

DISCUSSION

L. P. FOLDES: *A note on individualistic explanations.*

I agree broadly with Harsanyi's critique of functionalism, and with his view that the rationalistic and individualistic method used in economics and game theory should be more widely applied in the social sciences. Nevertheless I think it is important that the current enthusiasm for decision and game theory – which I share – should not blind us to the limitations of this method, or lead us to assert dogmatically that it is the only method proper to social science. In particular, I do not believe that, in the present state of knowledge, this approach provides a firm basis for 'a general theory of social behaviour based on some concept of rational choice'¹.

My remarks will be arranged as follows. I shall first discuss the explanation of *individual* behaviour in rationalistic terms, and point out circumstances in which this method is not particularly fruitful. Next, I shall consider whether *social* regularities should invariably be explained in individualistic terms, and suggest some instances to the contrary. Finally, I shall discuss in particular whether the methods of game theory are adequate for sociological explanation, with special reference to social norms, and shall suggest that in this respect also Harsanyi has claimed too much.

1. *The rationalistic explanation of individual behaviour.* In this section, I shall discuss the scope of the rationalistic method of explaining the behaviour of a single individual in a given situation. In particular, I shall criticise the so-called (methodological) principle of rationality, i.e. the doctrine that the social sciences should use only this method of explaining individual behaviour.

According to Harsanyi's definition, 'the rationalistic approach explains people's behavior in terms of their personal objectives,

¹ Harsanyi, p. 309.

as well as the strategies and the information personally available to them' (p. 305). It is common to think of people as having various objectives or aims and choosing among them according to their preferences, but for present purposes it is convenient to follow the usage of economics and regard the concepts of aims and preferences as equivalent. Now clearly the scope of the rationalistic method depends on the meaning assigned to this concept. Rather than discuss alternative formal definitions, I shall indicate the main points by referring to the axiomatic treatment of preferences in economic statics, which is generally regarded as the outstanding example of a rationalistic theory.

Consider an individual in a situation of given type, with definite information as to the range of available actions. Let S denote the set of all actions which can ever become available in situations of this type, and suppose that at any particular trial a subset A is actually available. The individual's preferences are represented by a 'preference relation' R defined between pairs of elements of S , which is assumed to be reflexive, transitive and complete. It is now assumed that, if any subset A is made available, the individual will choose an element of A which is maximal for R , i.e. which is most preferred, (assuming only that A is so chosen that the required element exists)¹.

The assumptions about R may be regarded as expressing properties of coherence and stability in the pattern of preferences. These two concepts are not really distinct, but intuitively one may think of the reflexive, transitive and complete character of R as reflecting a degree of coherence in the pattern of choice, while stability corresponds to the fact that R is independent of the particular

¹ See G. Debreu, *Theory of Value*, pp. 7–8. A binary relation R is said to be *reflexive* if xRx for all $x \in S$; it is *transitive* if xRy and yRz imply xRz , for all $x, y, z \in S$; it is *complete* if, for all $x, y \in S$, either xRy or yRx . A relation with these properties is called a complete preordering. For a preference preordering, the formula xRy is read 'x is not preferred to y' or 'y is preferred or indifferent to x', and the formula ' xRy and yRx ' is read 'x and y are indifferent'. Finally, x is a *maximal element* for R on a subset $A \subset S$ if $x \in A$, and if $y \in A$, $y \neq x$, yRx implies xRy ; intuitively, if x is most preferred and y is preferred or indifferent to x , then the two must be indifferent.

subset which is available to the actor at a given trial, and also – if the theory is one which envisages repeated trials¹ – from the order in which subsets are presented.

These definitions may be regarded as providing a 'basic' concept of preference in economics, but clearly there is no absolute definition. For a given pattern of observable behaviour, the properties which may be assumed for R will vary with the definition of the set S ; for instance, the less finely the alternative actions are classified, the stronger the postulates of consistency which can be imposed². More important, the properties of R which can be assumed will depend on the type of behaviour considered. Thus it is common for many purposes to add postulates which ensure that R may be represented by a continuous numerical (ordinal or cardinal) utility function on S . It would also be possible – though the point is rarely mentioned – to weaken the definition of preferences, for instance by sacrificing completeness or transitivity. Another very weak assumption would be to suppose simply that there is a subset B of S – the 'satisfactory' subset – from which the individual selects one element – or, equivalently, that he maximises the value of a numerical function which takes the value 1 on B and the value 0 elsewhere.

This brief discussion of formal models of choice in economics illustrates several important features of rationalistic explanation. First, there is no concept of aims which is defined by a fixed list of postulates, independently of the type of action considered and of the formulation of the set of available alternatives. As a general

¹ The general formulation includes the case where the elements of S are not single repeatable actions, but alternative mutually exclusive sequences of actions covering the entire period under consideration.

² To give a more precise, if somewhat extreme, example, consider a sequence of n choice experiments in which a set E_i of actions is made available at the i th experiment, and suppose that the chosen behaviour is $e = (e_1, \dots, e_n)$, where $e_i \in E_i$, $i = 1, \dots, n$. Now we may define the set S either as the union of the E_i or as their product set; in the first case the objects of choice considered in the model are actions in individual experiments, in the second case they are the sequences of n -tuples of actions. Clearly the behaviour e may be intransitive in the first model but not in the second.

definition, one may regard as aims or preferences any principle or procedure for selecting among alternative actions, on the basis of specified information; but this statement is subject to the qualifications which follow. The expression 'principle or procedure' is intended to suggest that the method of selection must possess a certain systematic character, a degree of coherence and stability.

Such a definition of preferences does not depend directly on any concept of subjective aims. This accords with the fact that the formal economic model of choice can be interpreted, if we wish, as a purely behavioural theory, by regarding the 'preference relation' R simply as defining a pattern of potential action, without reference to the subjective processes of reasoning and choice which lead to action¹. This approach has the advantage that it does not restrict the rationalistic method to action which corresponds to particular, clearly defined states of mind. The method is not confined in principle to action which is conscious, or calculated, or deliberately goal-oriented; and its use does not invariably presuppose that the aims attributed to the actor by the theory are, in some sense, his 'true' subjective aims, or that he can perform the

¹ Some economists have held that the preference relation R as defined in the text cannot be regarded as purely behavioural, because it allows for indifference, which does not correspond to any observable behaviour. This statement is not correct if a behavioural theory is defined as one which postulates regularities of behaviour without reference to subjective processes of choice. Even if it is granted that the theoretical relation of indifference does not correspond to an observable pattern of choice, it need not correspond to anything subjective; it may be regarded as a relation defined only in the formal theory, to ensure certain properties of continuity.

It may be added that, whatever view may be taken of the possibility of *observing* indifference, statements about indifference can certainly be made *testable* if the theory is suitably interpreted. For example, consider the standard case where S is the non-negative orthant of the n -dimensional commodity space and R is representable by a continuous utility function with the usual properties (including non-saturation). The assertion that x and y are indifferent is refuted if (say) we can find a point z which is dominated by x but is chosen over y – there are four symmetrical possibilities. The continuity assumptions ensure that it will indeed be possible to find a suitable point z if x and y are not indifferent. Thus R can be regarded as 'purely behavioural' even in the sense that all statements containing R can be tested by observation.

calculations needed for the theoretical derivation of the maximal element of R .

All this is not to say that economic theory should be regarded as purely behavioural; on the contrary, our conception of the subjective processes of reasoning and choice influences both our willingness to adopt a rationalistic approach – which would hardly be appropriate in the case of, say, nervous reflexes – and our selection of postulates. As soon as our interest extends beyond the mere testing of a given formal model to the criticism and formulation of models, our ideas concerning this subjective background, vague as they may be, must be regarded as an essential part of a rationalistic theory. At the same time, within the limitations thus set, the postulates concerning aims are selected for their explanatory value rather than their ‘psychological realism’.

The relationship between rationalistic and behavioural theories can therefore be summarised as follows. On the one hand, the part of a rationalistic theory which is directly testable in relation to a situation of given type can be interpreted as a purely behavioural theory; in this sense, restriction of the scope of a rationalistic theory leads, in the limit, to a behavioural theory. On the other hand, a behavioural theory can be interpreted formally as a rationalistic theory, since the regularity of behaviour which it postulates can be regarded as resulting from some suitably defined aims. But this procedure is subject to two reservations. First, it is objectionable if it makes the theory complicated and perhaps misleading, without adding to its testable content or establishing connections with other theories. Secondly, our theories of the subjective sources of certain actions may lead us to regard rationalistic explanation as inappropriate; having mentioned this point, I shall leave it aside from now on.

My conclusions as to the scope of the rationalistic method, and hence as to the validity of the rationality principle, are implicit in the preceding remarks. The rationalistic method is the more successful, the more diverse the actions which can be explained by a given hypothesis concerning preferences, the simpler the hypothesis, and the greater the degree of stability and coherence which can be assumed. The least favourable case is that of a theory

which relates to behaviour of such narrow range, or so incoherent, or so little dependent on stable characteristics of the actor that rationalistic explanation adds little but complexity; in this case at least purely behavioural explanation is appropriate.

As knowledge changes, the scope for the various kinds of rationalistic theories and for pure behaviourism is likely to change; indeed the need to explain different aspects of action by theories of different kinds is itself an important source of problems. In a sense, it is true that fruitful rationalistic theories represent a more advanced stage than behavioural theories, because they have many implications and can therefore attempt the unified explanation of a wide range of actions; but it is to be expected that at any given time there will be problems which can more usefully be attacked in other ways.

The following are offered as possible contemporary examples of such problems – but rather diffidently, because they raise specialised technical questions.

(1) The study of conditioned behaviour, specific to a well-defined range of situations, which follows closely the pattern which has been learnt and depends little on stable characteristics present before the beginning of the process (apart from characteristics common to all individuals). For many purposes, a wide range of social behaviour must be regarded as belonging to this category; it will be further discussed in the next section.

(2) Stochastic learning processes exhibiting regularities of behaviour which are capable of various explanations, some rationalistic and some not¹, which cannot readily be distinguished by experiment.

(3) Behavioural regularities which are thought to be the result of systematic miscalculation, such as the effects of ‘money illusion’ in economics².

¹ Cf. H. A. Simon, *Models of Man*, ch. 16.

² Supporters of rationalistic methodologies like those of Professors Hayek and L. v. Mises will probably feel that the approach which I have adopted is vague and misses all the important distinctions. There is no sharp distinction between actions which are rational, conscious, deliberate, goal-oriented – the criterion varies – and those which are not; no restriction

2. *Individualistic explanations of social processes.* The view that social processes should invariably be analysed in individualistic terms has been formulated in various ways, some more restrictive than others. A relatively weak – though somewhat ambiguous – requirement is that stated by Professor Popper, who writes that ‘all social phenomena, and especially the functioning of all social institutions, should always be understood as resulting from the decisions, actions, attitudes etc. of human individuals, and that we should never be satisfied by an explanation in terms of so-called ‘collectives’ (states, nations, races, etc.)’. He distinguishes this doctrine from the more extreme view of J. S. Mill, which he calls ‘psychologism’, whose ‘mistake . . . is its presumption that this methodological individualism . . . implies the programme of re-

of social science, and of the rationalistic method, to action of the first kind; no distinction between social science, which takes human action for granted and studies its social consequences, and psychology, which is concerned with the causes of action. I believe, on the contrary – though I cannot give detailed reasons here – that these distinctions are imprecise and largely irrelevant, and can stand in the way of useful work. Human action cannot be clearly divided into such categories as rational and irrational, conscious and unconscious, deliberate and automatic; on almost any definition, these are not dichotomies but distinctions of degree, which moreover refer to variations in complexes of variables rather than to movements along a single scale. In any case the introduction of such concepts raises difficult psychological questions at the very outset. Passing on to the next point, the restriction of social science to action which is conscious etc. is unduly narrow – again on almost any definition – and in any case the attempt to delimit social science rigidly from other fields of study should be avoided. A given action can be studied from many points of view – physiological, psychological, economic – whose scope is never sharply defined but does shift with changes of knowledge. As an example, the reader may wish to consider how the proposed distinction between psychology and social science would apply to the study of the effects of advertising on consumer behaviour. Admittedly one sympathises with the wish of economists like Hayek and Mises to keep the foundations of our subject clear of psychological questions; but unfortunately, as my earlier discussion indicates, the attempt to exclude all subjective factors leads instead to pure behaviourism, a conclusion which these writers would certainly find unacceptable.

The references are to L. v. Mises, *Human Action*, and F. A. Hayek, ‘Scientism and the study of society’, *Economica* N.S. 9 (1942) and 10 (1943).

ducing all social phenomena and all social regularities to psychological phenomena and psychological laws’¹. Popper explicitly takes into account the fact that human action is always influenced by social institutions, and that the institutions themselves cannot generally be explained without reference to other or earlier institutions. Harsanyi’s statement of the explanatory variables as ‘people’s personal objectives and interests, including their possible unselfish interest in each other’s well-being’ (p. 314) is perhaps rather narrower; he also has special views on the explanation of institutions by means of game theory, which will be discussed in the next section.

I shall start from the view, which accords closely with Harsanyi’s approach, that the use of the individualistic method must be justified by its explanatory (predictive) power, and shall leave aside all attempted justifications which invoke special epistemologies of social science². Now this power depends on various conditions,

¹ *The Open Society and its Enemies*, II, 2nd ed., p. 98.

² Both Mises and Hayek advance special arguments for methodological individualism, which it may be of interest to quote briefly. Mises writes, for example: ‘That there are nations, states and churches, that there is social cooperation under the division of labour, becomes discernible only in the actions of certain individuals. Nobody ever perceived a nation without perceiving its members . . . It is illusory to believe that it is possible to visualise collective wholes. They are never visible . . . We see a crowd, i.e. a multitude of people’ (op. cit. p. 43). On this I shall merely comment, first, that one cannot perceive an individual without perceiving his members, a line of thought which leads to absurd conclusions, and secondly that all science uses terms denoting unobservables.

The following quotations give Hayek’s main point: ‘The structure of men’s minds, the common principle on which they classify external events, provide us with the knowledge of the recurrent elements of which different social structures are built up and in terms of which we can alone describe and explain them.’ (op. cit. I, pp. 283–284). ‘. . . it is the concepts and views held by individuals which are directly known to us and which form the elements from which we must build up, as it were, the more complex phenomena’ (p. 286). I would comment that it is far from clear that my knowledge of my own mind, let alone that of another person’s, is more direct or more reliable than my knowledge of, say, the price of carrots in the local shop. Also, external events are not always classified on the same principles by different people, particularly if they belong to different socia

which may be satisfied in greater or lesser degree. In particular, it is important that the aims of individuals should be determined largely outside the system under study and should constitute the relatively stable and coherent elements in that system; otherwise it will be more interesting to regard them as determined than as determining factors. In any case, even if a field of study possesses excellent individualistic theories, there will generally be scope in a given state of knowledge for theories of other kinds, which may not be related in any particular way to the individualistic theories. I cannot attempt a general discussion here, but shall give examples of three kinds of theories which do not appear to fit well into the individualistic framework.

(1) Imagine that we are interested in the distribution of property in a rather static society, during a period extending at least over several generations, and suppose that this can be adequately explained by rules of primogeniture, endogamy, etc.¹ If these rules are rigidly enforced in a way which leaves the individual no effective choice (so that the set of available actions is reduced to a single element), then the introduction of personal preferences, though logically unobjectionable, will add nothing of interest to the theory. Alternatively, suppose that the rules are not so rigidly enforced as to eliminate effective choice, but that the preferences inculcated by a lifetime in the society are such that people generally choose to conform. The preferences are now determined endogenously, and to introduce them again adds little to the explanatory value of the theory. In either case the explanation of the distribution of property in terms of the rules seems more significant than an explanation in terms of individual actions. Of course, this conclusion depends on the assumption that the rules may be regarded for the purposes of the discussion as unalterable, or at least unaltered; and a change in the scope of the study to include the reasons for their survival or original adoption might call for a change of method.

structures; moreover any common principles are inextricably linked with the structure of language and culture – a conclusion which takes us no closer to Hayek's brand of individualism.

¹ I am indebted to Professor M. Ginsberg for suggesting this example.

(2) Sharper issues are raised by the important class of regularities which stem from group structure and, in particular, from conscious organisation. It is common in economic and political theory to treat even large organisations as single units, whose behaviour is in many respects comparable to that of individuals, and to explain certain processes in terms of the interaction of these units. This procedure is justified by the high degree of co-ordination achieved in some organisations; indeed it is not difficult to find instances of organisations whose behaviour in certain respects achieves a higher degree of coherence and stability than that displayed by individuals. It is unnecessary here to try to delimit the class of problems which should be treated in this way; the important methodological point is that the considerations which apply are closely analogous to those which determine whether it is appropriate to regard individuals as the basic units in terms of which a given social process is to be explained. Of course, the mechanisms which ensure coherence and stability are different in many respects, and the behavioural postulates which are appropriate must be expected to vary from one kind of organisation or cohesive group to another, and to differ from those which apply to individuals.

These considerations suggest that it may be entirely appropriate in certain cases to use theories in which the actors are groups such as nations, churches or large firms, provided that there is reason to believe that their behaviour will be sufficiently stable and coherent. Moreover, while it is clearly desirable to investigate the social mechanisms which ensure these properties, the use of the theories need no more await a complete understanding of the workings of organisations than the use of individualistic theories of social processes need await a complete account of the psychological mechanisms by which personal action is co-ordinated.

This discussion incidentally clarifies the proper scope for functionalism. The functional method, as I understand it, is essentially the rational method applied to social aggregates, and its value depends on the considerations which have been set out. Some of Harsanyi's criticisms can therefore be regarded as asserting that the method has been applied to aggregates which cannot be expected to behave

in a stable or coherent way, or alternatively that the mechanisms which might ensure such behaviour have not been investigated sufficiently to justify the conclusions which have been drawn.

(3) The last group of cases raises problems of a different kind. A wide variety of regularities has been discovered in economic data – cycles in the time series, constancy of distributive shares, functional relationships among aggregative quantities, stable frequency distributions – and these may serve as useful theories even where there are no adequate explanations of the relevant individual behaviour, or where no clear or complete connection has been established. A full discussion of these theories, and of their bearing on methodological individualism, would lead too far afield. I shall merely mention one among many possible examples; it is designed, in particular, to show that an economic theory which is not clearly related to rational individual behaviour need not be ‘purely empirical’, but can have a theoretical derivation of a different kind, based on probability theory.

The example is Pareto’s law of income distribution. If y denotes an individual’s annual income and $f(y)$ the relative frequency of that income in a given economy, then the law asserts that $f(y) = Ae^{-cy}$, where A and c are parameters characterising the economy. Originally the law was stated as a purely empirical regularity, but was regarded as interesting because of its agreement with certain observations. Later, it was possible to show that, if y is regarded as a stochastic process with certain transition probabilities for proportional changes, then $f(y)$ is the ‘steady state’ distribution. This derivation establishes connections with stochastic processes generating other economic variables, and greatly increases the predictions which can be obtained from the theory¹. It can be regarded, in a sense, as an individualistic derivation; nevertheless, it treats the individuals as entirely passive, as mere particles, and does not establish a connection with the ordinary rationalistic theory of individual decisions concerning the earning of income². Whether

¹ See, for example, B. Mandelbrot, ‘Stable Paretian functions and the multiplicative variation of income’, *Econometrica*, October 1961.

² It goes without saying that the treatment of given behaviour as a

or not such a connection can be established, the theory is interesting, testable, and useful.

The extent to which methodological individualism as a general rule of method is invalidated by these examples – assuming that they are accepted as instances of useful and interesting work – depends on the version of the doctrine which is considered. I shall merely comment on two general attitudes which may be adopted.

First, it may be argued that, although some of the theories admittedly use methods which are not individualistic, we ‘ought not to be satisfied’ with them until the connection with rational individual action has been fully established. This statement is too vague to offer any practical guidance; strictly speaking, it merely lays down a supposedly ‘ideal’ state which in most parts of social science will never be reached, or may be reached and destroyed by further research, (as has happened in important parts of economic theory). More generally, we ought in any case not to be permanently satisfied with our theories, and at any given time most of them give rise to a variety of problems. It is true that the search for connections between social regularities and individual behaviour is a source of important problems, but it is only one of many; for instance, it is also important to establish links among regularities at different levels of aggregation. The search for these connections is not necessarily the most urgent or the most fruitful work at a particular stage; in particular, its value depends largely on the quality of the individualistic theories which are available. It would therefore be incorrect in general to single out as particularly unsatisfactory those theories for which this work is relatively incomplete; and the injunction that we should not be satisfied with theories which are not fully individualistic certainly cannot be regarded as the main special rule of method of the social sciences.

Alternatively, it may be said that all the theories which I have cited as examples are quite capable of an interpretation which is consistent with methodological individualism, in a suitably wide version which admits institutional and statistical explanations. But

random variable need not be incompatible with the treatment of the same behaviour as rational; for example, consider the obvious explanation of the act of raising an umbrella.

on such an interpretation the doctrine would be almost empty; I say almost, not entirely, for two reasons. First, it can still be regarded as insisting that all social theories be cast in an individualistic form; but, as in the case of the rationality principle, this procedure is objectionable where it makes a theory more complicated, and possibly misleading, without adding significantly to its testable content. Secondly, methodological individualism presumably forbids the use of such concepts as collective personality, will, purpose and aims. As Harsanyi points out, this prohibition is largely redundant in relation to modern social science. In any case, to the extent that the questions connected with group minds, wills, etc. concern matters of substance rather than definition¹, they are not to be settled by an arbitrary principle of individualism, but in the usual way according to the testable consequences of hypotheses. If it were possible to define concepts of mind and will with sufficient precision, and if some social phenomena could not be adequately explained without postulating group mind and will, then there would be no objection in principle to such a procedure.

3. *Game theory and social norms.* In this final section I shall comment briefly on Harsanyi's special arguments for using the

¹ In this connection, it may be of interest to comment briefly on the following question, which turns out to be largely terminological: if explanations in terms of collectives such as organisations are to be allowed, is it permissible to ascribe aims to these collectives? Now, it is possible to distinguish between two common uses of such terms as 'group aims', namely 'aims which are formulated through a group' and 'aims which are related to group action in the same way as individual aims are related to individual action'. We leave aside the first meaning as unobjectionable, and further dismiss as trivial the case where 'group aims' are simply the aims shared by all members of a group. Turning to the main problem, we recall our definition of individual aims – any principle or procedure, possessing suitable properties of coherence and stability, for choice among alternative courses of action. Now, if we restrict ourselves to principles or procedures which operate to some extent in minds, then we can have group aims only if we postulate group minds. Alternatively, if we regard mind as inessential, then groups can have aims without minds – though on such a view the concept of aims would admittedly be redundant.

particular rationalistic and individualistic methods of game theory to explain practically the whole range of social phenomena, including social norms and even morality. Let me say at once that I consider his claims to be out of all proportion to the proven, and even to the potential, value of game theory.

While it is common ground that many social problems can usefully be analysed by means of game theory, Harsanyi claims excessive universality and definiteness of prediction. It has often been pointed out that even the 'classical' game theory, which follows the approach of von Neumann and Morgenstern, relies on strong assumptions which do not appear to fit most social situations; and despite its great intellectual importance, the range of social phenomena for which it has offered new and correct predictions is strikingly small. As is well known, this theory does not offer unique predictions for most games, other than zero-sum two person games; only if the possibilities for play are particularly restricted is a unique answer obtained. As to the Zeuthen-Nash theory, which Harsanyi has developed as a theory offering 'determinate solutions for all classes of games' (p. 309), it is not generally accepted as a general answer to either the predictive or the normative problem, and there is (so far as I am aware) little empirical evidence in its favour. Despite its theoretical interest, it can at present be regarded as no more than one of a number of alternative suggestions, which may be applicable in some cases. Since these matters have been discussed in the relevant literature, it is unnecessary to go into technical details here¹.

As a first corollary of these remarks, I believe that Harsanyi does not sufficiently recognise that the scope of individualistic explanation is far wider than that of game theory. Secondly, his view that every game has a unique solution leads him to exaggerate the extent to which social co-operation can be explained without reference to social norms, or alternatively the extent to which

¹ For a brief critique of Harsanyi's contributions, see R. L. Bishop's remarks in the *American Economic Review*, May 1965, pp. 467-469. General surveys of the literature are given in R. D. Luce and H. Raiffa, *Games and Decisions*, and R. L. Bishop, 'Game-theoretic analyses of bargaining', *Quarterly Journal of Economics*, November 1963.

social norms themselves can be explained in, purely rationalistic and individualistic terms. From the standpoint of 'classical' game theory, norms can be regarded as special rules of the social game which reduce the scope for incompatibility of mutual expectations, and hence the range of indeterminacy of the outcome. In fact, this is quite an interesting interpretation of the doctrine, which Harsanyi attributes to Durkheim and Parsons, that 'the model of rational behavior in itself cannot explain social co-operation . . . (without introducing) . . . *social norms* as a further not-analysed basic constituent'. (p. 313) I agree with Harsanyi that social norms call for explanation, and that this explanation need not be functional; but he goes too far when he states that 'the existence of these social norms themselves will admit of explanation in terms of people's personal objectives and interests' (p. 313) — if he means that the explanation can be given entirely in these terms. In general, it is not possible to explain norms without reference to other norms and institutions. If a given norm is regarded as the outcome of a game, other norms will generally figure among the rules; and even then the outcome may not be unique.

Although I am ill qualified to discuss the question of social status, it seems to me that Harsanyi's theory suffers from all the defects of the method which it is intended to illustrate. It is unduly rationalistic, since much status behaviour is conditioned and automatic rather than goal-oriented in any significant sense. The theory does not allow fully for the important part played by tradition in the determination of social status in most societies; even if there were a tendency for social status to adjust according to Harsanyi's principle of the 'balance of power', tradition might ensure that this mechanism operated so slowly that it would be of little help in explaining the relationships at a given time. Finally, Harsanyi talks as though the status structure could be determined from very limited, specific information, apparently without reference to traditions, institutions, etc. I can see no reason to suppose that a single formula can hope to replace detailed study of the circumstances and history of a given society¹.

¹ Apart from these defects, which are directly related to my methodological remarks, there are various other objections to the theory. Social

To conclude: the rationality principle, and the doctrine of methodological individualism, if interpreted restrictively, exclude from social science some valuable methods which are compatible with general criteria of good scientific work; if interpreted broadly, they permit virtually all methods, but are almost empty, apart from insisting that theories be cast in certain forms which in some cases are misleading and unduly cumbersome. The explanation of individual behaviour by rationalistic methods, and the individualistic explanation of social processes, are among the most powerful methods of social science. But the success of these methods depends on certain conditions — such as the coherence, stability and exogenous determination of preferences — and it may well be unrewarding to apply them to a particular problem at a particular stage of research. Wholesale condemnation of alternative methods is certainly out of place. Finally, game theory is more restricted in scope than the individualistic method generally; the view that every game has a unique outcome lacks support; and it is not usually possible to explain a given social norm as the outcome of a game unless other norms are included among the rules, and even then the outcome will not necessarily be unique.

J. C. HARSANYI: *Reply*.

1. I have no disagreement with many of Foldes' general statements about social-science methodology. I certainly share his view that fundamentally the use of the individualistic method in general — and the use of rational-behavior models in particular — 'must be justified by [their] explanatory (predictive) power'¹, rather than by appeals to introspection, etc.

Moreover, both of us seem to agree that, in terms of this funda-

status is not clearly defined; it is treated as if it were a well-defined quantity measurable along a single scale; the mechanisms relating conflict to the grant of status, and the connection between status and alternative forms of reward, are not clearly analysed; the fact that high social status frequently *restricts* the choice of associates and profession is ignored; and it is not recognised that status may be as much a source as a consequence of power.

¹ Foldes, p. 329.