

Can neuroscience inform economics? Rationality, emotions and preference formation

Nuno Martins*

The interaction between neuroscience and economics has gained much prominence recently, leading to the emergence of the new and expanding field of neuroeconomics. I will argue that, although there is much insight to be gained from the interaction between neuroscience and economics, the implications of recent developments in neuroscience and neuroeconomics for the deductivist methodology of mainstream economics, and its emphasis on prediction of events, have not been sufficiently addressed. In fact, much research on neuroeconomics has contributed to the formulation of deductivist models aimed at the prediction of events, when the more fruitful use of neuroscience in economics consists rather in the utilisation of its insights for the development of an explanation of social behaviour that moves beyond the mainstream deductivist methodology. The somatic marker hypothesis, developed by Damasio and others working closely with him, will be suggested as an alternative framework for conceptualising the emergence of social behaviour from a neurobiological substrate.

Key words: Neuroscience, Deductivism, Preferences, Open systems, Somatic marker hypothesis

JEL classifications: B41, D01, D87, I31

1. Introduction

Important advances in the study of the human brain have been made in the last few decades. These advances have been enabled by progress in brain imaging techniques, and clinical evidence gained through the study of patients with localised brain lesions. By bringing these advances from neuroscience to economics, neuroeconomics has provided a richer conception of the human agent. However, in so doing, the main proponents of neuroeconomics have not questioned the essence of the methodology of mainstream

Manuscript received 5 October 2008; final version received 23 February 2010.

Address for correspondence: Portuguese Catholic University, School of Economics and Management, Rua Diogo Botelho 1327, 169-005 Porto, Portugal; e-mail: nmartins@porto.ucp.pt

* Portuguese Catholic University, Porto, Portugal. For helpful comments on earlier drafts the author is most grateful to the anonymous referees of this journal, and to the participants of the following events where previous versions of the paper were presented: Cambridge Realist Workshop; 2007 Oxford Poverty and Human Development Initiative Conference; 2007 Conference of the Human Development and Capability Association.

© The Author 2010. Published by Oxford University Press on behalf of the Cambridge Political Economy Society. All rights reserved.

economics, which can be best characterised as a commitment to mathematico-deductivist models, as Lawson (1997, 2003) argues. The fruitfulness of the approach adopted in neuroeconomics, where neuroscience is used to complement the mainstream models of economic agency, will be critically scrutinised. I will argue that neuroscience can be more fruitful if used to describe the neurobiological structures from which social behaviour emerges, and not to predict behaviour using mainstream economic models, and that the somatic marker hypothesis, developed by Damasio and others working closely with him, can be most useful in this regard.

Also, research on the links between neuroscience and economics has been concerned with diverse topics like preferences, trust or decision making under risk and uncertainty, and no conception of the economic agent that successfully integrates these different branches of research has been provided yet. I will argue that the somatic marker hypothesis can be most helpful in order to achieve an integrated conception of human behaviour.

2. Rational choice theory and mainstream microeconomic theory

There are two dominant approaches in contemporary rational choice theory, which in turn underpin mainstream microeconomic theory. In the first of these approaches, we start by defining a set of axioms, from which a preference ordering is obtained. This preference ordering reflects self-interest, and can be represented by a utility function. It is also assumed that actual behaviour, and ‘rational’ choices, will be driven by such preference, which is the ‘rational’ preference.

In the second approach, instead of starting from a set of axioms from which a preference ordering that explains choice is obtained, we start from observed choices instead, and infer an underlying preference ordering that is consistent with those choices. This is the approach that underpins Paul Samuelson’s (1947) theory of revealed preference since, in this approach, an underlying preference ordering is inferred from observed behaviour, and rational behaviour is defined as any type of behaviour that is consistent with the revealed preference ordering.¹

Sen (2002) notes that even though these approaches have opposite starting points, both are committed to the postulates that there exists a single and complete preference ordering that characterises rational behaviour, and that actual behaviour mimics rational behaviour so defined. Furthermore, not only does this unique preference ordering drive choice, it is also supposed to reflect the chooser’s well-being, as Sen (1982) argues.

Sen (1982, 2002) criticises mainstream rational choice theory, and the mainstream microeconomic theory grounded on the latter, for failing to recognise that human behaviour cannot be described in terms of a single complete preference ordering only. Sen (2002) notes that human behaviour may be driven by motivations other than self-interest, such as social commitment, moral imperatives and conventional rule-following, and argues that not all of these motivations can be described by the same preference ordering. Furthermore, Sen argues that preference orderings need not even be completely specified. Limited information, value conflicts, or the need to act before the judgemental process has been made, may lead to incomplete preference orderings.

Moreover, authors like John Elster (1983), Martha Nussbaum (2000) and Sen (1982, 1985, 2002) have also noted that preferences often get adapted through time. But in mainstream economics, preference orderings are typically taken to be stable, or if

¹ See Vivian Walsh (1996) or Amartya Sen (2002) for a discussion.

adaptation is acknowledged as a possibility, no mechanism is provided to explain how adaptation takes place.

Sen (2002, p. 42) argues that the assumptions on rationality that characterise mainstream economics are made in order to ensure the possibility of prediction of events. In fact, Sen considers the prediction of events to be the main goal of mainstream economics, and argues that the notion of rationality that underpins mainstream economics is subsidiary to this goal.

The problems identified by Sen spring from a more general problem that underpins mainstream economics in general, and not just microeconomic theory: the use of a methodology that presupposes 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. Open systems are those in which such constant conjunctions do not always occur. The models of rational behaviour criticised by Sen 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 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 allows for the predictability of events that Sen mentions. Mainstream 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 that 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 rely upon closed system regularities, which are obtained either by assuming a complete preference ordering that can be represented by a utility function, or by other mathematico-deductivist techniques which enable prediction of actual behaviour (e.g., by supposing that human agents engage in constrained optimisation or maximise some objective function). Even when uncertainty is considered, it is modelled by assuming that agents know the probability distribution of the various possible scenarios (leading to a conceptualisation of uncertainty which contrasts with Keynes' 1936 conception, for example, in which uncertainty consists in the absence of knowledge of the probability distribution of future events), or resorting to other assumptions that facilitate mathematico-deductivist modelling.

In this sense, economic theory is characterised by a concern with deductivist modelling of actual behaviour (within which the explanation of behaviour in terms of a complete preference ordering is one possible form). What is essential to mathematico-deductivist models is the use of a framework that presupposes closed systems, and because it presupposes the constant conjunctions that characterise closed systems, can be used (successfully or unsuccessfully) for the prediction of events, regardless of whether it is ever used as such.

3. Neuroeconomics and the mainstream methodology

Neuroeconomics positions itself as a discipline that goes beyond mainstream economic theory by taking into account the role of different motivations, and does not explain

behaviour solely in terms of a single complete preference ordering. Camerer *et al.* (2005) argue that the tendency to neglect the role of emotions and a variety of motivations in mainstream economics can be explained in part by the scepticism that authors who had a central role in the development of mainstream economics (e.g., William Stanley Jevons, 1871) expressed about the possibility of studying human feelings in any way other than through the observable effects of those feelings. They argue that recent developments in neurobiological research enable us to overcome some of those limitations, and propose the advancement of the new and expanding field of neuroeconomics in this regard.² Neuroeconomics is thus described as a new field that aims at a conception of the human agent where the latter is driven by a multiplicity of motivations.

As Glimcher *et al.* (2008, p. 7) note, there are two main trends in neuroeconomics, one concerned with using ‘brain-imaging as a tool to both test and develop alternatives to neoclassical/revealed preference theories’, and another trend concerned with using ‘economic theory as a tool to test and develop algorithmic models of the neural hardware for choice’. This second trend has provided much insight to the neural functioning of the human brain, and to the neural correlates of such concepts as preferences or subjective utility, whereas the first trend has been concerned with using knowledge from neuroscience in order to improve economic models and economic predictions. In fact, Camerer (2007) argues that neuroeconomics, by providing a more complete characterisation of human motivations, will also enable a significant improvement of economic predictions.

However, a question that could be asked is whether this type of interaction between neuroscience and economics, where the added value of neuroscience consists in the improvement of economic predictions, really enables a conception of the human agent that goes beyond the mainstream approach to rationality and behaviour. In much of the research under the heading of ‘neuroeconomics’, especially in the type of research concerned with improving economic predictions, the way in which neuroscience has been used in order to inform economics consists essentially in using knowledge about the human brain, and about the role of emotions, in order to formulate models of behaviour that assume different preference orderings, which correspond to different brain states.

Nevertheless, since the models that are developed are still deductivist models, aimed at prediction of events, a complete preference ordering, or other closure conditions, are still assumed for each brain state. For example, Bernheim and Rangel (2005) develop a model where a ‘hot’ (emotional) state is described by a preference ordering that is different from the preference ordering that underpins a ‘cold’ (rational) state, and Romer (2000) distinguishes between ‘feeling-based mechanisms’ or ‘thinking-based mechanisms’ when proposing the formulation of deductivist models.

But for each brain state (e.g., ‘hot’ or ‘cold’ states, based either on ‘feeling-based mechanisms’ or ‘thinking-based mechanisms’), there is a complete preference ordering that enables the prediction of human action or other similar conditions that presuppose a closed system. Each brain state is described using the mainstream deductivist methodology (assuming, for example, single, complete and unchanging preferences for each brain state when defining payoff-functions or utility functions in general), presupposing closure conditions that are a necessary condition for a deductivist methodology.

Hence, much of the research on ‘neuroeconomics’ does not really lead to a change in the dominant mainstream paradigm, since it is still concerned with the prediction of actual events that Sen identifies as the main aim of rational choice theory and (micro)economic

² For a summary of which see Glimcher *et al.* (2008), Sanfey *et al.* (2006) or Camerer *et al.* (2004, 2005).

theory, which in turn is grounded on a vision of 'economic theory' that, as Lawson argues, presupposes closed systems, so that deductivist modelling can proceed. In these cases, there is not an abandonment of the mainstream deductivist methodology, but rather an expansion of this methodology.

Of course, knowledge of the role of emotions in changing a given brain state will no doubt help in improving economic predictions. Furthermore, neuroeconomics has provided many contributions to the study of how neural mechanisms influence preferences, choice and subjective well-being. But the enrichment of economics using neuroscience is constrained, rather than facilitated, by the belief that the knowledge achieved in neuroscience must be combined with the mainstream deductivist methodology. We find an extensive use of the mainstream deductivist methodology when modelling human behaviour in what Glimcher *et al.* (2008) call the first trend of neuroeconomics, and the implications of recent developments in neuroscience for the mainstream deductivist methodology have not yet been addressed within neuroeconomics in general.

In fact, 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. But 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), exact prediction of behaviour will not be possible, albeit the correct identification of the activated structures will often give us a good idea of the tendencies at play and the potential outcomes of the 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-deductivist techniques. The different parts (for example the 'modules') of the human brain, however, cannot be insulated as natural structures, not only because this would cause severe brain damage, but also because the activity of most parts depends also on the activity of other parts, and thus 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 neurobiological 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 (on which see Damasio, 1994, 2008). But even these cases do not present a situation of controllable closure conditions in the same sense that the experimental manipulation of physical or chemical structures does.

As will be argued here, recent developments in neurobiological and neuropsychological research seem to support an open system conception of social reality, in which the exact prediction of the outcome generated by actual behaviour may not be possible—albeit the identification of tendencies and dispositions caused by the underlying brain structures is not impossible. But the most fruitful contribution that neuroscience can bring to economics is in its descriptive potential and not on its predictive ability. That is, neuroscience can inform economics by providing a detailed study of neurobiological structures and dispositions, and of the psychological motivations and social structures that emerge from them, rather than by being combined with the mainstream methodology in order to attempt to predict events.

The conflicts that may emerge between motivations, and the preference orderings that are generated as the outcome of underlying structures and dispositions, can then be addressed in terms of a theory formulated in terms of underlying structures and

dispositions. In fact, if preferences are recognised to be adaptive, the concept of preference does not provide (itself alone) a solid ground for a theory of behaviour, for preferences must themselves be explained in terms of the structures and dispositions that cause their adaptation.

So the use of neuroscience in economics could be much more fruitful if aimed at the identification of neurobiological structures, and the psychological and social structures that emerge from the neurobiological substrate, rather than if aimed at the formulation of complete preference orderings for each brain state or of deductivist models in general. The human brain is already sufficiently complicated, and the attempt to provide complex mathematical models (which presuppose closed systems) in a context of an open system brings more confusion, rather than clarification, to the study of the human brain and of human behaviour.

4. The somatic marker hypothesis as a unifying theory

A crucial question to address now is how social behaviour emerges from underlying neurobiological structures. An answer to such a question requires empirical analysis of the particular structures under scrutiny. The work of neurologist Antonio Damasio, and his somatic marker hypothesis, will be used here as a more specific description of the underlying structures and dispositions behind human behaviour.

A question that could be raised concerns the use of the somatic marker hypothesis as a framework of analysis. The reason for the use of this framework springs from the fact that the somatic marker hypothesis is a very broad hypothesis that, due to its abstract nature, can encompass other more specific theories within its framework. As Bechara and Damasio (2005) suggest, the somatic marker hypothesis postulates that somatic signals bias our mental representations of competing options by attaching different emotional charges to them, thus helping in the computation of various possibilities by emphasising some possibilities above others and increasing the speed of decision making. The result is an improvement of the efficiency of our decision-making process. But the specific mechanisms through which this happens are still a matter of further debate.

The somatic marker hypothesis can be seen as a general framework of analysis, which enables the conceptualisation of decision making as a process where reasoning and decision making (i) depend on many levels of neural operation, (ii) are supported by underlying processes (like attention, working memory and emotion) and (iii) depend on the availability of knowledge about situations, actors, options and outcomes, which is stored in dispositional form.³

In fact, one of the challenges of the current research in neuroscience and neuroeconomics is how to integrate its various insights into a coherent explanation of human motivation and behaviour. A possible way to overcome this problem is to provide a framework for explaining economic decisions that has the potential to include various streams of research in neuroscience, and in neuroeconomics in particular, within a broader conceptual scheme. Hence, the somatic marker hypothesis seems an appropriate starting point with which other branches of research can be fruitfully combined.

One of the key insights provided by the somatic marker hypothesis is that higher order processes like human reasoning and decision making depend upon lower level bioregulatory processes. Hence, the hypothesis can accommodate the research on different

³ For further elaboration, see Bechara *et al.* (2000, pp. 295–6) or Bechara and Damasio (2005).

bioregulatory processes (including emotions), and on their influence in reasoning and decision making within its conceptual framework, where the brain is a layered (and structured) entity constituted by different levels of bioregulatory processes, in which older and simpler bioregulatory processes and dispositions give rise to newer and more complex processes.

Because bioregulatory processes may generate different brain states and motivations, which will not necessarily correspond to the same preference ordering (and may indeed lead to conflicting preferences or incomplete orderings), the key unifying principle for a theory of human behaviour must be defined in terms of the underlying structures and dispositions that cause actual behaviour (such as the somatic markers and the structured set of bioregulatory processes that emerged during a long evolutionary process) and not in terms of preference orderings, closed systems and the mathematical functions that represent the latter. Besides, the adaptation of preferences can itself be explained in terms of changes in the somatic state. I will now describe the somatic market hypothesis in more detail.

5. The human brain and bioregulatory processes

The human brain is the result of a long evolutionary process, during which different structures evolved and enabled the existence of diverse patterns of response to changes in the internal functionings of the organism, and to variations in the external environment that the organism faces.⁴

Damasio (2003) notes that during this evolutionary process the brain used older and simpler mechanisms in order to execute more complex functions. For example, brain circuits used to perform complex functions such as emotions resort upon brain machinery involved in simpler bioregulatory states. The brain structure evolved thus in a way that enables it to perform bioregulatory responses of varying degrees of complexity, from the control of the organism's metabolism to complex bioregulatory responses such as emotions.

Damasio (2003) divides the bioregulatory processes of human beings into three different levels. At the most basic level there are processes that are coordinated by the autonomic nervous system, such as metabolic processes of the human body (necessary for the maintenance of life through the transformation of energy sources), the responses of the immune system and the basic reflexes (including 'fight or flight' responses). Regions like the brain stem and the hypothalamus are key areas in the coordination of these types of bioregulatory processes.

However, these basic bioregulatory processes are not enough for complex organisms like human beings to find energy sources. Thus, they give rise to another level of bioregulatory processes of increasing complexity, through which the organism responds to possible failures in its internal functionings, such as the lack of energy sources to perform the more basic bioregulatory processes mentioned above. More complex bioregulatory processes include pleasure or pain behaviour and the basic impulses (e.g., hunger, thirst or exploratory behaviour), through which the organism responds to the needs of the underlying systems.

At a further level of analysis, the organism's responses include bioregulatory processes termed as emotions. Emotions are bioregulatory processes that are triggered when regions of the nervous system such as the amygdala, the cingulate cortex and the ventromedial

⁴ On the study of the brain as a result of an evolutionary process, see also Paul MacLean (1990).

prefrontal cortex are stimulated through a sensory cortex. When stimulated, the latter regions trigger emotional responses, which are then executed through the hypothalamus, the basal forebrain and the brain stem, causing changes in the bodily state. Afterwards, the emotional response of the organism, and the changes in the bodily state, are registered in the somatic sensorial cortices (including the insula), which permanently map the state of the organism. Damasio (1994, 1999, 2003) makes a distinction between the bodily responses associated with the execution of emotions, and their mapping by our somatic sensorial (including the insular) cortices. Only the latter does he designate by the word 'feeling', leaving the word 'emotion' to denote the bioregulatory process that causes the 'feeling'.

Damasio (2003) also distinguishes primary emotions from secondary (or social) emotions. Primary emotions include emotions such as fear, anger, sadness, happiness, surprise and disgust, and are bioregulatory responses that are similar across all human beings as long as their neurobiological structures are intact. Primary emotions occur when, after a stimulus, a given brain area (like the amygdala or the cingulate cortex) is activated, and triggers regions such as the hypothalamus, the basal forebrain and the brain stem, which in turn execute the corresponding bodily responses. Again, these responses are important for the organism's survival. For example, the amygdala triggers the emotion of fear when in the presence of threatening stimuli (which are first assessed by association cortices and high-order cortices), and the execution of the emotion is then processed through the hypothalamus, the basal forebrain and the brain stem.⁵ This emotion helps the organism to react to dangerous situations, contributing to its survival.

Secondary emotions include emotions like sympathy, compassion, embarrassment, shame, guilt, pride, jealousy, envy, gratitude, admiration, indignation and despise. Secondary emotions are derived from primary emotions, for they depend on the activity of both the brain machinery involved in primary emotions, and another brain area, namely the ventromedial prefrontal cortex, which is sensitive to contextual and social categorisation of external objects and situations. It is through the ventromedial prefrontal cortex that brain regions like the amygdala are afterwards (and hence indirectly) activated, and the bodily responses are then executed through the hypothalamus, the basal forebrain and the brain stem.⁶

The emergence of a more complex structure of bioregulatory processes that draws upon more primitive structures of bioregulatory processes is a common feature that appears at many levels of the organisation of the organism. The hypothalamus and the brain stem, which are responsible for the most basic bioregulatory responses of the autonomic nervous system, are the means through which primary emotions will be executed. In a similar way, secondary (or social) emotions will use the biological machinery of primary emotions, which is triggered by the ventromedial prefrontal cortex. At a further level of complexity, we find planning and reasoning, which are crucial for survival, but again rely on lower level mechanisms, as will be discussed in the next section.

6. Knowledge, memory and somatic markers

Knowledge of objects, facts, persons and situations must be kept active for some time in order to be processed. The latter role is undertaken by the prefrontal cortex, which is the

⁵ On the role of the amygdala and the emotion of fear, see also Joseph LeDoux (1996).

⁶ For a comparison of the roles of the amygdala and the ventromedial prefrontal cortex in decision making, see Bechara *et al.* (1999).

main site for our working memory. The dorsolateral region of the prefrontal cortex plays a crucial role in the cognitive manipulation of knowledge, with its inferior areas being associated with object memory, while its superior areas are associated with space memory (Bechara *et al.*, 2000).

Bechara and Damasio (2005, pp. 339–40) name the innate or learned stimuli which trigger the amygdala as ‘primary inducers’, while the entities generated by the recall of primary inducers—represented in our working memory—are ‘secondary inducers’ (with the ventromedial prefrontal cortex being the critical neuroanatomical structure for secondary inducers). So the working memory plays also a key role in the triggering of somatic states from secondary inducers.

Cognitive operations are supported by processes like working memory, emotions and attention. Once more, note how the principle of using older mechanisms to perform more complex functions applies to cognitive activities as well, for the latter depend on supporting processes such as attention, working memory and emotions, and the associated brain machinery.

The ventromedial region of the prefrontal cortex establishes a linkage between knowledge of objects and situations, and bioregulatory processes, including the emotions associated with such objects and situations. By containing dispositional linkages between factual knowledge and bioregulatory states, the ventromedial prefrontal cortex enables us to learn and anticipate an association between a given social situation and a bioregulatory state. The neural dispositions of the ventromedial prefrontal cortex, which respond to mental representations (like the representation of a given social situation) by triggering emotional responses, are the ‘somatic markers’ in Damasio’s (1994) terminology.

For Damasio (1994, p. 174) ‘somatic markers are a special instance of feelings generated from secondary emotions’, which ‘have been connected, by learning, to predicted future outcomes of certain scenarios’. They are ‘somatic’ because they originate in bioregulatory states of the body, and they ‘mark’ a mental image according to the associated feeling. The somatic marker hypothesis postulates that the way in which somatic signals bias our mental representation of competing options, by attaching differential feelings to them, improves the efficiency of our decision-making process.

Thus, our emotions and feelings (more precisely, the brain circuits responsible for emotional triggering and execution and for the mapping of emotions) also influence rational decision-making. The ventromedial prefrontal cortex associates positive or negative bioregulatory states (including positive and negative emotions) to our factual knowledge of given objects, persons and situations, and biases our decision-making towards those decisions associated with positive bioregulatory states.

This emotional bias, which makes some objects and situations more salient than others, is crucial for the efficiency of decision-making, since without the help of emotions our brain could not process all the existent information. So emotions, rather than being a disturbance of rational thinking, are in fact constitutive of, and essential to, rational thinking. Damasio (1994) argues that the brain can bypass the complete execution of the bodily response associated with an emotion when the ventromedial prefrontal cortex directly activates the somatic sensorial cortex (including the insula), which monitors our bioregulatory state, without actually causing the corresponding changes in our body. This mechanism makes decision-making a more efficient process.

Some dispositions, such as those implied in the most basic bioregulatory systems, are mostly innate. Damasio (1994, p. 104) argues that innate knowledge is based in dispositional representations in the hypothalamus, the brain stem and the limbic system

(including the cingulate gyrus, the basal forebrain and the amygdala, although the latter also contains dispositions that change, especially in the first years of life). But other dispositions are mostly acquired through our particular experiences and social interaction, such as (i) the dispositions of the ventromedial prefrontal cortex, which establish links between social situations and the associated somatic sensorial maps of bodily states (i.e., the somatic markers), (ii) our dispositional knowledge about ourselves and the facts of our life and (iii) dispositions located in higher-order cortices in general.

Since the brain areas involved in primary emotions contain mostly innate dispositions, they remain relatively more stable throughout our adult life than the ventromedial prefrontal cortex, which relies more on acquired dispositions. Since the ventromedial prefrontal cortex categorises external objects and situations (including social contexts), our reaction towards social situations, and our secondary emotions, can change through our lifetime more easily than primary emotions. The fact that our dispositional knowledge about facts, objects and social situations also changes introduces an additional variability to this process.

7. Consciousness and moral behaviour

Damasio (1999) argues that the principle of using older mechanisms and dispositions to perform more complex functions can also be applied to the explanation of our consciousness. He distinguishes between three levels: the proto-self, the nuclear self and the autobiographical self. The proto-self consists in the neural patterns that regions like the brain stem, the hypothalamus (which regulates the organic biochemistry), the basal forebrain and somatic sensorial cortices produce to map our organism as we interact with the environment.

The maps that the proto-self uses are designated by Damasio (1999) as ‘first order maps’. However, the proto-self does not have the capacity to perceive itself. Changes in the first order maps of the proto-self are registered in ‘second order maps’, which monitor the relation between the organism and the object, which causes changes in the organism’s first order maps (that is, in the proto-self). Second order maps are produced by regions like the thalamus, the cingulate cortex and the superior colliculus, and contain non-verbal images of the way in which the organism is changed by the object. These second order maps produce what Damasio (1999) calls our ‘nuclear consciousness’, which gives us our sense of being here and now (interacting with an object at a given time), but provides no sense of our past, nor any capability to anticipate our future actions.

As the nuclear self interacts with objects, higher order cortices (including the frontal and temporal cortex) and subcortical nuclei (including the amygdala) store dispositional knowledge about these objects (including facts, experiences and past situations), which can be reactivated as memories through convergence zones that group together to disperse information. Damasio (1999) designates this set of past experiences of the individual, which can be reactivated at any moment as images and motor responses, as the ‘autobiographical self’. However, for these memories to be processed, the dispositional neural patterns that support our images of objects must be kept active for some time, so that the brain can process it as if the object were again present to us— while engaging secondary inducers. As noted before, our working memory, located in the dorsolateral prefrontal cortex, maintains representations of the memorised objects long enough for them to be processed.

The way in which the ‘self’ interacts with other human agents depends upon dispositions that enable the simulation of somatic sensorial states in our body when we observe other people in particular situations. Damasio (2003) designates the capacity for such simulation

as ‘empathy’. The mechanism through which we empathise with others is the following: when observing another person in a particular situation, certain regions, like the prefrontal and pre-motor cortex, send direct signals to the somatic sensorial regions that map our body state (where the insular cortex seems to play a crucial role), representing the action or emotion that the other person displays as if it were our own.

The neurons that perform this function are designated as ‘mirror neurons’, since they ‘mirror’ the other person’s state as if it were our own body state. But the emotion that would correspond to this somatic sensorial state is not necessarily executed in our body. Therefore, the brain state that is generated through mirror neurons, in the somatic sensorial regions that map our body state, need not correspond to the actual state of our body, and is thus a simulation.⁷

Empathy is essential to social cognition, and to the social interaction of ourselves with other agents. In fact, it is through empathy that another level of complexity arises, namely the level of socio-psychological structures, which shape social activity, and lead us from neuroscience to the field of social theory. Damasio (2003) names this further level of complexity as the ‘moral’ level. The somatic marker hypothesis is a most useful framework for understanding the neurobiological capacities and dispositions that enable the emergence of social and moral behaviour as another level of analysis that is not reducible to neurobiological causation. In fact, much research on neuroeconomics has addressed topics such as empathy and reciprocity too, providing insights that can be integrated within the overall framework provided by the somatic marker hypothesis. This topic will now be addressed in the next section.

8. Social behaviour and social cognition

One example of an important development in neuroeconomics, which can be integrated within Damasio’s explanation of how social activity emerges from neurobiological structures, concerns the role of trust in rational deliberation. In mainstream game theory, cooperation is often explained as the outcome of rational (strategic) deliberation. However, much research in neuroeconomics has shown how the release of molecules like oxytocin through the pituitary gland of the hypothalamus is involved in affective bonding, maternal behaviour and trust, leading agents to engage in cooperative behaviour even when ‘rational’ (strategic) deliberation would lead to a non-cooperative response (see Kosfeld *et al.*, 2005, or Zak *et al.*, 2005).

Damasio (2003) notes that the release of molecules like oxytocin through the pituitary gland of the hypothalamus is also crucial for survival, since it directs human beings towards group solidarity, thus protecting them from external threats (for human beings are more likely to survive in group than in isolation). Damasio (1994) also notes that the concentration of serotonin in regions like the ventromedial prefrontal cortex and amygdala has been found to be positively correlated with cooperative behaviour and the inhibition of aggressive behaviour. The interaction of serotonin with other systems and neurotransmitters (like dopamine, norepinephrine or acetylcholine) is also an important part of this process (see Damasio, 1994, pp. 76–8).

Knoch *et al.* (2006, p. 831) also found that the right dorsolateral prefrontal cortex is ‘crucial in implementing fairness behaviours by overriding self-interested impulses’, which

⁷ On the role of mirror neurons and the activation of the insula in social contexts, see Wicker *et al.* (2003) and also Gallese *et al.* (2004).

means that reciprocity relies not only upon a positive emotional state towards others induced by emotion-executing regions like the hypothalamus, but also on the capacity of higher order structures like the right dorsolateral prefrontal cortex to exercise top-down (or executive) control. Here, we have different brain systems, which generate conflicting motivations, and the outcome (e.g., whether the agent cooperates or not) will be conditional on the somatic state of the agent.

Social cognition in general depends not only upon what was termed above as ‘empathy’—the activation of our somatic sensorial regions through ‘mirror neurons’, which make us simulate in our somatic sensorial cortex the emotional states we observe in other agents—but also on the ability to ‘mentalise’. To have the ability to mentalise is to be able to represent the psychological perspective of another person, that is, to make attributions about the mental states of another person, such as about the desires, beliefs or intentions of another person (see Amodio and Frith, 2006; Singer and Fehr, 2005). The medial frontal cortex (including the anterior cingulate cortex), together with the temporoparietal junction, the superior temporal sulcus and the temporal poles, have been found to play an important role in this process.

The neurobiological substrate that enables ‘mentalising’, together with the neurobiological substrate that enables ‘empathy’, described above, constitutes a key element to the explanation of how social and moral behaviour emerges. A most helpful step in developing the framework of the somatic marker hypothesis would in fact be to integrate the research on ‘mentalising’ and ‘empathy’ with the research on somatic markers in order to explain the interplay between the human agent’s somatic markers and the social structures that emerge. This leads to a conception of the human agent that contrasts with the conception of instrumental rationality driven by self-interest that has been a central element of mainstream economic theory.

9. Self-interest, prudence and rationality

The conception that arises in the somatic marker hypothesis is one in which changes in somatic states, and emotions, play an important role in decision making and in shaping social behaviour. Ashraf *et al.* (2005) argue that neuroeconomics, and behavioural economics in general, point towards a return to Adam Smith’s (2002) conception of an human agent driven by a variety of moral sentiments. It is certainly true that there are many connections between Smith’s analysis of moral sentiments and a neuroscience-informed study of emotions. In fact, notions addressed in neuroscience, such as ‘empathy’ and ‘mentalising’, are very similar to ideas presented by Adam Smith (2002), such as Smith’s notion of ‘sympathy’. The use of neuroscience in order to explain how social behaviour emerges from emotions and sentiments ultimately leads back to a descriptive approach that underpinned Adam Smith’s (2002) contribution, and was later abandoned by mainstream economics.

However, it is also true that the more fruitful framework to develop such connections will not be one in which the preference orderings that are generated by each brain state are assumed to be complete orderings, or any framework that assumes a closed system that can be represented through deductivist models aimed at prediction. The interplay between competing motivations, sentiments and emotions, an essential aspect of Smith’s analysis, is a crucial aspect that would be lost given the constraints imposed by deductivist modelling.

The way in which mainstream economists attempt to make Smith’s conception compatible with deductivist modelling consists in assuming that a particular motivation,

namely self-interest (which is characterised in terms of a single complete preference ordering), is the more important motivation for Smith's analysis.⁸

The motivation of self-interest, which has been traditionally the key motivation for mainstream economics, and believed by many to be the key motivation behind Smith's analysis, can be fruitfully integrated within a broader conception of human agency—as it was also the case in Smith's analysis, as Sen (1987) notes, in contrast with the mainstream interpretation of Smith. In fact, research on neuroscience and neuroeconomics has shown that there is a complex interplay, and sometimes conflict, between self-interest and competing motivations.

One example of conflicting motivations, in which self-interest plays a key role, can be observed when there is a conflict between the pursuit of pleasure on the one hand, and rational deliberation (be it moral reasoning, or any rational consideration of long-term goals) on the other hand. The main site for planning and rational deliberation is our prefrontal cortex (which, remember, also relies on supporting mechanisms such as the 'somatic markers'). However, rational deliberation often conflicts with motivations such as pleasure-seeking, which in turn depends upon specific systems, such as the dopaminergic system.

The dopaminergic system has been found to play an important role in motivations generated by the desire of reward, through the transfer of dopamine from the ventral tegmental area and the substantia nigra to the nucleus accumbens of the basal forebrain. The anticipation of a reward, produced by the dopaminergic system, can thus create a conflict between rational deliberation and pleasure-seeking (see Berridge and Robinson, 1998, 2003).

Bechara and Damasio (2005) argue that the dopaminergic activity that takes place in the striatum biases somatic states covertly (and hence the dispositions involved are non-conscious), by changing the threshold for neuronal firing in structures such as the insula, somatic sensorial cortices S1 and S2, the amygdala and the ventromedial prefrontal cortex—unlike other systems, which bias the somatic states in an overt way, such as the serotonergic activity that (Bechara and Damasio, 2005, suggest) takes place through the anterior cingulate. Thus, decision-making (and subjective well-being) may be affected at a non-conscious level, biasing what is often taken to be a 'rational' decision.

Smith's notion of prudence, which Sen (1987) discusses extensively, arises from various motivations concerned not only with self-interest, but also with an ability to restrain self-interest, and includes not only the ability for reason and understanding, but also for what Smith (2002) calls 'self-command'. The ability for reason, understanding and self-command that Smith and Sen discuss relies much on the suppression that regions from the neocortex exert on impulses generated in lower level regions. The ability for restraining from impulses generated by pleasure-seeking behaviour depends on other important mechanisms, which act in a complex way in combination with other emotions.

Brain circuits that are important for the suppression of pleasure-seeking include not only regions in the neocortex, such as the dorsolateral prefrontal cortex, but also regions associated with the emotion of fear. Remember that fear is triggered in the amygdala, and executed by regions such as the hypothalamus, the brain stem and the basal forebrain. Somatic sensorial regions map the body state caused by the emotion of fear, generating our feeling of the emotion (see Damasio, 1994, 1999, 2003). The execution of this bioregulatory process depends upon the sympathetic system (which in turn relies upon neurotransmitters such as norepinephrine and acetylcholine, amongst others).

⁸ See Sen (1987) for a discussion.

Preference orderings generated by the activation of the emotion of fear will give more prominence to safety concerns, comparatively to preference orderings that are generated without such an emotional stimulus. This can lead to a conflict between a motivation that is driven by safety concerns and other motivations where safety is not a prominent concern, such as reward-seeking (e.g., when attaining a reward requires a dangerous action).

Although Smith (1776) analysed extensively the impact of agency driven by self-interest, for Smith self-interest is a motivation that is always present in an interplay with other motivations, where this interplay is not an accidental perturbation of a well-defined preference ordering that reflects self-interest, but is in fact constitutive of all human action, even self-interested action. For example, emotions are not a perturbation of rational behaviour but, in fact, are essential to decision making, as Damasio argues in his somatic marker hypothesis.

In fact, if we focus on self-interest, abstracting from other motivations, we need not consider that the other motivations we are abstracting from do not play a role in self-interested behaviour too. Abstraction (as defined in Lawson, 1997, 2003) consists in focusing on one aspect while leaving other aspects aside momentarily, but without supposing that other things left aside are not playing a role in what is observed. Isolation, on the other hand, consists in picking up a part of reality while supposing that other things left aside are not playing a role in the part of reality we are analysing. But when we assume isolation in a part of reality that is in fact connected to other parts of reality, we are fictionalising (making a false assumption) rather than abstracting.

Mainstream economists take Smith's self-interest in isolation from other motivations, while assuming a single complete preference ordering that describes self-interest, or other fictional assumptions, in order to obtain closed systems, which render deductivist models applicable. Much research on neuroeconomics essentially accepts the (false) isolation of competing motivations, so that deductivist modelling can proceed, although with the caveat that each complete preference ordering is conditional on a given brain state.

The conception of human agency we arrive at in the light of recent neurobiological evidence contrasts not only with the postulate that a single complete preference ordering is sufficient to explain behaviour, but also with the postulate that various complete orderings act in isolation, and leads to an open system conception of reality, which contrasts with the presupposition of closed systems that characterises mainstream economics.

10. Concluding remarks

The use of neuroscience in economics has consisted in many cases in the employment of knowledge from neuroscience in order to obtain more sophisticated models so as to achieve prediction of events. Although neuroeconomics takes into account the existence of various motivations and preference orderings, in many cases each of these preference orderings is taken to act in isolation, within models which differentiate between a 'rational' state and an 'emotional' state, so that closure conditions are artificially generated (in fact, falsely posited).

This approach is inconsistent with a perspective where rationality cannot be seen as completely separated from emotions, as in the somatic marker hypothesis. According to the somatic marker hypothesis, decisions will be biased depending on the emotional charge associated with each option, where we do not have rational choices on the one hand, and emotional choices on the other hand, but rather different degrees of rationality and of (various types of) emotionality in every circumstance. In fact, emotions are essential to the exercise of reasoning, since if emotions did not make some options more salient, our brain

could not compute all available options, as Damasio argues. Emotions are constitutive of, and not an absence of, rationality.

Thus, our mental representations and our emotions (and feelings) are strongly interconnected, and causally interact in both directions. Even the emotions that support mental representations are themselves interconnected in a complex way. As Damasio (2003) notes, social (or secondary) emotions, for example, are a result of a complex interplay between primary emotions and representations of social situations.

Emotions and sentiments are a complex combination of other bioregulatory processes (including other emotions), which cannot be taken to act in isolation. We do not have a case of isolated brain circuits, but rather a conception where the interrelations between brain circuits and the emerging motivations are constitutive of each motivation that arises—we have a case that Lawson (1997, 2003) terms ‘internal relations’, when the relations between several entities are constitutive of those entities. While it is true that not all trends of neuroeconomics engage in deductivist modelling, it is also true that the implications of the research in neuroscience and neuroeconomics for deductivist modelling have been left unaddressed, even within neuroeconomics. The explanation of the emergence of co-operation, social structures, and social behaviour in general, has been constrained by the failure to realise the limitations of deductivist models for explaining choice, behaviour and well-being.

The picture that emerges within a broad conception of human motivation, informed by knowledge of the neurobiological substrate from which social behaviour emerges, is one in which we can thus distinguish different motivations, which rely upon different brain systems and circuits, that are not totally independent and which contain neural dispositions that may change during our lifetime. These motivations play a central role in our choice process and in our subjective representations of well-being. The neurobiological evidence about these different motivations contradicts the mainstream use of methods that presuppose closed systems.

However, the existence of these conflicting motivations does not mean that there cannot be an integrated and systematic theory of human behaviour. Rather, it means that such a theory of human behaviour should be formulated not in terms of complete and stable preference orderings, or any other assumptions that enable the formulation of a closed system, but rather in terms of the underlying structures and dispositions that cause actual behaviour—and cause the adaptation of preferences too. Concepts such as ‘somatic marker’, and the explanation of the brain in terms of the structured set of bioregulatory processes that were developed during a long evolutionary process, constitute a promising starting point for a research programme aimed at developing such a systematic theory of human behaviour.

In fact, different motivations mentioned here, such as physiological motivations, reward-seeking motivations, trust, empathy or security, all have a common denominator: they influence choice and behaviour by changing the underlying somatic state of the chooser, and even rationality is supported by similar mechanisms. The network of somatic markers identified by Damasio and colleagues can be seen as an underlying structure, or set of dispositions, in terms of which actual behaviour can be understood. The network of somatic markers, located in the ventromedial prefrontal cortex is, of course, connected to a larger hierarchical structure of bioregulatory processes, which support cognition and motivation, from which social and moral behaviour emerges.

Although much research remains to be done in order to develop the hypothesis suggested here, the more general form of the overall framework can already be sketched.

The structure of bioregulatory processes, which supports cognitive and motivational processes, is a hierarchical structure in which lower level processes and mechanisms support higher level processes. This hierarchical structure results from the fact that during a long evolutionary process, the brain has used older and simpler mechanisms in order to perform more complex functions. From these mechanisms emerges yet another level—the social realm. This hierarchical structure can accommodate more specific theories concerning specific subsystems that constitute it.

Social theories, which address the emergent social realm, can also benefit much from an understanding of the neurobiological basis of social and moral activity. In fact, the use of Damasio's somatic marker hypothesis within economic analysis points towards a return to Adam Smith's (2002) approach which, like Damasio's, aimed at identifying a set of moral sentiments (and social emotions) as the central concepts for a theory of social and moral behaviour.

Bibliography

- Amodio, D. M. and Frith, C. D. 2006. Meeting of minds: the medial frontal cortex and social cognition, *Nature*, vol. 7, 268–77
- Ashraf, N., Camerer, C. F. and Loewenstein, G. 2005. Adam Smith, behavioral economist, *Journal of Economic Perspectives*, vol. 19, 131–45
- Bechara, A. and Damasio, A. R. 2005. The somatic marker hypothesis: a neural theory of economic decision, *Games and Economic Behaviour*, vol. 52, 336–72
- Bechara, A., Damasio, H., Damasio, A. R. and Lee, G. P. 1999. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making, *Journal of Neuroscience*, vol. 19, 5473–81
- Bechara, A., Damasio, H. and Damasio, A. R. 2000. Emotion, decision making and the orbitofrontal cortex, *Cerebral Cortex*, vol. 10, 295–307
- Bernheim, D. and Rangel, A. 2005. Addiction and cue-triggered decision-making processes, *American Economic Review*, vol. 95, 314–46
- Berridge, K. C. and Robinson, T. E. 1998. The role of dopamine in reward: hedonics, learning, or incentive salience, *Brain Research Reviews*, vol. 28, 308–67
- Berridge, K. C. and Robinson, T. E. 2003. Addiction, *Annual Review of Psychology*, vol. 54, 25–53
- Camerer, C. F. 2007. Neuroeconomics: using neuroscience to make economic predictions, *Economic Journal*, vol. 117, C26–42
- Camerer, C. F., Loewenstein, G. and Prelec, D. 2004. Neuroeconomics: why economics needs brains, *Scandinavian Journal of Economics*, vol. 106, 555–79
- Camerer, C. F., Loewenstein, G. and Prelec, D. 2005. Neuroeconomics: how neuroscience can inform economics, *Journal of Economic Literature*, vol. 34, 9–64
- Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, G.P. Putnam and Sons
- Damasio, A. R. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, Harcourt Brace and Company
- Damasio, A. R. 2003. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*, New York, Harcourt Brace and Company
- Damasio, A. R. 2008. Neuroscience and the emergence of neuroeconomics, pp. 209–14 in Glimcher, P. W., Camerer, C., Fehr, E. and Poldrack, R. (eds), *Neuroeconomics: Decision Making and the Brain*, Amsterdam, Elsevier
- Elster, J. 1983. *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge, Cambridge University Press
- Gallese, V., Keysers, C. and Rizzolatti, G. 2004. A unifying view of the basis of social cognition, *Trends in Cognitive Sciences*, vol. 8, 396–403
- Glimcher, P. W., Camerer, C., Fehr, E. and Poldrack, R. 2008. *Neuroeconomics: Decision Making and the Brain*, Amsterdam, Elsevier

- Jevons, W. S. 1871. *The Theory of Political Economy*, first edn, London and New York, MacMillan and Co
- Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*, London, MacMillan
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. and Fehr, E. 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex, *Science*, vol. 314, 829–32
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. and Fehr, E. 2005. Oxytocin increases trust in humans, *Nature*, vol. 435, 473–6
- Lawson, T. 1997. *Economics and Reality*, London, Routledge
- Lawson, T. 2003. *Reorienting Economics*, London, Routledge
- LeDoux, J. E. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, New York, NY, Simon & Schuster
- MacLean, P. D. 1990. *The Triune Brain in Evolution: Role in Paleocerebral Function*, New York, Plenum
- Nussbaum, M. C. 2000. *Women and Human Development: The Capabilities Approach*, Cambridge, Cambridge University Press
- Romer, P. M. 2000. Thinking and feeling, *The American Economic Review*, vol. 90, 439–43
- Samuelson, P. 1947. *The Foundations of Economic Analysis*, Cambridge, MA, Harvard University Press
- Sanfey, A. G., Loewenstein, G., McClure, S. M. and Cohen, J. D. 2006. Neuroeconomics: cross-currents in research on decision-making, *Trends in Cognitive Sciences*, vol. 10, 108–16
- Sen, A. K. 1982. *Choice, Welfare and Measurement*, Oxford, Blackwell; Cambridge, MA, MIT Press
- Sen, A. K. 1985. *Commodities and Capabilities*, Oxford and New York, Oxford University Press
- Sen, A. K. 1987. *On Ethics and Economics*, Oxford and New York, Basil Blackwell
- Sen, A. K. 2002. *Rationality and Freedom*, Cambridge, MA, The Belknap Press of Harvard University Press
- Singer, T. and Fehr, E. 2005. The neuroeconomics of mind reading and empathy, *American Economic Review*, vol. 95, 340–5
- Smith, A. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*, London, Methuen
- Smith, A. 2002 [1759]. *The Theory of Moral Sentiments*, Cambridge, Cambridge University Press
- Walsh, V. 1996. *Rationality, Reproduction and Allocation*, Oxford, Clarendon Press
- Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. and Rizzolatti, G. 2003. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust, *Neuron*, vol. 40, 655–64
- Zak, P. J., Kurzban, R. and Matzner, W. T. 2005. Oxytocin is associated with human trustworthiness, *Hormones and Behaviour*, vol. 48, 522–7