Explaining institutions: A defence of reductionism

MARTIN VAN HEES
University of Twente, Department of Public Administration, Enschede, The Netherlands

Abstract. New institutionalism is an approach to the study of social events that is becoming increasingly popular. Different lines of new institutional research can be distinguished. In one line of research it is argued that the social sciences should transcend reductionist modes of explanation, in particular the methodological individualism exemplified by the theory of rational choice. In this view, reductionism does not permit the endogenous treatment of institutions and thus cannot account for important aspects of the social and political context in which agents act. In this paper the meaning of the terms ‘methodological individualism’ and ‘reductionism’ is discussed and illustrated by a description of some of the central assumptions of rational choice theory, in particular game theory. It is argued that the claim that reductionism should be transcended is unwarranted: reductionism is perfectly compatible with the new institutional concerns. First of all, the various institutional dimensions that have been distinguished in the new institutional literature can be systematically described in terms of a game-theoretic model. Furthermore, the dynamic models of game theory can be (and often already have been) used to explain the emergence of institutions

1. Introduction

In the earlier days of its development, two important spokesmen for new institutionalism said that this approach to the study of social and political phenomena ‘is far from coherent or consistent; it is not completely legitimate; but neither can it be entirely ignored’ (March & Olsen 1984: 734). These words still seem to be highly relevant. New institutionalism cannot be ignored. It purports to reshape our thinking about topics as divergent as local governments, legislative processes, public law, the origins of the state, international cooperation, bureaucratic policy making, and so on. The approach has not only influenced the study of politics and public administration, but it also has had an influence on such related disciplines as law, organization theory, economics and sociology. Furthermore, the stream of new institutional publications reveal that the number of scientists adopting this approach is still increasing. However, although new institutionalism cannot be ignored, speaking of the new institutional approach would suggest a greater consensus among its followers than can in fact be found. There is not only a wide diversity of opinions about the essentials of one of the central notions (if
not the central notion) of the paradigm, namely that of an institution, but also with respect to the appropriate way of carrying out institutional analyses (DiMaggio & Powell 1991).

The picture must not be drawn too pessimistically, though. The existence of common themes is evident. It is almost trivial to say that new institutionalists are first and foremost interested in the role that institutions play in our understanding of social and political events. Broadly speaking, this interest in institutions can be seen to result in two lines of research. The first research line is one in which institutions are described and analyzed in increasing detail. Since institutions determine the way individuals act, the detailed specification of the characteristics of the institutional context in which individuals act leads to a better understanding of that human behaviour. An important example of the sophisticated analysis of the institutional setting which this line of research yields is given in the context of the analysis of legislative processes. One of the central results of the formal theory of voting is the ubiquity of instability under the majority rule (McKelvey 1976). Shepsle has argued convincingly that the abstract model of the decision-making process on which this negative result rests contradicts the actual practice of legislative decision making in which the outcome is to a large extent determined by rules of jurisdiction and amendment control (Shepsle 1979, 1986). Taking account of those rules yields a more detailed description of the legislative process. Such a less abstract description of the institutional context may provide an explanation for the existence of political stability, that is, it may show that, under these particular institutional arrangements, individuals will adopt strategies that do form an equilibrium.1 In this first type of institutional analysis, new institutionalism can be seen as supplementary to conventional theories explaining human behaviour; it provides a more detailed description of the context in which the behaviour takes place. However, the institutional arrangements are still defined exogenously, that is, they form the given setting in which individuals perform actions.

A second line of research within new institutionalism emphasizes the importance of endogenous treatments of institutions. In this view, institutional arrangements should not only play a role insofar as they form the context in which agents act; the institutional context should itself be subject to explanation. Explanations of the existence of institutions cannot be reduced to the behaviour of individuals because the actions, goals, preferences, beliefs of individuals are themselves characteristics of the institutional setting. According to these theorists, new institutionalism should depart from those conceptions of political life which are ‘inclined to see political phenomena as the aggregate consequences of individual behavior, less inclined to ascribe the outcomes of politics to organizational structures’ (March & Olsen 1989:

In this paper we examine the claim that the concern with endogenous institutional variables entails the transcendence of reductionist accounts of social and political phenomena. We shall argue that the claim cannot be sustained. In fact, we argue that there are very strong reasons in favour of a reductionist analysis of institutions. First, we show that the dimensions new institutionalists sometimes merely implicitly ascribe to institutions can be systematically distinguished and analyzed in terms of the characteristics of the reductionist models offered by rational choice theory, and in particular by game theory. The new institutional concern with contextual factors thus can be (and often already has been) dealt with by a reductionist framework. Second, we argue that rational choice theory not only provides very useful tools for the study of individual behaviour within an institutional context, but that it is also perfectly capable of analyzing decision making about institutional arrangements. As long as at least some exogenous variables are allowed, the required endogenous treatment of institutional variables can be realized in a reductionist framework.

The plan of this paper is as follows. In Section 2 we discuss the meaning of the terms ‘methodological individualism’ and ‘reductionism’. After having made clear what we understand by these terms, we describe in Section 3 some of the essentials of game theory, the textbook example of a theory based on the principle of methodological individualism. In Section 4 we turn to new institutionalism and examine what new institutionalists capture under the term ‘institutions’. In Section 5 it is argued that new institutionalism, insofar as it is characterized by a concern with the endogenous treatment of institutions, does not go ‘beyond’ reductionism at all. The new institutional emphasis on the importance of the institutional setting should, in our perspective, not be seen as an argument for removing the restrictions allegedly set by the reductionist scheme of explanations used in rational choice theory. It should be interpreted in the same way as the new institutional concern with institutions has been treated in the first new institutional line of research distinguished above, to wit, as an impetus leading to the refinement of reductionist explanations, not as the driving force of a new non-reductionist paradigm.

2. Methodological individualism and reductionism

One of the perennial controversies in the history of the social sciences is between proponents and opponents of methodological individualism. Almost
every social scientist has in one way or another, sometimes ad nauseam, been confronted with the issue. Furthermore, almost every researcher takes (knowingly or unknowingly) a stand in the debate. The position one adopts is not without importance since it has profound consequences for the way one studies social phenomena and, consequently, for one’s understanding of the nature of social events. Like many other important concepts, the notion of methodological individualism (MI) is used in different and contrasting ways (Lukes 1973; Bhargava 1992). These differences, however, have not always been made clear and, as a result, the debate is sometimes confused. Proponents and opponents of MI dispute about different issues, and defenders of MI disagree no less about what to defend than opponents do about what to attack. Yet, for the purpose of a proper appraisal of the idea, it should be clear what kind of interpretations are possible and which type of MI is under consideration. To introduce some order, let us first consider the distinction between the so-called ontological and explanatory variants of MI (Bhargava 1992).

Ontological interpretations of MI rest on the claim that no social entities exist other than those that can be defined in terms of individuals and their properties. In the ontological view, it is nonsensical to speak about social entities as having a nature of their own. In this interpretation, MI is the ‘denial that there are such undefinable group properties or such superentities’ (Brodbeck 1968: 283) or the view that ‘the ultimate constituents of the social world are individual people’ (Watkins 1968: 270). However, since assumptions about the nature of reality or the status of social phenomena need not have any direct relation to the way we do research (Friedman 1968) and since the new institutional critique of reductionism and MI is first and foremost about methodological issues we shall restrict ourselves to those forms of MI that focus directly on methodological issues, leaving aside the ontological interpretations.

Explanatory variants of MI have as their core elements ideas about the proper way of explaining social events. These ideas boil down to the statement that social events should ultimately be explained in terms of individuals and their properties. However, since there are different types of explanation within the social sciences, there are also different opinions about what counts as valid instances of these strands of MI (Bhargava 1992). We shall restrict our attention to what can be called ‘mainstream’ MI: the view according to which social phenomena should be explained in terms of theories of individual action. In this view ‘the methodological individualist claims that all true theories of social science are reducible to theories of individual human action, plus boundary conditions specifying the conditions under which persons act’ (Nozick 1977: 353).
Classic instances of this form of MI are contractarian political theories. For instance, Thomas Hobbes is often mentioned as one of the founding fathers of MI on the basis of the political theory presented in *Leviathan*. In Hobbes’ theory, the sovereign state results from the actions and deliberations of individuals. When confronted with the hardships of the state of nature, individuals decide to make a contract to institute the Leviathan and thereby leave the state of nature. The emergence of the state as a social phenomenon is explained in terms of a theory about the way individuals act in a state of nature. The state is an intended result of that behaviour. Other examples of mainstream MI are theories that explain social phenomena in terms of unintended consequences of the behaviour of individuals. In these explanations, a social phenomenon is a by-product of individual actions. For instance, beach congestion on a sunny afternoon is the unintended consequence of individuals deciding, independently of one another, to take a swim. Or, to formulate a less trivial example of the same phenomenon, the environmental crisis with which modern society is confronted can be considered to be the unintended result of such isolated human actions as driving a car, reading a book, eating meat, keeping pets, and so on (Hardin 1968; Schelling 1978). 3

An explanation of a social phenomenon in line with MI is always in terms of some general theory about individuals. 4 The theory can be applied only if the boundary conditions under which it operates are known. Some of these conditions have to do with the constraints within which choices are made. For instance, if we explain the behaviour of bureaucrats on the basis of some budget-maximizing principle, we need to know what the range of actions of the bureaucrats is, what the consequences of the various actions in terms of the bureaucrats’ budgets are, whether there are any competitive pressures, etc. It may well be the case that policy x would maximize the bureaucrats’ budgets although it is not adopted by them: the policy may, for some other reason, not be feasible. Thus, constraints effectuate and shape the choices made by individuals. Let us for the moment refer to all more or less permanent determinants of the behaviour of individuals as the institutional setting or simply as the institution (we give a more detailed specification of the features of an institutional setting in Section 4). This setting determines the boundary conditions under which the general theory of human behaviour is to be applied. If the social event to be explained is itself an institution, then we can say that institutions are explained in terms of a general theory of individual behaviour within an institutional setting. Depending on the scope and depth one intends the analysis to have, one can also try to explain the institutions constraining individual actions. However, in order to do so along the lines of MI, one again has to define the institutional setting within which the emergence of those institutions are explained. To explain a particular
phenomenon one invokes a general theory in combination with information about the institutional setting. That institution can also be explained in terms of a general theory and some other institutional setting. In order to obtain an even more detailed explanation one may explain this other institution in terms of yet another institution, etc.\textsuperscript{5}

Thus, we see that MI in fact embraces the view that social events can never be explained solely in individual terms. An explanation along the lines of MI always requires a combination of a general theory of human action plus a specification of the boundary conditions under which the theory is effective. These boundary conditions need not be about individuals, relations between individuals or properties of individuals, but may well be social phenomena. One of the essentials of MI, however, is that any explanation of these boundary conditions, or, more generally, the institutional setting, should itself be in terms of a general theory of individual human behaviour.

MI is a specific form of reductionism.\textsuperscript{6} Reductionism demands that an event at level \( x \) should always be explained on the basis of a general theory about phenomena occurring at a lower level of analysis. Reductionist explanations abound in contemporary research. To give but one example, in the theory of international relations it is quite common to conceive of national states as agents to which motives can be ascribed without further reducing those motives to something on the level of the individual (say the preferences of the members of the cabinet). For instance, phenomena such as the emergence of institutional arrangements are then explained on the basis of the motives of the states participating in those arrangements (Axelrod & Keohane 1985). As with MI, the demand for an explanation in terms of a general theory on a lower level of analysis does \textit{not} imply that higher level phenomena may not appear in the explanation. On the contrary, boundary conditions should be stipulated that specify the constraints within which (individual or collective) agents act. Thus, the structure of other reductionist explanations is similar to that of MI, but the levels of analysis may differ.

3. \textbf{Methodological individualism illustrated: Game theory}

A representative type of MI is formed by those theories in which individual rationality plays an important role. Later we shall relate the new institutional concerns to the reductionist analysis of institutions offered by rational choice theory, in particular game theory.\textsuperscript{7} In order to do so, we outline in this section some of the essential features of game theory.

Game theory is that part of rational choice theory that focuses particularly on the strategic aspects of decision making; it is concerned with the inter-dependence of human behaviour. A game-theoretic analysis of a decision
situation proceeds in two steps. First, the characteristics of the decision situation are described. One way of doing so is in terms of a game in normal form. It consists of the following ingredients:

1. a set of individuals;
2. for each individual a set of strategies available to him or her;
3. a procedure that assigns an outcome to each possible combination of individual strategies (each ‘play of the game’);
4. the individual preferences regarding the set of possible outcomes.

The specification of a game belongs to the descriptive part of a game-theoretic model, in which the assumption of individual rationality already plays a role. It is assumed that the individual preference relations are orderings, i.e., are complete and transitive. In other words, an individual can compare any two alternatives with each other (completeness), and if a person happens to prefer $x$ to $y$ and $y$ to $z$, then the person also prefers $x$ to $z$ (transitivity).

A game-theoretic model not only contains a descriptive part, but also a solution part. In this part the postulate of individual rationality plays a more prominent role. Conjectures are made about what outcomes will be likely if the game is played by rational individuals. To do so a solution concept is employed. Well-known solution concepts are Nash equilibrium, dominance solvability, natural outcome, and so on. A solution concept is, essentially, a set of conditions that a theorist imposes on strategy combinations. The conditions are presumed to reduce the set of all possible plays of the game to those that can actually be expected to occur if individuals act rationally. In other words, each of the various solution concepts can be interpreted as a possible ‘translation’ of the principle of individual rationality. For instance, a Nash equilibrium of a game is defined as a play in which none of the individuals can secure a more preferred outcome if the others stick to their strategies. It expresses the view that rationality means that each individual will use his or her best response to the other person’s actions.

It should be emphasised that different game theorists adopt different solution concepts; there is not one particular solution concept that is considered to be superior to all others. The reasons are well known. First, for almost any solution concept one can define games in which the concept does not work: in that particular game there is no strategy combination satisfying the conditions defined by that particular solution concept. If one nevertheless wants to predict the outcomes of such games one must invoke some other solution concept. Second, there are often situations conceivable in which the particular solution concept yields too many possible solutions. The set of possible plays is then barely reduced and, as a result, the predictive power of the model is limited. Third, like all translations, translations of the postulate of individual rationality in terms of a particular solution concept can sometimes be disput-
ed. One can often construct examples in which the particular concept leads to outcomes that are counterintuitive.\textsuperscript{9}

Despite these difficulties, the basic form of a game-theoretic explanation of a social event should be clear. It is completely in line with MI. Individuals make choices within a set of boundary conditions. The general theory of human action stipulates that individuals act rationally, that is, the preferences of individuals have certain formal characteristics and the individual actions will correspond to some specified solution concept. If the explanation demands, the boundary conditions within which individuals act can also be explained in game-theoretic terms. For instance, suppose we are trying to give a game-theoretic account of why the defence expenditures are of amount $x$ at time $t_2$. Suppose, furthermore, that the range of actions of the politicians who make decisions about the budget is restricted. Parts of the budget are fixed due to prior commitments. If we are primarily interested in the changes in the budget compared with some earlier time period, say $t_1$, then a possible explanation might be in terms of the strategic choices made by politicians between $t_1$ and $t_2$. On the other hand, if we want to explain the whole array of defence expenditures, then we also have to explain those parts of the budget that have been decided upon a long time ago, say at $t_0$. Although such an explanation would be more elaborate than the one focusing on the decisions made at time $t_1$ alone, there is no reason to believe that such a more elaborate explanation could not be carried out with the help of game-theoretic models. In fact, such models have existed as long as game theory; these are models in which a \textit{game in extensive form} describes the situation at hand.\textsuperscript{10} A game in extensive form consists of a decision tree in which the nodes represent points in time at which choices are made. The game models the decision process as a sequence of individual choices. Since, at each node, new boundary conditions may arise (determining which individuals can choose, what choices individuals can make, to what nodes in the tree those choices may lead, etc.), individuals can be said to make decisions about, among other things, boundary conditions. A game in extensive form consists of a decision tree in which at each node at most one individual makes a choice. To model decision making about boundary conditions one can also make use of so-called \textit{simultaneous games}, that is, games consisting of decision nodes (‘subgames’) at which several individuals make their choices simultaneously (Shubik 1982: 47; Van Hees 1995).\textsuperscript{11}

It is often the case that such more elaborate games can be reduced to a game in normal form by defining a strategy as a sequence of choices made by an individual at different points in time. One can, for instance, model decision making about the defence expenditures at time $t_2$ as a decision process in which individuals make decisions both at time $t_0$ and at $t_1$. The strategy
EXPLAINING INSTITUTIONS: A DEFENCE OF REDUCTIONISM

adopted at time $t_0$ is then simply a description of actions taken at time $t_0$ and $t_1$ (‘Always vote against defence expenditures’ or ‘Vote yes for making of commitments at time $t_0$. Decide against additional changes in the budget at $t_1$ if commitments are actually made at $t_0$. Approve additional changes if commitments are not made’. And so on.)

Before we consider the question of how such a purely reductionist framework as game theory can contribute to our understanding of institutions, we return to new institutionalism and examine what new institutionalists have understood by the term ‘institutions’.

4. New institutionalism and institutions

To assess the claim that the new institutional concerns imply the transcendence of reductionism, we have to make clear what it is that new institutionalists conceive to be the central object of their study – what they have in mind when they speak about institutions. Regrettably, new institutionalists use the term in different and sometimes inconsistent ways (Ostrom 1986a: 4). We shall not try to give a systematic account of the various types of definition employed by new institutionalists, nor shall we try to distil some kind of basic definition. The state of the art does not seem to justify such analytic rigour yet. Instead, we shall draw a rather impressionistic picture of what new institutionalists refer to when they speak about institutions. We distinguish three dimensions – structural, procedural, and behavioural – which can be seen to play a role in new institutional interpretations of the notion of an institution.

The structural dimension of institutions refers to all those phenomena that persist over a period of time and that form the constraints within which individuals act and interact. For instance, political parties can be construed as institutions in this structural sense. They determine a politician’s range of actions. Some of the actions the politician might want to take may not be feasible, given his party’s platform, for instance. Another example of such structural aspects is the legal system of a society. It determines the admissibility of courses of action and thereby restricts the range of actions of individuals. Structural aspects are not always defined explicitly. Important forms of informal constraints, for instance, are those that influence a person’s preferences: the subtle ways in which the opinions of friends, family, teachers, etc., influence our thinking about and evaluation of the actions we can take. Other types of important informal constraints are, for instance, the roles individuals are assumed to be playing, or the channels through which communication between individuals take place.

The procedural dimension refers to mechanisms which systematically transform the actions taken by individuals into particular results. In a leg-
islative context, for instance, the procedure may be the method used to amalgamate votes into a collective outcome. Thus we can consider the method of majority voting as an institution in the procedural sense. Procedural aspects refer not only to the mechanism through which decision making within committees takes place, but also to the way committees are mutually related and to the effect that these mutual relations have on the way decisions are made. The long legislative road through subcommittees, committees, legislative chambers etc. forms an example of such procedural aspects. We need not think of voting rules alone, however. The mechanism can also be a legal rule attaching sanctions to violations of obligations.

Finally, some of the definitions offered by new institutionalists are framed in terms of behavioural regularities or patterns of human behaviour. For instance, Rowe states that ‘social institutions are in fact nothing more than agents rationally following rules of action, and being believed by others to do so’ (Rowe 1989: 5; see also Schotter 1981). We shall call these patterns or regularities the behavioural dimension of institutions. Examples of behavioural institutional aspects are conventions, which are regularities in conduct that emerge spontaneously when individuals in a group are engaged in an ongoing relationship with each other. Such conventions facilitate processes of mutual adjustment and learning, thus reducing the complexity of the social situation.

It should be clear that these three dimensions are closely related. Consider systems of norms, for instance: norms obviously belong to the institutional setting in which action takes place. In fact, each of the three types of institutional aspects is relevant to the study of norms. A norm like ‘you shall refrain from doing x’ has structural aspects (‘I may not do x and therefore my range of actions is restricted’), procedural aspects (‘if you do x you will be sanctioned’) and behavioural aspects (‘people tend not to do x’). Note, furthermore, that the rule-oriented character of institutions plays a crucial role in each of the three dimensions. The structural aspects of an institution do not refer to temporary constraints – they delimit the range of possible actions over a longer period of time. Procedural aspects translate individual actions systematically into a social outcome. Finally, the behavioural dimension refers to patterns of behaviour, not to incidental actions.

Not every definition of institutions offered by new institutionalists is framed in terms of all of these dimensions (nor need it be so from a nominalistic point of view). New institutionalism is less characterized by a consensus about the proper way of identifying and defining institutions than by an overall concern to go beyond ‘institution-free’ explanations. The thrust of the new institutional argument is that the mainstream approaches within the social sciences do not pay sufficient attention to the institutional setting in which social events are
embedded. As we have seen, this means for some authors that an explanation of a social event should not take the institutional framework as a set of exogenously given independent variables, but should take that framework as one of the things which itself is in need of explanation. We shall now turn to the question of whether this does indeed imply that reductionism should be abandoned.

5. New institutionalism and reductionism

In the previous sections we have given the central outlines of the methodological positions usually referred to as methodological individualism or, more generally, reductionism. Furthermore, we have described some fundamental concepts of a reductionist approach *par excellence*: game theory. Finally, a sketch was given of the various dimensions new institutionalists have ascribed to institutions. We are now able to address the question of whether new institutionalism should transcend reductionist modes of explanations, as has been claimed by some new institutionalists.

To do so, we first examine the definition and analysis of institutions in the reductionist explanations offered by game theorists. Each of the dimensions ascribed to institutions by new institutionalists can be described in game-theoretic terms. First of all, consider the *structural dimension*. As we have seen, it refers to all those constraints within which individuals act. What are those constraints? We distinguish three types: those relating to the *individuals* partaking in the decision process; those relating to the *strategies* available to the individuals making choices; and those relating to the *preferences* of the individuals. Constraints determining which individuals participate can have a wide variety of forms. They can be physical (‘Mr X is ill and therefore does not participate’), legal (‘Mr X does not have the right to make this type of contract’), moral (‘Mr X should not interfere’), etc. Furthermore, there are constraints demarcating the sets of individual strategies, that is, determining what an individual can and cannot, or may and may not do (Van Hees 1995). Again a multitude of forms exists: examples are legal arrangements like power-conferring rules, mechanisms influencing the amount of information an individual has or the individual’s capacity to deal with information, the communication channels in society, etc.12 Finally, there are constraints related to the preferences of an individual. One can think, for instance, of mechanisms of socialization as processes of preference formation.

The relation of the *procedural dimension* to a game is obvious. It is defined by the *procedure* through which the individual strategies are linked with outcomes.13 The notions to which the first two institutional dimensions are related are precisely the factors of which the descriptive part of a game-
Theoretic model is composed: procedures, individuals, strategies, preferences. Since a game consists of these four components, we can also say that structural and procedural aspects of institutions are described by the game (whether in normal or extensive form).14

The third dimension of institutions, the **behavioural dimension**, is related to the way individuals act. As we have seen, assumptions about the way individuals act belong to the solution part of a game-theoretic model. Game theorists have always been keenly aware of the close relationship between the solution part of a game-theoretic model and the institutional setting of a society. The founding fathers of game theory themselves stated that the solution part expresses the ‘accepted standard of behaviour’ or the ‘established order of society’ (Von Neumann & Morgenstern 1944: 41).

Thus each of the various institutional dimensions can be interpreted in game-theoretic terms. The structural and procedural dimensions are defined in terms of the descriptive part of a game-theoretic model – the specification of the game – whereas the behavioural dimension is captured by the solution part of the model. There seems to be no ground for the claim that the confines of game-theoretic models are too narrow to permit the analysis of institutions. On the contrary, the components of a game-theoretic model correspond nicely with the various aspects that new institutionalists have attributed to institutions. Game theory can be used to model the various dimensions of institutions in a coherent and systematic way.

Game theory can also be used to explain the existence of institutions and thus to provide an **endogenous** treatment of institutions. As we saw in Section 2, explanations along the lines of (explanatory) methodological individualism are in terms of the interplay between a set of boundary conditions and a theory of human behaviour. If the explanation demands, those boundary conditions which can be said to be of **first order** can themselves be subject to explanation. The only limitation thereby is that boundary conditions of the first order should also be explained through the combination of a general theory of human action and a set of boundary conditions. The latter conditions – the **second-order** conditions – can also be formulated endogenously. Now it becomes obvious that the claim that new institutionalism goes beyond methodological individualism and reductionism, because it treats institutional variables endogenously, is based on a naive and distorted picture of those methodological positions. The claim would only be justified if reductionism precludes explanations in which no conditions of second or higher order occur. However, we have seen in our exposition of reductionism that this is not the case. Reductionism does not preclude the explanation of the institutional settings within which individuals (or, for that matter, groups of individuals)
Hence, new institutionalism is, insofar as it is characterized by a concern with higher-order explanations, perfectly compatible with reductionism.

First, consider the structural aspects of an institution. As we have seen, a game is an abstract way of describing a decision situation. It may well be the case that the decision situation at hand is one in which the objects of choice themselves are games; the decision situation should then be modelled as a game consisting of a sequence of subgames, that is, a simultaneous game. The play of the first subgame determines the subgame that will be played next, and so on. Each subgame contains a specification of individuals and strategies. A model thus results which can be used to explain the emergence of structural aspects of institutions: players and strategies. Take, for instance, intraparty decision making about the division of labour among members of parliament. The division of labour not only determines which strategies each representative has (person $x$ is or is not allowed to speak on behalf of his party) but also which representatives do and do not participate in the various stages of the process of parliamentary decision making (person $x$ is or is not a member of this particular subcommittee).

It may seem less obvious that preferences can also be treated endogenously. Although the importance of preference formation processes has often been emphasised, preferences usually belong to the exogenous variables of rational choice models. This does not mean, however, that it is not possible to study these processes within a rational choice framework, nor that such research has not yet been carried out. In fact, at least three routes can be explored. First, approaches exist in which an individual has several sets of preferences that mutually influence each other (Thaler & Shefrin 1981; Kuran 1991). For instance, in their analysis of the notion of self control, Thaler and Shefrin make use of a model of intertemporal choice in which an individual is seen both as a farsighted ‘planner’ who wants to realize his long-term preferences and as a myopic ‘doer’ who is only concerned with short-term gains. Since the planner’s long-term preferences are a function of the doer’s short-term preferences, the planner can satisfy his preferences by, among other things, modifying those short-term preferences. A second route of analyzing endogenous preferences is in terms of models of incomplete information (Cohen & Axelrod 1984; Gerber & Jackson 1993). In these models individuals change their preferences if it becomes clear that their earlier preferences were based on incorrect beliefs concerning the characteristics of the decision situation. Finally, one can explain preferences on the basis of the other structural institutional aspects. Assuming that different types of players have different types of preferences, one can use models that explain the selection of players to account for the existing preferences. If, for instance, risk-averse politicians are more successful than their more risk-taking colleagues, then – given the
appropriate setting – politicians of the latter kind will disappear from the political arena.

Hence, each of the structural aspects of institutions – constraints determining the set of individuals, the strategies and the preferences – can in principle be treated endogenously in game theory. Now consider the procedural dimension. By focusing on the procedures used to link the various subgames of a game, one sees that the model also enables one to study the selection of a procedure for making decisions – it permits the analysis of constitutional decision making. Indeed, constitutional decision making has always been an important subject for rational choice theorists (Buchanan & Tullock 1962; Eichberger & Pethig 1994). Finally, not only the descriptive part of a game-theoretic model can be subject to choice; the same can in principle be said of the solution part. For instance, among game theorists it is well known that specific relationships between principles of rationality and patterns of choice can be considered as resulting from processes of adaptive behaviour in a dynamic context. During the play of a game, conventions may arise concerning the way the play should be played (Taylor 1976; Schotter 1981).

Thus we conclude that each of the institutional dimensions can be defined endogenously in a game-theoretic model. Adopting game-theoretic models does not entail anything about the permitted level of analysis of the independent variables. A game-theoretic explanation – whether in terms of a game in extensive or normal form – is not automatically restricted to the first, most basic order. There is only one limit to what game theory can do in this respect: in the end game theory should have at least some exogenously given institutional factors on the basis of which the other institutional variables are explained. There should be at least one initial point at which the game starts and hence there should also be at least one exogenously given specification of individuals, strategies, procedure, and preferences. Although game-theoretic explanations are not necessarily of the first order, they must be of some order. The fact that game-theoretic explanations should in the end take at least some institutional factors as exogenously given, can only be considered to be a drawback if it is claimed that chains of explanations can always be extended up to a point at which no further institutional variables are needed, that is, to some rock-bottom point at which only individually related phenomena play a role. However, this claim is not only rather dubious, it also takes reductionism to its extremes: social phenomena should in the end be explained only in terms of individuals, their relations and their properties. At this basic level references should no longer be made to the institutional context.17 In other words the new institutional rejection of methodological individualism as adopted by rational choice theorists can only be justified if an even more rigorous form of methodological individualism is adopted. Insofar as new institutionalism
is able to go beyond the reductionism of rational choice theorists, it is led to embrace a more radical form of reductionism. It is highly improbable that new institutionalists are willing to justify, let alone adopt, such an extreme position.

6. Conclusion

It cannot be emphasised enough that we do not want to dismiss the new institutional concern with institutions *per se*. Underlying new institutionalism is the basic assumption that the study of institutional arrangements contributes considerably to our understanding of social life. For this reason, the analysis of human behaviour within an institutional context and the explanation of the existence of that setting itself are important. In our opinion, it is indeed a truism that both the analysis and explanation of institutions should play an essential role in political science and in the related disciplines. The quest for ever more detailed explanations of the way individuals make decisions within institutional contexts is therefore an important and valuable enterprise. Hence, phenomena such as the inclusion and exclusion of individuals, the formation of preferences, the distribution of power, the choice of rules and procedures, the emergence of patterns of behaviour, and so on, are important topics, the study of which should play an essential role in any of the social sciences.

We started our exposition of new institutionalism by distinguishing two lines of new institutional research. In the first line, new institutionalism is seen as a demand to focus in greater detail on the institutional framework in which human behaviour occurs. It thus forms a welcome contribution to existing modes of explanation within the social sciences, including rational choice theory (Ostrom 1991). In the second line of research it is claimed that any mode of explanation that will try to do justice to the importance of institutional arrangements should abandon reductionist modes of explanations. We have concentrated on this second line of research. In our assessment of its claim we started with an exposition of a specific form of reductionism: methodological individualism. It was shown that this form of reductionism does not entail that social events should be explained in individual terms alone. Methodological individualism requires not only a general theory of human action but also a specification of the boundary conditions under which that theory applies. These boundary conditions need not be defined exclusively in individualistic terms; they may well describe institutional arrangements. The institutional arrangements that form the boundary conditions can themselves be the objects of explanation within a reductionist framework. In that event they are seen as arising from the combination of a theory of human behaviour to a set of second-order boundary conditions. These second-order
conditions can be explained on the basis of conditions of an even higher order, and these higher-order conditions can themselves also be explained in terms of conditions of a yet higher order. It follows that reductionism permits the endogenous treatment of institutions. This conclusion is illustrated by the reductionist approach adopted by rational choice theorists. First of all, the notion of a game-theoretic model provides a coherent and systematic account of the various institutional dimensions distinguished in the literature. Furthermore, game theory provides an excellent tool for the endogenous treatment of institutions. Indeed, game-theoretic models in which decisions about institutional arrangements are made form an important part of game theory. This is, of course, not to say that new institutionalism is nothing but a particular form of rational choice theory, or that any social theorist interested in the analysis of institutions should embrace the rational choice framework. Rational choice theory does have its problems. For instance, we have already touched on the existence of situations in which the standard solution concepts lack predictive power because they yield too many equilibria or no equilibrium at all. However, the fact that the theory is confronted with these problems does not imply that they cannot be solved within the rational choice framework.

What we do hope to have established is that, for the question of what institutions are, for how institutions emerge, and for how they effect individual behaviour there are good reasons for adopting the reductionist framework offered by rational choice theory. The study of institutions does not imply the necessity of new modes of scientific explanations or new paradigms.

Acknowledgments

This research was sponsored by the Foundation for Law and Government (REOB), which is part of the Netherlands Organization for Scientific Research (NWO). I thank Ad van Deemen, Bernd Hansjürgens, Henk van der Kolk and three anonymous referees for their comments.

Notes

1. See also Riker (1980) and Ordeshook (1980).
2. We use the term ‘rational choice theory’ to refer in general to those formal theories of decision making in which the assumption of individual rationality plays an important role. The theory encompasses game theory, social choice theory and public choice.
3. These examples describe unintended consequences or external effects which are negatively evaluated. Explanations in which a phenomenon is seen as a positive external effect are usually referred to as invisible hand explanations. Cf. Nozick (1974; 18–22).
4. This does not mean that we restrict ourselves to the so-called deductive-nomological mode of explanation. Intentional explanations, for instance, also invoke general theories. Cf. Elster (1983).
5. We shall ignore the question of whether this process can be extended *ad infinitum*.

6. As a consequence, the reductionist position does not commit one to MI. If we explain an institution on the basis of some of its internal, not individually related parts, then we are reductionists but not necessarily individualists.

7. We limit ourselves to a description of some of the theory’s characteristics. For an appraisal of the predictive power of models of rational choice theory, see Green & Shapiro (1994) and Schofield (1994).

8. Although this is often taken to be a drawback of the theory, one can question whether a theory of rational choice can have any room for a solution concept which is intuitively acceptable and which always yields a unique solution. Can we legitimately say that rational individuals still have a *choice* when it is always clear beforehand what the outcome of a decision process is?

9. For instance, Kreps (1990: 416) contains such a type of argument against the Nash equilibrium.

10. The concept of a game in extensive form derives from Von Neumann & Morgenstern (1944).

11. The notion of a *nested game*, which is a type of game introduced in Tsebelis (1990), is essentially the same as that of a simultaneous game.

12. For a game theorist the distinction between the first two types of structural constraints is not always clear-cut. For instance, to say that Mr X does not participate in the choice process can be modelled in two different ways. Mr X can be assumed not to be a member of the set of individuals making decisions (he is not a player), but he can also be modelled as a player with only one strategy, viz. staying passive.

13. Procedures not only play an important role in game theory. *Social choice theory* is that part of rational choice theory that focuses entirely on the procedures through which individual preferences are amalgamated and translated into outcomes. The theory contains detailed studies of procedures and examines such problems as the structural properties of procedures, the compatibility or incompatibility of different procedures, the strategic possibilities which procedures induce, etc.


15. Another example of the endogenous treatment of the structural dimension of institutions is formed by models in which entry in political contests is studied. Cf. Shepsle (1991).

16. Such models are becoming increasingly important, especially in the so-called evolutionary game theory, i.e., in game theory applied to problems of evolutionary biology. See Weibull (1995).

17. Lukes (1973: 122) adopts such an interpretation in his attack on methodological individualism. In his opinion methodological individualism ‘is a prescription for explanation, asserting that no purported explanation of social (or individual) phenomena are to count as explanations (…) unless they are couched *wholly* in terms of facts about individuals’ (italics added ± MvH).

References


*Address for correspondence:* Martin van Hees, University of Twente, Department of Public Administration, P.O. Box 217, 7500 AE Enschede, The Netherlands

Phone: +31 53 4894553; Fax: 31 53 4892255; E-mail: m.v.b.p.m.vanhees@bsk.utwente.nl