Rational and moral action: a critical survey of rational choice theory

de Jonge, J.P.R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER IV

STRATEGIC CHOICE

1. Introduction

In rational choice theory attention gradually shifted from issues of individual rationality to the problems of collective and strategic rationality when theorists became more receptive to the interdependencies of individual choices. After many years of negligence economists embraced game theory in the late 1970s.

Besides the interest in the interdependencies of individual actions there is also a quite different reason for the growing popularity of game theory and that is that non-market applications of rational choice theory required a general analytical framework without markets and market prices. Nash's theory of non-cooperative games provided such a framework. "Today economists can define their field more broadly as being about the analysis of incentives in all social institutions." (Myerson, 1999, 1068).
The interdependencies of individual actions brought old questions to the fore: will cooperation emerge in a world of self-regarding agents without a benevolent dictator and how are the numerous voluntary transactions coordinated? In game theory cooperation and coordination problems are modeled in very diverse ways in order to answer the question whether (unique) solutions to such problems in various situations can be thought to emerge spontaneously. Usually one departs from a state of nature to explain the emergence of institutions (rules/arrangements). This point of departure betrays a commitment to reductionism; one wants to prevent the explanation of the emergence of institutions by means of institutions.

From the point of view of generating a general analytical framework the task of game theory is defined in a different way. It is to "redefine economics as being the study of rational competitive behavior in any institution of society." (Ib., 1080) From this perspective the task for economists (rational choice theorists) is to "identify the game models that yield the most useful insights into economic problems." (Ib., 1080)

Game theory is thus a means to different ends. Maybe it is useful to make a distinction between a micro- and a macro-oriented application of game theory. In this chapter I will be mainly focused on the macro-oriented application of game theory. Therefore, I will use game theory as a theoretical instrument to explain how a spontaneous order can evolve, how rules regulating human behavior emerge without conscious human design and why some rules (or complexes of rules) are maintained while others have disappeared. In this orientation of game theory the emergence of social rules (or institutions) is explained in terms of solutions to games. Are there unique rational solutions for the complexities that arise from the interdependencies of human actions, or are we merely watching self-perpetuating patterns of human behavior? In this chapter these questions will not be addressed explicitly. Most of the time I will be concerned with the more technical aspects of game theory.
In the next two sections I explain some general characteristics of game theory: the information players can dispose of and the rationality assumptions. Thereafter I present a typology of games. In the section that follows I discuss solutions to games with an emphasis on the solution of non-cooperative games and I wind up this chapter with some concluding remarks about game theory: its relation to evolution theory and to other parts of decision theory.

2. How to play games

Game theory is defined as the formal analysis of rational behavior in conflict situations (Van Damme and Heertje, 1994, 937); it describes the behavior of mutually dependent, egoistic actors. (Bianchi and Moulin, 1991, 183) The adjectives ‘conflict’ and ‘egoistic’ are, in my opinion, in need of some moderation. Coordination games, for example, can as well be characterized by lack of information concerning mutual expectations, as by conflicting ends. Not all conflict games are zero-sum games, and actors are not a priori incapable of cooperation in order to obtain a collective optimal result. Their behavior is occasionally selfish, and occasionally it can more adequately be described as self-regarding (non-tuistic).

In a strategic situation, actors cannot disregard each other. Each individual actor has to take the choices of the other actors into account. How are social interactions modeled in game theory? There are numerous games, but three aspects of the game situation are relevant for all: the amount of information, the definition of rationality and the type of games actors are involved in. I start with the first two aspects and discuss the type of games in the next section.

In an ordinary game with complete information, the structure as well as the rules of the game is known. (Brandenburger, 1992; Lyons, 1992; Hargreaves Heap and Varoufakis, 1995) The players know the type of the game; they are informed about the number of players, the feasible set of action alternatives for each player, the payoffs of every choice combination (they know the preferences of the other
players) and the possibilities for entering into coalitions. Every player has this information at his disposal and he knows that the other players have it too. In addition some rationality postulates have to be formulated to ensure that games have (unique) solutions.

In the case of incomplete information there is either uncertainty about the strategic possibilities of the other players or about the payoffs or about both. When both the opportunities and the payoffs are unknown, we have a case of uncertainty. In game theory this is rarely the case. Usually the strategic possibilities are known, but the information about the (expected) payoffs is lacking. When players are confronted with such situations, they will act on expectations. "One of the most important problems for game theory is precisely to decide what expectations intelligent players can rationally entertain about other intelligent players' behavior. This may be called the problem of mutual "rational expectations". " (Harsanyi, 1965, 450) Harsanyi has, of course, offered a solution. The expectations can be modeled, he suggests, in the form of the types of players that are conceivably represented by his opponents. The uncertainty about the payoffs is transformed into uncertainty about the type of players one deals with. All one needs to assume is that each player has some expectation as to the type of player represented by each of his opponents.  

In a two person game player K assumes for instance that her opponent R can be represented by the alternatives types (R1, R2,..,Rm), each type representing a different combination of preferences, willingness to take risk, capabilities and so on. She will assign probabilities to the alternative hypotheses that R has the type R1, R2,.., Rm, respectively (different types of K will assign different sets of probabilities to types of R). The other player will act in a similar way regarding K. By using such probabilistic models, games with incomplete information are transformed into games with complete information, because the players know all the parameters that define the game. Of course it will be a game with imperfect information, because even though the players know the basic parameters, they will not know the actual type of the other player. But unlike games with incomplete information, those with imperfect information are readily accessible to game theoretic analysis. (Harsanyi, 1992)
With regard to the rationality postulate it is assumed that all players act rational and that all know this. This means that actor A in his considerations with regard to the choice of an action-strategy assumes that all players take everything they know about the game into account when they deliberate about their strategy. The ‘common knowledge of rationality’ (CKR) applies, i.e., that every player assumes that his fellow players are rational and that they will select strategies that maximize their payoffs. The CKR implies that each player holds separately consistent beliefs about the other players’ choices and beliefs. But in order to reach equilibrium there must be mutual consistency and hence each player must hold true beliefs about what the other player will do. Therefore, what a player expects a fellow player to do has an immediate effect on what is rational for him or her to do. This rationality assumption more or less excludes that a player has to act on the expectation that another player will select a strategy that he himself would regard as irrational under the given circumstances. (see Harsanyi, 1965, 451) The fixing, in this manner, of the beliefs that players have about each other is of crucial importance for game theory. Actors cannot be surprised by unexpected moves. Why this is crucial, is easy to see. Actor A knows that actor B will try to assess what A intends to do and selects a strategy that is to his knowledge the best answer to A’s strategy. A has to anticipate B’s reaction, and B will realize this, and so on.79 This rules out that strategies are chosen which are inconsistent with CKR. Any strategy that conforms to this axiom is said to be rationalizable.

3. Which games to play

Concerning the type of games, we can, in a first approach, distinguish two main types: coordination games and cooperation games. Coordination problems arise in cases of

---

79 This assumption has been adopted to avoid an endless regression that would result from the assumption that each player tries to imagine what the other players are likely to do.
multiple equilibria. The need for cooperation arises in games in which the equilibrium-solution is Pareto-inferior.

There are clear differences between coordination problems and cooperation problems. It is a very different thing to coordinate on one equilibrium out of many or to improve on a Pareto-inferior equilibrium. In coordination games agents have to form concordant mutual expectations, in cooperation games agents have to decide whether they are willing to select cooperative strategies. Secondly, the conditions for a solution differ widely. In pure coordination games the main option for every agent is to coordinate. The interests of the agents converge, because each kind of equilibrium is better than none. In pure cooperation games interests are divergent, because it is rational for every single agent not to cooperate regardless what the other agents do. So the main option of every individual agent is not to cooperate. On the micro level many situations demand for their representation mixed type games, because coordination and cooperation presuppose each other. Often a solution to a coordination problem is inconceivable without some cooperation, and the solution of a cooperation problem demands some coordination.

The distinction between coordination and cooperation games can be combined with the distinction between cooperative and non-cooperative games. Cooperative games are games where players can make fully binding and enforceable commitments; in non-cooperative games this is not the case. This second distinction does not contradict the existence of (pure or predominantly) coordination and cooperation problems, but shifts the attention towards the conditions for their solution. In cooperative games, players can agree on any possible combination of strategies since they can be sure that any such agreement will be kept. In contrast,

---

80 One can imagine situations where agents have to cooperate in order to coordinate (the stag hunt game could provide an example). The assurance game is an example of a type of game where agents have to coordinate (the collecting of their contributions) in order to cooperate and thereby to realize the preferred collective goal.
in a non-cooperative game the only viable spontaneous equilibria are Nash-equilibria, because the only self-enforcing equilibria are the combinations of strategies where every player's strategy is the best reply to all other players' strategies. Nash-equilibria, given by the combination of strictly dominant strategies can, however, be collectively inferior.

To combine both distinctions, we can take notice of Schelling's classification. Schelling distinguished in his seminal publication pure conflict games from pure collaboration games. The former pictures zero-sum situations, the latter situations in which the players win or lose together, having identical preferences regarding the outcome. He named these pure collaboration games coordination games. Mixed type games are, in his terminology, nonzero-sum games. These are games in which the relation between the players is ambivalent because of the mixture of mutual dependence and conflict. He called this type of games plainly mixed-motive games or bargaining games. (Schelling, 1976 (1960)) In my reading of his account a pure conflict is not the absence of coordination, but the absence of bargaining as in for instance a zero-sum game as playing chess.

Hampton gives a similar notion, although she does not refer to Schelling. She proposes to define a coordination game as any game without cooperation problems. Agents are either solely occupied with coordination problems or have already agreed to cooperate and so they do not need to discuss that they should cooperate, but only how they do it (as for instance

---

81 This distinction between cooperative and non-cooperative games is waning, because when one assumes binding agreements, one has to know what makes agreements binding and this requires a non-cooperative approach. (Hargreaves Heap and Varoufakis, 1995, 38) In this sense all games are considered to be non-cooperative games, although some appear to be cooperative if it turns out that agreements are considered by the agents involved to be binding.
in the assurance game). A consequence of this proposal is that only conflicts create cooperation problems. (Hampton, 1987) From this point of view a conflict game is equivalent to a game with an unsolved cooperation problem. The distinction between coordination and conflict is therefore not useful. One would be fully justified to talk about coordination games with conflict because players value the equilibria differently (as in the battle of the sexes game). Therefore, I propose to make a distinction between coordination and cooperation games that runs analogously with the distinction between cooperative and non-cooperative games. In coordination games, agents only gain if they succeed in coordinating their actions. Therefore they have an incentive to commit themselves to an agreement. Coordination games satisfy the criteria of cooperative games. And while not all games that describe cooperation problems are non-cooperative games (this was the point Hampton was making), surely non-cooperative games are exclusively concerned with cooperation problems. The reason is that in situations that demand cooperation one can reap gains (for instance by unilaterally maintaining trade barriers) or forgo costs (as free riders) by not cooperating, whereas in situations that picture coordination problems one can only gain by collectively agreeing on a solution.

4. Solutions to games

Social institutions are looked upon as (stable) equilibrium outcomes of strategic games. (Schotter and Schwödiauer, 1980) A convention, for example, is the solution of a coordination game, a norm the solution of a cooperation

---

82 In an indirect way Vanberg also supports this classification. He distinguishes between conventions, as solutions to coordination problems and moral rules, as solutions to recurrent problems in which there is a potential conflict between common interest and individual interest (see, Vanberg, 1994, 245). The latter kinds of problems are cooperation problems in my terminology.
One of the ambitions of game theory is to investigate whether and in what way social rules can arise in various games. In each strategic interaction there are several strategic possibilities. As a consequence numerous solutions are possible. The question arises whether a unique (and stable) solution can arise spontaneously, or whether (institutional) restrictions have to be imposed on strategic choices, in order to successfully predict the solution. As a point of departure I take it that an equilibrium-solution, far from being a consequence of individual rationality, arises from certain restrictions on agent’s expectations. When it is possible to reasonably assess the choice of a strategy in this way, we call this choice rationalizable.

I will first discuss solutions to non-cooperative games. What is the rational (equilibrium-) solution? To be able to answer this question we need a solution concept, a rule of behavior that players can use to select their best strategy in the knowledge that other players will apply the same rule. There are several solution concepts and the choice of a concept will depend on: first, the type of the game; second, the question

---

83 Solution concepts usually apply to games with two players. When the number of players increases the strategic complexity growth exponentially. Aumann and Dreze confess that when a game is not a two-person-zero-sum game (2PZS) none of the equilibrium theories tell the players what to expect. "Both CKR and CPA are needede [in a 2PSZ game], without either one, the result fails." (Aumann and Dreze, 2008, 73). Three players are already enough for entirely new phenomena and structural properties to arise; in particular it is possible for players to form coalitions. Usually introducing coalitions will reduce a game with more than two players to a game with two players.

84 "Nash equilibrium is of central importance in studying norms of strategic behavior, (..)." (Aumann and Dreze, 2008, 80) I use social rules as a general term for both conventions and norms. I assume that conventions only put restrictions on the possible strategies, while norms might also change the preferences of the agents, thereby changing the pay-off structure. Not all stable equilibria represent norms, however. Playing a dominant strategy is a solution concept and can therefore create a stable equilibrium. But it cannot be called a social norm.
whether it is played once only or repeatedly (by the same players); and third, whether the strategic choices are taken simultaneously or sequentially. I will start with games in which the choices are made simultaneously and that are played once only.

When the mistrust among players is deeply rooted they will choose a strategy that guarantees a certain payoff regardless what their opponents do. In that case one chooses the highest secure payoff. This approach is called the minimax or the maximin strategy. It secures the best of all the worst outcomes, while avoiding the risk of being found out by a rival. An obvious rule is that dominated strategies are (first) rejected. A strategy that is inferior in comparison to another strategy, whatever the other player might do, will be eliminated. When, after the elimination of dominated strategies, there still remain several strategic possibilities, then one will select one which is rationalizable on the basis of the beliefs an agent has about the beliefs of the opponent which are consistent with the information about the game. Player R knows that player K has some opinion about what R will do and will select a strategy that K thinks is the best answer to the strategy of R. This is common knowledge. Rationalizable strategies are based on such common knowledge (CKR). A rationalizable strategy does not need to be a unique strategy. In a game several strategies can be rationalizable. Nash-strategies are the only rationalizable strategies that confirm the expectations on which they are based. The Nash-equilibrium counts as the only sustainable outcome of rational strategic choices in the absence of externally imposed agreements. In the prisoners’ dilemma, for instance, there is a unique Nash-equilibrium; defecting is

85 I will only give a rough sketch of the main issues in the discussion about solution concepts. Readers, interested in more details, are referred to the literature (for instance; Hargreaves Heap and Varoufakis, 1995)
the best answer to defecting. For a long time it was thought that the Nash-strategy was the unique rational action in strategic decision situations. Nowadays, this opinion has changed. The most important reason is that several Nash-equilibria turned out to be possible (in the 'game of chicken', for instance, there are two sustainable equilibria). The most important question is how can one deduce which one will be implemented. All kinds of attempts have been made to reduce the number of possible Nash-equilibria. Selten introduced the construction of a "trembling hand", to create the possibility that a player makes an unexpected move, which does not have to make his opponents believe that he is irrational. This creates the so-called "perturbed" game. In perturbed games also dominated strategies can be played, although the chance is negligible. This minimal possibility can lead the players to a unique equilibrium outcome. This is called the "trembling hand perfect equilibrium".

The 'mixed strategy' was originally seen as a tool for the solution of games with several Nash-equilibria or when no combination of (pure) strategies corresponded to a Nash equilibrium. The mixed strategy is a probabilistic combination of (pure) actual strategies in situations in which a player has no clear preference for strategy K1, K2,...,Km; for all are potential Nash-equilibria strategies (given probabilities). What does one do when one does not have a preference between options? One randomizes! The idea that strategic choices were determined by chance was, however, a worrisome idea because there is only one pair of randomisations which is consistent with each player not knowing what to do and still arriving collectively at a symmetrical Nash equilibrium (Nash equilibrium mixed strategy: NEMS), which awards equal rewards to each player. (see Hargreaves Heap and Varoufakis, 1995, 73) But why should each player play NEMS? Aumann suggested that the probabilities in which strategies are mixed should not be interpreted as the individual's probability of selecting one pure strategy or the other, but be thought of as the subjective belief that K holds about what R will do and vice versa. In
this situation equilibrium is a relation that holds between beliefs (about probabilities) and not between strategies. The only chance that a unique "Nash equilibrium mixed strategy" will result is that the subjective beliefs each player holds about what the other will do are consistent. We must assume a "consistent alignment of beliefs" (CAB). Preconditions to that end are: a "common prior assumption" (Rizvi, 1994a) and the stochastic independency of guesses (Brandenburger, 1992). The last condition says that the estimates of player R are independent of the strategic choices of the players K, L, M etc. and vice versa. The first condition says that the players employ the same definition of the situation in terms of the dispersion of probabilities for all states of the world. Novel information is processed by means of 'Bayesian updating' (what, in fact, is only allowed in the case of closed systems). The idea behind all of this is the idea that players with the same information will draw the same conclusions (the Harsanyi doctrine); that is, that rational players with the same information never can "agree to disagree" (Aumann).

86 The assumption of CKR is sufficient for the creation of a Nash-equilibrium. But the creation of a unique Nash-equilibrium demands the assumption of a CAB. When, in mixed strategies, there are more than two players, the knowledge requirements increase significantly and involve the CKR, which on its turn assumes the common prior assumption (CPA) and the common knowledge of conjectures. "(..), if common knowledge of conjectures is required for a Nash equilibrium to be well founded, this seems to be too much knowledge for players to have. They must all know, know that each knows, know that each knows that each knows (and so on) what each conjectures each other's play to be." (Rizvi, 2007, 308)

87 The common prior assumption asserts that players have the same beliefs prior to any information. Difference in beliefs can be attributed to difference in information. But when agents reveal it, this will lead others to revise their beliefs. (see Rizvi, 2007, 310) This process is supported by the assumption of perfect symmetry, which says that agents will follow identical (symmetrical) rules of behavior (whether they follow the same principles of rational behavior or are subject to the same psychological laws. (Harsanyi, 1956, 149) This requirement is the condition of perfect symmetry of the games participants. This requirement literally annihilates the players' individuality and turns them all into an undifferentiated homogeneous agent.
Another defense of NEMS comes from Harsanyi. He showed that NEMS is a useful addition to the Nash project because it enables the Nash equilibrium concept to be applied when there is no Nash equilibrium in pure strategies. "Harsanyi defends NEMS for the reason that we rarely know for certain what type of player we are playing with. (..) Thus the Bayesian equilibrium concept often turns out to be the most appropriate equilibrium to use and it so happens that when the doubt in these games shrinks to zero NEMS emerges as a Bayesian solution." (Hargreaves Heap and Varoufakis, 1995, 77) The construction of the Bayesian equilibrium does require both CKR and CBA and in this sense the defense is very similar to Aumann's solution. The difference is that in the former example the selection of a pure strategy is psychological determined, whereas Harsanyi selection depends on the selection of the type of player.

Does the solution concept of games in which players make their strategic choices one after another differ from the solution concepts in which players have to decide simultaneously? If they don't know the choice of their opponent, then the situation is not different from the one in which they have to decide simultaneously. But when they do know the decision of the other player, how does this information change the equilibrium solution of the game?

When players choose sequentially, the order of the moves is important. The player who will make the last move is able to optimize on the basis of the choices of the other players. In sequential games there usually is a chain of decisions: first R has to make a decision, then K has to decide which move she will make, then R again etc. The player, who has to make the first move, will of course take into account what the reaction will be of the other player on his decision. According to game theory the player whose turn it is will reason backwards from the last segment of the game (or 'subgame') to the segment where he at present must decide on his move ('backward induction').

A segment or a subgame starts at a stage of the game where the player whose turn it is, knows what has happened
previously. The general opinion is that a rational player, faced with the choice to make a stand for a collective interest or to choose his own interest, will always defect in the last game, because it has no consequences for her. The player who has to decide in the last but one subgame is aware of this and will defect right away, because he knows that there is no sense in making a cooperative move in the expectation that his opponent will reward this move in the next (and last) subgame. This reasoning can be stretched and the only "subgame perfect Nash equilibrium" is to defect right from the start. This is known as the "backward induction paradox". The intuition behind this solution is that only strategies that are the best replies to each other in each segment of the game are rational, and such strategies demand backward induction. The only exception is a situation in which the chain of decisions is infinite and nobody knows what will be the last subgame. In such a situation everything is possible.

To conclude this exposition on non-cooperative games, I will introduce an important distinction, namely, the distinction between games that are played once only and games that are played repeatedly. When games are repeated with the same players, then the number of strategic possibilities increases considerably. When players know that they will have to deal with each other in the future, they might not be interested in the question how to behave in each separate case, but how to

---

88 The 'paradox of backward induction' is that rational players will defect right from the start, although this is, intuitively, not plausible, given the higher payoffs that can be realized in subsequent subgames. Sugden and Pettit (1989), therefore, think that conditional cooperation (tit-for-tat) is a more obvious strategy. When the number of subgames is large enough, a tit-for-tat strategy is rationalizable.

89 This strategy results in a unique Nash-equilibrium in each subgame, because the combination of 'CKR' with 'backward induction' introduces the 'CAB'.
act with an eye on future events. This knowledge will create some kind of a ‘supergame’ in which all kinds of new options come into being. Players can punish or reward, and all kinds of learning- and reputation-effects can occur. How a player judges these options will depend on the way he balances rewards now and in the future. When a player attaches sufficient worth to rewards in the future, he will, according to some theorists, always prefer cooperation. (Schick, 1992) This leaves unimpeded that many Nash-equilibria are possible. In infinite iterated or indefinite iterated games almost every strategic combination is a Nash-equilibrium (this is known as the "folk-theorem"). In order to know whether a unique equilibrium will, in due course, be established, theorists have taken refuge in evolutionary game theory (EGT). An EGT model can be viewed as the analysis of a dynamic process that describes how populations of players adjust their behavior over time in response to the pay offs that arise from repeated interaction. Successful behavior prevails not only because market forces eliminate inferior one's, but also because agents abandon the latter and imitate the former.

The equilibrium concept most frequently used in evolutionary game theory is the "evolutionary stable strategy" (ESS). The basic idea behind this strategy concept is that it is a strategy (I) which, when used among some population, cannot be 'invaded' by another strategy because it cannot be bested. So when a population uses a strategy I, 'mutants' using any other strategy J cannot get a toehold or expand among that population. Let us define the expected utility for a player for using strategy I when the other uses strategy J as E(I, J). Strategy I is an evolutionary stable strategy when the following two conditions hold: first; it must be at least as good a reply to itself as any other strategy, otherwise people will drift away from its use (E(I, I) > (J, I)) and secondly, I must be strictly better than J when playing against I, or, when this condition does not hold, I must be better when playing against J than J is when playing against itself (E(I, I) > (E(J, I) or E(I, J)) > (E(J, J)). The second condition must preclude a successful invasion of another strategy. Formulated
differently, 'the first condition is simply the definition of a symmetrical Nash-equilibrium and the second a refinement which is similar in spirit as the 'trembling hand'. (Lyons, 1992, 126) Each ESS is a Nash-equilibrium, but not each Nash-equilibrium is an ESS. In iterated games tit-for-tat is an equilibrium strategy, for it is as least as good against itself than any other strategy and it is better than the strategy of defecting than this strategy when playing against itself. Conditional cooperation, therefore, could come into being 'spontaneously' and survive in an evolutionary process.

Strictly speaking, evolutionary game theory is not about individual decision-making. Instead of focusing on the calculating individual, as does traditional game theory, evolutionary game theory focuses on the survival properties of strategies themselves in environments in which they have to compete with each other. Individuals are interesting merely as carriers of behavioral patterns. Essential are the way genes replicating themselves through the survival of individuals who are the carriers of genes. Sugden criticizes evolutionary game theory for surplanting biological theories about natural selection to the economic domain. In the domain about decision-making we need theories about learning and imitation; theories that teach us what gets replicated and how. Only then can we apply theories of natural selection to economics. "[But] within evolutionary game theory, surprising little work has been done to investigate how learning and imitation actually work. (..)" Sugden, 2001, 123) An adequate evolutionary game theory would include a theory of how ideas of salience evolve. But evolutionary game theory allows no space for salience in its explanations. "What we see is not so much a challenge to the theory of rational choice as a superficial restyling of it". (Ib., 127)

To complete this search for solution concepts, I will briefly discuss solutions for cooperative games; specifically, coordination games. In contrast to cooperation problems, solutions to coordination problems can be established relatively easily. As agents have an incentive to coordinate
their activities, as is indicated by their preferences, and as conventions are said to be self-enforcing, one would intuitively predict that conventions emerge spontaneously. It is only when communication is not possible and multiple coordination equilibria are imaginable, that the question arises how a specific equilibrium is selected. Usually, three clues are mentioned which could give rise to a particular solution: focal points, rules of thumb and precedents.\footnote{Lewis mentions also agreement as a means to solve a coordination problem. This solution is recommended when agents are involved in a single unrepeated coordination problem. (Lewis, 1969)}

Of course, solutions will not always become conventions. Conventions do not arise instantaneously. They are accepted and established as successful solutions to recurrent coordination problems. Only when agents jointly accept a procedure, can it become a convention. "Once the parties jointly accept a principle of the right sort, they will finally have positive grounds for expecting future conformity." (Gilbert, 1990, 18) As a convention has the character of a (positive) merit good, it is self-reinforcing for it becomes more effective (valuable) as more agents conform to it.

5. Game theory: some conclusions

Game theory has turned out to be a very useful tool to elucidate coordination- and cooperation problems. But it also confronts us with numerous theoretical problems. Whereas conventions may be relatively easy to establish, the issue is much more complicated in the case of the development of behavioral rules which must facilitate cooperation.

Decision processes to cooperate require usually small groups in which norms are shared and reciprocity is vital. (Olson, 1971) The explanation of the spontaneous development of social rules to regulate cooperation problems by means of non-cooperative game theory, therefore, is not promising (see Bates, 1988; Elster, 1989; Field, 1984; Hechter, 1990; Rutherford, 1994) The main problem is that there are
many equilibrium strategies and the selection of one of these equilibria already presupposes norms. This problem manifests itself also in evolutionary game theories (see Binmore en Samuelson, 1994).

In short: in simple games the conclusions are discouraging, because defecting appears to be the dominant strategy. In complex games there is confusion as to the question what the right analysis is, because there are several rationalizable strategies.

The tit-for-tat strategy that is offered as a solution for iterated prisoners' dilemma games is basically the conditional cooperation that counts as the solution for the assurance game. (see Sen, 1967) The solution to a prisoners' dilemma, therefore, would be the transformation of this game into another game. It is not very likely that this will happen spontaneously, because it demands a counter-preferential choice in the short term and a change of the preference ordering in the longer term. A solution is more likely to result from the intervention of a third party that has the capacity to impose sanctions and thereby is able to change the payoff matrix. Spontaneous solutions are only likely to be established when players submit to pre-existing norms which induce them to cooperate. This is a setback for the attempt to employ a reductionist' strategy starting from a state of nature.

In the course of integrating game theory into his overall project, Harsanyi has carried out several reductions and generalizations. A striking example is his reduction of a game with incomplete information to the status of a special case of a theory with complete, but imperfect, information. Uncertainty has become a strange kind of uncertainty in his approach to game theory: all eventualities are known, including their probabilities. Harsanyi and other theorists in fact employ some kind of a rational expectation hypothesis in the guise of the concept of the 'common alignment of beliefs'. This hypothesis will materialize, Harsanyi suggests, because all players use the same (rational) theory and draw the same (rational) conclusions. This is the so-called Harsanyi/Aumann doctrine that has been mentioned before. Rational
people cannot disagree. There can be no false mutual expectations. Ignorance may be possible, but not disagreement. This is a controversial doctrine. It is based on the idea that the beliefs that players cherish are converging. But a convergence of beliefs is unlikely when players are attracted by different theories or do not draw the same conclusions from the same theory. In the Harsanyi/Aumann doctrine the expectations of the actors have been upgraded into (intersubjective) knowledge. This again is in fact the formulation of an equilibrium condition. Game theory, therefore, is becoming ambivalent.

But what counts in its favor is that it takes the interdependencies of agents as point of departure. This is a very welcome complement of choice theories that depart from isolated, individual choices. Game theory could be considered as a more general theory of individual choice-behavior. It has been most successful at the micro level. Many game theorists are engaged in experimenting with all kinds of game models in order to find behavioral regularities that would be helpful to explain social problems of all kinds. It has been less successful on the macro level, the level that is discussed in this chapter.

In its present form there are many arguments that weigh against macro game theory. Take, once again, the assumption of the 'consistent alignments of beliefs'. Given this assumption rational behavior is in fact incompatible with attempts to confound the opponent, or to deceive, or to eliminate him and so on. And for the same reason a conflict becomes a paradox. Because, when a conflict threatens, the players can predict the result, because they command the same rational theory and, therefore, they must be able to negotiate this result (or a better one) (Varoufakis, 1991). The only way that conflicts can arise is when players make

---

91 CAB depends on a conception of internal consistency that applies across agents. Given the individualistic presuppositions of game theory (rational choice theory, in general) a criterion of inter-agency consistency is inconceivable. This suggests that CAB is just an equilibrium condition.
mistakes ('trembles') or when the communication is disturbed. Conflicts are only a fleeting thing. Systematic errors are excluded because the players are rational and communication problems can be solved. 'Another troubling point is that what is rational for actor A to do depends on the probabilities he assigns to the actions of actor B. But at the same time it is assumed that A knows that B is as rational as he is, which means that the probabilities he assigns to B's actions depend on what he himself thinks would be rational. This kind of circular reasoning turns the 'common knowledge of rationality' into another equilibrium condition.' (Sugden, 1991, 783). Solutions in game theory seem to be dependent on strong equilibrium conditions. One that pictures idealized agents (with rational expectations).

The question, moreover, is whether the other assumption, the CAB assumption, is compatible with decision theory. In Savage’s approach, subjective probabilities are defined in terms of an agent’s willingness to take bets on an act’s occurrence. "This approach cannot work unless we are free to construct artificial acts (the hypothetical betting options) by assigning suitable consequences to the event whose probability we wish to measure. If, as Savage intended, events are understood as states of nature, then we do indeed have this freedom. But it is not clear that we have the same freedom if events are identified with strategic choices of rational agents [and agents have to assign subjective probabilities to their opponent’s strategies]." (Hollis and Sugden, 1993, 29)

---

92 Take the case in which players differ both in beliefs (probabilities) and values (utilities), and therefore the ranking of their preferences. When we assume that there are only two players and two acts and we impose only a weak Pareto condition and we ask the question: is it possible to construct a shared preference within a Bayesian rationale? Then the answer is that no such compromise exists. Only autocratic solutions (in which one of the agents imposes his preferences on the shared preference) conform to the weak Pareto condition. There is no room for a Bayesian compromise (see Seidenfeld et. al., 1989). The ‘solutions’ in game theory are in fact equilibrium conditions.
The relation between decision theory and game theory is, therefore, troublesome. "The central problem seems to be this: any theory of rationality has to make a distinction between a realm of agency and a realm of causation. If we apply this distinction to games, a player has to see his strategies both as acts (for himself) and as states (for his opponents); To treat behavior both as willed and as naturally caused leads to incoherence". (Sugden, 1998, 159) This incoherence is a consequence of Savage's theory that requires that states be defined independent of acts. A solution would be to take each player's beliefs about the other player(s) as given.

Recently an evolutionary approach to human behavior has gained currency. The capacity of species to adapt themselves to their environment is significant for their chances of survival. The concept of rationality, it is suggested, should be replaced by the term adaptability as a reference to the capacity of the human cognitive processes to enhance the chances for survival and procreation. Social animals may evolve gene structures which facilitate the learning of cooperative behavior, for when cooperation has significant advantages for group survival, the group will adopt mores and laws which condition people to behave as if they are maximizing collective or social utility.

The 'evolutionary turn' in game theory was welcomed because it seemed to offer an escape from the problem of equilibrium selection. But till now it has not been very successful in solving problems. Therefore, one wonders why other non-evolutionary ways of tackling this problem were not followed. Sugden refers to the solution that Schelling offered. "(...) Schelling analyzed the equilibrium selection problem and showed how (...) equilibrium selection depend on shared conception of prominence (...) which allow individuals' expectation to converge on particular equilibria, or focal points." (Sugden, 2001, 117) However, these concepts were never integrated in the formal structure of game theory. Why not? One reason, suggested by Rizvi, is that there was a big difference in style. Schelling developed his concepts with a minimum of mathematical tools. His concepts rely on
induction "whereas the dominant approach in game theory is technical, mathematical and deductive". (Rizvi, 2007, 300) Schelling himself supported this argument by pointing out that game theory had not developed into an interdisciplinary and applied branch of social science, but had remained 'at the mathematical frontier'. (cited by Sugden, 2001, 117)

Innocenti thinks that Schelling remained a relative outsider because he had a quite different view on agents playing games. "A fundamental assumption of standard game theory is external symmetry, according to which players are all alike outside the game and any differentiating characteristic has to be embedded into the game's mathematical form." (Innocenti, 2007, 410) In contrast, Schelling assumes heterogeneous players. "There is no way to define deductively how players' beliefs converge because this process is the outcome of individually specific processes. (...) As a consequence, the game solution is not the result of an instantaneous convergence to equilibrium but the final step of the dynamic adjustment process of the beliefs of heterogeneous players." (Ib., 424)

There are three ways to involve evolutionary theories in rational choice theories: as an improvement on game theory (evolutionary game theory); as a supplement to decision theory (decision procedures are based on the same principles as evolutionary processes, they evolve via blind variation and selective retention); and, third, as residues of our evolutionary past they manifest themselves in the cognitive and affective aspects of our decision processes.

Personally I do not believe that evolutionary theories can contribute a lot to our understanding of the development of human society. The human species is a product of both evolution and development and I think that these processes should be kept apart. Evolution is a process that stretches itself over millions of years. In this perspective the development process is a short-term process, in which the gene-structure of the human species remained unchanged. But it is in this short-term process that human institutions - in amazing varieties- evolved. The development from a state of
nature to a society with laws, conventions and norms can be explained without reference to evolution theory. Evolutionary theory is, moreover, of no help for explaining how social institutions are supported by moral beliefs and why people think that they ought to keep to these social rules. "The belief that one ought to follow a convention is the product of the same process of evolution as the convention itself". (Sugden, 1989, 87)

I am aware that many people use the term evolution as an equivalent for development. But evolution is associated with natural selection and the survival of the fittest and therefore is suggestive of the impression that the existing institutions are the best, otherwise they would not have survived. The term development does not excite this false impression and, therefore, is to be preferred. I will only occasionally refer to evolutionary theories.

---

93 Such an argument could strengthen evolutionists in their belief that the biological evolution established our capacities once and for all and that these evolved capacities facilitate the processes of cultural evolution.

94 There is also another idea that is fueled by evolutionary theory and that is the view that evolutionary theory excludes agency. Darwin identified selection, variation and heredity as the three basic conditions for natural selection to bring about evolution. Dawkins suggested that genes rather than the organisms are the real ‘agents’ in evolutionary theory. The organism is just the ‘vehicle’ of the selfish genes. The image that Dawkins is fostering is that of a straightforward chain running from genes to neurophysiologic processes to overt behavior. Anyway, evolution refers to the human specie not to individuals.