
Journal of Economic Methodology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjec20

# The philosopher in the scanner (or: how can neuroscience contribute to social philosophy?)

Francesco Guala [a] & Tim Hodgson [b]
[a] University of Milan, Italy
[b] University of Exeter, UK

Version of record first published: 10 Jun 2010.

To cite this article: Francesco Guala & Tim Hodgson (2010): The philosopher in the scanner (or: how can neuroscience contribute to social philosophy?), Journal of Economic Methodology, 17:2, 147-157

To link to this article: http://dx.doi.org/10.1080/13501781003756527

*Journal of Economic Methodology,*
Vol. 17, No. 2, June 2010, 147–157



# The philosopher in the scanner (or: how can neuroscience contribute to social philosophy?)

Francesco Guala[a]* and Tim Hodgson[b]

[a]*University of Milan, Italy;* [b]*University of Exeter, UK*

Analytical philosophy has been challenged by experimental approaches that make use of, among other things, cognitive science methods. In this paper we illustrate the benefits of merging philosophy with neuroscience, using an example of research in the foundations of social science. We argue that designing novel experiments to answer specific philosophical questions has several advantages compared to relying passively on neuroscientists' data. In this particular case, the data redirect attention towards topics – such as inductive reasoning – that are relatively overlooked by mainstream social neuroscience.

**Keywords:** coordination; induction; neuroscience

**JEL Codes:** B40; C73; C90



## 1 Naturalism and experimental philosophy

Naturalism is one of the great research programmes of contemporary philosophy. Naturalists typically follow Quine's (1952) rejection of the analytical–synthetic distinction that in the heyday of logical positivism grounded the separation of philosophical from scientific research. They claim not only that such separation is conceptually untenable, but that the answer to many philosophical questions is more likely to be found using scientific methods of inquiry, than the *a priori* approach of standard analytical philosophy.

In spite of such declarations, however, surprisingly few naturalists since the 1950s have put their research where their mouth is. Quine himself was primarily a logician, and references to science in his writings are limited to general remarks concerning theoretical physics and biology. Most philosophers of the next generation, while becoming more fluent and sophisticated in the interpretation of science, have also been reluctant to leave the armchair and head towards the experimental laboratory. As Jesse Prinz has nicely put it, as things stand 'empirical philosophers are theoreticians' (2008, p. 205).

There are signs, however, that the situation is beginning to change. Over the past few years many philosophers have been engaged in empirical projects that make use of scientific data to answer philosophical questions. This new style of 'empirical' or 'experimental' philosophy is on the rise, and its emergence has provoked annoyed reactions within the discipline.[1] In this paper we shall defend experimental philosophy by example, and briefly illustrate how neuroscience can contribute to the philosophy of social science. We shall argue that there are benefits for philosophers not only to *rely* on scientific results, but also to *do* research on the neural basis of social behaviour. By engaging

*Corresponding author. Email: francesco.guala@unimi.it

ISSN 1350-178X print/ISSN 1469-9427 online

DOI: 10.1080/13501781003756527
http://www.informaworld.com

directly with neuroscience, philosophers will disengage in part from scientists' research agenda and contribute to re-direct it towards topics that are important for their discipline.

Neuroeconomics is a good model in this respect: one important innovation in this area has been the introduction of experimental designs and theoretical models that have increased the variety of tools used by contemporary neuroscientists.[2] At the same time, economists have pushed neuroscientists to face questions that fall in their (i.e. economists') core domain of interest. This sort of collaboration, we argue, should be a model for the philosophy of social science too. By relying passively on the results of science, philosophers will miss the opportunity to influence scientists' agenda and fail to answer questions that have been traditionally central in their discipline.

The example that we shall explore in this paper concerns the role of inductive reasoning in the emergence and resilience of social conventions. While social neuroscientific research currently emphasizes the importance of mind-reading and emotion for social behaviour,[3] less attention has been paid to inductive inference and rule-following. In the next section we outline a model of social interaction introduced three centuries ago by David Hume, one of the fathers of contemporary social philosophy. We describe how this model can be turned into an experimental protocol and tested in the neuroscience laboratory (sections 2 and 3). Section 4 is devoted to two prominent theories of decision in current neuroscience that prima facie could be used to explain behaviour in Hume's game. Section 5 explains why these theories fail to account for the data, and proposes a different explanation based on inductive propensities and attitudes towards risk. Section 6 concludes with general remarks and methodological discussion.

## 2 Hume's game

David Hume is one of the great precursors of the game-theoretic analysis of social institutions. In a famous paragraph of the *Treatise of Human Nature*, he compares social cooperation with the action of two rowers in a boat:

> Two men, who pull the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other. Nor is the rule concerning the stability of possession the less deriv'd from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it. On the contrary, this experience assures us still more, that the sense of interest has become common to all our fellows, and gives us a confidence of the future regularity of their conduct: And 'tis only on the expectation of this, that our moderation and abstinence are founded. (Hume 1740, Part II, Section II)

In modern terminology, Hume models society as a sequence of coordination games, and highlights the role of habits and customs to select among the many possible equilibria of the game of life (Lewis 1969; Sugden 1986). Identifying Hume as the philosopher of coordination, however, does not give him full credit for his sophistication. Hume was aware of the problem posed by free riding in complex societies where interactions with strangers are frequent and anonymous. In fact, his view of social life may be more accurately characterized as a sequence of coordination games occasionally interrupted by 'temptation games' that offer the opportunity of gaining at the expense of other players. As he writes in the *Essays*,

> All men are sensible of the necessity of justice to maintain peace and order, and all men are sensible of the necessity of peace and order for the maintenance of society. Yet [ . . . ] such is the frailty or perverseness of our nature! It is impossible to keep men faithfully and unerringly in the paths of justice. Some extraordinary circumstances may happen, in which a man finds

| | Left | Right |
|---|---|---|
| Left | 1, 1 | 0, 0 |
| Right | 0, 0 | 1, 1 |

Coordination round (C)

| | Left | Right |
|---|---|---|
| Left | 1, 1 | 2, 0 |
| Right | 2, 0 | 1, 1 |

Temptation round (T)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C→ | C→ | C→ | C→ | C→ | C→ | C→ | C→ | T→ | C→ | C→ | C→ | T→ | C→ | C→ | C→ | T→ | C→ | C→ | C→ | T |

Figure 1. Sequence of coordination and temptation rounds in Hume's game.

> his interests to be more promoted by fraud or rapine, than hurt by the breach which his injustice makes in the social union. (Hume 1777, Part I, Essay V)[4]

Consider the game in Figure 1. Over a sequence of coordination rounds, two players have the opportunity to create a convention (Left or Right). Over time, however, the sequence is interrupted by 'Temptation rounds' offering the possibility of deviation to one player (Row) at the expense of the other (Column). Let us suppose that only the row player (the 'Potential Deviant') is aware of the imminent change of payoffs before a Temptation round. The other player (Column) is taken by surprise, and offers Row a free-riding opportunity.

Although the Potential Deviant faces an incentive to breach the convention at each Temptation round, she is also potentially vulnerable to sanctions. Column may withdraw cooperation following a breach of convention: a costly 'trigger strategy', in game-theoretic jargon, which may deter deviations in the early rounds of the game. At the very last round, however, even this threat becomes ineffective: the game will end and the players will never meet again. At this point, the Potential Deviant has no selfish incentive to stick to the convention that has evolved thus far.

When Hume's game is played in the laboratory, compliance with the convention turns out to be remarkably robust. Starting with roughly 60% of compliance, there is a tendency to breach the convention more frequently as the end of the game is approaching. But almost 40% of Potential Deviants cooperate even in the last round, when the 'shadow of the future' has completely disappeared.

So conventions tend to acquire normative power in Hume's game – they bias players' choices in such a way as to increase conformity, in spite of individualistic incentives to deviate. But why do experimental subjects give up material gains and conform to a rule of conduct that has evolved in the early part of the game? What proximate mechanisms make conventions robust to disturbances and even changes in the structure of payoffs?

## 3 Learning in the brain

To answer these questions, we designed an experiment that implemented the essential features of Hume's game. Two subjects communicated by means of a computer network, and played a repeated game like the one in Figure 1. Right at the start one of the subjects was selected randomly and invited to lie in the fMRI scanner. The screen of her PC was reflected via a mirror, and her decisions were transmitted using two buttons on a remote control. The subject's brain was scanned repeatedly before she made her decisions (during the 'decision period'), and when she received feedback about the other player's moves (the 'outcome period').

Subjects on average reached coordination in two rounds. Once a convention was established, most of them kept choosing it unproblematically, and made money by simply sticking to the rule. They applied the inductive principle that – other things being equal – the other player will continue to behave as she has done until now, and she expects us to do the same.

Magnetic resonance imaging gives us the opportunity of seeing inductive reasoning at work in the human brain. A plausible candidate for the role of inductive mechanism is the *reward circuitry* implicated in Pavlovian and instrumental learning.[5] Learning about rewards involves crucially the so-called *dopaminergic system*. Dopamine is a neurotransmitter – one of over 100 molecules that regulate the transmission of signals between neurons, by increasing or decreasing the probability that a receptor will 'fire' upon receiving a signal from a transmitting neuron.

The dopaminergic pathway is implicated in processes of *instrumental learning*, whereby a behaviour that leads to a rewarding state is reinforced by mere repetition (e.g. Schultz 2000). The neural basis of instrumental learning has been studied extensively both in animals and human subjects. An important feature of instrumental learning is the formation of expectations of reward before executing an action that is believed to generate a valued state of affairs. These expectations of reward influence action, through complex interactions between limbic regions of the brain (especially the corpus striatum, the thalamus and the amygdala) and parts of the frontal cortex (especially the dorsolateral prefrontal and the orbitofrontal cortex).

The analysis of BOLD signals during decision periods shows that conformity and deviance are both associated with the anticipation of reward in Hume's game: conformists display more activity in the ventral region of the striatum than deviants during decision periods. The latter (who breach the convention) in contrast display higher activity in the

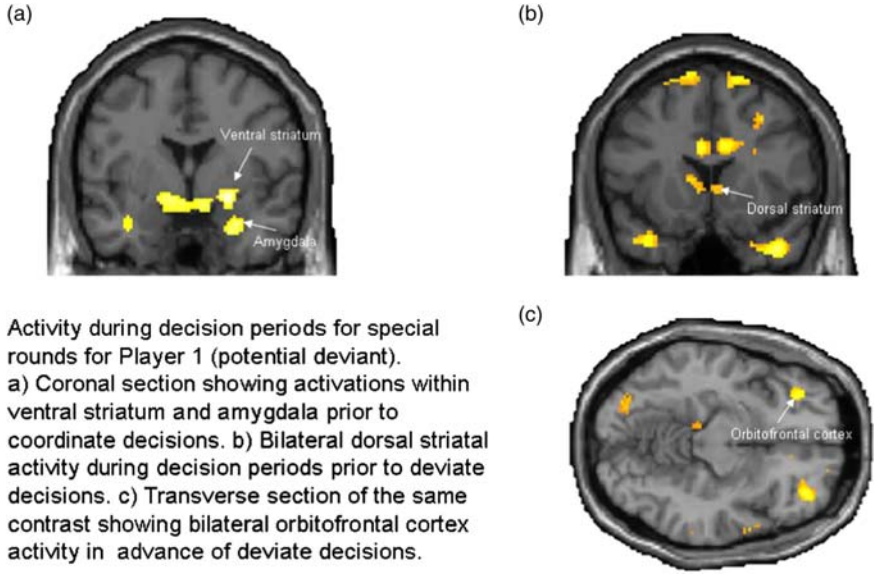


Figure 2. Differential activations during decision periods, temptation rounds. The figures show BOLD contrasts after the variance attributable to other factors (such as the number of preceding coordinating rounds, whether subjects had met the other participant or not, etc.) has been accounted for.

dorsal area. These differential activations are depicted in Figure 2. Other activations of interest in conformists' brains take place in the dorsolateral prefrontal cortex (DLPFC), an area that mediates between emotive impulse and the pursuit of long-term, abstract, or impersonal rewards (Greene, Nystrom, Engell, Darley, and Cohen 2004; Hare, Camerer, and Rangel 2009), and the amygdala, a limbic region of the brain integrated in the dopaminergic pathway (Phelps 2006; Seymour and Dolan 2008). Deviants, in contrast, exhibit increased activation of the orbitofrontal cortex (OFC), a region that is still poorly understood but is probably implicated in learning and decision.

The amygdala is involved in the processing of positive and negative emotions, in particular fear, and is known to interact with frontal areas of the brain in social cognition. It is well established for example that the amygdala plays a role in the recognition of trustworthy faces (Adolphs, Tranel, and Damasio 1998). Clinical studies suggest that it may have an alarm function, screening between safe and potentially threatening social situations. Post-mortem and fMRI studies, for example, show that subjects with 'Williams Syndrome' suffer from anomalies at the level of amygdala-frontal interaction that impair their social inhibition mechanisms (Meyer-Lindenberg, Mervis, and Berman 2006). As a consequence, they manifest excessive friendliness towards strangers and reduced capacity to identify threatening social stimuli (such as angry facial expressions).

High amygdala activity is also associated with low transfers in economic trust games,[6] and its suppression using synthetic oxytocin provokes an increase of trusting behaviour (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, and Fehr 2008). Interestingly, oxytocin does not change subjects' beliefs in the probability of a positive outcome, nor does it simply boost their desire to benefit others. Damping amygdala activity, rather, seems to reduce subjects' anxieties regarding social betrayal. For all these reasons, the amygdala has been identified as a crucial link in the mechanism connecting emotions with social decision making in humans.

## 4 Decision and emotion

Emotions are a hot topic in cognitive science and behavioural economics. Their exact role in decision-making, however, is still highly controversial. In such circumstances just picking some theory 'off-the-shelf' for philosophical consumption may be tricky. To give an idea, we shall here discuss two accounts – the somatic marker hypothesis and altruistic punishment theory – that have gained wide currency in contemporary social neuroscience.[7]

Antonio Damasio's 'somatic marker hypothesis' (SMH) is probably the best known theory of emotional decision-making among philosophers, neuroscientists, and the larger public. SMH defines emotions as changes in bodily and brain states triggered by the occurrence or the recollection of specific objects and events. These changes help rational decision-making by short-cutting calculative processes that would otherwise be long and cognitively expensive. Emotional states are associated with events or objects by means of 'somatic markers' that quickly and subliminally code the valence of decision outcomes (for example, 'how it feels like' to gain or lose a sum of money). When such somatic markers are unavailable, decision-making becomes long, expensive, and often just 'bad' – in the sense that subjects are unable to do what is in their best interests.

Through the study of brain-damaged patients, Damasio and his collaborators have identified the amygdala and the prefrontal cortex as two key brain areas where the triggering and processing of emotions takes place (cf. e.g. Damasio 1994; Bechara and Damasio 2005). According to SMH, the amygdala receives information about 'primary

inducers' (the objects or events that cause emotions) and sends out signals that activate somatic states. Like the corpus striatum, the amygdala learns associations between events, and over time shifts its reaction backwards to anticipate the occurrence of emotionally charged situations. The orbitofrontal cortex (OFC), according to Damasio, works as a repository for representations of emotional signals sent from the amygdala and other limbic regions of the brain.[8] It is important in particular for the production of 'secondary inducers' – for example the memory of a scary event or the image of a counterfactual dangerous situation – that generate emotional responses similar to those associated with real objects and events.

This division of labour between amygdala and OFC is important for Damasio's interpretation of the Iowa Gambling Task, a decision-making experiment that constitutes the main source of evidence for the SMH (cf. Damasio 1994; Bechara, Tranel, Damasio, and Damasio 1996). In this task subjects choose repeatedly between four options that deliver monetary gains or losses with different (but initially unknown) frequencies. Two options (A and B) deliver large gains occasionally interrupted by large losses, but in the long run produce a net loss for the subject. The other two options (C and D) deliver instead small wins and small losses which in the long run result in a net gain for the subject. Patients with amygdala damage and patients with orbitofrontal damage after initial sampling fail to shift their behaviour progressively towards C and D (as normal subjects instead do). However, they seem to understand theoretically that A and B are 'bad' choices. This paradox is explained, according to Damasio, by the absence of emotional signals (somatic markers) that subliminally identify A and B as options to be avoided.

The Iowa Gambling Task is only superficially simple, and on deeper inspection it is not obvious that its results support Damasio's SMH (see e.g. Dunn, Dalglaish, and Lawrence 2006; Colombetti 2008). Setting these worries aside, however, we shall focus here on the relation between Damasio's theory and behaviour in Hume's game. Following SMH, it is tempting to interpret amygdala activations in Hume's game as evidence of an emotional component in conformist behaviour. Conformist players stick to the convention because they are afraid of the losses associated with deviance.

But what sort of losses? Deviating from a convention in a Temptation round brings higher immediate rewards, in monetary terms. Of course conformists may be initially worried about the *future* losses that may result from lack of coordination. But this sort of concern should become less relevant as the game proceeds, and should not arise at all in the last and final round, when the shadow of the future has disappeared. The increased activation of OFC in deviants, moreover, is scarcely compatible with the idea of the OFC as a store of representations of emotional events.

Another possibility is that the brain may subconsciously associate breaches of convention with sanctions (for example reproaches, or ostracism) that in real life are typically inflicted on social deviants. This thesis (the internalization of norms hypothesis) has been recently defended by supporters of so-called 'strong reciprocity' theory (e.g. Fehr and Gachter 2002; Gintis 2003) and prima facie would help explain conformism in Hume's game too.

The punishment hypothesis, however, is also problematic. Again, there is the problem of accounting for OFC activations. But even more seriously, a recent study by Li, Xiao, Houser, and Montague (2009) suggests that OFC-amygdala activity is more prominent when *no* punishment mechanism is in place. Li and colleagues report brain activations in a trust game played with and without sanctions. When punishment is available, typical reward areas such as the parietal cortex are activated in the trustee's brain. This suggests that under the threat of sanctions subjects calculate the pros and cons (in monetary terms)

of reciprocating by returning some of the money they have received. In the absence of punishment threat, in contrast, there is activation of the very same areas associated with conformist behaviour in Hume's game – including amygdala and OFC. If pro-social behaviour with punishment exploits altogether different neural circuits, conformism in Hume's game is unlikely to result from the internalization of punishment threats.

## 5 Caution in the brain

If the amygdala is involved in detecting social danger, we must conceive the latter in broader terms than the punishment theory does. The amygdala is almost certainly involved in the evaluation of probabilistic prospects. As in standard decision theory, it is important to distinguish between *risk* and *uncertainty* here: neural studies suggest that the evaluation of risky prospects (where *objective* probabilities are known) takes place in more evolved, 'higher' regions of the brain such as the medial prefrontal cortex (Knutson, Taylor, Kaufman, Peterson, and Glover 2005). In contrast, the evaluation of ambiguous and uncertain prospects (where objective chances are unknown and actors must rely on subjective estimates) relies on emotional signals coming from limbic regions of the brain. Hsu, Bhatt, Adolphs, Tranel, and Camerer (2005) for example report significant increases in amygdala-OFC activation in tasks with ambiguous and uncertain prospects (where the number of winning cards in a deck is unknown) compared with risky tasks (where the number of cards is known).[9]

Various studies suggest that under uncertainty the OFC and the amygdala work together to learn associations between stimulus and reward. Hampton, Adolphs, Tyszka, and O'Doherty (2007) report fMRI scanning of two patients with amygdala lesions, while engaged in a switch/stay task that is in many ways similar to our experiment. Subjects had to choose between two buttons, each one delivering a monetary gain or loss with a given (but unknown) probability. The probabilities were not stable throughout the game. On the basis of observed outcomes, the subjects had to learn to switch buttons when the probabilities changed (i.e. when pressing a button became on average more profitable).

Normal subjects displayed greater activity in the OFC during switch trials compared with stay trials. Interestingly, subjects with amygdala damage in the experiment of Hampton and colleagues had anomalous (enhanced) OFC activity compared to normal subjects. Behaviourally the amygdala patients had problems 'staying', and switched buttons too frequently. This suggests that in normal subjects the amygdala damps the OFC impulse to deviate from a rule that has been followed until now. While the OFC seeks new opportunities, the amygdala acts as the 'moderator' or conservative advisor in our brain.[10]

Notice that there is no social learning in these experiments – subjects are reasoning about the properties of natural (non-intentional) systems. So it is significant that amygdala and OFC activations are inversely correlated in Hume's game too. Collectively, these data support the hypothesis that OFC and amygdala play 'opportunistic' and 'moderating' functions across a variety of decision and learning tasks, of social and non-social nature alike. Breaching the convention in Hume's game brings a certain immediate reward, but at the same time introduces an element of uncertainty in the application of a conventional rule. So it is not surprising that the thought of violating a rule triggers a mechanism of ambiguity aversion. Emotions are involved in social decision-making in a very specific way: they bias decision-makers by signalling that they are entering a 'grey' area where the usual rules do not apply.

It is not difficult to account for this alarm system in evolutionary terms. The breakdown of social customs is an extremely important source of uncertainty for homo

sapiens – one of the most important ones, perhaps, in terms of fitness. In simpler organisms the amygdala probably evolved as an automatic 'fight or flee' mechanism. In our social ancestors it became an alarm system devoted to the detection and inhibition of socially ambiguous situations. The human brain then may result from a compromise between two desiderata: maximizing the advantages of Machiavellian reasoning and exploratory behaviour for the exploitation of new opportunities; but also maximizing reliability and predictability for the sake of social coordination. These desiderata often pull in opposite directions, and behaviour in Hume's game probably reflects different ways of coping with this tension. Conformity with social norms and conventions is enhanced by a neural 'braking system' that guarantees a degree of stability in spite of changes in incentives and uncertainties in the payoff structure.

## 6 Concluding remarks

The idea that inductive reasoning is important for social ontology may strike some philosophers as old news. Inductive propensities play a central role in so-called 'dispositional' approaches to rule-following, for example (see Kripke 1982), and Lewis (1969) explicitly recognizes that conventions of coordination rely on a 'brute' human tendency to extrapolate from past behaviour. And yet, little progress has been made until now in articulating the mechanisms that sustain our propensity to follow social rules.[11] By uncovering the specific ways in which induction matters, social neuroscience provides an insight that cannot be obtained by *a priori* speculation.

Current science however does not help either: standard theories of decision-making cannot explain adequately the neural evidence obtained with Hume's game. Damasio's SMH and punishment theory, as we have seen, encourage the interpretation of amygdala activations as anticipations of future losses, but cannot explain the correlation between OFC activity and deviation from the convention observed in our experiment. Our data and the data of previous experiments support an alternative story: the OFC and the amygdala may play the role of 'switch' and 'stay' mechanisms, respectively, that control our response to stimuli in an uncertain environment. Conformism, according to this account, is caused by a natural aversion to uncertainty, rather than by the anticipation of monetary losses or moralistic punishment.

It seems obvious to us that these results have *both* scientific and philosophical interest. Nevertheless, when we present this research in front of a philosophical audience we are often asked why philosophers should do this kind of work. Notice that the objection is not always addressed to the *content* of this research – it is not necessarily a worry about the philosophical relevance of neural evidence or neuroscientific theory.[12] It is rather a worry about *philosophers* doing this research, instead of delegating to scientists and then relying on their discoveries. This is a sensible concern, at a time when specialization has made it increasingly difficult to learn more than one trade, and relying on the work of the experts seems wiser than getting one's hands dirty with data collection and interpretation. Still, this attitude is unsatisfactory: the data collected by scientists are rarely aimed at answering the specific questions philosophers are interested in, and stretching the evidence to cover such questions is often a dangerously speculative exercise.

We have tried to illustrate this point focusing on two accounts which have gained wide currency among neuroeconomists. As we have seen, analysis of Hume's game delivers insights that are partly inconsistent with both SMH and punishment theory. Relying passively on existing research prevents the exploration of phenomena (such as the emergence of social conventions) that have been traditionally of great interest for

philosophers. It also encourages philosophers to borrow theories developed in different contexts, to explain phenomena for which these theories may not be best suited.

Since any account of the emergence of conventions and norms based on current neuroscience is bound to be speculative to a large extent, one is best advised to run experiments that are especially designed for this task. As we have shown the payoffs are considerable. Of course we cannot claim that our account of the resilience of conventions in Hume's game is definitely correct, and we do not claim that the evidence is entirely at odds with current views about the role of emotions in social decision-making. (If it were, in fact, we would be rather worried.) But we do believe that studying new settings offers original insights that cannot be gained by merely scanning the existing scientific literature. Via these interdisciplinary collaborations naturalistic philosophy will become more scientific in the years to come – not just in the usual sense of relying on scientific results but in the genuine sense of making use of scientific methods of investigation to tackle philosophical questions.



**Acknowledgements**

The fMRI studies mentioned in this paper were funded by ESRC/MRC grant RES-000-22-2392. We thank the editors for their suggestions, and take responsibility for all the remaining mistakes.

**Notes**

1. See Knobe and Nichols (2008) for examples, critical assessments, and defences of experimental philosophy.
2. Cf. for example Glimcher (2004).
3. See e.g. Adolphs (2003), Singer and Fehr (2005), Frith (2007).
4. Similar passages can be found in Hume (1740, Part II, Section II).
5. In 'Pavlovian' learning, the experimental subject is conditioned to associate an independent stimulus with the passive receipt of a reward. In 'instrumental' learning, in contrast, the subject must *act* to bring about a state of affairs that may include a reward.
6. In a trust game one player (the 'investor') sends a sum of money to another player (the 'entrepreneur'); the money is doubled by the experimenter, and the 'entrepreneur' has the opportunity of sending back part, none, or all of the money to the 'investor' (see Berg, Dickhaut, and McCabe 1995).
7. Several other theories identify emotions as crucial cogwheels in the mechanics of decision-making, see e.g. Frank (1988), Blackburn (1998), Loewenstein (2000).
8. The terminology used by neuroscientists to classify areas in the frontal part of the brain is not always uniform. Damasio's group use interchangeably the terms 'ventro-medial prefrontal' and 'orbitofrontal' to denote the part of human cortex situated right behind and above the eyes' sockets (roughly corresponding to Brodmann areas 10, 11, and 47). In this paper we shall use the second term (OFC) by convention.
9. Interestingly, they also report increased activation of dorsal striatum under risk than under ambiguity/uncertainty. This is similar to the dorsal striatum activity observed for deviants during decision periods in Hume's game. See also de Martino, Kumaran, Seymour, and Dolan (2006).
10. Rudebeck and Murray (2008) report convergent evidence from lesion studies of monkeys. The hypothesis that the OFC (or part of it) is involved in task reversal is supported by several experiments (e.g. Fellows and Farah 2003, 2005) and is consistent with so-called 'gateway' theories of the prefrontal cortex (Burgess Simons, Dumontheil, and Gilbert 2007).
11. See also Sugden (1998) on this point.
12. This second, more fundamental worry is quite common among analytical philosophers, but discussing it here would be impossible given the limitations of space. We focus instead on the worries of naturalistically minded philosophers who already recognize the relevance of scientific evidence for philosophical research.

## References


Adolphs, R. (2003), 'Cognitive Neuroscience and Human Social Behaviour', *Nature Reviews Neuroscience*, 4, 165–178.

Adolphs, R., Tranel, D., and Damasio, A.R. (1998), 'The Human Amygdala in Social Judgment', *Nature*, 393, 470–474.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008), 'Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans', *Neuron*, 58, 639–650.

Bechara, A., and Damasio, A. (2005), 'The Somatic Marker Hypothesis: A Neural Theory of Economic Decision', *Games and Economic Behavior*, 52, 336–372.

Bechara, A., Tranel, D., Damasio, H., and Damasio, A. (1996), 'Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex', *Cerebral Cortex*, 6, 215–225.

Berg, J., Dickhaut, J., and McCabe, K. (1995), 'Trust, Reciprocity, and Social History', *Games and Economic Behaviour*, 10, 122–142.

Blackburn, S. (1998), *Ruling Passions: A Theory of Practical Reasoning*, Oxford: Oxford University Press.

Burgess, P.W., Simons, J.S., Dumontheil, I., and Gilbert, S.J. (2007), 'The Gateway Hypothesis of Rostral Prefrontal Cortex (Area 10) Function', in *Measuring the Mind: Speed, Control, and Age*, eds. J. Duncan, L. Phillips and P. McLeod, Oxford: Oxford University Press, pp. 217–248.

Colombetti, G. (2008), 'The Somatic Marker Hypothesis, and What the Iowa Gambling Task Does and Does Not Show', *British Journal for the Philosophy of Science*, 59, 51–71.

Damasio, A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Vintage.

De Martino, B., Kumaran, D., Seymour, B., and Dolan, R.J. (2006), 'Frames, Biases, and Rational Decision-Making in the Brain', *Science*, 313, 684–687.

Dunn, B.D., Dalglaish, T., and Lawrence, A.D. (2006), 'The Somatic Marker Hypothesis: A Critical Evaluation', *Neuroscience and Biobehavioral Reviews*, 30, 239–271.

Fehr, E., and Gachter, S. (2002), 'Altruistic Punishment in Humans', *Nature*, 415, 137–140.

Fellows, L.K., and Farah, M.J. (2003), 'Ventromedial Frontal Cortex Mediates Affective Shifting in Humans: Evidence From a Reversal Learning Paradigm', *Brain*, 126, 1830–1837.

——— (2005), 'Different Underlying Impairments in Decision-Making Following Ventromedial and Dorsolateral Frontal Lobe Damage in Humans', *Cerebral Cortex*, 15, 58–63.

Frank, R. (1988), *Passions Within Reason*, New York: Norton.

Frith, C.D. (2007), 'The Social Brain?', *Philosophical Transactions of the Royal Society B*, 362, 671–678.

Gintis, H. (2003), 'The Hitchhiker's Guide to Altruism: Gene-Culture Coevolution and the Internalization of Norms', *Journal of Theoretical Biology*, 220, 407–418.

Glimcher, P.W. (2004), *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*, Cambridge, MA: MIT Press.

Greene, J., Nystrom, L., Engell, A., Darley, J., and Cohen, J. (2004), 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment', *Neuron*, 44, 389–400.

Hampton, A.N., Adolphs, R., Tyszka, M.J., and O'Doherty, J.P. (2007), 'Contributions of the Amygdala to Reward Expectancy and Choice Signals in Human Prefrontal Cortex', *Neuron*, 55, 545–555.

Hare, T.D., Camerer, C.F., and Rangel, A. (2009), 'Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System', *Science*, 324, 646–648.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005), 'Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making', *Science*, 310, 1680–1683.

Hume, D. (1740), *A Treatise of Human Nature*, Oxford: Oxford University Press.

——— (1777), *Essays: Moral, Political and Literary*, Oxford: Oxford University Press.

Knobe, J., and Nichols, S. (eds.) (2008), *Experimental Philosophy*, Oxford: Oxford University Press.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005), 'Distributed Neural Representation of Expected Value', *Journal of Neuroscience*, 25, 4806–4812.

Kripke, S. (1982), *Wittgenstein on Rules and Private Language*, Cambridge, MA: Harvard University Press.

Lewis, D.K. (1969), *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

Li, J., Xiao, E., Houser, D., and Montague, P.R. (2009), 'Neural Responses to Sanction Threats in Two-Party Economic Exchange', *Proceedings of the National Academy of Sciences*, 106, 16835–16840.

Loewenstein, G. (2000), 'Emotions in Economic Theory and Economic Behaviour', *American Economic Review*, 90, 426–432.
Meyer-Lindenberg, A., Mervis, C.B., and Berman, K.F. (2006), 'Neural Mechanisms in Williams Syndrome: A Unique Window to Genetic Influences on Cognition and Behaviour', *Nature*, 7, 380–393.
Phelps, E.A. (2006), 'Emotion and Cognition: Insights From Studies of the Human Amygdala', *Annual Review of Psychology*, 57, 27–53.
Prinz, J. (2008), 'Empirical Philosophy and Experimental Philosophy', in *Experimental Philosophy*, eds. J. Knobe and S. Nichols, Oxford: Oxford University Press, pp. 189–208.
Quine, W.O. (1952), 'Two Dogmas of Empiricism', in *From a Logical Point of View*, ed. W.O. Quine, Cambridge, MA: Harvard University Press, pp. 20–46.
Rudebeck, P.H., and Murray, E.A. (2008), 'Amygdala and Orbitofrontal Cortex Lesions Differentially Influence Choices During Object Reversal Learning', *Journal of Neuroscience*, 28, 8338–8343.
Schultz, W. (2000), 'Multiple Reward Systems in the Brain', *Nature Reviews Neuroscience*, 1, 199–207.
Seymour, B., and Dolan, R. (2008), 'Emotion, Decision Making, and the Amygdala', *Neuron*, 58, 662–671.
Singer, T., and Fehr, E. (2005), 'The Neuroeconomics of Mind Reading and Empathy', *American Economic Review*, 95, 340–345.
Sugden, R. (1986), *The Economics of Rights, Cooperation and Welfare*, Oxford: Blackwell.
——— (1998), 'The Role of Inductive Reasoning in the Evolution of Conventions', *Law and Philosophy*, 17, 377–410.