

- On positivism and 'scientific' approaches to social explanation more generally, see Kuhn (1970), Hempel (1965, 1966), Hanson (1958), Halfpenny (1982) and Chalmers (1990).
- On the philosophical origins of behaviouralism, see Carnap (1936, 1950), Schlick (1974) and Ayer (1971).
- For a useful explanation of some of the terms used in these studies, see Lacey (1976).
- For justifications of quantitative approaches to the analysis of empirical evidence in the social sciences, see Blalock (1964, 1969, 1970, 1972) and King (1989).
- For a recent summary of the ways in which qualitative data can be employed within the 'behavioural-scientific' approach, see King *et al.* (1994).

Chapter 3

Rational Choice

HUGH WARD

The essence of rational choice theory is that 'when faced with several courses of action, people usually do what they believe is likely to have the best overall outcome' (Elster 1989a: 22). I will argue that rational choice is an indispensable part of the toolkit of the political scientist, because there are important political phenomena which it can partially explain. Nevertheless I do not claim that rational choice theory is free-standing (cf. Almond 1990). It needs other perspectives to help explain why individuals have the interests they do, how they perceive those interests, and the distribution of rules, powers and social roles that determines the constraints on their actions. First I briefly sketch how rational choice methods have developed over the last 40 years. Then I explain what rational choice modelling involves and examine the epistemological underpinnings of the method. I elaborate my argument that rational choice is best regarded as a toolkit rather than as an approach by considering the array of criticisms that have been made of it. Finally I consider some recent developments in rational choice, partly to show how rational choice theorists have responded to the criticism.

The development of rational choice theory

Rational choice arose as part of the behavioural revolution in American political science of the 1950s and 1960s that sought actually to examine how individuals behaved, using empirical methods (see Chapter 2). It has arguably become the dominant approach to political science at least in the United States. However, rational choice draws on the methodology of economics in contrast to behaviouralists who drew on sociology or psychology (Barry 1970). Anthony Downs (1957; cf. Downs 1991) was the pioneer in the application of rational choice theory to electoral behaviour and party competition and his work revolutionised electoral studies (reviewed in Hinich and Munger 1997). The individual votes for the party which, if it got into office, is expected to yield them the highest utility. Parties are assumed to be motivated solely by the desire for office, competing for votes by changing their policy platforms.

From Downs's pioneering work rational choice has flowered in a variety of directions. Mancur Olson (1965) showed that self-interested individuals would not always take part in collective action to further a shared goal. For example, why do so many of us continue to act in ways which harm the environment even though we know what we are doing is anti-social? A plausible explanation is that we feel changing our own ways will have little or no impact on the overall problem and there are major financial and other costs associated with living differently. The result is a collective action failure in which rational self-interest leads to everyone being worse off (Hardin 1969). His work constitutes a fundamental critique of pluralism and orthodox Marxism, which both assume that a shared interest is sufficient for political mobilisation to occur. It has generated empirical work in areas as diverse as the study of social revolutions (for example, Popkin 1979) and cooperation between states over such problems as the deterioration of the global environment (for example, Sandler 1997).

Game theory deals with situations where others' choice of strategy affects your best choice and vice versa. It has led to important developments in collective action theory, enabling us to explain why collective action failures can sometimes be avoided if the number of individual decision-makers is small (Axelrod 1984; Taylor 1987). Extensive use has been made of game theory to model nuclear deterrence, arms races, disarmament and other phenomena important to international relations specialists (Nicholson 1989; Powell 1999). It has also been crucial to attempts to explain the formation of legislative coalitions (Riker 1962).

The sub-field of social choice theory developed when economists asked whether any satisfactory and broadly democratic way could be found of aggregating the preference of individual citizens so as to arrive at a social ranking of alternatives. An example of such a procedure is to use simple majority rule, ranking x above y if x can gain more votes than y . This method has long been known to lead to a paradox where there are multiple alternatives (Maclean 1987). The key theorem, first proved by Kenneth Arrow (1951), is that no satisfactory democratic method of aggregation exists, so that the problem is not peculiar to simple majority rule. This result has led to further fundamental questions being asked about democracy (Sen 1970). For some authors, results like Arrow's, together with related results about tactical voting and agenda manipulation (Farquharson 1969; Gibbard 1973), call into question the idea that democracy is the implementation of the popular will represented by a social preference ranking (Riker 1982).

The central theme of the public choice sub-field is that the intervention of democratic governments to repair market failures often creates more problems than it solves. One argument is that the combination of the self-

interest of bureaucrats in maximising their budgets and bureaucratic control over information on the cost structure of state provision of public goods results in their over-provision, at the expense of the citizenry (Niskanen 1971). Another important theme is rent-seeking, organised interests successfully lobbying for monopoly or quasi-monopoly powers and subsidies from states, with consequent erosion of market efficiency and slower economic growth (Buchanan *et al.* 1980; Olson 1982; North 1990). The literature on the political business cycle, based on ideas about pocket-book voting ultimately deriving from Downs's work (Goodhart and Bhansali 1970; Kramer 1971), suggests that the search for electoral success through the manipulation of the economy leads to economic instability and a higher-than-optimal level of inflation (for example, Nordhaus 1975). The normative thrust of public choice theory is towards constitutional limitation of the size and autonomy of the state and disengagement from corporatist entanglements. As filtered through neo-liberal think tanks, public choice was crucial to the development of Thatcherism and Reaganomics (Self 1993).

The intellectual roots of postwar developments run back through microeconomics and welfare economics, nineteenth-century liberalism and utilitarianism, and the work of classical political economists like Adam Smith, to the work of authors like John Locke and Thomas Hobbes. If rational choice theory owes intellectual debts to the liberal tradition, it has made repayments by suggesting lines of analysis and argument. For example, John Rawls's influential work (1972) grounds the idea that, within constraints set by the equal distribution of liberties and certain rights, it is just for society to maximise the well-being of the least well-off members of society. The argument is that individuals who (hypothetically) did not know what social position they would occupy, and are thereby impartial, would rationally accept a social contract embodying a principle protecting themselves against the case in which they turned out to be one of the worst off.

We can now see that rational choice is useful both to those trying to explain political phenomena and to those whose orientation is normative. Moreover, it is far from being the case that rational choice is necessarily wedded to the conservative agenda of its public choice variant, although this is the dominant orthodoxy among US rational choice theorists: while Rawls reaches broadly social democratic conclusions, the method has also been used to take Marxian political economy and the Marxist critique of capitalism further (for example, Roemer 1988). This is possible because what you get out of a rational choice model depends on what you feed in by way of assumptions, and the questions you pose. In the next section I describe the assumptions of the mainstream variant of rational choice theory in more detail.

- If correctly applied it ensures that propositions actually follow logically; so the method can be used to see if a logically coherent basis for widely believed conclusions can be constructed.
- It goes beyond inductively derived correlations to provide a mechanism linking independent and dependent variables, running through the actions individuals take.
- It provides a unified framework of explanation across different fields of the social sciences and across sub-disciplines, allowing cross-fertilisation of ideas and a viewpoint from which common patterns can be seen across diverse phenomena.
- Even in circumstances in which action is irrational, it provides a standard against which action can be judged and indicates variables that might lead to departures from rationality. (Mansbridge 1990b: 20)

Rational choice theory takes individuals' preferences, beliefs and feasible strategies as causes of the actions they take (Little 1991: 39–67). Related to this, rational choice theory is typically seen by commentators as accepting the principle of methodological individualism – that 'bed-rock' explanations of social phenomena should build upwards from individuals' beliefs, strategies and preferences (for example, Almond 1990: 123). That is, rational choice is claimed to be reductionist, aiming to explain things in terms of the properties of individual 'social atoms'. Shortly I will dispute the claim that rational choice is, or indeed can be methodologically individualistic. If it were, it would be committed to an ontology in which only individuals 'really' exist, so that social structures, institutions, roles, norms and the rest of the paraphernalia of sociology are, at best, convenient shorthand ways of talking about individuals. Notoriously this position has been taken by some extreme neo-liberals like Hayek and has been influential in the development of the New Right (Self 1993). This leads to the thought that rational choice has an in-built conservative bias, as reflected by its public choice variant.

Almost all games that are remotely realistic representations of reality have more than one equilibrium, generating the problem of coordination of beliefs. In order to play rationally, players must have a common conjecture that one particular equilibrium will come about. If such a common conjecture is not present, even if players choose strategies corresponding to some equilibrium there is no reason in general that their strategies will be best replies to each other, for different players may focus on different equilibria when picking their strategies. Now a common conjecture is an intersubjective, not an individual fact (Bicchieri 1993). For example, in a two-player game it takes the form 'A and B believe the equilibrium *e* will eventuate; they know that each other believes this; they know that they know this; and so on'. Clearly

such a pattern of beliefs form a system that cannot be reduced to beliefs of any one player considered as an isolated 'social atom'. Contrary to the folk wisdom in political science, it is far from the case that game theory is the exemplar of methodological individualism. Rather it demonstrates the incoherence of reductionism.

As I have already noted, to specify a rational choice model, you need to specify the rules of the game – roughly what players can and cannot do, and what they do and do not know. In practical application this amounts to providing a stylised representation of players' roles and powers. For instance, rational choice models of the relationships between chairs of Congressional committees and the individuals on the floor of the House and the Senate in the USA take as given the rules governing chairs' ability to control the agenda and often postulate that members of specialist committees know more about the consequences of bills in their domain than do ordinary Congressmen/women (Shepsle and Bonchek 1997). Models of this sort do not just comprise facts about individuals; they also include 'institutional facts' about 'rules and roles'. In ontological terms, they partly comprise taken-as-given socio-structural elements. I return to the institutional turn in rational choice theory, which foregrounds this fact (Hall and Taylor 1996: 942–6), below.

Most rational choice theorists are committed to some form of empiricist epistemology (see Chapter 1) – minimally that there are facts about the world, discernible through observation, that are 'independent enough' of the theory under test to be potentially able to refute it, in the sense of showing that some underlying assumption made is false (Nicholson 1983: 40–3). As we shall see below, there is a major controversy about whether rational choice succeeds in empirical terms. In any case it is by no means clear that a user of rational choice theory logically must be committed to empiricism – in any form. Some critics of empiricism try to achieve an understanding of a particular individual's actions from within their own frame of reference, denying that the social sciences should look for general laws well-supported by evidence. Rational choice is not inconsistent with such an interpretive, qualitative enterprise, because it can also be a way of investigating the meaning of others' actions, enjoining us to look at the individuals' desires and beliefs, picturing these as leading to intentions and actions (cf. Hindess 1988: 59).

Criticisms of rational choice theory

In recent years a stream of critical commentary on rational choice theory has appeared in edited volumes and monographs (for example, Barry 1970; Green and Shapiro 1994; Harrison 1994; Hirschman 1995; ...).

1991; Self 1993; Zey 1992). In order to provide a route-map I examine four modes of criticism: (a) the critique of those who wish particularly to emphasise bounded rationality; (b) the sociological critique, which centres on the way rational choice theory appears to downplay social structure and holistic modes of explanation; (c) the psychologists' argument that individuals often do not act rationally in the standard sense and are motivationally and psychologically complex; (d) the critique from mainstream political science, based on the implausibility of the assumptions made and the predictive failures of the model.

Bounded rationality

If nothing else, one would expect rational choice theory to be able to give an unambiguous account of what it means to behave rationally. However, it has failed to do this. While there is general agreement among game theorists that some equilibria do not make sense, there is little consensus on how to 'refine' the equilibrium concept so as to narrow down the alternatives (Hargreaves-Heap *et al.* 1992; Morrow 1994). First, the existence of multiple equilibria reduces the predictive power of the model. Second it is not possible to define what rational action is unless a theory exists about how players coordinate their expectations on a common conjecture, and it is not clear that standard game theory has such a theory (Bicchieri 1993; Johnson 1993). While coordination of this sort might seem merely a technical issue, it is actually foundational for politics. For instance, Cox (1997) shows that many features of the way that electoral systems operate are explained by the need for like-minded groups of voters and political elites to coordinate their behaviour on a favourable equilibrium.

Some rational choice theorists feel that the mainstream model makes highly implausible assumptions about the rational capacity of individuals. Herbert Simon's work (1982, 1985; March 1986) on bounded rationality has been particularly influential. In the face of limited information, limited time and limited cognitive capacity to process information, Simon envisages individuals using heuristics built into standard operating procedures as a shorthand guide to getting a satisfactory result. While some see rational action as only possible on the basis of rationality-held beliefs (for example, Elster 1989a), for Simon action is procedurally rational if it is based on beliefs that are reasonable given the context the actor is in. Decision-makers carry on with what they are doing until the pay-off drops below a satisfactory level; then search until they find another option that is satisfactory. Such a pattern of decision-making will tend to give rise only to incremental policy change, which may be normatively defensible

Etzioni 1967).

There has been a recent resurgence of interest in bounded rationality in political science. For example, Ostrom (1997; Ostrom *et al.* 1994) calls for second-generation models of collective action based on bounded rationality. Game-theoretical models fail to explain patterns observed both in experimental settings and in fieldwork. An alternative approach would acknowledge that natural selection has left humans open to learning solutions to collective action problems and may also have selected for a degree of altruism (Gintis 2000: 237-83). Individuals develop a range of heuristics to deal with the problem of when and when not to reciprocate. They rely on communication and others' reputations for trustworthiness, and they internalise norms of appropriate behaviour which there are intangible costs to violating.

Where do the routines, standard operating procedures and heuristics emphasised in the literature on bounded rationality arise? One way is that players copy the methods of those who are more successful, possibly because of chance discovery of a good heuristic. Emulation of this sort has strong analogies to natural selection (Van Parijs 1981). Routines 'evolve' in repeated game-like interactions between players who 'carry' them as a sort of 'cultural genetic code', as relatively successful ones spread through the population of players by copying. Evolutionary game theory, which developed first in biology but can also deal with social evolution of this sort, shows that given enough time evolutionary pressures ensure that, to the observer, it will look as if players are using routines that form part of a Nash equilibrium, although not all Nash equilibria can be reached in this way (for example, Gintis 2000: 148-236). The reason is that if players' routines are not a best reply to what others are doing, they will eventually emulate others who are more successful. For instance it can be shown that in relatively stable environments, parties adapting their competitive strategies will behave in much the same way that Downs predicted, converging on an equilibrium in the political centre-ground (Kollman *et al.* 1992). If the predictions are the same as those of the standard models, some will ask 'why bother?' (for example, Friedman 1953). Beside being more realistic, the pay-off from an evolutionary approach may turn out to be that it can explain coordination on one equilibrium where there are several in the game, a problem I have suggested that is difficult to overcome in standard game theory. In evolutionary game theory, which equilibrium is reached generally depends on the starting point and dynamics of the process, because of path-dependence. While opinions differ as to the significance of evolutionary explanation in political science (cf. Dowding 2000; John 1998), it seems to me to have considerable potential.

② The sociological critique.

Sociologists often claim that individual behaviour is largely a function of social structures. Choice is illusory for individuals and the rational choice approach based on individual choice is, therefore, unhelpful (for example, Hindess 1988). For instance, some argue that the Downsian approach is inferior to an account of voting in terms of the individual's position in the social structure. Social class, geographic location, gender, consumption and production location and religion, among other variables, all have known correlations, of greater or lesser strength, with voting behaviour (for example, Harrop and Miller 1987). In fact, in voting, as in other domains, an individual's structural location typically does not completely account for what he or she does. Neo-institutionalists often emphasise the way in which the institutional structures of government shape the worldview of politicians and bureaucrats, mould their preferences and define the options they consider when making policy choices (for example, March and Olsen 1984). Much of the time individuals follow rules rather than making choices. Classic case studies such as Allison's study of the Cuban Missile Crisis demonstrate that organisational structures are very important, but that individual decision-makers such as President Kennedy still had enough autonomy profoundly to influence the outcome (Allison 1971). In general, it seems implausible either that individuals are fully autonomous or that their actions are determined completely by social structure (Hollis 1977).

Social structure

Even if social structure determines the individual's feasible set of strategies, their beliefs and preferences, rational choice can add to the explanation by making predictions when the rational course of action is non-obvious. This is especially likely to be the case where there is strategic interdependence of decision-making. For example, it is not at all obvious under what circumstances liberalisers within the existing regime and moderates among reformers will be able to sideline hard-liners and revolutionaries in order to achieve a peaceful transition to democracy. Przeworski (1991) approaches this issue using game theory, showing how outcomes vary with the pay-offs and beliefs of these social blocs. The conclusion is that the only plausible scenario for a peaceful transition is that liberalisers put a greater probability on their former allies, the hard-liners, being able successfully to repress a revolution than the opposition do; so the opposition actually push for change leading the liberalisers to cave in rather than face a blood-bath, forsaking their former allies. While this

capacity of rational choice to reach surprising conclusions from widely accepted 'structural facts' about social blocs and their interests.

Because they favour methodological individualism, many rational choice theorists argue that social structures do not provide basic elements of explanations in the social sciences; rather we need to explain the structure using rational choice theory (for example, Elster 1989a). This can be illustrated by looking at the debate between Theda Skocpol, who tried to explain social revolutions without using voluntarist forms of explanation like rational choice theory, and rational choice theorists like Michael Taylor. Skocpol (1979) used the comparative case study method to isolate a set of sufficient structural conditions for a social revolution, three of which were: external stress upon the state in the geopolitical arena; breakdown in the state's ability to maintain internal order; and strong community structures among peasants. Taylor's point is that these structural factors can be seen as the result of decisions taken by individuals: decisions to attack another power; failures by state decision-makers to invest enough resources in social control; and decisions made by the state which bolstered peasant communities in order to provide a bulwark against classes antagonistic to the monarch (Taylor 1989). Taylor's critique is a powerful one, but structural factors surely shaped the decisions to which he alludes. To take one obvious example: why were there states rather than some other form of rule?

In my view it is a practical impossibility for rational choice theorists to eliminate taken-as-given structural factors from any application of rational choice. As I suggested above, these enter models as the rules of the game. Beside being incoherent, it simply is not practicable to reduce explanation only to facts concerning individuals, as methodological individualism demands (Lukes 1977). I do not believe that practising rational choice theorists typically exhibit much desire to squeeze out structure; rather they often seek to illuminate how choices are made within structures, the agenda sometimes stretching to the consideration of how rational choices reproduce or transform structures. As such, rational choice can form part of a structuration approach. I return to this point when considering rational choice accounts of political institutions below.

Norms

The general points made here about social structure also arise in relation to norms. Drawing on the work of founding fathers of sociology like Durkheim, many sociologists emphasise norm-driven behaviour, with social norms understood as deriving from society's need for system integration. While recognising the possibility of anomic and dysfunctional

angles. First, norms can be seen as conventional forms of behaviour that solve coordination problems. It does not matter whether everyone drives on the left or whether everyone drives on the right: both solutions are equally good from the perspective of preventing collisions *so long as everyone expects others to follow the convention regularly*. In a game-theoretic representation, both solutions would be equilibria, so the coordination problem is another aspect of the multiple equilibrium problem alluded to above. Following the work of the philosopher Lewis (1969), it can be shown that such norms can evolve through trial-and-error learning, in a similar manner to the evolutionary games discussed above (Sugden 1986).

Second, rational choice theorists have seen norms as *injunctions to 'be good'* – to behave in a way that maximises *social welfare* when there is a collective action problem, so that the individually rational thing to do is to free-ride (for example, Ullmann-Margalit 1977; Gauthier 1986). For example, one way to explain why individuals bother to register a vote despite the infinitesimal chance that this will influence the outcome of a national election in favour of their preferred party is to say that citizens gain pleasure out of doing their citizen's duty. In deciding whether to vote they set this incentive against the costs of voting (Riker and Ordeshook 1968; cf. Aldrich 1993). The general implications of this line of thinking are: that people are more likely to conform to norms when this has low costs; and that they do not conform unreflectively. For example, people vote because it is low-cost; they do not participate much in other ways in democracy because it is high-cost. Many sociologists start with the collective action problem when explaining norms (for example, Parsons 1937), though they typically use a functional analysis to explain norms in terms of the all-round benefits they bring and see norms as operating through socialisation rather than through incentives, something that many of them feel is a better description of how norms influence action (Elster 1989b: 106–7). While equating norms with moral incentives to 'right action' is common enough among rational choice theorists, some argue that this is inadequate because it ignores the corrosive effects of self-interest on collective action: those who do not adhere to the norm, or pay no part of the price of enforcing it, may, nonetheless, benefit if others conform (for example, Taylor 1987: 29–30). There are also empirical problems that I return to below.

Ideologies

Yet another variation on the basic sociological critique concerns ideologies. Ideologies can be seen as structures of belief, assigning meaning

meaningfulness to the individual (for example, Winch 1958). Many sociologists would argue that action can only be seen as rational or irrational within the context of a particular system of meaning, or discursive formation. In addition, action often cannot be interpreted from an instrumental perspective. Indeed, symbolic and ritual action are crucial to politics (Edelman 1964). Individuals' identities are formed in complex social processes in which discourses form and re-form, giving only limited autonomy to individual human subjects. Processes of identity formation of this sort are crucial to belief and preference formation, again suggesting that important elements of the rational choice model are given by discursive social processes unamenable to rational choice methods.

These criticisms are certainly significant, but the same counter-arguments apply. It is widely recognised in recent work on ideology and discourse that there is some individual autonomy from ideological determination; and ideological structures arise and are reproduced and transformed as the result of individual action, some of which is instrumentally rational (Norval 2000). To expand on this, individuals often combine elements of one or more ideologies in novel ways with a view to instrumentally furthering an interest, and this can have profound political effects. Moreover, while the view that ideology is false consciousness has been rejected by some on the grounds that there are no universal standards against which to judge the truth, ideology is now often defined as a system of belief that serves the interests of some and disadvantages others (Norval 2000). Party competition can surely be illuminated by this idea. For example, Thatcher's Conservatism drew on liberalism and traditional strands of Conservatism and was, to a degree, a deliberate construct. Few will deny that it is associated with increasing inequality in British society.

Can rational choice do any more than this to illuminate how ideological structures change? I believe it can, as the work of William Riker on the manipulation of issue dimensions in democracies illustrates (Riker 1982). Drawing on formal results from spatial theories of voting and elections (for example, Ordeshook 1986), Riker shows that politicians may destabilise majorities by inserting extra issue dimensions into the debate and may solidify majorities by encouraging the separate consideration of issues. While Riker sees such strategies as expressions of elite self-interest and anti-democratic, others have seen them as forms of statecraft which may be conducive to the general good (Nagel 1993). One way to develop Riker's argument – not necessarily the way Riker would have developed it himself – is to suggest that behind the manipulation of issue dimensions there lies the construction or mobilisation of ideologies which 'organise in' or 'organise out' certain questions and the interconnections between them.

Riker's argument makes very transparent how such ideological movements may be linked to the electoral fortunes of parties and the legislative fortunes of policies.

Rules and conventions

Organisational sociologists argue that, even if collective actors go through processes of deliberation with a view to achieving given ends, the processes are liable to be strongly influenced by: rules and conventions used to categorise problems; paradigmatic filters biasing the use of incoming information; limited efforts to search for available solutions; pressures to appear consistent, even at the cost of failures of goal attainment; the upgrading of means into ends in themselves; and other organisational pathologies (March and Olsen 1984; Hindess 1988). Often decisions are emergent, result from conflicting strands of deliberation, or from inaction due to problem avoidance or internal political gridlock. While many of these phenomena have their analogues at the individual level (see the next section), they pose a challenge to the way rational choice theory typically treats collective actors.

It is often held that rational choice pictures individuals as isolated social atoms – autonomous sources of social causality in the social process. In contrast, the focus of much sociology is upon individual interrelatedness. It is not that relationships exist between fully constituted individuals: rather relationships modify individuals' identity in crucial ways. The atomistic picture painted by rational choice theory is said to be in line with other individualistic ideologies which support the social status quo by denying the existential reality of social groups, communities, social classes, and even societies. At the same time, forms of political action which affirm individuals' social identity and which are not based on self-interest are denied the validating stamp of rationality (Benn 1976; Sen 1970). The very concept of rationality which rational choice theory celebrates is said to be historically and culturally specific to capitalist societies. Its logic is said to drive out other rationalities and forms of understanding, especially any notion of rationality which problematises the goals to which action is orientated (Dryzek 1990). In short, the rational choice picture of the political world is a distorted reflection of a reality only approached in capitalism, generating forms of understanding of the political realm which prevent all but shallow criticism of the social status quo (MacPherson 1970).

It seems to me that rational choice theory need not be committed to viewing individuals as isolated social atoms any more than it is committed to seeing them as self-interested: rational choice modelling starts with

notion that instrumental rationality first arose with the capitalist market economy is surely historically indefensible: as one mode of human action it has always been important outside the immediate family circle (for example, Sahlin 1972: 191–204). I showed above that there is an element of social inter-relatedness, concerning common conjectures, that cannot be squeezed out of rational choice explanations – unless it be by invoking a *social* process of evolution.

3 The psychologists' critique

Psychologists typically argue that individuals' motives need not reflect self-interest: envy is important and is incompatible with self-concern; and drives such as revenge, guilt and greed may exist, whether or not they are consciously acknowledged. Critics have been especially worried by the exclusion of altruism from most rational choice models of politics (for example, Lewin 1991; Mansbridge 1990a). They argue that the empirical evidence suggests that individuals frequently act altruistically in political life. For example, while individuals' personal economic expectations may influence the way they vote, there is considerable evidence that the general state of the economy also matters, suggesting that voters are often also concerned about the well-being of others (for example, Sears and Funk 1990). When individuals act in accordance with social norms, there also often seems to be some sacrifice of self-interest.

Normatively orientated rational choice is not wedded to the self-interest assumption. For example, social choice theory makes no assumptions about the motives which lie behind individual preferences, being concerned only with the problem of how they might be aggregated so as to make a choice for society. Rational choice theorists interested in explaining political phenomena have always been aware that altruism is important (for example, Downs 1957: 29). Their position has often been that applications of rational choice should be confined to those areas where self-interest dominates: a conclusion supported by some of their critics (for example, Green and Shapiro 1994). For example, Olson suggested that his theory of collective action would apply best to economic interest groups and not to philanthropic ones (Olson 1965: 64–5). The question then becomes how much room such a self-denying ordinance would leave rational choice theorists in which to operate.

One way around the problem of altruism is to suggest that individuals get pleasure out of others' happiness. It is not difficult to model such a phenomenon in terms of positive utility interaction between individuals (for example, Collard 1978). Margolis's model also allows for change in the relative weight attached to self-interest and others' interests, more weight being placed on self-interest

80 advocate much more extensive use of this sort of modelling (Mansbridge 1990c; Gintis 2000), but it raises methodological issues that I will return to below.

It has become clear that some forms of altruism can be disguised forms of self-interest. Biologists have pointed out that, because kin share genetic material, self-sacrifice in favour of close kin might increase the chances of copies of your genes surviving. So, kin altruism might be selected for in evolutionary processes. In addition, it may pay in evolutionary terms to help another now in the expectation that they will help you in the future, so that reciprocal altruism may also have an evolutionary basis. Game-theoretic collective action theory has done much to clarify the conditions under which such reciprocal altruism may occur in contexts where it is consciously entered into, as well as where it is selected for in processes of social evolution: 'nice' actions must be conditional on others being 'nice' in the past, with punishment for those who were 'nasty'; the interaction must not have a definite time limit; the individuals should not be too short-term-orientated; and short-term benefits from being 'nasty' should not be too high (Axelrod 1984; Taylor 1987; Frank 1992; Gintis 2000).

Many psychologists regard synoptically rational decision-making, approaching the ideal of the mainstream rational choice model, as relatively rare (for example, Rosenberg 1991; Janis and Mann 1977: 21-3; cf. Wittman 1991). Beside the cognitive limits emphasised by authors like Herbert Simon, emotions and unconscious drives make the level of detachment necessary for the synoptic approach highly unlikely in many settings (Elster 1989a). Decisions are often made more on the grounds of consistency with past actions, reduction of strains within the individual's belief system (cognitive dissonance) or normative orientation than through a calculation of the most efficient means to given ends. The norms that the individual adheres to and the affective orientations they have may prevent feasible options being considered and relevant information being obtained, as well as biasing decision-making away from what is instrumentally rational (Etzioni 1992).

Decision conflicts occur when individuals can find no alternative that simultaneously satisfies all their goals. This creates problems for normative decision theory (Levi 1986) and it also tends to generate behaviour which is irrational. Decision conflicts are a source of stress. Whatever course of action is chosen there appear to be losses; there are simultaneous opposing tendencies both to accept and to reject a course of action (Janis and Mann 1977: 45-6). Decision conflicts also lead to vacillation attempts to avoid making a choice at all and forms of apprehensiveness which tend to lead to poor decision-making (Janis and Mann 1977). Regret about past decisions, made when decision conflicts were not resolved, may immobilise the

81
speculative, and perhaps unconscious, rationalisation of the idea that the chosen alternative is the best – is liable to occur if a choice is made at all (Janis and Mann 1977: 91-3). Where commitments to an existing path of action are strong, individuals 'bolster', carrying on in the same way, and freeze out consideration of other alternatives, even if they are aware that to do so is not necessarily desirable (Janis and Mann 1977: 15). Case studies of areas like foreign policy decision-making suggest that such pathologies are probably widespread in political life (Janis 1972).

New information is often not dealt with in a neutral way. Rather, it is fitted into existing patterns of belief and often ignored if it cannot be so construed. For instance, there is 'anchoring bias' or insufficient adjustment of initial probability estimates in the light of new information (Tversky and Kahneman 1982: 14-18). Individuals' focus of attention is very important to explaining their behaviour, for relevant and important aspects of reality are typically ignored (Simon 1986: 31). Individuals rely on a number of heuristic principles and limited data to estimate risks and these commonly result in mishandling of risk estimation. These problems are crucial to explaining decision-making in areas like foreign policy (Jervis 1976).

There are widespread, systematic and fundamental deviations in behaviour from the predictions made by the expected utility model (Hargreaves-Heap *et al.* 1992). For example, alternative descriptions of decision problems often give rise to different choices, even though they are the same from the perspective of the conventional approach (Tversky and Kahneman 1986: 73-9). While people will accept big downside risks to protect their status quo, they are averse to taking risks to improve their position, so their attitude towards risk varies depending on the framing effect of the status quo (Hargreaves-Heap *et al.* 1992: 38). Rather than holding subjective probability estimates which are analogous to objectively derived estimates of risk, individuals often have diffuse and ill-defined feelings about uncertainty and avoid ambiguity about the true risks they face (Einhorn and Hogarth 1986: 43-7). The desirability of options may affect perceptions of the chances of occurrence, as in the phenomenon of wishful thinking; or the probability of occurrence may affect their perceived desirability, as in the phenomenon of sour grapes (Einhorn and Hogarth 1986: 42; Elster 1989a: 17-20). Drawing on the above critique, prospect theory – an alternative to the mainstream theory of expected utility maximisation in the face of risk and uncertainty – has been extensively formalised and tested empirically, including in the field of international relations (Farnham 1994).

The idea that we are inhabited by multiple, conflicting selves seems to be able to account for a number of observable forms of irrational behaviour,

if only in a metaphorical manner (Elster 1985b). The idea has a very long history in philosophy and has been important to psychology, not least because of the work of Freud. Violations of the transitivity assumption fundamental to all the mainstream models of decision-making are common. This can be connected with the idea that individuals have 'multiple selves' who see decisions from different points of view, leading to the impossibility of acting rationally in the conventional sense (Steedman and Krause 1985). While there may be a meta-preference ranking which tells us which self should dominate in a particular context (Sen 1977), decision conflict may be due to inner conflict between different selves. Quattrone and Tversky argue that unconscious self deception – implying the idea of one self deceiving others – may account for why individuals go to the voting booth at elections (1988). The self-deception comes in believing that if you vote others like you will be encouraged to vote too, making it instrumentally rational to vote yourself. Recently Grafstein (1999) has attempted to rebuild collective action theory around this point. Weakness of will can be thought of as involving the inability of the 'higher self' to control impulsive urges, including the delay of immediate gratification in order to enjoy higher future pay-offs. The idea that we have both an instrumentally rational, self-interested self and a socially orientated, norm-driven self provides one way of thinking about the individual tensions generated when self-interest collides with doing what is normatively right.

The evidence reviewed in this section suggests that the mainstream models of decision-making will often be descriptively inaccurate and will make correct predictions only in more limited domains of application than some rational choice theorists believe. Of course, it can still be claimed that the mainstream models provide a standard of rational behaviour against which actual behaviour can be compared; and that some decision-making will approximate the mainstream model. Paralleling the arguments for the bounded rationality approach, there is a strong case for a more descriptively accurate model of the way in which individuals deal with information and uncertainty.

The critique from mainstream political science

Many political scientists claim that rational choice has a poor empirical record. Green and Shapiro (1994) are prominent exponents of this theme, their work occasioning furious counterblasts by rational choice theorists (Friedman 1996). They argue that the desire of rational choice theorists to generate a universally applicable model of politics leads them to evade and

This allows authors to evade falsification by, for example, altering the assumptions made about individuals' motives away from self-interest, as has been done to explain the facts about voter turnout in terms of normative pay-offs or mass collective action in terms of pay-offs associated with affirmation of political identity, among other moral motivations.

Rational choice does not require that individual preferences reflect self-interest: so long as preferences are well-defined, generating a partial ordering of the options, modelling can proceed. But when it moves away from the self-interest assumption, as it has done in some areas as the work of first-generation authors has been developed, it risks becoming untestable and vacuously tautological: individuals acted the way they did because they derived *some* benefits, we know that this must be true because we see them in action and some combination of self-interest and altruism will always give the right prediction (cf. Barry 1970: 19–23). Green and Shapiro overexaggerate the dangers, though. The keys here are to: make firm assumptions about the relative importance of the two motives in the particular empirical context concerned, so that the model is falsifiable; and look at other possible explanations of empirical anomalies that arise rather than favouring the rationality assumptions by making further *ad hoc* modifications to the motivational model (Ward 1996).

Green and Shapiro also claim that rational choice models are largely *post hoc*: modellers typically 'explain' a well-known phenomena in terms of a model derived, at least in part, through inductive reasoning from the observed cases, then use those cases to 'test' the model. They see the empirical claims of the rational choice approach as exaggerated, arguing that various alternatives provide better explanations, depending on context. While they focus solely on a limited field of study, American politics, Walt (1999) makes similar points about a number of well-known rational choice contributions to international relations (IR). These critiques ignore the many novel hypotheses and examples of rigorous empirical testing to be found in the literature. To take one example, Laver and Shepsle's (1996) predictions, derived from game theory and spatial modelling, about the polar importance of strong, centrally located parties to forming government coalitions, are both novel and well-supported empirically.

Green and Shapiro do not compare rational choice with other approaches (cf. Barry 1970): while rational choice may be poorly supported empirically in some applications, the same is true of approaches starting from, say, political culture or social structure. Indeed it is hard to think of any approach to political science that has not been repeatedly called into question by observed facts. A less naive methodological position would be to ask whether rational choice is *worse than other approaches* with respect to its empirical success in areas where there is overlap with their concerns;

rational choice be written off in empirical terms.

To summarise, in many areas of application the mainstream rational choice model is descriptively implausible, yet individuals do make somewhat rational decisions relative to reasonably well-defined goals. To stick to the mainstream approach is to put further development of rational choice theory in a straitjacket. Thus, there ought to be concerted attempts further to develop and to apply non-mainstream variants of the model, allowing for: bounded rationality; choice under uncertainty incompatible with the expected utility approach; and non-egoistic and 'moral motivations'. For instance, in collective action theory as applied to social movements the way forward clearly lies along the path of: examining a wider range of motivations (for example, Opp 1986; Chong 1991); contextualising individuals' perceptions of the likely efficacy of the movement and their own actions in ways that allow slippage from 'objective' information (for example, Dunleavy 1991); and examining action from the perspective of bounded rationality (for example, Ostrom 1997). Only then will collective action theory start to reflect the facts uncovered by the numerous empirical studies that show the limitations of the conventional approach, while suggesting that rationality has an important place in explanations (for example, Whiteley *et al.* 1994; Jordan and Maloney 1997; Finkel and Muller 1998). Methodologically, though, we need to be careful not to be overprotective of rational choice when it does not work well, even in modified form.

Developments: where is rational choice going?

Rational choice is not static but is an active research programme, responding to outside criticism. One change that is difficult to quantify, but easy to discern through reading top-rated journals, is that rational choice theorists increasingly are concerned to bring data to bear rigorously to test hypotheses derived from their models, as opposed to using anecdotes or finding single confirmatory instances. In part this is a response to the sort of criticism that Green and Shapiro advance. Here, I deal in a little more depth with two important trends in rational choice: the attempt to deal with unequal spread of information and the communication of information; and institutional rational choice, which has formed a significant part of the 'rediscovery' of the importance of institutions in political science (see Chapter 14).

A frequent, although ill-informed, criticism of rational choice is that it assumes every player knows everything that is relevant to making a

uncertainty, as we saw above; but what information existed was assumed to be equally spread. Starting in the 1980s economists and game theorists started to model games of incomplete information in which some players have *private information*, that is, they know things that other players do not that are relevant to their decisions. These ideas have now percolated into political science (Morrow 1994). This opens up the possibility of modelling communication of information, including *false* disclosure of private information. Very often the information that is private concerns one side's preferences. 'Signalling' of what type of preferences a player has may also be modelled, information about this being communicated by moves the player makes in the sequence of the game that others can infer would only be made in equilibrium if the other was of a certain type. Signalling involves others modifying, or 'updating', their original beliefs about the other player, typically by reducing the range of possible types they might be, often to a single type. So aspects of learning are dealt with. Again, 'signals' of this sort may be 'bluffs' in which a player 'pretends' to be something they are not, to get a better pay-off through misleading others. The idea of equilibrium discussed above has to be modified to allow for these complications: in essence players' beliefs each time they move must support the actions they intend to take as being rational; and beliefs have to make allowance for logically correct inferences from the observation of past moves of others, combined with 'prior beliefs' at the start of the game.

Despite technical problems, this framework is increasingly finding application in political science and international relations. One application concerns communication between opinion leaders and citizens. One of the central ideas of rational choice theory is that it may be irrational to be fully informed when information is costly, as it is in most political contexts. Individuals will often rely on cues taken from ideologies as an economising device when making political decisions: for instance they vote for the party which has the ideological label suggesting it is most likely to serve their interests, because learning about parties' actual programmes is time-consuming, if not costly in monetary terms (for example, Downs 1957). Voters also rely on opinion leaders for information (Downs 1957; Popkin 1991 is a useful survey). The issue becomes whether citizens in a democracy can expect to get reliable information. Some argue that private information will only be transmitted among players who believe they have similar interests, so it is unlikely that information will be shared among legislators, for example (Austen-Smith 1990). However, there are incentives for opinion-formers to be trustworthy and to give reliable information, even to those who do not share similar underlying interests to their own. For instance, there are low-cost ways for voters to check information

which often deter false information being provided (Lupia and McCubbins 1998). Voters can use the media and independent political commentators and the endorsements of interest groups they belong to or share concerns with as checks on information. By using a framework derived from games of incomplete information, Lupia and McCubbins (1998) do much to illuminate the conditions under which opinion leaders are deterred from sending false signals and the conditions under which voters will trust opinion leaders. So rational choice theory can engage constructively with the issue of false beliefs among mass publics.

The EU has been a major growth area for a second development that is worthy of comment – institutional rational choice theory (Hall and Taylor 1996; Weingast 1996). In brief, the focus is on how institutional rules constrain rational action and how and why rules arise. So, for instance, rational choice theorists have tried to show how EU decision rules affect the way that member states, the Commission and the EU Parliament attempt to get EU legislation as close as possible to their ideal; and how the power of the actors varies with the decision rule in force (see, for example, Tsebelis and Garrett 1997; Hix 1999 for a lucid discussion of the controversy over this approach). The key thing to notice is that this work takes *the details* of EU decision rules and institutional roles, in all their complexity, very seriously, whereas earlier work on voting in the social choice tradition proceeded as if simple majority rule with unrestricted possibilities of amending legislation was in place, despite the fact that this is empirically uncommon. Application of institutional rational choice to the EU raises issues crucial to debates within the 'new institutionalism' between rational choice theorists and those influenced by sociology and history. While the sociological and historical approaches to institutions allow at least some room for rational action within institutionally given rules, they also argue that institutions shape the preferences and discursive frames of reference through which actors see the world (Hall and Taylor 1996). While neo-functional theories of the EU postulate just this sort of moulding and reformulation of world-views and preferences, broadly towards a Eurocentric perspective, intergovernmentalism is much closer to institutional rational choice in assuming that governments pursue security and economic power, members of the EU Parliament pursue reelection, and Commission bureaucrats pursue more power (Hix 1999).

The view among rational choice theorists that 'institutions matter' was first developed in relation to the US Congress. A number of authors were able to show how the rules governing the interaction of the President, the Courts, the bureaucracy, the Congressional committees and the floor of the House and Senate alter outcomes away from what they might have been under simple majority voting on the floor of a single chamber, 1996

and Shepsle and Bonchek 1997). Others have called for context-specific versions of rational choice that take the institutions of other countries, like the UK and Western European nations, seriously (Tsebelis 1990; Dowding and King 1995), opening up the possibility of a comparative politics based on institutional rational choice.

Social choice theory set political scientists a puzzle: it led to the conclusion that there would be very little stability in democratic politics, because majorities are inherently unstable where there are open possibilities for amending legislation and vote-trading (for example, Arrow 1951; Riker 1982). Authors began to explain specific institutional features of Congressional decision-making as *functional responses*, institutions producing 'structure-induced equilibria' where there might otherwise have been chaos under majority rule (Shepsle and Bonchek 1999). For instance, the committee structure of Congress, where committees deal with particular issue dimensions *in isolation* from other issues, may stabilise majorities. In similar vein, Keohane (1984) shows how international regimes of cooperation – sets of rules, roles and convergent expectations over a particular area of interaction between nations – may help solve collective action failures at the international level, by facilitating cooperation through time, and Ostrom discusses the sets of institutions, often at the local level, that can best help prevent over-exploitation of environmental resources by rational, self-interested players (1990). All these authors are well-aware that there is an explanatory gap between saying that a set of institutions *functions* in a way that constrains self-interest or coordinates behaviour and the expectation that the institution will arise among rational players: just as in the related case of norms, discussed above, there may be incentives to free-ride by not contributing to institution-building or by cheating on institutions. So, for instance, many international regimes governing the environment are highly ineffective: nations sign the treaties but do not implement them (Young and Levy 1999). Moreover, if institutions structure equilibria and preferences vary over outcomes, there may be second-order conflict over what institutions to choose (Riker 1980). In response the second prong of the rational choice approach to institutions is to try to explain how rational players might have individual incentives to stick with them – the 'endogenous' approach to institutions (Weingast 1996).

How do economic institutions arise? How do they structure transactions in the market-place? Can institutional politics lead to inefficient economic structures (for example, North 1990)? Sened's discussion of the origins of the institution of private property (1997) is a useful expositional focus. Generally, institutions involve a new set of rules, grafted onto an underlying game, that alter the set of equilibria. To explain an institution, thought of in these terms

the new set of rules; and these rules are an equilibrium of the institutional choice game. Sened's account of the origins of private property says that the state was a decisive coalition in the game of choosing economic rules. It had an incentive to impose private property whenever this increased its net revenue, allowing for an increased tax take due to increased economic efficiency and the state's cost in enforcing rights. Notice that Sened does not attempt an explanation of property rights that 'bootstraps' upwards from an assumption of homogenous rational actors, like some social contract theorists do. Rather a certain, historically given, initial distribution of powers and interests is assumed and then rights are inferred as equilibrium outcomes of play. In short, Sened's is a *structuration story* about the origins of institutions. Paralleling the point I made about social structure and methodological individualism above, I regard this sort of argument as both more historically plausible and more methodologically satisfactory.

Conclusions

Rational choice offers a valuable set of tools to political science. As I have shown, rational choice theory can help illuminate how structures arise and are transformed, but I cannot conceive of any rational choice model which does not introduce some premises about social structure from outside. Thus, rational choice theorists ought to give limited acknowledgement to the sociological critique, recognising that methodological individualism and fully reductive explanations are impractical. Its status is more akin to that of statistical techniques which are appropriate for certain types of data; it is not a stand-alone paradigm for understanding the whole of the political sphere.

Rational choice theory can be put to use by a wide range of social scientists operating within very different paradigms because the results derived depend so crucially on ideas about structure imported from elsewhere. But, like any other tool, it leaves its mark on the work: scholars using rational choice to develop some underlying perspective on society may well come to different conclusions from those using other methods. What is nice about the two developments I covered, games of incomplete information and institutional rational choice, is that the problems dealt with and many of the underlying assumptions are much closer to those that interest mainstream political scientists and political theorists than the first generation of rational choice models; but quite new insights emerge. While these insights might be empirically false, they challenge other forms of scholarship by their rigorous derivation and formulation.

- Non-technical introductions to rational choice include Laver (1997), Maclean (1987), Shepsle and Bonchek (1997) – this is of particular relevance to those primarily interested in the USA.
- At the intermediate level there are Riker and Ordeshook (1973) and Dunleavy (1991) – this text is an excellent example of how a sceptical and empirically orientated political scientist can make use of rational choice alongside other approaches.
- Two books with in-depth technical coverage but virtually no critical commentary are Mueller (1989) and Ordeshook (1986).
- For those who wish to take game theory further there are Luce and Raiffa (1989), Rasmusen (1989) and Morrow (1994).
- From the numerous critical surveys available the most useful are Barry (1970), Hindess (1988), Mansbridge (1990a) and Monroe (1991).