



## Cross-talk in economics and neuroscience

Sacha Bourgeois-Gironde, Carl Schoonover

► **To cite this version:**

| Sacha Bourgeois-Gironde, Carl Schoonover. Cross-talk in economics and neuroscience. Revue d  
| Economie Politique, Editions Dalloz, 2008, 118 (1), pp.35-50. <ijn\_00432665>

**HAL Id: ijn\_00432665**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_00432665](https://jeannicod.ccsd.cnrs.fr/ijn_00432665)**

Submitted on 16 Nov 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cet article est disponible en ligne à l'adresse :

[http://www.cairn.info/article.php?ID\\_REVUE=REDP&ID\\_NUMPUBLIE=REDP\\_181&ID\\_ARTICLE=REDP\\_181\\_0035](http://www.cairn.info/article.php?ID_REVUE=REDP&ID_NUMPUBLIE=REDP_181&ID_ARTICLE=REDP_181_0035)

---

## Une hybridation de l'économie et des neurosciences a-t-elle un sens ?

par Sacha BOURGEOIS-GIRONDE et Carl SCHOONOVER

| Dalloz | Revue d'économie politique

2008/1 - Volume 118

ISSN 0373-2630 | pages 35 à 50

---

Pour citer cet article :

— Bourgeois-Gironde S. et Schoonover C., Une hybridation de l'économie et des neurosciences a-t-elle un sens ?, Revue d'économie politique 2008/1, Volume 118, p. 35-50.

---

Distribution électronique Cairn pour Dalloz.

© Dalloz. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# Une hybridation de l'économie et des neurosciences a-t-elle un sens ?

Sacha Bourgeois-Gironde\*  
Carl Schoonover\*\*

La neuroéconomie est une extension récente de l'économie comportementale qui cherche à mettre en évidence les activités et les mécanismes cérébraux qui sous-tendent les comportements normaux ou déviants par rapport à la théorie économique de la rationalité. Gul et Pesendorfer [2005] ont formulé une critique sérieuse contre la validité épistémologique de la neuroéconomie. Une de leurs critiques fondamentales est basée sur l'incommensurabilité des concepts utilisés en économie et en neurosciences. Nous répondrons à cette critique par l'exemple, en envisageant différentes directions d'ajustement possibles entre les deux disciplines. Cette hybridation disciplinaire peut mettre en relief l'unité des sciences comportementales.

**neuroéconomie - unité des sciences comportementales - valeur - rationalité - émotions**

## *Cross-talk in economics and neuroscience*

Neuroeconomics is a recent extension of behavioural economics which aims at uncovering the brain mechanisms and activities that mediate regular and anomalous economic behaviour. Gul and Pesendorfer [2005] have launched a critique against the neuroeconomic research programme, based on what they argue is the incommensurability of the theoretical constructs employed by each respective discipline. To respond to their argument we envision and illustrate several "directions of instruction" between neuroscience and economics, and provide counterexamples to their critique. This disciplinary cross-talk suggests that neuroeconomics may play a crucial conceptual and methodological role in fostering the unity of behavioural sciences.

**neuroeconomics - unity of behavioural sciences - value - rationality - emotions**

*Classification JEL : D87*

## 1. The "mindless economics" objection: incommensurable disciplines and constructs

Neuroeconomics, a hybrid composed of experimental methods in neurosciences and economics, can be understood as an extension of behavioural

\* Institut Jean-Nicod (CNRS-EHESS-ENS)

\*\* Columbia University

economics. It is predicated upon the hypothesis that in order to faithfully account for human economic behaviour, one needs to understand the neural underpinnings of this behaviour. In other words, since biology generates behaviour, the subject of behavioural economics, an understanding of the underlying biological mechanisms will afford a more accurate, and finer-grained, explanation of this behaviour. Topics of interest in neuroeconomics include choice under uncertainty, social games and fairness, the tension between consumption and saving, preference construction, and decision-making. Several surveys summarize the various lines of research and recent results in this emerging field (Camerer and al. [2005], Rustichini [2005], Braeutigam [2005]). Three major converging trends has given rise to the neuroeconomics research programme: the availability of new brain-imaging tools, such as functional magnetic resonance imaging (fMRI), which measure neural activity underpinning cognitive events; the recognition that neural systems encode economically relevant factors such as uncertainty and value, as well as the identification of brain systems underlying economically relevant behavioural patterns such as impulsivity, utility discounting or altruism; and finally, the rise of behavioural economics as a discipline that relaxes the strict rationality assumptions of classical economics in order to generate more psychologically realistic models of economic behaviour.

In spite of the optimism of many of its heralds, this programme has been subjected to a series of critiques launched against its rationale and epistemological coherence. Some find grounds for optimism in the potential for overlap between economics and neuroscience, specifically because they believe that it is possible to construct methodological and formal bridges between the two fields. A pointed critique takes this very point to task: "Economics and psychology address different questions, utilize different abstractions, and address different types of empirical evidence. Neuroscience evidence cannot refute economic models because the latter make no assumptions and draw no conclusions about the physiology of the brain. Conversely, brain science cannot revolutionize economics because the latter has no vehicle for addressing the concerns of economics" (Gul et Pesendorfer [2005]). Gul and Pesendorfer do not assess the contributions of neuroeconomic research; rather, without denying that neuroeconomics might afford interesting findings, they question their relevance to descriptions and models of the behaviour of the economic man, and reject the possibility that mind/brain data can unseat classical assumptions that underlie economics. In this article we reconsider the issue of the relevance of this kind of data to economics. Specifically, we outline several promising "directions of instruction", or ways in which the findings and models in one field can contribute to development in the other. In sum, we seek to put pressure on Gul and Pesendorfer's negative claim that there is no vehicle through which one field can address the concerns of the other, by presenting a general view of possible interactions between them.

In this paper, we seek to address the following four key epistemological issues:

i) What are the "directions of instruction" between neuroscience and economics? Should we use economic models to model what happens in the brain, or rather neural data to analyze, critique or refine economic models

and theories? Clearly, the latter is the focus of Gul and Pesendorfer's – and our – discussion, but the term “neuroeconomics” is used at times to denote the former direction, which renders discussions of the field somewhat ambiguous. A more general way of formulating the question is to ask whether neuroeconomics is a branch of neuroscience or of economics. To some authors like Sanfey [2006] the two-tiered contribution is acknowledged with no further epistemological fuss.

ii) What is the role of new data about brain processes in refuting or confirming economic theories? One of neuroeconomics' highest objectives is to find neural correlates of economic concepts that, when taken together with mathematical constructs and behavioural data would either corroborate or invalidate them.

iii) Can brain data enable neuroeconomists to discriminate between theories that make differing predictions?

iv) Can neural facts provide alternative foundations to economic models? Some ambitious neuroeconomists do not only point to brain data as confirming their favourite models but as a broader justification of their particular views on economic behaviour.

In addressing these issues we seek to respond to Gul and Pesendorfer's critique by providing examples of fecund cross-disciplinary overlaps that result from the use of common theoretical constructs. Taken together, they plead in favour of the unity of behavioural sciences.

## 2. Economics and neuroscience may share theoretical constructs

### 2.1. Analysis of brain-functioning in decision terms

One way of highlighting the gap between the respective explanatory levels of neuroscience and economics is to emphasize the contrast between, on the one hand, algebraic utility models that may approximate human decision behaviour at a macro-level and, on the other hand, biologically-inspired models that focus on overt decision behaviour. There are nonetheless several ways one may bridge this gap. One is to relinquish some of the fundamental rationality assumptions that economic theories make. For instance, one may abandon the notion that behaviour is a “top-down” process that conforms to ideal axioms of rationality, and replace it with a complex “bottom-up” system of coordination and self-organization from which recognizable patterns of behaviour emerge. The idea here would be to bridge the gap between neural patterns on the one hand, and behaviour and economic theoretical constructs on the other. Neuroeconomics then would be

come a science of levels of behavioural self-organization, applicable to either brains or markets, and would borrow its formalizations from complex systems theory rather than classical economics models. Clearly this would constitute a shift away from the classical concerns of economists and psychologists in interpreting human behaviour. In particular, beyond relaxing certain axiomatic rationality assumptions in order to approximate actual behaviour, it would result in altogether forsaking top-down assumptions and replacing them with physical laws of organisation and coordination. This approach was originally inspired by Hayek [1952] and can be found in V. Smith's own appraisal of the relevance of neuroeconomics (Smith [2003]).

In a similar vein, an approach favoured by Sanfey *et al.* (Sanfey [2006]) seeks to combine joint insights from economics and neuroscience in order to gain a model of the functional organization of the brain and to acquire data on neural correlates of economic behaviour. This notion of mutual benefit for the two disciplines diverges from a traditional epistemological quest for an independent foundation of a new discipline, and instead defines neuroeconomics as a by-product of possibly converging endeavours. The gist of the proposal is that the brain is a hierarchical structure encompassing different subsystems cooperating or competing for the control of overt behaviour. This forms the picture of a brain as a firm (Brocas et Carillo [2006]). One of the main functional contrasts here is between automatic processes and controlled and reflective ones. Sanfey's hierarchical model envisions this contrast in terms of cooperation and competition between executive control systems and emotional structures. Dual systems vary along several dimensions – such as deliberative and emotive, or slow and fast, or cognitively flexible and cognitively rigid – and capture a diverse array of contrasted observations on decision behaviour (Kahneman [2003]). High-level executive and deliberative processes engage anterior and dorsal lateral regions of the prefrontal cortex while low-level and automatic psychological processes rely on the limbic system. The fact that different neural structures mediate different behaviours relevant to decision-making, invites us to revise, in turn, our understanding of how we make decisions, such as unitary models in terms of expected utility (choice between different outcomes) or discounted utility (choice extended over time).

We should disambiguate here between the application of economic models in order to understand the brain's decision-making processes, and the more general notion that economic theory may offer a method for understanding brain function. Camerer [2005] notes: "neuroscience is shot through with familiar economic language – delegation, division of labor, constraint, coordination, executive function – but these concepts are not formalized in neuroscience as they are in economics. There is no overall theory of how the brain allocates resources that are essentially fixed (e.g. blood flow and attention). An 'economic model of the brain' could help here".

## 2.2. The search for neural correlates of value

Some recent single-cell physiology experiments are particularly promising examples of how economic concepts can mesh with the methodology of biology. Paton and his colleagues [2006] trained monkeys using a “reinforcement learning” paradigm so that the animals would learn to associate abstract images with a reward (juice) or penalty (an irritating air puff directed towards their eyes). Thus, when the image was presented before the penalty or the reward, the monkeys were able to anticipate the outcome and behave accordingly. After training, the monkeys know to begin licking a juice bottle when presented with a “good” image, and blink when presented with a “bad” one, in order to avoid the noxious stimulus.

As the monkeys engaged in this behavioural task, the authors recorded single neurons in the amygdala, an area of the brain that is associated with reinforcement learning. They found that these neurons were reliably predictive of the rewards and penalties in the task, that is, the temporal dynamics of their electrical activity correlated with the type of payoff that the monkey could expect. Some neurons displayed a burst of electrical activity immediately after the display of a “bad” image; others displayed a similar burst immediately after the display of a “good” one. The authors then performed a control experiment in order to lend support to the hypothesis that these neurons effectively carry representations of value. They inverted the rewards and penalties associated with each image and, as expected, found that the monkeys adapted their behaviour to the novel circumstances, and came to associate the old “bad” image with the positive reward, and vice-versa. Crucially, the authors observed that the electrical activity in these neurons was modulated, reflecting the novel scenario. For instance, neurons that had displayed a high level of activity after a “good” image was presented, no longer displayed this activity after the switch. Furthermore, the modulation of this neuronal activity fit the temporal dynamics of monkey's behavioural adaptation.

These experiments demonstrate that there are representations of positive and negative value encoded at the level of single cells in the monkey amygdala. Other recent experiments (e.g. Platt et Glimcher [1999], Roesch et Olson [2004], Doris et Glimcher [2004]) have shown different types of value representation in different areas of the primate brain. Platt et Glimcher's study went so far as to vary the relevant variables relevant to the computation of an expected value (probability of gain and size of gain) and found strict correlations in the neural dynamics. In sum, these studies borrow, to varying degrees of complexity, notions inherent to behavioural economics in order to determine the optimal behaviour of the animals, and then establish neural correlates for the variables that are necessary to making decisions on a trial-by-trial basis. It should be noted that while these types of studies depend in part on behavioural economics, and employ its formalizations to some degree, they do not aim to challenge, refine, or develop economic theories. Rather, they harness notions from economics in order to develop neurobiology itself, introducing into the biological sciences concepts that are typically the domain of economics, like value and utility. The studies

then present a clear counter-example to Gul and Pesendorfer's contention that the two disciplines can't mix because they employ different abstractions.

### 3. Brain data may help refine, confirm or refute economic theories and concepts

Modern economics is based of psychological concepts that date back to the end of 19th century or the beginning of 20th century. Knight's distinction between risk and uncertainty is one case in point. Even more fundamentally the substitution of ordinal utility for cardinal utility is a founding act of the modern axiomatization of rational economic behaviour, and, perhaps, a step away from psychological realism. Neuroeconomics can bear on these initial conceptual orientations by elucidating fine-grained neural facts that would deepen and sophisticate our conceptualization of risk and uncertainty and by documenting how humans assess the value of a good without comparing it to others.

Ellsberg [1961] devised a famous distinction between subjects who display contrasting attitudes when facing risky situations (where they do in fact employ the principles of probability theory) versus uncertain situations (where they do not, and simply manifest a kind of "uncertainty aversion"). Since its inception Ellsberg's paradox has attracted a fair amount of attention from decision-theorists (Gilboa et Schmeidler [1989]) and, more recently, from practitioners of neuro-imaging (Rustichini [2005]). Huettel and his colleagues (Huettel [2006]) recently detected individual differences in brain activation depending on a subject's preferences, and aversion to risk and ambiguity. People who prefer ambiguity demonstrate increased activity in the prefrontal cortex while those who prefer risk have increased activity in the parietal cortex. This, taken together with the fact that, in our societies, subjects manifest varying degrees of "uncertainty aversion", provides grounds for re-examining the standard model of rationality. The neural mechanisms that explain the varied attitudes toward uncertainty are a particularly fruitful avenue of neuroeconomics research.

Dayan and Yu point out an intuitive contrast between "expected uncertainty", which refers to the fact that our knowledge of the environment is based on noisy signals (we actually expect not to know all outcomes in every situation), and "unexpected uncertainty" which points to the unsteadiness of the environment due to the fact that the identity of relevant signals may change when the context changes (Dayan et Yu [2003]). In the ecological situations in which we face environment variability, we switch between one sort of uncertainty and the other. We are sometimes unsure of which type of context we are currently in. These different forms of uncertainty are represented and manipulated in different ways and one might expect that they engage different neural systems. Dayan and Yu developed a model of



how two distinct neuromodulatory systems – acetylcholine and noradrenaline – encode uncertainty and how their interaction might predict whether we should abide by a prior belief in the reliability of a cue or abandon it. We tend to adopt different attitudes depending on the form of uncertainty we have encoded in relation to a given environment. In particular, we either stick to our beliefs or search for novel strategies and opportunities. This dichotomy has been described as “exploitation versus exploration”. It certainly captures a basic distinction between the ways we make investments: stick to familiar sources of rewards (professional, affective) or seek new sources in a more or less reliable environment. This contrast is well-suited to support the design of ecological protocols in neuroeconomics by mixing field studies and measurements of brain-activity. For instance, some neuroeconomists (see Lo [2005]) have recently focused on the study of stock-markets from the perspective of neuromodulation.

In the same way that they wish to differentiate between different kinds of attitudes towards uncertainty by elucidating their neural underpinnings, neuroeconomists hope investigate the neural substrates of utility – the economic theoretical construct *par excellence*. Cardinal utility is a form of utility which is measured in arbitrary units of pleasure; each individual may assign different values of the utility of a given good. Since it is difficult to compare the valuation of the same goods across subjects, and the valuations of different goods made by individual subjects, economists have resorted to using ordinal utility and ranked preferences, in which the individual subjective valuation of a good is deprived of sense. Stuphorn [2006] rightly points out, however, that from a psychological and neuronal point of view, it would be interesting to distinguish between assignment of value on one side, and comparison of values on the other side. Padoa-Schioppa and Assad [2006] have identified a population of neurons in the orbitofrontal cortex that assigns values to economic goods and represents its subjective utility independently of the action needed for its acquisition. The suggestion is that the orbitofrontal cortex might contain a map for subjective cardinal utility. In their experiments the authors had macaque monkeys choose between two types of juice offered in different amounts. In some trials the monkey could choose between one drop of grape juice (which monkeys prefer) and one drop of apple juice. In other trials, the monkey chooses between one drop of grape juice and two drops of apple juice, and so on. The authors noted a trade-off point between kinds of juices, at around one drop of grape juice for three drops of apple juice. With this trade-off in mind, Padoa-Schioppa and Assad then searched for correlations between neural activity in the orbitofrontal cortex and the choice of one option over the other. They found that some of these cells represented a subjective valuation of the reward, which, crucially, were not influenced by the physical properties of the reward (type of juice et volume), or by the action by which it was to be obtained, suggesting that they may serve as a cardinal utility map. The findings touch upon foundational concepts in economics. Indeed, should we consider economic choice as one between goods rather than as the result of comparisons between consequences of actions undertaken to acquire those goods? Second, does that choice rely on the “direct” assignment of values to available goods, or comparisons between different goods? The facts in the brain bear directly on the principles of economics.

## 4. Brain data may provide grounds for deciding between competing economic theories

In addition to confirming or refuting economic concepts, one may employ brain data in order to discriminate between two theories that offer differing predictions in a given situation. Suppose that we were to decide between two theories, (A) and (B), the former based on a linear conception of expected utility, and the latter taking into account the loss aversion phenomenon. Given a simple lottery offering the choice between (1) a 0.1 probability of obtaining \$1000 and (2) a 0.5 probability to obtaining \$200, theory (A) would predict that the brain data would look identical in either case, assuming that the cognitive mechanism for evaluating (1) and (2) is identical (a plausible assumption, since both cases involve the same operation). On the other hand, theory (B) would predict that brain data might look quite different in either case, under the assumption that evaluating the lotteries would implicate different activation levels of areas in the brain that mediate emotions, or whichever centres are responsible for risk-aversion. Detecting differential activation of known neural systems would provide grounds for deciding in favour of theory (B), although of course if no differences were detected, one could not draw any conclusion from the data. This approach differs from the attempt to directly correlate, for instance, behaviours implicating loss aversion or linear expected utility with a particular neural event. This method rather would consist in taking advantage of neural observations – whatever their direct relevance to the economic concepts under scrutiny – in order to discriminate between different theories encompassing those concepts.

An instance of this kind of methodology may be found in De Martino and colleagues' [2006] study of the neural correlates of framing-effects. They investigated the neural pattern associated with decisions between small, sure amounts of money and higher, riskier ones in experimental conditions wherein the description or framing of outcomes was manipulated. The descriptive factor that was manipulated had an affective value: the sure option was either phrased in terms of the sum to be kept or the sum to be deducted from a fixed amount of money. The different options, while phrased differently, would result in the same outcome. The study revealed that subjects chose the sure option more often when it was phrased positively ("keep 20 out of 50"), than when it was framed negatively (lose 30 out of 50). Brain imaging showed that the activity of the amygdala – a region traditionally associated with the control of emotions and the mediation of immediate impulsive reactions – was correlated with this bias in the decision process. The decomposition in elementary stages of this evidence yields a twofold process of an immediate emotional response to frames used to present the prospects and, a tendency to accept or avoid that prospect. Different activities in another brain region, the anterior cingulate cortex – associated with the interpretation and solving of cognitive conflicts – was correlated with the

congruence between the chosen action (avoid or accept the prospect) and the initial emotional reaction. De Martino's results, then, would offer an indirect confirmation of Kahneman and Tversky's Prospect Theory's prediction according to which subjects tend to avoid risk to secure a sure gain and take risks to avoid possible losses (see Kahneman and Frederick [2006]). Moreover, as we will see more broadly in section 6, instead of focussing on neural correlates of theoretical constructs (e.g. "loss aversion", "framing effect") this research embeds these notions in a network of patterns of behaviour and brain activity in which is sought not only a refutation or confirmation of a particular theory, but an *explanation* of it.

## 5. Neuroscience may drive the development of new economic theories and models

Social neuroscience bears on the neural mechanisms underlying cognitive processes such as "mentalizing", which is the ability to represent one's and others' beliefs, intentions and desires and which makes socialization and social exchange possible. Since these cognitive mechanisms are the bread and butter of our behaviour in games, crossover between studies of neurophysiological function and of complex social patterns would appear to be immediately relevant (see cautionary words in Blakemore *et al.* [2004]). Certain foundational questions have yet to be worked out. What does it mean for the social scientist to know that a certain brain area is implicated in a certain task that bears a social connotation? Moreover, even though specific neurological abilities are involved in cognitive capacities such as anticipating others' reactions, what foundational benefit may game-theorists obtain from this fact? These questions are addressed by authors who promote the crossover between social neuroscience and behavioural game theory. Singer and Fehr [2005] justify this endeavour by noting that "economists still know little about what enables people to put themselves in others' shoes and how this ability interacts with their own preferences and beliefs; (and) social neuroscience provides insight into the neural mechanism underlying our capacity to represent others' intentions, beliefs and desires". Neural mechanisms, such as that of mirror-neurons (Rizzolatti [2004]) – which are said to enable us to encode others' actions in the same manner that we encode our own – might provide novel foundations for game theory. Yet it is still unclear whether the impact of this sort of data on economics will be foundational, conceptual or merely illustrative.

Singer and Fehr, whose explicit aim is to assess the implications of such brain data for economics, find reasons for introducing these findings into game theory. In a Prisoner's Dilemma sequence, wherein the anticipation of the other's behaviour is crucial to determine one's own moves, it is important to discriminate between one's own beliefs about the other player's profile (whether she is a reciprocator or not) and one's own understanding

of her current feelings towards oneself (whether she wishes to reciprocate or not independently of her profile). This last ability specifically requires the ability to empathize. Seemingly paradoxically, given that the normal solution of the Prisoner's Dilemma predicts defection, Singer's research shows that people who have higher empathy are better predictors of others' behaviour, hence tend to be better players in a Prisoner's Dilemma (Singer [2004]) sequence. The crucial distinction here is between an abstract solution in game-theory, and players' actual aptitude to behave successfully in real-life games.

In personal interactions we employ rules that have emerged in our cultural and moral contexts. Fairness, reciprocity and mutual consent are, in part, neuronally-encoded norms that define what is sometimes called the social brain. Two lines of study in neuroeconomics can be pursued here: to observe the social brain at work in experimental economics settings, and to capture subjective reactions when prevailing social and moral norms are violated by individuals. McCabe and his colleagues (McCabe [2003]) have discovered that players cooperate far more than predicted by traditional game-theory in two-person anonymous single play interactions, indicating that trust is not only the result of learning in repeated interactions but is a primary behaviour we exemplify in one-move cooperation games. Second-movers who cooperate in these environments do so by reading intentions in moves, and naturally forgo defection incentives to reward trusting offers to cooperate. Moreover, trusting first-movers anticipate such behaviour as likely. Thus, one may posit that "mind-reading" neural mechanisms are connected to compliance with reciprocity norms.

The second line of study of the social brain focuses on when social and moral norms are violated. Transgressions of these norms by others and by oneself give rise to specific and neural activities that resemble each other in part. De Quervain and his colleagues have investigated the neural basis of so-called altruistic punishment of defectors, which they count as a key element in explaining cooperation among individuals who are not genetically related (de Quervain [2004]). They found that when a subject feels the urge to punish a defector and can do so effectively, both the caudate nucleus and the thalamus, which are part of the reward system in the brain, display relatively high activity. In other words, punishing the guilty is rewarding for humans. Conversely Berthoz and her colleagues were able to show that an anticipation-of-punishment system was engaged when subjects violated moral norms (Berthoz [2006]). They uncovered significant bilateral amygdala activation, which they interpret as being related to one's anticipation of possible punishment as the consequence of one's immoral behaviour. This hypothesis meshes well with De Quervain's study on punishment of defectors in trust games, suggesting a correspondence between cooperative and defective behaviours on the one hand, and on the other, reward and punishment systems in the brain.

According to Fehr and his colleagues, the emerging knowledge concerning the brain mechanisms that underlie altruistic behaviour and punishment of defection give grounds for reconsidering the foundations of game-theory. Behavioural game-theory should account for the subjects' deviation from a self-centred strategy. One apparent limitation of this line of research,

though, is that economics is not only concerned with personal exchange structures but also with impersonal markets and institutions – contexts in which considerations like altruism and empathy are obviously less relevant. However, even given the impersonal nature of these contexts, it might prove fruitful to consider the issues of trust and trustworthiness.

## 6. Brain data may enter a functional network in accounting for human behaviour

Behavioural deviations from the unitary utility-maximizing model of decision-making suggest that different systems compete for overt behaviour, only a subset of which can be interpreted in terms of an explicit and effortful attempt at maximizing utility. Another part may be understood as the interaction between automatic emotional processes and executive control. Some authors try to explain the use of heuristics in reasoning and decision-making in terms of the optimal use of brain resources, and associate them with emotional processes (Muramatsu et Hanoch [2005]). But even if heuristics – fast and frugal procedures – may rely, to some extent, on brain-processes associated with emotions, they don't necessarily present any sort of psychological unity, let alone of functional convergence at a neuronal level. Indeed, various emotions themselves are associated with different encapsulated brain structures whose interaction with executive function in the brain is a good predictor of certain behavioural patterns which seem particularly relevant to economists. (e.g. impulsivity, cognitive biases, etc.) The view that several specialized neural sub-systems either cooperate or compete for the control of overt behaviour certainly contributes to undermine classical unified views of decision-making. What emerges is a multiple-systems view of neural structures in the brain, each of which can be studied in its own right as encoding relevant parameters in economically relevant behavioural patterns.

How could a mix of encapsulated neural systems govern an individual's rationality? There appears to be a profound gap between neural observations and rationality attributions. That said, many philosophers and psychologists from the 1980s on (Cohen [1981], Stich [1986]) have discussed the question rationality-attribution from the standpoint of behavioural evidence. In analyzing the principle of charity, they contend that if we were able to reconstruct local principles that subjects apply in their cognitive endeavours we should prefer to make our rationality-attributions according to those local heuristics. These authors postulate that norms of rationality – however local – are defined by what we can call their cognitive anchoring and adequacy. If people are able to maintain representations of the rules they follow in cognitive tasks such as reasoning and decision-making, they cannot be described as being essentially irrational. Conversely the set of norms by which we judge individual behaviour should be cognitive adequacy in the

sense that those norms may in principle be the object of individuals' explicit representations.

Behavioural economists have borrowed a dual process theory of cognition from experimental psychology, which holds that two kinds of cognitive operations – intuitive fast rigid processes on the one side and analytic controlled slow flexible processes on the other side – compete for the control of overt behaviour (Sloman [1996], Evans [2003], Kahneman [2003]). But confusion arises when the fast/slow and the emotion/cognition dualities are conflated. The contribution of automatic and sometimes unconscious processes and the contribution of emotions in our definition of rational agents must be distinguished. Automatic processes and heuristics sometimes lead to suboptimal behavioural responses but they can be overridden by controlled processes which make explicit to the agent which rules should be followed in order to improve her performance. We should not judge the agent as lacking cognitive competence so long as she displays the ability to inhibit erroneous intuitive or impulsive behaviours and is capable of optimizing her behaviour. An understanding of the neural underpinnings of the mechanisms of cognitive inhibition and post-error behavioural adjustments would afford a "broad" view of rationality.

A popular hypothesis bears on the contribution of emotions to reasoning and decision-making. Damasio's Somatic Marker Hypothesis (SMH) (Damasio [1994]) proposes that emotions-based signals in the brain, which arise from the body, are integrated in brain regions dedicated to higher cognitive functions, in particular the ventromedial prefrontal cortex, in order to regulate decision-making in complex individual or social situations. Those signals are emotional somatic markers which influence processes of reasoning and decision-making in a conscious or in an unconscious way. Emotions interfere with higher cognitive processes and introduce biases in the selection of appetitive or aversive modes of behaviour. Thus, they shape our attitudes, in particular in situations that present features such as risk, temptation, and social exchange, which, again, are of primary interest to neuro-economists.

Most of the empirical evidence in support of SMH has been channelled through an experimental paradigm known as the Iowa Gambling Task, which creates a situation in which there is a trade-off between immediate and postponed rewards. Two decks of cards are associated with rewards; the first deck leads to immediate rewards, but long-term losses; the second deck leads to lesser immediate rewards, but long-term gains. Damasio has extensively documented the fact that patients who suffer from lesions in the ventromedial prefrontal cortex will systematically pick cards in the first deck even when they have already experienced repeated losses. Although several rival hypotheses could explain this pattern of behaviour, Damasio concludes that what was specifically affected by the non-tunnelling of emotive somatic markers through the damaged brain area was a sort of "memory of the future" – namely the capacity, when facing a new but similar situation, to remember our anticipation of the right thing to do after a series of errors. Several criticisms have been aired against the SMH. Maia and McClelland [2004] discriminate between the conscious knowledge that subjects may obtain in each situation they face (one deck is more favourable than the

other) and their ability to reverse their behaviour. They argue that SMH does not adequately answer the question of why patients with a ventromedial lesion can eventually report the difference between the decks but still persevere in their suboptimal behaviour. SMH would argue that that cognitive processes unaccompanied by somatic markers fail to yield optimal behavioural regulation, conscious knowledge of a situation being behaviourally inert. An interesting research topic for neuroeconomics, then, would be to understand how emotional and cognitive reactions to the infringement of optimal behaviour lead to behavioural adaptation in economically relevant situations: is cognitive adequacy alone sufficient for post-error behavioural adjustment or is emotional marking a necessary condition? Neural evidence can play a discriminatory role in addressing this issue on the sources of rationality.

## 7. Conclusion : the unity of behavioural sciences

Gul and Pesendorfer's critique is founded on what they characterize as the fundamental separation between economics and neuroscience, in terms of the scope and the incommensurability of the theoretical constructs employed by each field. "Neuroeconomics imports the questions and abstractions of psychology and re-interprets economic models as if their purpose were to address those questions. The standard economic model of choice is treated as a model of the brain and found to be inadequate. Either economics is treated as amateur brain science and rejected as such or brain evidence is treated as economic evidence to reject economic models". On the contrary, we have sought to illustrate different ways in which neural data and economic concepts or theories could relate quite naturally to one another.

To fully respond to Gul and Pesendorfer's scepticism, however, it is not sufficient to merely exhibit such "directions of instruction" between neuroscience and economics. One would need, more fundamentally, an a priori principle to think that possibly overlapping constructs and models between the two fields may be epistemologically coherent. In other words one would need what Gintis [2007] labelled an integrative framework for the behavioural sciences. The notion that an organism adapts to an environment in order to increase its pleasure, plays a foundational role in this framework: "the fitness of an organism depends on how effectively it makes choices in an uncertain and varying environment." For simpler organisms the environment is primitive and information is distributed in a decentralized way onto sensory terminals. In the case of more sophisticated organisms, the brain is understood as having evolved in a complex manner so as to become a decision centre: "The distinguishing characteristic of the human brain lies in its power as a decision-making mechanism". In this framework, neuroeconomics, the study of brain mechanisms that mediate choice in a variety of

complex environments, may play a pivotal role in the integration of diverse behavioural approaches.

## References

- BECHARA A., TRANEL D., DAMASIO H., DAMASIO A. R. [1996], Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, 6(2), p. 215-225.
- BERTHOZ S. *et al.* [2006], Affective response to one's own moral violations. *NeuroImage*, (in press).
- BLAKEMORE S.-J., WINSTON J., FRITH U. [2004], Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Science*, 8, p. 216-222.
- BRAEUTIGAM S. [2005], Neuroeconomics: From neural systems to economic behaviour. *Brain Research Bulletin*, 67(5), p. 355-360.
- BROCAS I., CARILLO J. [2006], "The brain as a hierarchical organization", CEPR D.P. 5168, working paper.
- BUSH G., LUU P., POSNER M. I. [2000], Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Science*, 4, p. 215-222.
- CAMERER C., LOEWENSTEIN G., PRELEC D. [2005], Neuroeconomics: How neuroscience can inform neuroeconomics. *Journal of Economic Literature*, 43, p. 9-64.
- COHEN L. J. [1981], Can Human Irrationality Be Experimentally Demonstrated? *Behavioral and Brain Sciences*, 4, p. 317-331.
- DAMASIO A. [1994], *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Avon Books.
- DAYAN P., YU A. [2003], Uncertainty and learning. *IETE Journal of Research*, 49, p. 171-182.
- DE MARTINO B. *et al.* [2006], Frames, biases, and rational decision-making in the human brain. *Science* 313, p. 684-687.
- DE QUERVAIN D., FISCHBACHER U., TREYER V., SCHELLHAMMER M., SCHNYDER U., BUCK A., FEHR E. [2004], The neural basis of altruistic punishment. *Science*, 305, p. 1254-1258.
- DORIS M. C., GLIMCHER P. W. [2004], Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, 44, p. 365-378.
- ELLSBERG D. [1961], Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75, p. 643-699.
- EVANS J. St. B. T. [2003], In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 10, p. 454-459.
- FREDERICK S. [2005], Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. 19(4), p. 24-42.
- GIGERENZER G., SELTEN R. (Eds). [2001], *Bounded rationality: The Adaptive toolbox*. Cambridge, The MIT Press.
- GILBOA I., SCHMEIDLER D. [1989], Maximum expected utility with non unique prior, *Journal of Mathematical Economics*, 18, p. 141-153.
- GINTIS H. [2007], A Framework for the unification of the behavioral science, *Brain and Behavioral Science*.



- GUL F., PESENDORFER W. [2005], The case for mindless economics. Working paper.
- HAYEK F. [1952], *The Sensory order: An Inquiry into the foundations of theoretical psychology*. Chicago, IL: University of Chicago Press.
- HUETTEL S., STOWE C., GORDON E., WARNER B., PLATT M. [2006], Neural signatures of economic preferences for risk and ambiguity. *Neuron*, 49(5), p. 765-75.
- KAHNEMAN D. [2003], Maps of bounded rationality: Psychology for Behavioral Economics. *American Economic Review*, 93(5), p. 1449-1475.
- KAHNEMAN D., FREDERICK S. [2006], *Trends in Cognitive Sciences*, 11(2), p. 45-46.
- LO A., REPIN D., STEENBERGER B. [2005], Fear and greed on financial markets: A clinical study of day-traders. *Journal of Investment Consulting*, 7, p. 21-44.
- MAIA T., MCCLELLAND J. [2004], A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa Gambling Task. *Proceeding of the National Academy of Sciences*, (101), p. 16075-16080.
- MCCABE K., RIGDON M., SMITH V. [2003], Positive Reciprocity and Intentions in Trust Games. *Journal of Economic Behavior and Organization*, 52, p. 267-275.
- MURAMATSU R., HANOCH Y. S., COHEN J. D. [2006], Neuroeconomics: cross-currents in research on decision-making. *Trends in Cognitive Sciences*, 10(3), p. 108-116.
- O'DOHERTY J. P. [2004], Reward representations and reward-related learning in the human brain: insights from neuro-imaging. *Current opinions in neurobiology*, 14, p. 769-776.
- PADOA-SCHIOPPIA C., ASSAD J. [2006], Neurons in orbitofrontal cortex encode economic value. *Nature*, 441, p. 223-226.
- PATON J. J., BELOVA M. A., MORRISON S. E., SALZMAN C. D. [2006], The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, 439, p. 865-870.
- PLATT M. L., GLIMCHER P. W. [1999], Neural correlates of decision variables in parietal cortex. *Nature*, 400, p. 233-238.
- REDISH D. [2004], Addiction as a computational process gone awry. *Science*, 306, p. 1944-1947
- RIZZOLATTI C., CRAIGHERO L. [2004], The mirror-neuron system. *Annual Review of Neuroscience*, 27, p. 169-192.
- ROESCH M. R., OLSON C. R. [2004], Neuronal activity related to reward value and motivation in primate frontal cortex. *Science*, 304, p. 307-310.
- RUSTICHINI A. [2005], Neuroeconomics: Present and future. *Games and Economic Behavior*, 52, p. 201-212.
- RUSTICHINI A. [2005], Emotion and reason in making decisions. *Science*, 310(9), p. 1624-5.
- SANFEY A., LOEWENSTEIN G., McCLURE S., COHEN J. [2006], Neuroeconomics: cross-currents in research on decision-making, *Trends in Cognitive Sciences*, 30, p. 108-117.
- SIMON H. [1976], From "Bounded" to "Procedural" Rationality. In MASTEN S. et WILLIAMSON O. (Eds), *Transaction Cost* 2005). Emotions as a mechanism for boundedly rational agents: The fast and frugal way. *Journal of Economic Psychology*, 26(2), p. 201-221.
- SINGER T., FEHR E. [2005], The neuroeconomics of mind reading and empathy, *American Economic Review*, 95(2), p. 340-345.

SINGER T., SEYMOUR B., O'DOHERTY J., KAUBE H., DOLAN J. D., FRITH C. [2004], Empathy for pain involves the affective but not sensory component of pain. *Science*, 303, p. 1157-1162.

SLOMAN S. [1996], The empirical case for two systems of reasoning. *Psychological bulletin*, 119, p. 3-22.

SMITH V. [2003], Constructivist and ecological rationality in economics. *American Economic Review*, 93(3), p. 465-508.

STICH S. [1984], Could Man Be an Irrational Animal? *Synthese* 64, p. 115-35.

STUPHORN V. [2006], Neuroeconomics: Cardinal utility in the orbitofrontal cortex? *Current Biology*, 16, p. R591-R593.