical Decision-Making", *Scandinavian Political Studies* 7, 155-173.
Steinbruner, John D. 1974. *The Cybernetic Theory of Decision. New Dimensions of Political Analysis.* Princeton, N.J.: Princeton University Press.
Straffin, Philip D. 1982 (1977). "Majority Rule and General Decision Rules", pp. 316-324 in Barry & Hardin 1982.
Sugden, Robert. 1986. *The Economics of Rights, Co-operation and Welfare.* Oxford: Basil Blackwell.
Taylor, Michael. 1976. *Anarchy and Cooperation.* New York: Wiley.
Taylor, Michael. 1987. *The Possibility of Cooperation.* Cambridge: Cambridge University Press.
Tullock, Gordon. 1976. *The Vote Motive.* With a British Commentary by Morris Perlman. London: The Institute of Economic Affairs, Hobart Paperback No. 9.
Tversky, Amos & Kahneman, Daniel. 1986 (1981). "The Framing of Decisions and the Psychology of Choice", pp. 123-141 in Elster 1986b.
Vickrey, William. 1982 (1960). "Utility, Strategy, and Social Decision Rules", pp. 343-353 in Barry & Hardin 1982.
Walzer, Michael. 1983. *Spheres of Justice. A Defense of Equality and Pluralism.* New York: Basic Books.
Williams, Bernard. 1981. *Moral Luck. Philosphical Papers 1973 – 1980.* Cambridge: Cambridge University Press.
Williams, Bernard. 1985. *Ethics and the Limits of Philosophy.* London: Fontana Press.
Zagare, Frank C. 1984. *Game Theory: Concepts and Applications.* Beverly Hills: Sage.