Neuroeconomics, Rationality and Preference Formation: Methodological Implications for Economic Theory

Nuno Martins*
Portuguese Catholic University
Faculty of Economics and Management
Rua Diogo Botelho, 1327
4169-005 Porto, Portugal
E-mail: nmartins@porto.ucp.pt
Telephone number: +351917729069

03 June 2007

Preliminary draft – please do not quote

Abstract:

Recent advances in cognitive neuroscience research suggest that different preference orderings and choices may emerge depending on which brain circuits are activated. This contradicts the microeconomic postulate that one complete preference ordering provides sufficient information to predict choice and behaviour. Amartya Sen argued before how the emergence of a complete preference ordering may be prevented by the existence of conflicting motivations, but does not provide an explanation of how the latter are formed and how they impact on choice. I will examine and develop Sen's critique of mainstream microeconomic theory resorting to recent developments in the study of neurobiological structures.

Keywords: Sen, neuroeconomics, rationality, preference ordering, closed system

JEL classifications: B41, D00

*Assistant Professor at the Faculty of Economics and Management of the Portuguese Catholic University, Porto, and researcher at CEGE. CEGE – Centro de Estudos em Gestão e Economia – is supported by the Fundação para a Ciência e a Tecnologia, Portugal, through the Programa Operacional Ciência e Inovação (POCI) of the Quadro Comunitário de Apoio III, which is financed by FEDER and Portuguese funds. For extremely valuable comments I am most thankful to Sabina Alkire, Tony Lawson, and the participants of the Workshop on Multidimensional Comparisons, organised by the Oxford Poverty and Human Development Initiative, Department of International Development, University of Oxford. Of course, remaining mistakes are my sole responsability

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

Ciência.Inovação 2010 Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

1. Introduction

Amartya Sen (1987, 1997, 2002) criticises the behavioural foundations of economic theory in several contributions. Sen argues that due to the existence of competing motivations (or 'reasons for choice'), one unique preference ordering is not sufficient for describing human behaviour (unless, by chance, all motivations provide the same preference ordering).

However, Sen has not provided an explanation of how different motivations impact on choice. Some elements for such an explanation can be found in recent neuroscience research. One of the key insights achieved in the neurosciences is the modularity of the human brain, which, in its weak version, means that not all brain circuits become activated when executing a response to a given circumstance (see Camerer, Loewenstein and Prelec (2004)). Moreover, the same set of stimuli may generate different behavioural responses depending on which brain circuits are activated. So if the hypothesis of modularity is true, different brain circuits can lead to different choices over the same menu of choice, depending on which brain structures and circuits are activated in the act of choice. Hence, there would be various (possibly conflicting) preference orderings at play.

Furthermore, even if a particular brain circuit could act as relatively insulated from others, with the activity of such brain circuit bringing about a stereotyped pattern of choice, a unique and complete preference ordering would result only if such circuit could act as what will be termed here as a closed system. I will argue that neurobiological research does not provide evidence to support the existence of closed systems in brain circuits (or at least the continued existence of stable closed systems). It will also be argued that Sen's view ultimately leads to an open system conception of reality – that is, to a conception of reality where regularities of the form 'if event X then event Y' are not ubiquitous – and that Sen's perspective is supported by recent neurobiological research.

In section 2 I will present a brief description of the two main approaches to rational choice theory. In section 3 a sketch of Sen's critique of those approaches will be presented. In section 4 I will address the role of prediction in rational choice theory, and in section 5 I will systematise Sen's understanding of rationality and behaviour. In section 6 I will develop Sen's perspective drawing upon recent findings about neurobiological structures. I will then present Tony Lawson's distinction between open and closed systems in section 7, and argue that the approaches to rational choice Sen criticises (and their emphasis on predictability) can be justified only when in the presence of closed systems. In section 8 I will discuss Sen's distinction between maximisation and optimisation, and argue that the former entails an open system conception of reality, while the latter entails a closed system conception of reality.

2. Rational choice theory

Amartya Sen (2002:225) identifies two dominant approaches to rational choice theory, which he designates as the 'internal consistency' approach, and the 'self-interest pursuit' approach, respectively. The 'internal consistency' approach aims to explain behaviour by finding regularities in observed behaviour that would enable us to assess its consistency without reference to anything other than (or external to) observed behaviour – hence the name 'internal consistency', for consistency properties are internal to the choice function that describes behaviour.[1]

The 'internal consistency' approach has been a basis for the theory of 'revealed preference' in economics. In this theory, instead of starting from assumptions on preferences (from which choices are then derived), we start by observing choices, and then work out which preferences[2] are consistent with those choices, by checking whether the agents' choices do or do not violate certain axioms of revealed preference[3] – and hence we infer the underlying preference that is revealed in choice. Sen quotes Ian Little as a supporter of this approach:

> "The new [Samuelson's revealed preference] formulation is scientifically more respectable [since] if an individual's behavior is consistent, then it must be possible to explain the behavior without reference to anything other than behavior" (Little 1949: 90; quoted in Sen 2002: 124)

Another approach to rational behaviour Sen discusses is the 'self-interest pursuit' approach. In the 'self-interest pursuit' approach, it is assumed that one motivation, viz., self-interest, dominates all other motivations in economic agency. So 'rational' behaviour, insofar as it is economically relevant, will consist in the pursuit of self-interest. This assumption overcomes the problem of disentangling the effects of different motivations, by ruling out motivations other than what is defined as 'self-interest'. It is also often assumed that 'self-interest' can be represented by a complete preference ordering. This approach provides a basis for the application of utility theory in microeconomic analysis, through the concept of a utility function. The utility function is supposed to represent the chooser's preferences, and constitutes a tool of analysis for explaining how preferences determine choices.[4]

The notion of 'utility' is used in a somewhat ambiguous way in these two approaches. Sen identifies three common interpretations of the concept of utility: 'happiness', 'desire-fulfilment', and 'choice'. The 'self-interest' pursuit approach seems to suggest one of the first two interpretations, for 'happiness' and 'desire-fulfilment' are commonly regarded as the aim of self-interested behaviour. The ‘internal consistency’ approach, on the other hand, looks like a case where utility is taken in the third interpretation mentioned, since in such an approach one starts from observed choices – and utility, welfare and preferences are then inferred from observed choices.

3. Sen’s critique of traditional rational choice theory

Sen notes how internal consistency of choice is neither a sufficient nor a necessary condition of choice. It is not sufficient because “[a] person who always chooses the things he values least and hates most would have great consistency of behavior, but he can scarcely count as a model of rationality” (Sen 2002: 20). Sen argues that internal consistency of choice is not a necessary condition for rationality either, for there may be actions that are rational (according to our common understanding of rationality) but where the axiomatic conditions of consistency of behaviour would not obtain (for elaborations, see Sen 1987, 2002).

Sen also points out how the internal (intrinsic) psychological structure of the individual may be affected by conflicting motivations, values or goals, each of them corresponding to a different ordering, and interacting in a way that precludes the emergence of an internally consistent preference ordering. Furthermore, external (extrinsic) factors may influence choice as well. An example is the case that Sen (1997, 2002) calls “menu-dependence”. Changes of menu, for example by introducing a new element in the menu, may change our attitude towards the other elements of the menu, thus changing the preference ordering that ranks the other elements of the menu. This “menu-dependence” effect violates the axiomatic conditions of internal consistency of choice, which require that orderings must be independent from external conditions. It is thus very unlikely that one ordering only will consistently represent the joint effect of non-congruent motivations, and their interplay with the external environment.

Furthermore, orderings may not even be complete. In many circumstances, some options may not be ranked vis-à-vis each other in any way, and hence internal consistency will not obtain, for different choices would arise in the same circumstances. Incompleteness may characterise not only an ordering that describes a given motivation but also any ordering that describes the 'actual' behaviour that arises out of competing motivations. Limited information, value conflicts, or the need to act before the judgemental process has been made, may undermine the possibility of consistently acting according to a complete preference ordering.

Sen concludes thus that the 'internal consistency' approach fails conceptually: it is "foundationally misconceived", for "[w]hat counts as "consistency" is basically undecidable without taking some note of the motivation of the chooser." (Sen 2002:20). Moreover, it fails substantively, for, Sen argues, the conditions of 'internal consistency' may not even obtain.

Sen has also criticised the 'self-interest pursuit' approach, for ignoring motivations for behaviour other than what is defined as self-interest. Sen (1982, 1997) explains that there are situations where our personal choices may not correspond to the ones that would increase our personal welfare. Sen advocates that social rules which arise out of commitment can lead to a difference between personal choice and personal welfare, and that the word 'preference' is often (ambiguously) used with these two meanings (both to denote the ranking of options in terms of the welfare they provide, and the ranking of options according to our choices). However, there are situations where a certain choice would increase our welfare (and thus would be the choice that 'self-interest pursuit' would lead us to), but social or moral constraints prevent us from choosing the welfare-optimising option. Thus, choice and self-interest may not coincide, as it is posited in the self-interest pursuit approach. So the ‘self-interest pursuit’ approach is also unsatisfactory.

4. The role of prediction

Sen notes how the ‘internal consistency’ approach and the ‘self-interest pursuit’ approach are often conflated. In the ‘self-interest pursuit’ approach, whatever is defined as being our preferences (represented by a utility function which, remember, is also taken to reflect our welfare) determines which option will be chosen. So, in the latter approach, preferences, utility, welfare and choice are represented by the same ordering, and to define preferences and utility according to any of the first two interpretations of utility (viz., ‘happiness’ and ‘desire-fulfilment’) would lead us to identify the latter with the third interpretation of utility (namely, ‘choice’).

Analogously, in the ‘internal consistency’ approach, whatever choice is made is supposed to be the utility (and welfare) optimising choice – in fact, such a choice only ‘reveals’ the ‘preference’ that led to it in the first place – so the same identity between all concepts follows. Hence, in both approaches, welfare, choice and preference imply the same ordering (and rank) of options. Furthermore, in both approaches, this ordering must be a complete ordering, from which all other results about choices are derived.

Sen notes how even though the two approaches are “fundamentally different” (Sen 2002:226) they are confounded in this way in standard microeconomic theory because self-interest and utility are defined through the binary relation of revealed preference[5]. This is why both these approaches lead to the same results, and to an identification of choice with preferences and welfare.

The assumption that rational behaviour can be described through a complete preference ordering, combined with the usual assumption that rational behaviour mimics or approximates actual behaviour, enables the use of rational choice theory for predictive purposes. As Sen argues:

> “[T]he use of “rational choice” in economics and related disciplines is very often indirect, particularly as a predicting device for actual behavior, and this can often overshadow the direct use of rationality. That indirect program is geared to the prognostication of actual behaviour by *first* characterizing rational behavior, and *then* assuming that actual behavior will coincide with rational behavior, or at least approximate it. In this indirect use, the idea of rationality plays an intermediating role in taking us to predictive analysis via the presumption of rational behavior (combined with a view – typically a simple view – as to what makes a behavior rational).
>
> A substantial part of the immediate appeal of this approach of “prediction via rationality” lies in the tractability – and perhaps the simplification – that this procedure may provide” (Sen 2002: 42, emphasis in original)

In short, it in not a particular interpretation or use of concepts like ‘choice’, ‘preference’, ‘welfare’, ‘utility’ and ‘self-interest’ that is central to mainstream economic theory, for these concepts are often conflated (or used in an ambiguous way) in mainstream economic theory, but rather the modelling of human behaviour in terms of exact regularities in order to enable prediction of events. The conception of rationality used in mainstream economic theory is thus subsidiary to the goal of predictability of events.

5. Summarising Sen’s conception of rationality and behaviour

After criticising the accounts of rationality given by traditional rational choice theory, Sen provides his own conception of rationality. According to Sen, rationality is neither the permanent exercise of self-interest pursuit, nor does it mean choice (internally) consistent with some set of axioms. Rather, rationality is the discipline of subjecting competing options, goals and values to scrutiny:

> “Rationality is interpreted here, broadly, as the discipline of subjecting one´s choices – of actions as well as of objectives, values and priorities – to reasoned scrutiny. Rather than defining rationality in terms of some formulaic conditions that have been proposed in the literature (such as satisfying some prespecified axioms of “internal consistency of choice”, or being in conformity with “intelligent pursuit of self-interest”, or being some variant of maximizing behavior), rationality is seen here in much more general terms as the need to subject one’s choices to the demands of reason.” (Sen 2002:4)

Sen’s (1987, 1997, 2002) view of the human agent is one in which the latter is driven by a multiplicity of motivations other than self-interest, which not only may reflect different preference orderings, as they may not even lead to a complete ordering. For Sen, rationality does neither mean that one of these motivations (like self-interest) must dominate all others, nor does it mean that actual behaviour (or any of our competing objectives, goals or values) can be described by a complete preference ordering. Rationality, as the discipline to scrutinise our actions, objectives, motivations, goals and values, does not mean the conformity of behaviour to any preference ordering at all, but rather the possibility of revising and changing preference orderings.

In the paper “Maximization and the act of choice”, Sen (1997) clarifies the differences between the competing motivations, goals and values that we can rationally scrutinise. Sen argues that whilst motivations such as sympathy or reputation can be viewed as part of the person’s welfare or ‘utility’ (if sympathy affects our welfare, and if the reason to maintain a reputation is to attain a higher level of welfare in the future), other motivations such as conventional rule following, social commitment and moral imperatives, are motivations that “drive a wedge” between personal choice and personal welfare. Hence, the motivation of self-interest is not enough to account for all relevant ‘reasons for choice’. This also means that one has to be careful when using the term ‘preference’, stating whether one means actual choices or personal welfare. Sen writes:

“The economic theory of utility, which relates to the theory of rational behaviour, is sometimes criticized for having too much structure; human beings are alleged to be ‘simpler’ in reality. If our argument so far has been correct, precisely the opposite seems to be the case: traditional theory has *too little* structure. A person is given *one* preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behaviour.” (Sen 1982: 99, emphasis in original)

Elsewhere, Sen also argues that:

“[Rational Choice Theory] has denied room for some important motivations and certain reasons for choice, including some concerns that Adam Smith had seen as parts of standard “moral sentiments” and Immanuel Kant had included among the demands of rationality in social living (in the form of “categorical imperatives”).” (Sen 2002:28)

Now, whilst Sen criticises the philosophical and methodological postulates of mainstream microeconomic theory, providing an account of an alternative methodological frameworks, he does not offer a full-fledged theory of rationality and behaviour. Sen notes how competing motivations interact in the human mind, possibly precluding the emergence of a unique and complete preference ordering. However, an explanation of how these multiple motivations interaction is not supplied.

A promising route for a theory of choice and behaviour lies in the study of the structure of the human brain, in order to understand how the latter generates competing motivations, and how the latter impact on behaviour. In the next section I shall draw in recent developments in neuroscience, in order to understand how competing motivations are generated by neurobiological (and neuropsychological) structures, and how they interact.

6. Brain modularity, motivations and preference orderings

Recent developments in the neurosciences (examples are Damásio (1994, 1999, 2003), or LeDoux (1996)) have provided much insight about the neuropsychological structures that cause motivational states. There has been a considerable amount of research in economics attempting to incorporate neurobiological and psychological research in economic theory (see, for example, Romer (2000), Camerer, Loewenstein and Prelec (2004), Bernheim and Rangel (2005), or Kahneman and Tversky (1979)). An analysis of how competing motivations generate conflicting preference orderings can benefit greatly from these developments in the study of the human brain.

One of the features of the human brain that has been noted in this literature is the modularity of the human brain, a hypothesis suggested by Jerry Fodor (1983), and developed in the context of economic theory by authors like Camerer, Loewenstein and Prelec (2004). Modularity, in its weak version, means that that not all the different regions and brain circuits become activated in order to execute different responses to different circumstances. If the brain is composed of different systems and circuits which may or may not be activated under different circumstances, and which become, to some degree, specialized in particular functions, it seems reasonable to question whether some particular motivations (and corresponding preference orderings) dominate when different brain systems are activated. For example, if self-interest, social commitment, moral imperatives, rationality, and other psychological factors like emotions are found to depend on different systems, it seems reasonable to postulate that the activation of the preference orderings corresponding to each of these will be conditional on which brain systems are activated.

Now, modularity would not preclude the possibility of describing behaviour in terms of a complete preference ordering, or even predicting actual behaviour, if we could attribute a different preference ordering to each brain circuit, and know the relative contribution of each brain circuit to each concrete action. Equipped with this knowledge, we could then predict actual behaviour conditional on which brain circuits (and corresponding preference orderings) are activated. Paul Romer (2000), for example, distinguishes between feeling-based mechanisms and thinking-based mechanisms, and constructs models that separate both mechanisms. Bernheim and Rangel, (2005)) also distinguish between a "cold state" when agents follow rational behaviour, and a "hot state" where the agent's choice is dominated by non-rational factors (visceral influences, instincts and emotions, for example), and construct different models for each brain state.

However, the interaction between the different brain circuits (or 'modules') may be far more complicated than what is presupposed in these models. Empirical evidence supports a weaker form of the hypothesis of modularity, one in which the different brain systems are not completely independent of each other, and where the influence of one system on another can become manifest in many ways and degrees. Brain systems responsible for emotion processing, for example, have been found to play a prominent role in decision-making, with the degree of its influence being different in different circumstances, and modulated through additional systems or subsystems.

Emotions are a bioregulatory response that is triggered in regions like the amygdala, the cingulated cortex, and the ventromedial prefrontal cortex. After being triggered, emotions are then executed by regions like the hypothalamus, the brain stem and the basal forebrain. Somatic sensorial regions map the body state caused by the emotion, generating our feeling of the emotion (see Damásio, 1994, 1999, 2003).

Rationality could *prima facie* seem to depend on entirely different systems. Our working memory, which is essential for keeping our images of objects active for some time in order for planning activities to take place, is located in the dorsolateral prefrontal cortex, a brain region that has not been found to be involved in significant emotional processing. However, as Damásio notes, our brain creates associations between emotions and feelings on the one hand, and the decisions undertaken during the execution of those emotions and feelings on the other hand.

Thus, our mental representations and our emotions (and feelings) are strongly interconnected, and causally interact in both directions. On one direction, our mental representations trigger emotions and feelings. More precisely, the ventromedial prefrontal cortex responds to mental representations, for example the representation of a given social situation, by triggering emotional responses that will be executed through the hypothalamus, the brain stem and the basal forebrain, and mapped as feelings in the somatic sensory cortices. The neural dispositions of the ventromedial prefrontal cortex which enable these emotional responses are the “somatic markers”, in Damásio’s (1994) terminology.

Somatic markers, in turn, will make our decision-making be biased towards the options that the brain systems responsible for emotions rank as positive, or rewarding. Thus, on the opposite direction, our emotions and feelings (in particular the brain circuits responsible for emotional triggering and execution) also influence rational decision-making.

Now, preference orderings may be influenced by systems responsible for emotion processing in such a way as to contradict central postulates of mainstream economic theory, while producing behavioural effects similar to those discussed by Sen. As an illustration, remember the effect that Sen (1997) names as “menu-dependence”, i.e., the fact that the choice between two or more elements may change due to the inclusion of a new element in the menu of choice, which modifies our attitude (and so our preference ordering) towards the former two or more elements. As Sen notes, the effect of “menu-dependence” would violate some axioms of internally consistent behaviour (for example basic expansion consistency or basic contraction consistency) because, if behaviour were to be consistent, the relative rankings of the other elements of the menu vis-à-vis each other could not change with the inclusion of a new element. Also, “menu-dependence” would undermine the possibility of obtaining a complete preference ordering through the use of revealed preference axioms.

But from a neuropsychological perspective, “menu-dependence” seems to be a likely scenario in many instances. For example, the human amygdala has been found to be especially stimulated when reacting to threatening situations (see LeDoux 1996), triggering the emotion of fear, which is executed by the hypothalamus, the brain stem and the basal forebrain. Presumably, preference orderings generated by the activation of the emotion of fear will give more prominence to safety concerns in the act of choice, comparatively to preference orderings generated without such an emotional stimulus. So the inclusion of a new element in the menu of choice that somehow triggers a fear emotion would generate a different brain state, and thus activate a different preference ordering, which could in turn rank all the other elements in the menu in a different way, producing the effect Sen named as “menu-dependence”.

Furthermore, Damásio (1994) distinguishes between two types of neural dispositions: innate dispositions, which remain relatively stable during adult life, and acquired dispositions, which may change during our lifetime. This signifies that at least some of the brain regions responsible for the triggering of emotions can change their neural dispositions through experience, and thus any preference ordering they may generate would change over time. Damásio notes how the ventromedial prefrontal cortex, one of the brain regions that triggers emotions, can change its neural dispositions over time. The ventromedial prefrontal cortex is the main site for the emotional categorization of social objects and situations. For example, in a threatening social situation, the ventromedial prefrontal cortex activates the amygdala, which in turn leans to the execution of the associated bodily responses through the hypothalamus, the basal forebrain and the brain stem. The possibility of change of the neural dispositions of the ventromedial prefrontal cortex means that emotions associated with social commitment and social values can change over time, and thus modify our preference orderings. So preferences are not fixed, but rather adaptive, as Sen also notes.

Also, social behaviour is affected by our interactions with other agents in another way. Human agents simulate in their somatic sensorial cortex the emotional states they observe in other agents. When observing other agents in a given situation, neurons designated as “mirror neurons” (which, Damásio (2003) suggests, are located in the prefrontal and premotor cortex) activate the somatic sensorial regions**.** These somatic sensorial regions, which map our body states, simulate a neural state that reproduces a feeling similar to the one (we believe) the agent we observe is experiencing – this feeling is designated as ‘empathy’, and is very similar to Adam Smith’s (1759) notion of ‘sympathy’, which is central to Sen’s own conception of behaviour. The insular cortex seems to be one of the key somatic sensorial cortex involved in this simulation of the other agent’s body state as if it were our own body state – on this, see Wicker, Keysers, Plailly, Royet, Gallese and Rizzolatti (2003), and also Gallese, Keysers and Rizzolatti (2004). Again, the activation of these circuits will affect our preference ordering, which will have the tendency to mimic the other agents’ brain states, and a choice made from the same menu will be conditional on the situation of other agents. In traditional rational choice theory, however, this possibility, which again violates the axiomatic conditions of internally consistent behaviour, is not contemplated.

A scrutiny of some stylised facts about neurobiological research seems thus to give support to Sen’s hypothesis that the human agent is driven by multiple and conflicting motivations, which cannot be all represented by the same preference ordering. Also, the interaction of these motivations need neither generate a unique preference ordering, nor even internally consistent behaviour, and raises the question of whether is it possible to construct a model which represents adequately these different motivations, along the lines of Romer (2000) and Bernheim and Rangel (2005). Can our understanding of neuropsychological structures enable the formulation of a model that successfully predicts actual behaviour? Doing so would require identifying the different neuropsychological structures behind choice, and quantifying the relative contribution of each structure (and corresponding motivation) for each choice act. In the next section I will address this prospect, by discussing the conditions of possibility for the identification of neuropsychological structures, and for the prediction of actual behaviour.

7. Open systems and closed systems

The traditional accounts of rationality Sen criticises entail a *closed system* conception of the social realm. According to Tony Lawson (1997), closed systems are systems in which constant conjunctions of the form "whenever event X then event Y" occur. The models of rational behaviour Sen criticises presuppose the existence of closed systems, otherwise the predictability of actual behaviour that such models aim for would be impossible. Following Lawson (1997), I will name the mode of explanation where regularities of the form 'if event X then event Y' are a necessary condition as deductivism.

The use of rational choice in order to obtain deductivist models where there is only one possible (rational) choice for human agents seems in fact to be the central characteristic of most mainstream economic theory, and is in line with its emphasis on the predictability of events that Sen mentions. Notice that economic theorists are often open to competing explanations of human action, allowing not only for different conceptions of 'preference', 'welfare', 'utility' and 'self-interest', but also for models of human behaviour which include such notions as social rules (e.g., evolutionary game theory models where strategies are sometimes interpreted as social rules, and other models of social interaction). But it is a common characteristic of mainstream economic models that whatever explanation is provided for behaviour, the latter must consist in closed systems regularities, which are obtained either by assuming a complete preference ordering that can be represented by a utility function, or by other mathematico-deductive techniques which enable prediction of actual behaviour (for example, by supposing that human agents maximise some objective function).

In this sense, economic theory is characterised by a concern with deductivist modelling of actual behaviour (of which the emphasis on a complete preference ordering is one possible form), and not by an analysis of how underlying motivations impact on choice. As Lawson argues:

> "It is clear that the various features of the 'economic theory' project traditionally held as essential, along with the more recently proposed revisions of perspective, can easily be explained once the unquestioning, uncritical, orthodox adherence to the deductivist mode of reasoning is acknowledged. Given the revealed flexibility of that project at the level of substantive premises, including axioms as well as assumptions, and its apparent *in*flexibility at the level of its *mode* of explanation, I suggest that an adherence to deductivism in the context of attempting to understand social phenomena be recognised not merely as fundamental to, but actually constitutive of, the 'economic theory' project. (Lawson 1997: 103, emphasis in original)

This application of the deductivist mode of explanation requires some methodological presuppositions. Lawson argues:

> "A presumption of the universal applicability of the deductivist mode of explanation must ultimately rest upon an adherence to a metaphysical thesis that is referred to here as *regularity determinism.* According to this thesis […], for every economic event or

state of affairs y there is a set of events or conditions x_1, x_2, …, x_n, such that y and x_1, x_2, …, x_n are regularly conjoined under some (set of) formulation(s)" (Lawson 1997:98).

In the context of the 'economic theory' project, Lawson identifies two conditions that must hold for the dedutivist mode of explanation to proceed: *intrinsic closure* and *extrinsic closure*. Lawson defines intrinsic closure as the condition that:

"[…] any individual (or set of individuals) is so intrinsically constituted or organised that under any repeated, completely specified, or isolated, set of states or conditions of action x_1, x_2, …, x_n, the same outcome for y is guaranteed to follow" (Lawson 1997:98).

Intrinsic closure implies imposing two constraints: "*intrinsic constancy*: that the internal, or intrinsic, structure of any (delineated state of any) individual of analysis be constant"; and "*reducibility*: that the overall outcome event, for any state description, be reducible to the system conditions obtaining" (Lawson 1997:98).

The extrinsic closure condition, on the other hand, is defined as requiring "that only the explicitly elaborated conditions x_1, x_2, …, x_n have a systematic, non-constant, influence on the outcome event y in question" (Lawson 1997:99).

Now, note that even if a given neurobiological (or neuropsychological) structure were identified, the activation of such a structure would deliver predictable results only under closure conditions. If the act of choice is influenced by multiple structures in the context of an open system (where we are unable to discriminate the relative contribution of each particular structure or substructure), on the other hand, the exact prediction of behaviour will not be possible, albeit the correct identification of the structures at play will often provide us a good idea of the tendencies at play and potential outcomes of an act of choice.

In the natural sciences, natural systems are insulated in experimental situations (that is, in closed systems that are artificially created) so that underlying structures and mechanisms are identified, and their individual contribution to events quantitatively measured and modelled through mathematico-deductive techniques. The different 'modules' of the human brain, however, cannot be insulated as natural structures, and the relevant conception of reality in neural analysis is that of an open system. Patients with brain lesions provide the closest situation to an insulation of a particular neurological structure that we can find. This is why some of the most important findings concerning the functioning of particular neurobiological structures were gained through the clinical study of these patients. But even these cases do not present a situation of controllable closure conditions in the same sense that the experimental manipulation of natural structures does.

The prediction of actual events that Sen refers to as one of the central features of rational choice theory and (micro)economic theory is grounded on a vision of 'economic theory' which presupposes closed systems – both intrinsic closure and extrinsic closure – so that deductive modelling can proceed. Sen's critiques to traditional rational choice theory, on the other hand, are based on a vision of rationality and choice that is not compatible with this deductivist project, and rather point to an open system conception of social reality, a conception which seems to be in line with recent developments in neurobiological and neuropsychological research. In Sen's view, behaviour does not conform to exact regularities (or to complete preference orderings), and rationality just means that goals and values can be revised at any moment. Therefore, the exact prediction of which outcome will be generated by actual behaviour may not be possible (albeit the identification of general tendencies and dispositions caused by underlying structures is not impossible).

8. Maximisation and optimisation

A question now is how choice can be possible if the orderings that arise may be incomplete. Sen (1997) makes a distinction between optimisation and maximisation that may help us to understand this. For Sen, optimisation means finding a 'best' alternative amongst all the feasible options (where the optimal option is weakly preferred to all other feasible options in a set).[6] Maximisation, on the other hand, means to choose an element such that there is no other alternative in the set that is strictly preferred to the chosen 'maximal' element.[7]

So to optimise implies comparing and ranking all alternatives vis-à-vis each other, in order to find the element that is preferred to all other elements. But to maximise (in Sen's usage of the term) does not require comparing all elements. The 'maximal' element is any element such that there is no other element that is strictly preferred to the former. This can happen in two situations: either one realises that no other element is strictly preferred; or the preference ordering is not even defined, and thus there is no other element that is strictly preferred to the chosen element, hence the chosen element is 'maximal' as well.

So when we have some elements of a set that are not ranked vis-à-vis each other, and so the preference ordering is incomplete, there will exist 'maximal' elements nevertheless – the set of 'maximal' elements will be non-empty. This terminology may generate some confusion, for Sen uses the term 'maximisation' in a very different sense than most economic literature. Sen explains:

> “The formulation of maximizing behavior in economics has often paralleled the modelling of maximization in physics and related disciplines. But maximizing *behavior* differs from nonvolitional *maximization* because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior.” (Sen 2002:159)

Maximisation, in Sen’s conception of the term, allows for a choice being made even with ‘incompleteness’ of preferences. Sen explains that this incompleteness may be tentative incompleteness, when some pairs of alternatives are not yet ranked though they may get ranked with more deliberation or information, or assertive incompleteness, when some pairs of alternatives are simply ‘non-rankable’. Sen’s view is that human agents are what he calls ‘maximisers’, while in orthodox microeconomic theory they are seen as what Sen calls ‘optimisers’.

In an optimisation exercise, there will be a complete ordering of options from which only one is the rational option (that will be chosen), and hence the prediction of which outcome will emerge out of actual behaviour will be possible. In such a case, the relevant conception of reality will be a closed system. But within a set of maximal elements (of which at least some are not ranked vis-à-vis against each other in any way), the choice of the human agent can be any from the set of maximal elements, and thus it will not be possible to have an exact prediction of a unique outcome.

Of course, it may happen, as a particular case, that only one maximal element exists, but the point to note is that whilst in mainstream economic theory human agents are always faced with a unique optimal outcome, in Sen’s conception the situation that is posited in mainstream economic theory becomes a particular case, which may or may not occur. In fact, remember that Sen does not argue that actual behaviour will never turn out to be representable by some preference ordering, but rather that one cannot assume *a priori* that a complete preference ordering will always describe human behaviour. Thus, the relevant conception of reality in Sen's account of choice as maximisation will be that of an open system, in which closed systems of the sort posited by mainstream economic theory can occur as particular cases. Recent neurobiological research is in line with this hypothesis, for it also suggests that closed systems in the sphere of human behaviour are not ubiquitous, and at best will be particular cases that may be approximated in some circumstances. It will be in those particular cases that deductivist models, and also models along the lines of Romer (2000) or Bernheim and Rangel (2005), will be particularly useful.

9. Concluding remarks

One can summarise Sen's critique of the behavioural foundations of economic theory and his own account of human behaviour in the following claims: (a) each individual possesses different motivations and 'reasons for choice' (commitment, values, conventional rule-following, moral imperatives, etc.); (b) many of these motivations and 'reasons for choice' (for example, conventional rule-following, social commitment or moral imperatives) reflect a type of social behaviour which cannot be reduced to, or explained only in terms of, individual self-interest or any sort of optimising behaviour (c) one preference ordering is not sufficient to describe each of the different goals, values, motivations and 'reasons for choice' of the individual agent; (d) competing motivations and external factors may not impact on behaviour in a constant and regular way (for what Lawson calls "intrinsic closure" and "extrinsic closure" may not hold), and thus the actual behaviour that arises out of competing motivations and 'reasons for choice' will not necessarily be congruent with one preference ordering only (as it is attempted in the "internal consistency" approach); (e) even if a preference ordering that describes behaviour exists, it will not necessarily be complete, it will most likely be incomplete due to limited information, unresolved value conflicts or the need to act before the judgemental process has been made; and (f) the capacity of rationality means that human agents always have the power to choose and act in a different way than they did, by following a different set of motivations and 'reasons for choice'.

Recent developments in the study of the human brain give support to Sen's perspective. The human brain is composed of different systems with different functions, which often generate different preference orderings depending on which brain structures are activated. However, these different structures are not completely autonomous, and typically generate a situation that is better described as being what Tony Lawson calls an open system. This not only contradicts the traditional assumption of (micro)economic theory that behaviour can always be described by one complete preference ordering, as it also raises serious doubts about the general possibility of describing human behaviour assuming the existence of exact regularities of behaviour that would enable the deductivist modelling of human agency to proceed. Regularities should rather be sought in the underlying social, psychological and neurobiological structures and dispositions that generate actual behaviour. Of course, this must not be taken to mean that the prediction of actual behaviour undertaken in mainstream economic theory is always impossible, but rather that the (deductivist) models used in mainstream economic theory represent particular cases only.

Notes:

[1] These regularities are analysed by checking whether they conform to certain axioms of 'internal consistency' of the choice function, such as the weak axiom of revealed preference, the strong axiom of revealed preference, basic contraction consistency, basic expansion consistency or binariness of choice. Basic contraction consistency of a choice function C(.) implies that, for any element "*x*", non-empty set S, non-empty set T:

$[x \in C(S) \wedge x \in T \subseteq S] \Rightarrow x \in C(T)$

Basic expansion consistency implies that, for any non-empty subset S_j of S

$[x \in \cap_j C(S_j) \; \forall \; S_j \text{ in a class}] \Rightarrow x \in C(\cup_j S_j)$

These are the two basic axioms of internal consistency of a choice function C(.). Axioms of revealed preference and binariness of choice are defined in terms of a preference relation. For elaborations see Sen (1982, 1993, 1997, 2002).

[2] Preferences are assumed to be complete, reflexive and transitive. Let R be a binary relation of preference, where $x\mathrm{R}y$ means that x is at least as good as y. Completeness requires that for every pair of alternatives, either $x\mathrm{R}y$ or $y\mathrm{R}x$ or both. Reflexivity means that $x\mathrm{R}x$. Transitiveness implies that $(x\mathrm{R}y \wedge y\mathrm{R}z) \Rightarrow x\mathrm{R}z$.

[3] One example is the weak axiom of revealed preference. The weak revealed preference binary relation $\mathrm{R_c}$ is defined by: $\mathrm{xR_cy} \Leftrightarrow \exists \mathrm{S}: [\mathrm{x} \in \mathrm{C(S)} \wedge \mathrm{y} \in \mathrm{S}]$.

[4] Notice that one may interpret axioms of 'revealed preference' not as conditions of internal consistency of choice or behaviour, but as a consequence of utility optimisation. In this case, the 'revealed preference' approach would be in line with the 'self-interest pursuit' approach, due to its grounds on utility optimisation.

[5] Taking the preference relation R as the primitive from which the (best) choice function is derived, as in the 'self-interest pursuit' approach, the best choice function is defined as: $B(S,R) = [x \setminus x \in S \wedge \text{for all } y \in S: xRy]$. Taking the choice function C(.) as the primitive, from which the revealed preference relation R_c is derived, one gets: $xR_cy \Leftrightarrow \exists S: [x \in C(S) \wedge y \in S]$, where S is a set and x and y are elements from that set. To define self-interest and utility as the binary relation of revealed preference leads to the binariness of the choice function C(S), that is, for every nonempty S, $C(S) = B(S, R_c)$, where $B(S, R_c) = [x \setminus x \in S \wedge \text{for all } y \in S: xR_cy]$.

[6] This 'best alternative' is given by $B(S,R) = \{x \setminus x \in S \wedge \text{for all } y \in S: xRy\}$.

[7] The 'maximal element' is $M(S,R) = \{x \setminus x \in S \wedge \text{for no } y \in S: yPx\}$. P denotes strict preference. In fact, note that since R is a binary relation of preference, where $x\text{R}y$ means that x is at least as good as y, then this includes two possible situations: either x is as good as (but not better than) y, which is denoted by $x\text{I}y$, or x is strictly preferred to y, which is denoted by $x\text{P}y$. Hence, $(x\text{R}y \wedge y\text{R}x) \Leftrightarrow x\text{I}y$, while $(x\text{R}y \wedge \sim(y\text{R}x)) \Leftrightarrow x\text{P}y$.

References:

Bernheim, D., and A. Rangel, (2005) “Addiction and Cue-Triggered Decision-Making Processes”, *American Economic Review*, 95, 314-346.

Camerer, C. F., G. Loewenstein and D. Prelec (2004), “Neuroeconomics: Why Economics Needs Brains”, *Scandinavian Journal of Economics*, 106, 555-579.

Damásio, A. R. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, G.P. Putnam and Sons.

Damásio, A. R. (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, Harcourt Brace and Company.

Damásio, A. R. (2003), *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, New York, Harcourt Brace and Company

Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, Mass., MIT Press.

Gallese V., Keysers C. and Rizzolatti G. (2004), “A unifying view of the basis of social cognition”, *Trends in Cognitive Sciences,* 8, 396-403.

Lawson, T. (1997), *Economics and Reality*, London, Routledge.

LeDoux, J. E. (1996), *The Emotional Brain: The Mysterious Underpinnings of Emotional Life,* New York, NY, Simon & Schuster, Inc.

Romer, P. M. (2000), “Thinking and Feeling”, *American Economic Review*, 90*,* 439-443.

Sen, A. K. (1982), *Choice, Welfare and Measurement*, Oxford, Basil Blackwell.

Sen, A. K. (1987), *On Ethics and Economics*, Oxford and New York, Basil Blackwell.

Sen, A. K. (1993), “Internal Consistency of Choice”, *Econometrica*, 61, 495-521.

Sen, A. K. (1997), “Maximization and the Act of Choice”, *Econometrica*, 65, 745-779.

Sen, A. K. (2002), *Rationality and Freedom*, Cambridge Massachussets, The Belknap Press of Harvard University Press.

Smith, Adam (2002 [1759]), *The Theory of Moral Sentiments*, Cambridge, Cambridge University Press.

Wicker B., Keysers C., Plailly J., Royet J-P., Gallese V., Rizzolatti G. (2003), “Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust”, *Neuron,* 40, 655-664.