

Eager for Fairness or for Revenge?

Altruism and Emotion in Neuroeconomics

CHRISTINE CLAVIEN

University of Lausanne

REBEKKA KLEIN

University of Heidelberg

Abstract

In order to understand the human capacity for altruism one requires a proper understanding of how people actually think and feel. This paper addresses the relevance that recent findings in neuroeconomics may have for the philosophical controversy between altruism and egoism, with particular emphasis on the importance of emotion in understanding altruistic motivation. After briefly contextualising and sketching the philosophical controversy, we survey the results of three interesting studies that provide stimulating clues for the debate. We focus our attention particularly on the 2004 study in neuroeconomics by Dominique de Quervain, Urs Fischbacher and colleagues, which contains an argument in favour of psychological egoism. On the basis of an emotional account of decision-making, we show that their analysis of the results – people seek fairness – may be questioned; we propose an alternative interpretation of the data – people seek revenge. Unfortunately, our ‘emotion-directed’ interpretation renders this study far less relevant for the debate over the possibility of psychological altruism than previously expected.

1 Three types of debate over altruism

One can distinguish three types of debate that make use of the notion of altruism, each of them employing different senses of the term.¹ The first type of debate centres on the notion of ‘*biological* altruism’; it is about explaining how behaviours that increase other organisms’ *Darwinian fitness* – measured in terms of expected number of offspring – at a cost to the actor’s own fitness come to be selected in evolution. Since the publication of Darwin’s *Origin of species* (2003), this question had been one of the main challenges to the theory of evolution. Thanks to the efforts of William Hamilton (1964) and others, this difficulty has been resolved.²

¹ We would like to thank Philip Kitcher for helping us frame the question in this way.

² More precisely, an expanded version of ‘kin selection’ theory, which refers to the broader notion of inclusive fitness – and could be labelled ‘discriminatory altruism’ – can explain all cases of biological altruism (see Hamilton 1970, Hamilton 1975, Lehmann, *et al.* 2006).

The second type of debate concerns ‘*behavioural* altruism’; it is about showing that ordinary people often fail to behave in the way predicted by the neoclassical economic model of human agency. Neoclassical economists support the ‘false consciousness’ assumption, according to which “we all like to think of ourselves as nice, caring, altruistic beings, but then when put in the appropriate circumstances – when money is at stake –, we just cannot help but act as the cynical agents postulated by economic models” (Guala 2005: 241). This way of conceiving human beings as rationally self-interested and always aiming to maximize profits has been powerfully challenged recently by work in the field of experimental economics. In a large number of studies, it was shown that ordinary people violate this paradigm; they are ready to contribute to others’ welfare and to the common good at their own expense, even where monetary incentive is at stake (Fehr & Fischbacher 2004a). In the context of this socio-economic debate, fitness measures do not enter the picture; altruism is defined in terms of individual costs and benefits instead of expected number of offspring. More formally, a behaviour is considered altruistic if it benefits other individuals or the common good at some cost for the agent *and* if the cost cannot be compensated in the future. A paradigmatic study case of behavioural altruism is called ‘altruistic punishment’. It consists in a disposition to punish unfair agents even if costly and if the punishment provides neither present nor future material rewards (Fehr & Fischbacher 2005, Fehr & Fischbacher 2003, Fehr & Gächter 2002).

Experimental economists’ research has immensely improved our understanding of human social behaviour. Much of the research is also confirmed by a growing body of data stemming from psychology (Gigerenzer 2008). It has a major impact on economics and more generally on future social sciences.

The biological and behavioural notions of altruism should not be confused with the psychological or philosophical notion of it. The former two focus strictly on a cost-benefit analysis of behavioural outcomes. In contrast, common ‘everyday’ use of the term does not refer to *outcomes* but to the subjective *motivation* of the agents. This leads us to the philosophical debate over altruism.

The traditional debate over the possibility of ‘*psychological* altruism’ centres on the nature of primary motives. The notion of *motive* is a broad category that includes different things, such as desires, intentions, or judgments. There are two sorts of motives: the *primary* motives³ that are usually conceived as the starting points of causal chains that lead towards action – they also are the driving force that last until actions have come about; the *instrumental* motives whose function is to help achieve the aims of the primary motives. Here is an example:

Raymond seeks pleasure [primary motive] → Raymond knows that if he does x, he will obtain pleasure [instrumental practical reasoning] → Raymond desires to do x [instrumental motive in order to achieve pleasure] → Raymond does x

If a primary motive is directed towards the needs and well-being of other individuals, it earns the label ‘altruistic’. If a primary motive is aimed at some personal benefit for oneself – as in Raymond’s example – it is considered ‘self-interested’.

‘Psychological altruism’ is the view according to which *at least some* actions are motivated by altruistic primary motives. On the contrary, ‘psychological egoism’ denies the possibility of primary altruistic motives. According to this latter view, human actions are *always* motivated by the expectation of some personal benefit, usually conceived of in terms of pleasure and avoidance of pain – hedonistic version – or such things as power, resources, or reputation.

Let us now consider some advances made in experimental economics, neuropsychology and neuroeconomics that seem to be relevant for the controversy over psychological altruism.

2 Some interesting data from neuroeconomics

Experimental economists are usually concerned with behavioural altruism – to avoid confusions in what follows we will use the term ‘b-altruism’ when referring to this notion. However, recently, some researchers have also become interested in the psychological aspect of this behaviour. In combining the experimental tools of economics with the neuroscientific tool of brain scanning, they have started to investigate the neurological underpinnings of behavioural altruism. The question we address in this paper is whether these studies are relevant for the philosophical debate over the possibility of psychological altruism. Hence, we will not discuss the most well-known and important aspects of these studies, namely their results relating to b-altruism and their consequences for our understanding of human social behaviour in economic settings. We are merely interested in the possible contribution of these studies to the philosophical debate over psychological altruism.

Two articles are of particular significance for our question: Fehr and Gächter, “Altruistic Punishment in Humans” (2002), and de Quervain, Fischbacher et al., “The Neural Basis of Altruistic Punishment” (2004). A complementary study in social neuroscience is also worth mentioning: Sanfey, Rilling et al., “The Neural Basis of Economic Decision-Making in the Ultima-

³ In the philosophical literature, ‘primary motives’ are usually called ‘ultimate motives’ (Sober & Wilson 1998: 217-22). However, in order to avoid confusion with the notion of “ultimate cause” as described in biology, we prefer

tum Game” (2003). These three articles investigate the motivational systems underlying b-altruistic punishment and there is hope that they provide relevant data for the philosophical debate. We will briefly present their results in this section.

In their 2002 paper, Ernst Fehr and Simon Gächter conducted an experiment with two tasks. The subjects were first asked to play a ‘public good game’. In such a game, all participants are free to contribute to a group project and once the group project is realized, every member of the group can benefit from it, even those who did not contribute. The results showed that subjects usually adopted a b-altruistic punishing behaviour; they sanctioned free-riders – those who benefit from the public good without contributing – even under conditions where it was costly for them and they could not expect any present or future material benefits. In the second part of the experiment, participants were asked to fill in a questionnaire designed to sound out the psychological motivation for b-altruistic punishment. Participants were presented with a hypothetical situation and asked to indicate the intensity of their anger and annoyance towards free-riders and cooperators. The experimenters hypothesized that strong negative emotions such as anger or disgust were the proximate mechanisms responsible for b-altruistic punishment. This could be confirmed by the participants’ responses: 84% of them indicated that they would feel a high intensity of anger if they were cheated by a free-rider. Fehr and Gächter concluded that negative emotions were the proximate mechanisms behind human b-altruistic punishment.

In a complementary study using brain-imaging as a research tool, Alan Sanfey, James Rilling and colleagues (2003) investigated the neural substrates of the cognitive and emotional processes involved in decision-making. They used an ‘ultimatum game’,⁴ where participants have to decide either to accept or reject an offer of money made by another player whose task is to distribute a sum of money between the two of them. If the offer is rejected, no one gets anything. Thus, the offer is called an ‘ultimatum offer’. The subjects of the experiment were asked to play this game while their brains were scanned with the help of functional magnetic resonance imaging methods (fMRI). The results of the study revealed that players confronted with unfair behaviour – unfair distribution of the money proposed by the other player – showed increased activity in the ‘anterior insula’, a brain area associated with negative emotional feelings. Moreover, the strength of the negative emotional response was correlated with the rejection rate of the unfair offer. In this context, rejection of a distribution offer can be interpreted as b-altruistic punishment, because a possible gain is sacrificed by the player in order to maintain a fairness norm. The results of the neuropsychological study are supported by the above men-

to use the former term.

⁴ The Ultimatum Game was first used as an experimental paradigm by W. Güth in 1980. At that time, fair distribution of outcomes in a cooperation game was considered ‘irrational’. Experiments using the Ultimatum Game draw attention to the crucial role of fairness and pro-sociality in economic behaviour.

tioned experiment on the public good game conducted by Fehr and Gächter (2002). As we have seen, the second part of their experiment consisted in asking the subjects what motivated them to b-altruistically punish free-riders; results pointed to the emotion of anger.

More recently, one can observe a slight shift in economists' arguments regarding altruism. Specifically, there is a tendency to neglect the motivational aspect of emotions and replace it with an understanding of motivation that is bound to the desire to maintain social norms. In their 2004 study, Dominique de Quervain, Urs Fischbacher, Ernst Fehr and colleagues introduced a neurobiological tool into the experimental design of a trust game followed by a punishment condition. In such a game, two players receive the same amount of money. The first player is asked to decide how much of his money to pass on to the second player – the trustee. All money passed is increased by a multiplication factor of two to four – depending on the game. The trustee then decides how much of this to return to the first player. She is allowed to keep all the money for herself, in which case she would show free-riding behaviour. As in Sanfey's experiment, the subjects were brain-scanned during the game while they learned that they had been cheated by another player and made the decision whether to punish b-altruistically. Observation of the subjects' neural circuit activation showed that a brain area linked to anticipation of reward – the 'caudate nucleus' – played a prominent role when people decided to punish. Subjects who exhibited stronger activation of the caudate nucleus were ready to incur more personal costs to punish a free-rider in comparison with subjects who exhibited low caudate activation.⁵ Moreover, it must be noted that the configuration of the game did not allow the punishers to expect future monetary gains in the course of the game; thus, punishment had to be b-altruistic. Experimenters interpreted the reward-directed mechanism underlying b-altruistic punishment in hedonistic terms – people seek satisfaction – and considered it to be a result of evolution.⁶

At first glance, the analysis of the results in terms of *hedonistic* motives is rather surprising because experimental economists are well known to argue in favour of *b-altruism*. In fact, there is no contradiction here because the concept of psychological egoism does not deny that actions caused by self-interested *motives* can have positive *effects* for others and unfavourable ones for

⁵ Similar results have been obtained in studies on the psychological and neural mechanisms underlying reciprocal cooperation. It was shown that activation of the brain areas linked with reward processing – including the caudate nucleus – positively reinforced reciprocity and helped resist the temptation to defect (Decety, *et al.* 2004, King-Casas, *et al.* 2005, Rilling, *et al.* 2002, Rilling, *et al.* 2004). However, these results are mildly relevant for the altruism debate because reciprocal cooperation can hardly be conceived as a case of psychological altruism. These studies do not help to answer the question whether people can act altruistically when it is costly for them and no future benefit is to be expected. It must be noted that most of these studies display economic games where the goal is to win as much money as possible. In such settings, subjects are inclined to cooperate because they are placed in a setting in which they are expected to behave in a way that produces long term benefits.

⁶ This paradigmatic way of conceiving motivation in terms of anticipated pleasure is to be found in many economic studies. De Quervain and colleagues report that people seek satisfaction in re-establishing equity. The same line of thought is applied to generous behaviour: for example, Haselhuhn and Mellers (2005) report a correlation between fair or cooperative behaviour and self-reported degree of pleasure derived from acting fairly. Tabibnia and

the agent – which is precisely the feature of b-altruism. This allows researchers such as Fehr or Fischbacher to argue in favour of psychological egoism while maintaining the b-altruistic aspect of the behaviour. As they put it themselves:

Thus, the punishment of defectors is an altruistic act in the biological sense⁷ because, typically, it is costly for the punisher and induces the punished individual to defect less in the future interactions with others. However, our results suggest that it is not an altruistic act in the psychological sense. (de Quervain, *et al.* 2004: 1257)

In fact, their analysis is more complex than that. The authors suggest that “many people voluntarily incur costs to punish violations of social norms” (p.1254) and they feel “satisfaction” and “relief” (p.1258) when they are able to punish norm violations. It is precisely this satisfaction that drives them to sustain social norms in their group. More formally, as we understand it, the causal chain implicitly favoured by the authors seems to be the following:

Understanding that the other player has failed to send money back → Understanding that this behaviour does not uphold the social norm of fairness → Feeling that the situation is unsatisfactory [ground of motivation] → Desire to get rid of this uneasiness [primary motive] → Desire to re-establish equity of payoffs as a means of relief from the unsatisfactory situation [instrumental motive] → Cost-benefit deliberation: re-establishing equity would be a source of satisfaction (a relief); punishment is a means to achieve it, but is costly → Choice to punish → Punishing act

In contrast with the 2002 study, the interpretation of the data provided in the 2004 paper seems to neglect the notion of anger. In the 2004 analysis, we find the idea of “relief” from an unwanted situation (p.1258), which indicates that people feel uneasy until they are able to punish. However, there is no mention of anger or any other fully-fledged emotion.⁸ Instead, an in-

colleagues (2008, 2007) make the same move and go as far as saying that it is regrettable that Sanfey *et al.* (2003) failed to report activation in the reward regions of the brain.

⁷ According to the definitions provided in the first section of this paper, the formula “biological sense” should be understood here as “behavioural sense”. Driven by a noble desire to draw interdisciplinary links, experimental economists sometimes fail to make important distinctions that would help to avoid unnecessary confusions (more on this in West, *et al.* 2007).

⁸ An emotion is not merely a feeling or a desire for pleasure. It is a complex phenomenon that cannot be defined by reference to a single feature. It is something that is experienced by an organism and has typical causes, expressions and action tendencies. In other words, an emotional episode – say of anger – is made up of three elements: (a) it responds to a typical class of circumstances – in the case of anger, insults or injuries; (b) it has typical bodily characteristics – neuronal activity, state of the endocrine system – and modifications of facial expression; (c) it involves typical action tendencies – in the case of anger, a readiness to commit a punishing act.

terpretation in terms of social norms – equity –⁹ and desire for pleasure is favoured. This shift may be due to the fact that economists consider the establishment of cooperation and other-directed behaviour through the enforcement of social norms to be a very special feature of *human* societies, in contrast with animal societies. As they write:

The ability to develop social norms that apply to large groups of genetically unrelated individuals and to enforce these norms through altruistic sanctions is one of the distinguishing characteristics of the human species. (de Quervain, *et al.* 2004: 1258)

On this occasion, the neuroeconomists are very much in line with the neoclassical tradition in economics and the social sciences; they elaborate the notion of human b-altruism precisely through a characteristic that allows them to posit a sharp distinction between humans and animals: the behavioural characteristic of norm compliance fits perfectly with the ideal of the uniqueness of the human being, whereas common emotions would not help to make human altruism unique. In their words:

Experimental evidence indicates that human altruism is a powerful force and unique in the animal world. [...] Human societies represent a large anomaly in the animal world. They are based on a detailed division of labour and cooperation of genetically unrelated individuals in large groups. (Fehr & Fischbacher 2003: 785)

3 An Alternative Interpretation

As we have seen, in their 2004 experiment, de Quervain, Fischbacher and colleagues used fMRI technologies in order to see what would happen in subjects' brains while they behaved b-altruistically. They showed that altruistic punishers' behaviour is driven by a hedonistic reward mechanism. However, nothing in the experiment allows us to determine which mental considerations – conscious or not – are primarily responsible for the choice to punish b-altruistically. At least two interpretations are possible.

The first goes back to our understanding of de Quervain and colleagues' paper. It relies on the idea that people act upon considerations relating to social norm compliance. They understand that they are confronted with a behaviour that does not uphold the social norm of fairness.

⁹ This tendency to concentrate the analysis of cooperative behaviour around the notions of fairness – or equity – and psychological reward is to be found in other studies (e.g. Tabibnia, *et al.* 2008, Tabibnia & Lieberman 2007). However, we would like to point out that even if fairness is obviously hedonically valued *in some contexts* – mainly

This situation makes them feel uncomfortable and incites them to seek more satisfactory outcomes. They think that punishing the free-rider would re-establish equity of payoff and relieve them of this unsatisfactory situation – maybe even cause satisfaction. Such an interpretation enables them to defend a hedonistic view. Actions apparently motivated by the desire to do justice are in fact motivated by self-interested considerations: the desire to re-establish equity is only an instrumental motive. Of course, this presents a powerful argument in favour of psychological egoism. Noble actions that seem to be driven by a desire for justice and fairness prove to be self-directed under close analysis.

However, there is an alternative and more plausible interpretation for the same data. If we rely on the two other studies mentioned in the second section (Fehr & Gächter 2002, Sanfey, *et al.* 2003), it seems clear that emotions – more precisely anger – have an important role to play in decision-making and should enter into the explanatory picture. Our suggestion is that subjects in a trust game become angry when another player in whom they trust does not fulfil their expectations. In trusting him, they expect him to cooperate and share the benefit he has gained thanks to their own trusting behaviour. As soon as they understand that this is not the case, they start feeling frustrated, they become angry, seek revenge and begin to think of possible retaliatory actions.¹⁰ Under this interpretation, the following causal chain would hold:

Understanding that the other player has failed to send money back → Understanding that this behaviour does not match with one's expectations → Feeling angry [ground of motivation] → Conceiving a desire for revenge [motive] → Cost-benefit reflection: punishment is a means to satisfy the desire for revenge, but it is costly → Choice to punish → Punishing act

The data presented in the 2004 paper do not help to decide which of the two competing interpretations is correct. However, there are good reasons to favour our interpretation. Firstly, an argument for simplicity can be proposed. Our interpretation is straightforward and does not imply abstract desires such as the willingness to re-establish justice. The subjects in the experiment are simply angry and frustrated and therefore seek revenge. Moreover, an interpretation in terms of norms does not match with the previous findings presented in the Fehr and Gächter and de Quervain *et al.* papers; it overlooks the predominant role of the emotion of anger as proximate mechanism for b-altruistic punishment.

economic contexts with the expectation of future win-win cooperation –, it is not clear whether this is the case in all possible contexts, and in particular those that allow for the occurrence of altruistically motivated behaviour.

¹⁰ A similar interpretation of de Quervain *et al.* paper has already been proposed by Brian Knutson (2004).

It is worth noting here that we are not trying to deny that social norms play a role in the decision to punish; for us, norms are not merely individual choices, but shared beliefs and expectations that influence behaviour by shaping the context of people's reactions.¹¹ In a social environment where norms of fairness are not widely held and applied, people may not expect fair behaviour from other agents. Thus, greedy behaviour may not cause anger and revenge reactions; subjects may even react negatively towards excessively generous offers! These behavioural patterns have in fact been observed in various studies (see Bahry & Wilson 2006, Henrich 2004). Overall, it seems that people's emotional and behavioural tendencies are highly context dependant.

4 Disappointments and hopes

As we shall now see, if we favour our interpretation of de Quervain and colleagues' study, it becomes clear that no relevant conclusion can be drawn concerning the possibility of psychological altruism. Indeed, if subjects in the experiment reacted out of anger and sought revenge, the experiment is no longer dealing with actions apparently motivated by the desire to carry out justice. It is simply concerned with the motivation underlying retaliatory actions. No advocate of psychological altruism would deny that motivation for those actions is self-directed. The altruism thesis does not assert that all actions are altruistically motivated; it maintains that at least some actions are.

More generally, there are good reasons to be sceptical about psychological attributions on the basis of brain scan. The possibility of teasing out selfish motivation from other directed motivation by looking at brain activation is controversial. This technology is immensely helpful in understanding how basic and highly modular machineries such as vision or smell work. But it is unclear whether we know enough about the physical mechanisms underlying people's thoughts – and even an organism's motivation – to allow for interesting results. A mental state or an emotional reaction is never located in one single brain region and each brain region is involved in more than one process (LeDoux 2002). Neuropsychologists are very well aware of the difficulty. For example, Golnaz Tabibnia and Matthew Lieberman admit that “we cannot confidently infer from the observation of increased signal in a region that activity in that region evoked one mental process rather than another” (Tabibnia & Lieberman 2007). However, observation of neural activity in particular regions of the brain sometimes enables one to dismiss some of the

¹¹ This understanding of norms is consistent with recent proposals for analysing social norms in philosophy of social science (see for example Bicchieri 2006: 8).

available hypotheses¹² or to strengthen some others. In the latter case, neuroimaging data could converge with interpretations elaborated in other sciences – such as psychology, philosophy or economics.

In brief, one should carefully avoid overstating neurological findings but without being over-suspicious. After all, a brief look back into history of science reminds us that new scientific discoveries can stun people at first and that it often takes a long time for the broader scientific community to admit and integrate new knowledge.

It is difficult to judge whether activation of the caudate nucleus is necessarily linked to self-directed thoughts – more precisely, to anticipation of reward for oneself. For the sake of the argument, let us take it for granted. Under this assumption, it seems to us that if brain imaging methods are to become useful in the context of the philosophical altruism versus egoism debate, the experimenters should scan the brains of subjects who appear to act without seeking reward. The particular setting of de Quervain, Fischbacher and colleagues' experiment does not fulfil this condition. Hence, it cannot serve the egoism thesis simply because no relevant type of action comes into play.

This is not to say that b-altruistic punishment cannot be used as a model for testing the possibility of psychological altruism. However, to make it relevant one should think of a more fruitful setting. An interesting alternative would consist in introducing a 'third party' condition:¹³ a third player, not directly involved in the game, observes the other players and is offered the possibility of investing his own money in order to punish the free-riders. Scanning the brains of those players may prove relevant for understanding altruism. Without knowing what happens in the brain of third party punishers, one might assume that they simply take pleasure in punishing others, whatever their deeds. As an illustrative example, consider the way detainees held in the Abu Ghraib prison in Iraq were treated by some of their prison guards. Alternatively, one might imagine that these anti-social punishers do not really discriminate between their own needs and welfare and those of others; they identify themselves with the betrayed players and, consequently, feel angry and want to punish in revenge, as if they had been the target of the free-riding behaviour. If a brain scan can show that the 'caudate nucleus' is not activated in relation with third party punishment, both of these hypotheses could be ruled out. B-altruistic punishers would more accurately be described as impartial observers and caretakers of justice and social

¹² For example, if a particular type of mental process is hypothesised and previous research in neurology indicates that activity of two or more brain regions underlie this type of mental process, one can test whether these particular brain areas are activated according to the hypothesis. If it is not the case, the hypothesis can be put into question.

¹³ It has already been shown that third-party punishment is a stable behavioural pattern. In one of their studies, Fehr and Fischbacher (2004b) observed that 60% of third-party participants punished, although they knew that their economic payoff could not be affected by norm violation and that punishment was costly for them and would yield no future benefit.

order, thus, altruists in the psychological sense. Hence, it could be worth investigating the neurobiological basis of this kind of behaviour.

Conclusion

Debates over biological and behavioural altruism seem to have more or less been settled thanks to excellent studies and recent theoretical achievements. This is not the case for the controversy over psychological altruism. The target of this paper was to discuss the possible contribution of neuroeconomics to this particular debate. After having reviewed some relevant studies in experimental and neuroeconomics, we have argued that these experiments are open to different interpretations regarding motivation. Our interpretation of the data differs from that proposed by the neuroeconomists. Rather than analysing decision-making in terms of self-directed calculation of future benefits, we favour an account of decision-making in terms of emotional motivation. Under this interpretation, it turns out that the controversy over the possibility of psychological altruism cannot be resolved with the help of these studies. This is not to deny in principle that experiments in neuroeconomics could prove relevant for the philosophical debate. However, more sophistication is needed; we have attempted to provide some hints for the direction of possible future research in this field.

* Many thanks to Philip Kitcher, Michel Chapuisat and Chloë FitzGerald for correction, advice, and comments on previous versions of this paper.

References

- Bahry, Donna L. & Wilson, Rick K. (2006), "Confusion or Fairness in the Field? Rejections in the Ultimatum Game under the Strategy Method", *Journal of Economic Behavior & Organization*, 60, pp. 37-54.
- Bicchieri, Cristina (2006), *The Grammar of Society : The Nature and Dynamics of Social Norms*. Cambridge ; New York: Cambridge University Press.
- Darwin, Charles (2003), *On the Origin of Species by Means of Natural Selection* (1859). Peterborough, Ont.: Broadview Press.
- de Quervain, Dominique J. F. Fischbacher, Urs Treyer, Valerie Schellhammer, Melanie Schnyder, Ulrich Buck, Alfred & Fehr, Ernst (2004), "The Neural Basis of Altruistic Punishment", *Science*, 305, pp. 1254-58.
- Decety, Jean Jackson, Philip L Sommerville, Jessica A Chaminade, Thierry & Meltzoff, Andrew N (2004), "The Neural Bases of Cooperation and Competition: An Fmri Investigation", *NeuroImage*, 23, pp. 744-51.
- Fehr, Ernst & Fischbacher, Urs (2005), "Human Altruism—Proximate Patterns and Evolutionary Origins", *Analyse & Kritik*, 27, pp. 6-47.
- Fehr, Ernst & Fischbacher, Urs (2003), "The Nature of Human Altruism", *Nature*, 425, pp. 785-91.
- Fehr, Ernst & Fischbacher, Urs (2004a), "Social Norms and Human Cooperation", *Trends in Cognitive Sciences*, 8, pp. 185-90.
- Fehr, Ernst & Fischbacher, Urs (2004b), "Third-Party Punishment and Social Norms", *Evolution and Human Behavior*, 25, pp. 63-87.
- Fehr, Ernst & Gächter, Simon (2002), "Altruistic Punishment in Humans", *Nature*, 415, pp. 137-40.

- Gigerenzer, Gerd (2008), *Gut Feelings : Short Cuts to Better Decision Making* (2007). London, New York: Penguin Books.
- Guala, Francesco (2005), *The Methodology of Experimental Economics*. Cambridge ; New York: Cambridge University Press.
- Hamilton, William D. (1964), "The Genetical Evolution of Social Behaviour. I & II", *Journal of Theoretical Biology*, 7, pp. 1-52.
- Hamilton, William D. (1970), "Selfish and Spiteful Behaviour in an Evolutionary Model", *Nature*, 228, pp. 1218-20.
- Hamilton, William D. (1975), "Social Aptitudes of Man; an Approach from Evolutionary Genetics", in R. Fox (eds.), *Biosocial Anthropology*, New York: Wiley, pp. 133-55.
- Haselhuhn, Michael P. & Mellers, Barbara A. (2005), "Emotions and Cooperation in Economic Games", *Cognitive Brain Research*, 23, pp. 24-33.
- Henrich, Joseph Patrick (2004), *Foundations of Human Sociality : Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- King-Casas, Brooks Tomlin, Damon Anen, Cedric Camerer, Colin F. Quartz, Steven R. & Montague, P. Read (2005), "Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange", *Science*, 308, pp. 78-83.
- Knutson, Brian (2004), "Behavior: Sweet Revenge?" *Science*, 305, pp. 1246-47.
- LeDoux, Joseph E. (2002), *Synaptic Self : How Our Brains Become Who We Are*. New York: Viking.
- Lehmann, Laurent Perrin, Nicolas & Rousset, François (2006), "Population Demography and the Evolution of Helping Behaviors", *Evolution*, 60, pp. 1137-51.
- Rilling, J. K. Gutman, D. A. Zeh, T. R. Pagnoni, G. Berns, G. S. & Kilts, C. D. (2002), "A Neural Basis for Social Cooperation", *Neuron*, 35, pp. 395-405.
- Rilling, J. K. Sanfey, A. G. Aronson, Jessica A. Nystrom, L. E. & Cohen, J. D. (2004), "Opposing Bold Responses to Reciprocated and Unreciprocated Altruism in Putative Reward Pathways", *Neuroreport*, 15, pp. 2539-43.
- Sanfey, Alan G. Rilling, James K. Aronson, Jessica A. Nystrom, Leigh E. & Cohen, Jonathan D. (2003), "The Neural Basis of Economic Decision-Making in the Ultimatum Game", *Science*, 300, pp. 1755-58.
- Sober, Elliott & Wilson, David Sloan (1998), *Unto Others : The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press.
- Tabibnia, Golnaz Satpute, Ajay B. & Lieberman, Matthew D. (2008), "The Sunny Side of Fairness: Preference for Fairness Activates Reward Circuitry (and Disregarding Unfairness Activates Self-Control Circuitry)", *Psychological Science*, 19, pp. 339-47.
- Tabibnia, Golnaz & Lieberman, Matthew D. (2007), "Fairness and Cooperation Are Rewarding", *Annals of the New York Academy of Sciences*, 1118, pp. 90-101.
- West, S. A. Griffin, A. S. & Gardner, A. (2007), "Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection", *Journal of Evolutionary Biology*, 20, pp. 415-32.