

Ce texte sera présenté au congrès de l'*European Society for the History of Economic Thought* qui aura lieu à Valencia en février 1999. L'auteur remercie Bruce Maxwell de ses très utiles commentaires ainsi que le CRSH et le Fonds FCAR pour leur aide financière.

### **Three concepts of rationality**

by  
**Maurice Lagueux**  
**Université de Montréal**

There is a general consensus among economists that the notion of rationality plays a central role in microeconomics. It is important to note, however, that they are far from agreement on the meaning of this notion. It would be difficult to lay out a set of well-defined concepts of rationality, but it might be useful to distinguish three quite different approaches around which economists tend to situate themselves when characterising this notion. I prefer to refer to three "approaches" rather than to three concepts, as I did in the title of the paper, because the concepts of rationality which correspond to these approaches cannot be isolated easily. Indeed, their unsettled definitions often overlap, and even when they purport to be related to opposing approaches, they are more or less interrelated. In any case, I propose to refer to these three approaches as "rationality-purposefulness", "rationality-efficiency", and "rationality-consistency". According to rationality-purposefulness, an action is rational if and only if it is oriented towards the satisfaction of the agent's purpose, as most Austrian economists would tend to say. According to rationality-efficiency, an action is rational if and only if it actually maximises a positively valued magnitude such as utility or profit, as many neoclassical economists would say, especially when adopting a traditional approach. By contrast, rationality-consistency states that choices are rational if and only if they are consistent as a set, which implies in particular that they are transitive.

In this paper, I would like to emphasise the differences between and the interrelations among of these three ways of characterising rationality, and, in addition, to challenge the privilege which tends currently to be granted to the third. But before going on this, let us briefly consider their respective genesis. The philosophers Thomas Hobbes and David Hume are usually credited as being the first to posit rationality as the basis of a theory of action. In these accounts, reason, what Hume describes as "the slave of passion", is mobilised to satisfy subjective goals. Classical economists like Turgot, Smith and Ricardo,

whose respective theories were developed in this intellectual context, had only to assume that wealth was such a subjective goal for those involved in economic life. Consequently, according to these economists, traders were led to manage their affairs in a way appropriate to an increase rather than a diminution of their wealth, just because it would be senseless to do otherwise. It is not clear whether the idea of rationality, which played such a role in these economists' theories, was conceived as rationality-efficiency or as rationality-purposefulness, let alone rationality-consistency. More precisely, this question had no relevance in their case. Because no attempt was made to derive precise price levels from an analysis of what is going on in an agent's mind, the question of the nature of this mental activity had not been raised.

This question appears only with the rise of Marginalism when it became significant to determine precise maximum levels to be reached and consequently to define a quantifiable entity to maximise that Jevons, influenced by Bentham's views, named "utility". It was then that rationality came to be seen implicitly as efficiency in attaining maximal utility. Utility was still a subjective entity, but because it was easily associated with something more objective, namely a sum of money, rational economic choice soon became itself closely associated with revenue or profit maximisation. Up to this point, our two first approaches in defining rationality are not clearly disentangled. If an agent succeeds in maximising a positively valued magnitude, it is normally because he or she had the intention or purpose to maximise it in order to reach satisfaction. Conversely, an action oriented towards the satisfaction of a purpose will at least tend to maximise any positively valued magnitude measuring the degree of that satisfaction. Given that, in both these views, reason is seen as an instrument used in order to reach a goal, they together fall under the general rubric of the "instrumental" view of rationality in spite of important differences between them. It is to these differences that I now turn.

In the case of rationality-efficiency, only those actions which actually maximise a pre-determined economic value (like utility, expected utility, profit, revenue or possibly happiness) are deemed rational. An action which fails to maximise such a pre-determined positive value can hardly be considered as equally rational. Otherwise, any action whatsoever could be rational. The beauty of the rationality principle is that it permits us to predict that agents (while free to do what they like) will actually do what is objectively the most advantageous for them. But if agents, due to ignorance or to lack of skill, are unable to determine their best interest, the rationality principle fails to be able to predict anything

from an analysis of their situation. Thus, in this context, rationality tends to be associated with efficiency which in turns requires quasi omniscience.

What I called rationality-purposefulness is quite a different animal. According to this view of rationality, an action is rational when it is oriented towards the satisfaction of any purpose regardless of what that purpose may be. Consequently, whether an action fails due to misinformation, say, is not a criterion of its irrationality. Note that this point of view is less counter-intuitive than the opposite one: after all, who could blame an investor for being irrational if, after having carefully considered every available piece of information which seemed relevant, he loses money due to a unpredictable move by a competitor which turns out to be damaging to his investment? Though such a view does not demand unrealistic conditions like omniscience, its problem is that any action, since it is by definition oriented towards the satisfaction of *some* purpose, will tend to be characterised as rational. Being infalsifiable by hypothesis, such a conception of rationality would hardly be useful for making predictions. In such a view, irrationality would not correspond to a lack of efficiency (due to misinformation or to anything else) but to a lack of purpose — that is to say to a situation in which behaviour would literally cease to be action. In such a view, the mechanical response of a computer, while perfectly efficient, will not be characterised as rational unless it is construed as the action of its programmer. The fact that Menger, by contrast with Jevons, referred to satisfaction rather than to utility, and that his whole analysis stressed the subjective character of decision oriented towards satisfaction, open the way for such an understanding of rationality, but it is 20th-century Austrian economists like Mises and Lachmann who have fully developed this idea of rationality-purposefulness

The third approach that I have referred to tends to evacuate all psychological dimensions from the concept of rationality. Ideally, rationality will no longer need to be associated either with the subjective decision-makers of the Austrians or with the objective quasi-mechanical utility-maximizers of the early neoclassicists. As is well known, a first step was taken with Pareto's indifference curves which allow one to eliminate dependence on cumbersome psychological representations of utility without giving up the mathematics that the concept of utility had made possible (cf. Hollis & Sugden, 1993, pp. 6-7). Thanks to the ranking of indifference curves, maximization became possible without getting involved in the characterisation of what is maximised. Indifference curves do, however, imply preferences, and the next and complementary step, taken by Samuelson, reduced these preferences in a way that they came to be revealed through choices instead of the inverse. With such an approach, rationality was no longer defined by efficiency in a

maximisation process, and still less by purposefulness in action. It was seen rather as nothing but sheer consistency in choice making. Preferences remaining unrevealed by inconsistent choices, the first axiom of revealed preferences implies that if a good X is revealed as superior to a good Y, then this good Y cannot in turn be revealed superior to good X. In other versions of this view of rationality, this axiom is extended to a requirement of transitivity, or of acyclicity: if A is revealed to be preferred to B and B to C, C cannot be revealed to be preferred to A. At first glance, such purely logical requirements seem to be a very modest price to pay for preferences revelation.

This consistency approach to rationality has often been presented as closely related to the more traditional instrumental view of the efficiency type. It was one of the contentions of Houthakker's classic paper of 1950 that a consistency approach is complementary to an approach based on utility maximisation. However, this approach to rationality has nonetheless been criticised for its independence from any particular objectives or goals ascribable to those who make the choices. As emphasised by Amartya Sen (New Palgrave, vol. 4, p. 70b) a person can do *in a perfectly consistent fashion* the exact opposite of what would be suggested by this person's objectives. Characterising as rational such a behaviour would be rather odd indeed. This illustrates fairly well a difference between consistency and instrumental approaches to rationality, but it is a somewhat different point that the one I would like to emphasise about the consistency view.

In his interesting entry on rational choice in the recent Elgar Handbook, Shaun Hargreaves Heap concisely explained why transitivity might be taken as a necessary condition of an efficient utility maximisation. He observes that intransitive preferences can, as he puts it, "lead to poverty" since, by hypothesis, they can bring a trader back to the same point after a succession of costly transactions aimed, in each case, at getting something which is deemed preferable. (p. 400) According to Hargreaves Heap, this state of affairs illustrates the close connection between the consistency and instrumental approaches. However, the eventual impoverishment of the trader in question can be associated with irrationality only if stability of preferences is assumed. Transitivity as such implies the stability of preferences. Changing preferences do not have to be transitive: I can prefer B to A and C to B and make the appropriate exchanges in consequence, but then start to be fed up with C and attracted anew by the hidden merits of A. In such a situation, intransitivity might even be required by rationality understood as adaptive behaviour.

Note that such a view about intransitivity and changing preferences is not so unorthodox. Neither Hume, nor classical economists, nor Austrians and not even early neoclassicists like Jevons required stability of preferences, even if the question was not explicitly discussed by them. Changes in taste have always been difficult to deal with, but the problem was never large enough to paralyse the whole economic apparatus before the introduction of the consistency theory of rationality, for which preferences can no longer be an exogenous datum to be taken just as they are. For Hume, reason has to adapt itself to serve fluctuating whims of its master, the passions. For Classical, Austrians and early Neoclassical economists, taste was treated as an exogenous datum to which the economic process had to adapt itself. After all, adaptation is normally perceived to be an essential feature of rationality. Why should rationality be so closely associated with a kind of consistency which implies immobility and rigidity?

Stability of preferences have rarely been challenged by economists, whereas consistency as such has sometimes been the object of discussion. I would like to raise a few questions about stability with the help of two relatively recent discussions of consistency. In a paper entitled "Why be Consistent?" (1985), Robert Sugden has clearly shown that consistency is not a necessary condition of rationality. He argues that an apparently innocuous rule like minimal consistency (according to which someone who chooses  $x$  from a set limited to  $x$  and  $y$  should not choose  $y$  from a larger set including  $x$ ,  $y$  and some other elements) is violated by behaviour which can hardly be characterised as irrational. Sugden illustrates his point with the case of regret. It seems rational, when choosing between two actions, to integrate into the evaluation the displeasure caused by regretting not to have chosen another action which was feasible. However, as a table in Sugden's paper clearly illustrates, depending whether a third action is available or not, the evaluation of the possible regrets to be included in the evaluation of the two first actions might suggest a choice of either  $x$  or  $y$ . Thus, the fact of choosing  $x$  over  $y$  when only  $x$  and  $y$  are available does not imply that  $y$  should not be chosen over  $x$  when other actions are made available in addition to these. This conclusion clearly denies that rational choice implies minimal consistency.

In fact, in rejecting the necessity of consistency, Sugden claims that rationality supposes adaptation: the best choice depends on the possibilities available in the environment. He illustrates his point in a more pleasant and straightforward fashion with the help of his so-called marriage game. Bill will choose not to marry Annie if there is nobody else who would marry her, but knowing that Charlie would, Bill (who is jealous)

chooses instead to marry her. Being single or married will be chosen in this case depending on the presence of another alternative in the set of possibilities. In Sugden's analysis, Bill's apparently contradictory yet quite rational decisions still suppose that preferences are ordered and stable: Bill knows in advance whether Annie has a mate in the offing and his preferences are fixed accordingly from the outset. But suppose that Bill doesn't know anything about the existence of Charlie and that, according to his assumed preferences, he decides not to marry Annie, but changes his mind when Charlie appears. In this case, one can say that his preferences are not stable without concluding that he is irrational for all that. However, an obvious objection to this argument is to suppose that Bill *has* stable preferences ranked in the fashion proposed by Sugden: being single is preferred to marriage with Annie which is itself preferred to the prospect of Annie marrying Charlie. However, this supposes that Bill knows in advance Charlie's feeling for Annie and his own hostility to their eventual marriage. But this possibility is excluded by my hypothesis. Alternatively, one might attribute to him a much more complex set of preferences which cover all possible states of the world including ones that he cannot even imagine. For example, it is quite possible that if an unknown person by the name of Johnny had appeared instead of Charlie, Bill would have preferred to see Annie marry this fellow rather than get engaged with her. It is quite possible that Charlie but not Johnny provokes Bill's jealousy. But is it conceivable that preferences for states of the world which are not known and possibly not imaginable be ordered in advance? If not, the above strategy of reducing variation in preferences to a case of pre-ordered preferences fails and, short of finding a more successful strategy, one must conclude that there is no reason to assume that preferences are stable. Therefore, a commitment to consistency in such a situation of changing preferences would be nothing but an irrational refusal of adaptation. Suppose, for example, I face a totally unexpected and unpredictable situation such that an action which was once totally inappropriate becomes highly recommendable. Then, it would be totally irrational to avoid making such an action for the sake of consistency.

A somewhat similar line of argument can be brought out with the help of an example that Alexander Rosenberg (1992, p. 119) used in a slightly different context. It is not irrational, Rosenberg observes, to express a preference for regular coffee over milk at breakfast, milk over decaffeinated coffee at lunch, and the latter over regular coffee at dinner. However, Rosenberg admits that, in order to save transitivity (though at the same time making empirical tests impossible), economists might construct artificial bundles of goods corresponding to each kind of possible succession of preferences and assume that consistent choices are made among such "goods". In any case, if we push this kind of

solution further and consider the purely fanciful choices of someone who prefers variety in life, transitivity (and stability) would be saved only by considering a whole life structured in a fanciful fashion as the bundle chosen. But in this limiting case, the very ideas of rational choice, of revelation of preferences, and of transitivity would become vacuous. Indeed, the most whimsical life chosen in a single shot and as a whole bundle would surely be the object of a stable preference, but it could not be compared with any other bundle in such a way that one of them is revealed superior. Without a variety of punctual choices to be compared, the ideas of revealed preferences and of transitivity would have no meaning at all. Either one admits that preferences are not stable, and then transitivity is no longer a condition of rationality, or preferences are artificially made stable, but, then, less and less room is left for choice and rationality themselves.

Economists have generally assumed that preferences are stable with the help of the qualification "taste remaining the same". They have rarely considered the problem raised by changing preferences. An oft-cited exception to this rule is Stigler and Becker's paper "De Gustibus Non Est Disputandum". However, far from questioning the stability of preferences as such, these authors claim that apparently changing preferences result from rational choices made on the basis of more fundamental preference functions. Therefrom, changing taste attributed to addiction, custom, advertising or fashion are as endogenously derived as any standard economic result. While extremely ingenious, Stigler and Becker's analysis cannot be of great help to solve the paradoxes of consistency. Indeed, the preferences which are supposed to be empirically *revealed* by choices made with alleged transitivity are not such highly theoretical, fundamental preferences but changing ones like those influenced by fashion. What we are left in as a result is a catch-22 situation: either choices, whose inconsistency is due to changes in fashion, are deemed to be irrational, but this contradicts the point of Stigler and Becker's paper which is precisely that such choices are rationally explainable as any other economic choice; or, one declares them rational on this ground, but these authors succeed in doing this thanks to a standard instrumental, but not to a "revealed preferences" approach to rationality.

These ambiguities of the notion of consistency render highly dubious the commonly accepted view that the instrumental approach of rationality, either of the "purposefulness" or of the "efficiency" type, is dispensable when one adopts the "consistency" view. I do not deny that the ideas of consistency and transitivity play a significant role in economic analysis; my contention is rather that the attempt to redefine, in terms of logical consistency, the concept of rationality, by excluding from it any reference to subjective purpose or to

efficient accomplishment of some goal, replaces a concept on which the very idea of economics is based with an artificial logical constraint which is completely disconnected from the most ordinary notion of rationality.