Is prosocial behavior egocentric? The "invisible hand" of emotions

Anne Corcos, Yorgos Rizopoulos

Économie et institutions

16 | 2011
Neuroéconomie et interactions sociales

Is prosocial behavior egocentric?

The "invisible hand" of emotions

Anne Corcos and Yorgos Rizopoulos

revues.org

Publisher
Association Économie et Institutions

Electronic version
URL: http://ei.revues.org/76
ISSN: 2553-1891

Printed version
Date of publication: 30 March 2011
Number of pages: 33-58
ISSN: 1775-2329

Electronic reference
Anne Corcos and Yorgos Rizopoulos, « Is prosocial behavior egocentric? », *Économie et institutions* [Online], 16 | 2011, Online since 31 January 2013, connection on 11 May 2017. URL : http://ei.revues.org/76

Is prosocial behavior egocentric? The "invisible hand" of emotions

Anne Corcos[1] **et Yorgos Rizopoulos**[2]

1. Introduction

Social interaction implies a set of behaviors that can notably be broken down into the terms reciprocity, loyalty, equity, altruism, benevolence and trust, or conversely, punishment and retaliation. This type of prosocial behavior raises a problem for conventional economic theory, according to which individuals, supposedly substantively rational, will pursue their strict self-interest, which, in most cases, will win out over all other considerations (Stigler, 1981). Admittedly, it might be rational to act in an "altruistic" or "cooperative" manner if the expected gains are higher than for adopting more individualistic behavior at first glance (Wintrobe and Breton, 1986; Kreps, 1990; Kreps and Wilson, 1982). In infinitely repeated games, the folk theorem suggests an infinite number of non-equilibrium strategies in which the fear of retaliation (i.e. punishment of players that defect) constitutes the basis for possible, but not systematic, cooperative behavior. Sometimes, this possibility is assimilated to "socially normative" behavior, but players are presumed to act always in pursuit of their strict material self-interest.[3] Although based on a hypothesis of bounded rationality, Williamson (1993) also considers cooperative behavior to be the result of a calculation in terms of comparative risks and rewards. The institutional context and the specific environment in which the transaction takes place are guides for this calculation, and cooperation emerges when interests converge. Conversely, when the possibility of mutual gains does not appear certain from the outset, the hypothesis of pure self-interest means that prosocial behavior appears inconceivable.

In any case, for a single interaction, standard economic theory offers no explanation for the emergence of such prosocial behavior. Yet a very large number of experiments based on the

[1] LEM (Paris II)-University of Picardy
[2] CRIISEA-University of Picardy
[3] In this type of game, it is always possible to achieve a higher gain than the minimax value if the discount factor is high enough (Fudenberg and Maskin, 1986).

ultimatum game, the dictator game and the trust game (Camerer, 2003; Henrich, 2001 and 2004; Roth *et al.*, 1991, for an overview), while not necessarily invalidating the postulate of strict economic rationality, at least raise a contradiction by highlighting the existence of altruistic and cooperative behavior. Recent discoveries in neuroscience have contributed new information on decision-making during social interaction. Referring notably to emotions, these discoveries pave the way to understanding why individuals do not systematically behave opportunistically in a single interaction.

The term "emotion" encompasses the sensation of pleasure of displeasure following the perception of a set of external or internal stimuli. It is the result of an interaction of subjective and objective factors that activate the neural and endocrinal systems. These systems generate autonomous responses that are detected by the hypothalamus, without involving conscious control. These responses may induce behavioral and somato-vegetative reactions described as *emotional experience*. In other words, adverse or pleasant stimuli prepare the body for behavioral responses (fight, aggression, etc.), whereas the limbic system intervenes in awareness and modulates behavior by adapting it. The behavioral reaction is thus modulated by individual reactivity (i.e. the personality), the environment of the situation and cognitive behavioral control.

Economists have only been interested in the role of emotion in decision-making for a relatively short period of time. Admittedly, some great authors have sporadically mentioned emotions as a parameter determining behavior. In the *Theory of Moral Sentiments,* Smith points out the importance of empathy, gratitude, resentment, deference and jealousy in social life. Bentham's utilitarism in the *Introduction to the Principles of Morals and Legislation* includes everything that is source of "happiness" in the community. Keynes considers the "animal spirits" that influence human behavior in his *General Theory of Employment, Interest and Money*. Simon's selection theory links altruism to bounded rationality. Further, Simon (1967) proposes a complex behavioral model in which emotions interfere with the motivational system that controls the hierarchy of objectives. More recently, Hirshleifer (1987), Frank (1988), Ménard

(1994), Casson (1997), and Nooteboom (2002)[4] lay stress on the importance of emotions. However, there was no overall examination of the pertinence or the limitations of the *homo economicus* paradigmatic trajectory until the contributions of Elster (1996, 1998), Loewenstein (2000) and Thaler (2000).

Elster's perspective is one of extended rational choice, which integrates social and natural constraints (including biological ones). He questions the failures of formal rationality and presents the possibility of integrating emotions into a cost/reward model. Attributing a value to emotional gratification or displeasure and calculating the marginal utility or non-utility of emotions makes it possible for him to build an indifference curve and to identify potential trade-offs. He also emphasises emotion as a factor that resolves the indetermination of a rational choice in contexts that are uncertain, complex and lack transparency. Thus, emotions would fulfill the *orientation of attention* function identified by Simon by limiting the scope of actions and enabling the "right decision" to be made.[5] Thaler adopts a more neutral approach by suggesting simply that economic analysis should gradually integrate a more emotional *homo economicus*. Conversely, the contribution of Loewenstein *et al.* (2001), relating notably to decision-making in situations of high uncertainty, introduces a more radical proposition. By distinguishing between anticipated emotions, expected to be experienced in the future, and immediate emotions, experienced at the moment a decision is made, Loewenstein highlights the fact that some behavioral effects are not subject to logical evaluation and in any case cannot be attributed to substantive reasoning.

From this standpoint, once a stimulus occurs, we could observe the following loop: emotion ➔ behavior ➔ emotion. The first

---

[4] Ménard (op. cit. p. 202) speaks of "emotional ties… that are based on the principles of equity and solidarity", Casson (op. cit. pp. 135-136) mentions "emotional reward" as a motivation, whereas Nooteboom (op. cit. pp.42-43) underscores the role of the affective as a foundation for trust and "the rationality of emotions", emphasising the psychological mechanisms that foster benevolent behavior. Citing the teachings of evolutionary psychology, he even affirms the existence of selection based on trust – said to be written into the genes – since the prehistorical era when survival depended on collective hunting and gathering activities (from this standpoint, for the vast majority of human beings, it would be psychologically unbearable to live in a state of continuous mistrust). On a more specific level, Frank and Hirshleifer point out the fact that emotions facilitate cooperative relationships by shifting the priority away from the consequences of decisions.

[5] For Elster, the perception of emotions reflects the author's intellectual trajectory and, logically, is open to debate. The same holds true, among other things, for the *intentionality* of an emotion (Elster, 1996, p. 1387), which amounts to highlighting only expected emotions depending on the resulting decision.

leg (emotion ➔ behavior) essentially involves the effect of immediate emotions (i.e. a reaction induced by an emotional state), whereas the second (behavior ➔ emotion) refers to expected emotions integrated into a deliberative process (behavior is oriented towards seeking a certain emotion).

Our objective is to understand how immediate emotions, the search for positive emotions or the avoidance of negative emotions can determine our actions, notably prosocial behavior. In an initial stage, the discoveries of neuroscience prompt us to qualify the interpretations typically made of these behaviors, be they calculative or normative. Prosocial behavior such as some acts of charity, cooperation, trust or punishment triggers gratification for those that exhibit such behavior. These decisions, which appear to be contrary to strict self-interest, are in fact sources of pleasure. They thus have egocentric determinants, without necessarily being the result of a deliberative process. This observation will enable us, in a second stage, to broaden the notion of rationality in order to reflect the emotional dimension of our actions. This is undoubtedly the price economists must pay to build a more realistic representation of decision-making based on the interactions between the deliberative and emotional systems.

The first section of this article is therefore dedicated to presenting the experimental results that highlight individual types of prosocial behavior. The explanations typically put forward to explain such behavior will then be discussed. In the third section, the egocentric determinants of prosocial behavior will be identified thanks notably to the contributions of neuroscience. Lastly, the rationality of such behavior will be discussed in light of the emotional experiences and the complexity of cerebral processes during decision-making.

2. Prosocial behavior: experimental results

Experimental economics studies the behaviors in social interaction using a range of games: the ultimatum game (UG), the dictator game (DG), the public goods game (PG) and the trust game (TG). The UG (Guth *et al.*, 1982) is a game in which two players must divide a monetary sum, S. The first player proposes how to divide the sum (α; S-α), wherein α is the portion he will keep, and (S-α) the portion for the second player. If the second player accepts this proposal, the sum is divided according to the terms laid out by the first player and the game is over; if the second player refuses, neither player receives anything and the game is over. Economic theory suggests that the second player will accept any division as long as he receives a sum equal to or greater than zero. Anticipating this, the

first player should rationally propose a division (S; 0), which constitutes the Nash equilibrium for this game.

The DG is a "degenerate" version of the ultimatum game. Player 1 has an endowment that he can choose to share with player 2 in any way he desires. Unlike in the previous game, player 2 has can not to refuse the division proposed by player 1. The sum is shared according to player 1's proposal and the game ends. Strictly speaking, this game relates more to decision theory than to game theory, as there is no actual interaction between the two players. Nevertheless, the theoretical equilibrium of this game would consist of player 1 keeping the full sum and giving nothing to player 2.

The PG illustrates the problem of financing a public good: each individual's self-interest lies in a public good being produced without having to contribute to its financing. Thus, the PG involves five or six subjects, each endowed with ten experimental tokens. The players can place all or a portion of this endowment into a common pot. They also have the possibility of keeping all their tokens. Each token kept is worth two euros to its holder, whereas each token invested in the common pot results in a payout of one euro to each group member. Given the game's rules, it would be rational to behave as a "free rider": each player's self-interest would lie in all the others contributing to finance the common pot, while not contributing himself. At equilibrium, no one would contribute to financing and the public good will not be produced.

In the TG (Berg *et al.*, 1995), two subjects take turns playing one after the other. Player 1 has an endowment and must decide whether he gives a portion of it to player 2. The amount given, if any, is multiplied (e.g. by a factor of 3x) by the experimenter before being transferred to player 2. Player 2 must then decide what amount to send back to player 1. The game ends, and the players keep the amounts resulting from their interaction. The subgame perfect equilibrium solution would consist of player 2 sending nothing back to player 1. Player 1 would anticipate this choice, and thus give nothing to player 2 in the initial period.

For each of these games, substantial multiethnic experimental data (Roth *et al.*, 1991; Henrich *et al.*, 2001; Cardenas and Carpenter, 2008) converge towards the observation that few players behave according to the teachings of standard economic theory. On the contrary, these games allow for a set of prosocial decisions and actions to be expressed. They reveal a large variety of behavior ranging from reciprocity to pure altruism or benevolence, as well as cooperation, trust, loyalty, or even costly punishment of opportunistic and inequitable behavior.

In the PG, Ledyard (1995) observes that subjects send between 40% and 60% of their initial endowment to the common pot, even though cooperation decreases as the game is repeated.

Remarkably enough, Andréoni (1988) shows that by changing game players, the initial cooperation rate (for the first iteration of the game) is relatively constant, at 40-60%. The contribution to financing the public good reveals individuals' tendency to cooperate and to go beyond the temptation to act opportunistically. In the TG (Berg *et al.*, 1995), players generally send a sum of money (around 50% of their endowment) whereas standard rationality would predict that they send nothing. This is undoubtedly the expression of a form of trust. Even though a strategic dimension may be at work in player 1's decision[6], player 2's decision to send back money calls upon considerations of reciprocity and loyalty.

On another level, the DG supplies an interesting measurement of the degree of benevolence and altruism of individuals. Players show a propensity to share their endowment with a stranger from whom they cannot hope to obtain any sort of gratification. In fact, the amounts given range from 20% to 47% of player 1's endowment, regardless of whether players come from developed countries (the US, Sweden or Russia) or developing ones (Africa and South America, Nigeria, Kenya, Ghana or Tanzania). In the UG, considerations of equity and justice also appear to be at work. The offerers (players 1) propose on average 40% of the total amount. The modal offer is an even 50/50 breakdown. A proposal deemed inequitable (i.e. equal to or less than 20%) is often punished (i.e. player 2 refuses the offer), which implies costs for both players and invalidates the theoretical prediction that a few euros are better than nothing at all (Camerer and Thaler, 1995; Blount, 1995; Camerer *et al.*, 2004). Furthermore, comparing the amounts offered in the UG and in the DG enables us to separate out the purely altruistic dimension of player 1 and the strategic dimension of sharing (Cox, 2004). In the UG, player 1 may be tempted to make an "equitable" proposal as he anticipates that an excessively unfair proposal will be refused. Conversely, this risk does not exist in the DG. As a result, we can consider player 1's gift in the DG to represent a measure of his altruism.

In general, there is a strong tendency towards rewarding players that have behaved well in the past, i.e. by showing trust or benevolent and cooperative behavior, while punishing those that "defected". Thus, in the TG and the UG, player 2, if treated fairly or entrusted with a generous amount by player 1, will tend to be generous in exchange or to accept player 1's offer. Conversely, many subjects are prepared to punish free riders by incurring the cost of such punishment themselves ("altruistic punishment"), even when interactions are anonymous, there is no possible reputational effect

[6] Player 1 may adopt a gift/payback perspective, sending a "sign" of his good will and counting on player 2's loyal behavior.

and the "righter of wrongs" does not personally suffer from the defection (Fehr and Gächter, 2000; Fehr and Fischbacher, 2003). "Fair punishment" behavior emerges and in fact proves to be quite robust. Many subjects will choose to punish, despite the non-nil cost for themselves, disloyal behavior in a single-game situation, in other words, without expecting any immediate or future payback (Fehr and Gächter, 2000, for monetary punishment; Masclet *et al.*, 2003, incorporating peer pressure). Furthermore, Ozgur *et al.* (2006) implemented a public goods game and showed that, when given the choice, players decide freely to play in an economy equipped with an institution that has the power to punish opportunistic individuals or defectors, rather than in an economy with institutions lacking sanctioning power. Such an attitude enables robust cooperation to emerge.

3. Explanations for prosocial behavior

These results reveal an inconsistency between the predictions of economic theory and observed behaviors. They emphasise that in addition to strict self-interest, other determinants enter into the decision-making process, notably motivations related to social interaction.

Various interpretations of prosocial behavior coexist, and these do not always converge and often reflect the author's theoretical and doctrinal prejudices.

A first approach asserts that individuals that adopt prosocial behavior are strictly selfish and motivated by their own pecuniary gains, but equipped with limited rationality (Binmore *et al.*, 1995). According to this approach, subjects are unable to identify their own self-interest immediately given their limited cognitive abilities, but, after a learning phase, there is convergence towards the subgame perfect equilibrium of standard theory. Internal criticism of these proposals are based on the fact that the experimental games are simple enough that a learning phase is not necessary (Fehr and Schmidt, 2005). Thus, the behavioral hypothesis must evolve to integrate other preferences.

In this respect, "social preference" models, while maintaining the hypothesis of substantive rationality, postulate that individuals' actions are not purely selfish. Prosocial behavior is considered to reflect another type of preference that could be integrated without difficulty as part of the utility function of rational individuals. The only difference compared to conventional approaches lies in the fact that the gains (and, generally speaking, the resources) of others, not just one's own gains, are part of the bundle (hence the altruism shown in certain game situations, for instance). In other words, a value could be allocated to these preferences, enabling them to be

ranked. For some individuals in certain circumstances, prosocial preferences will win. Thus, in the end, prosocial behavior would result from a deliberative process based on calculation.

The most well-known of these models are those that underscore the principles of reciprocity and equity. Equity enables contribution and retribution to be taken into account, with the notion of proportionality between the two (Homans, 1974; Adams, 1963). Reciprocal behavior can be considered the result of an equity norm. The equality principle, on the contrary, is aimed at ensuring that each individual receives an equal share (Deutsch, 1975) with an identical distribution of rewards to all. In light of the experimental results, we could pick out a priority rule: in a context in which neither of the players contributes to the endowment through any sort of effort, the equality norm appears to prevail. Conversely, in a configuration in which one or the other player supplies an effort (in the form of work, good will, etc.), the equity norm emerges.

Two main models illustrate individuals' preference for an equal distribution of gains due to a subject's discomfort when he has more, or less, than others. Fehr and Schmidt's model (1999), firstly, postulates that individuals' satisfaction depends not only on their endowment, but also on an envy coefficient, if they have less than others, or a guilt coefficient, if they have more. Secondly, the ERC (Equity, Reciprocity, Competition) model, put forth by Bolton and Ockenfels (2000), is similar to that of Fehr and Schmidt in its interpretation but formally different in that individuals are interested in the average and seek to be treated in a similar manner. They do not want to have less than others, but are also ready to sacrifice a portion of their endowment if they have more than the average. In both cases, individuals feel discomfort that prompts them to reject offers that they deem unfair in the ultimatum game.

Research into altruism (Andréoni, 1990) focuses on individuals that value charitable acts while revealing that only a minority of them behave in an unconditionally altruistic fashion, as the utility of charitable acts is tightly and directly correlated to the payments received by the partner. Charness and Rabin (2002) propose a model that combines a particular form of aversion to inequity and altruistic preferences, whereby individuals seek the social wellbeing of the group they belong to.

Intention-based models of reciprocity examine the reciprocal fairness based on the *intentions* attributed to others (Rabin, 1993; Dufwenberg and Kirchteiger, 2004; Falk and Fischbacher, 2006). These models are based on a theory of mind that refers to the capacity to envisage others' intentions, beliefs and desires. According to these models, individuals seek to put themselves in another's place in order to adapt their own behavior. They will make an equitable proposal if they believe that their partner has good

intentions, while their proposals will be less equitable towards partners whose intentions are deemed hostile. McCabe *et al.* (2003) thus set up a trust game with two possible scenarios, the first in which player 1 intentionally entrusts the other with a sum, and the second wherein his "trust" is "involuntary" (he has no choice other than to send a sum to player 2). This experiment showed that player 2 rewarded player 1's generosity much more when it was intentional than when it was "required" by the test protocol.

Another variant is proposed by "interdependent preference" models, which address the problem of resource allocation (payments) based on the other player's *type* ("good" or "bad"). The central question that these models attempt to answer, concerns differences in behavior shown by a single person. The explanation is apparently that a player's altruistic preferences will be stronger towards another altruistic player (Levine, 1998). In this case, each player's decisions will not depend solely on the actions of another, but also on the information transmitted via that action concerning the other individual's type. These models are particularly complex due to the interdependence of preferences.

Almost all versions of social preference theories mobilise conventional economic tools – except for intention-based reciprocity models, which apply psychological game theory – and start off by formulating a few hypotheses regarding players' utility function. Yet the question that may be raised involves the reasons behind these egalitarian preoccupations or sensitivity to others' intentions or personality. This is a missing link to understanding the prosocial tendencies of a large number of individuals. While taking into account prosocial motivations is a considerable step forward, we nevertheless need to cast light on the processes behind such behavior.

An alternative answer would be to ascribe such behavior to the strength of social norms. A social norm has three basic characteristics: it induces recurring, regular behavior; it is based on a shared social belief that is imposed on the individual; and lastly, it is applied with the use of informal sanctions (e.g. exclusion). Prosocial behavior would thus be caused by normativity, equivalent to an ethical value, entering into conflict with self-interest[7].

Social normativity approaches do indeed postulate *a priori* conformity to a norm – determined by the history of the social groups

[7] Some authors do assimilate the social norm into a repeated game involving reputational effects (Binmore, 1998). In their view, the decisions resulting from such a norm are no different from those leading to the long-term maximisation of one's own gains. Yet Fehr and Schmidt (2005) underscore that there is little chance in a one-shot game of individuals applying "normative" behavior adapted to repeated games.

in question – that is imposed exogenously on individuals and involves breaking with the strict principle of defending one's material self-interest. Yet while the norm is a crucial factor, there are formidable analytic problems in the explanations put forward for the search for reciprocity and for conditioning one's actions to those of others (Cordonnier, 1997). The limitations of oversocialised approaches are the mirror reflection of those of undersocialised ones (Granovetter, 1985). Firstly, because an action often precedes the formation of a shared representation (Weick, 1979, 1995) and nothing allows us to affirm that an abstract, general norm is imposed outside the context of the individuals and characteristics involved in a specific interaction. Secondly, sacrifice and pure altruism are not generalised principles of human behavior (Robin and Ruffieux, 1999; Bénamou and Tirole, 2006). Experimental results show that while a very large number of individuals adopt prosocial attitudes, this may change depending on the game rules, and other individuals nevertheless act selfishly.

It is interesting to note that some versions of the public goods game also observe negative "prosocial" attitudes: free riders, punished when their defection is revealed, seek revenge by counter-punishment behavior (Nikiforakis, 2008; Gächter *et al.*, 2008a). This is probably why more conventional explanations are reaffirmed, such as the advantages obtained from the free riders (Fowler, 2005), long-term gains resulting from punishment (Gächter *et al.*, 2008b) or incentives linked to reputational effects (Bénamou and Tirole, 2006).

In any case, the fact remains that these prosocial behaviors, either positive or negative, are irrational in light of the teachings of conventional economic theory, as they often lead to economic losses for the protagonists. This means that determinants other than the strict pursuit of material self-interest are involved in the decision-making process. Nor can these determinants be oversimplified to normative social motivations. In this sense, the frequency of prosocial behavior implies the existence of an emotion sparked by interaction with another individual. It is therefore reasonable to suppose that normativity derives from individual mental processes that orient attitude.

4. Prosocial *effects* but egocentric determinants

Taking into account emotional parameters enables us to better explain the *why* of irrational behavior from a strict economic standpoint. Neuroscience has highlighted the role played by the reward and punishment circuits, two major motivational systems for the individual. These circuits are at the core of our mental activity and play an important role in learning and memory. They motivate

action in order to satisfy instinctive impulses and avoid painful experiences. They orient all our behavior.

With regard to the reward circuit, we distinguish the mechanisms responsible for the emotion triggered by *obtaining* an award from those that enable future rewards to be *anticipated and expected,* controlling goal-oriented behavior. Rewards first involve a subjective feeling of pleasure and spark positive emotions. The reward circuit (or the punishment circuit) is then activated without anticipating or "seeking". In this case, behavioral reactions will not be controlled logically and will not result from deliberation. Rewards can also be apprehended as an end in themselves, thus triggering exploratory behaviors (Schultz, 2000). Lastly, the reward circuit (the dorsal striatum and in particular the caudate nucleus) is also activated when anticipating a reward (Delgado, 2007). Dopaminergic activity marks the divergence between the expected reward (based on experience) and the actual one. This process enables learning. Rewards act as positive reinforcement[8] to increase the likelihood of goal-oriented actions.

Currently, there is convergence towards the idea that the limbic system and in particular the amygdale bring information about the emotional value of stimuli and possible behavioral reactions to the orbitofrontal cortex. The latter uses this information to select behavioral responses according to their possible emotional effects, reward or punishment, via its projections to the dorsal prefrontal cortex (O'Doherty, 2004a and b; Ollat and Pirot, 2004).

The entire striatum is involved in integrating affective, motor and cognitive information. In particular, it is activated when an individual receives or anticipates a primary or secondary reward. King Casas *et al.* (2005) show that after a few repetitions of the TG, player 2's caudate nucleus, which had been activated after learning the amount sent by player 1, is then activated before player 1 sends an amount, illustrating a learning effect whereby player 2 anticipates player 1's behavior. The intensity of the activity of the caudate nucleus is also a good signal of player 2's return.

Some research (Delgado, 2007) has focused on the respective roles of the dorsal and ventral striatum (more involved in cognitive and sensomotor functions) in their relationship with reward (obtained or anticipated). The dorsal striatum appears to play the role of assessing possible consequences of actions and learning action/reward connections (Balleine *et al.*, 2007). According to Delgado *et al.* (2005a) and Haruno *et al.* (2004), the caudate nucleus is activated throughout the process of learning about the contingency

[8] Reinforcement is positive when the stimulus induces a pleasant effect and negative when it relieves an unpleasant (negative) sensation. In both cases, the stimulus leads to repetition.

between behavior and reward[9]. Once the reward is no longer informative, i.e. once it is correctly anticipated, the activity of the caudate nucleus declines. The authors deduce that "the human caudate nucleus is an integral component of a circuit involved in learning and updating current rewards with the purpose of guiding action that will maximize reward consumption" (Delgado, 2007, p. 80). The limbic system is also involved in selection information to be memorised. Conversely, the representation and identification of the reward is more a function attributable to the frontal cortex[10].

Thus, the dorsal striatum appears to be involved in triggering and programming emotional responses, but also in reward-oriented responses, notably while preparing, initiating or executing behavior. This therefore suggests two activation modes for the reward circuit, reflecting the two operating modes of the brain. Some decisions are made according to an emotional mode, impulsively, routinely; other decisions follow a deliberative mode and result from reflection. Each of the two modes, while activating different areas of the brain, interacts with the other.

Several research studies now establish the link between individual emotions and prosocial behavior. Notably, the reward circuit (i.e. striatum and insula) is activated in situations of altruism, charity, cooperation and fair punishment (Tabibnia *et al.*, 2007), as well as when an individual anticipates a monetary gain or obtains one unexpectedly (Schultz *et al.*, 1997; Kuhnen and Knutson, 2005).

The reward circuit is activated reflexively in cases of charitable behavior, which shows that individuals feel pleasure when donating to *charitable* associations dedicated to causes they support (Moll *et al.*, 2006). But altruistic considerations can also prove "rewarding". This is shown in a study by Harbaugh *et al.* (2007), which focused on public goods. Public goods financing via voluntary donations has two main determinants: "pure *altruism*", whereby donors value the production of public goods, and "impure altruism", motivated by the "warm glow effect" (i.e., the pleasure of giving). In the former, donors are concerned about the quantity of the public goods financed, regardless of the source of financing, whereas in the latter, the donor's satisfaction lies more in the donation itself than in the quantity of the public goods it generates. The study's authors shows that, to a certain extent, similar areas of the reward circuit are activated when a subject receives money, sees an association for

[9] More precisely, Haruno *et al.* (2004) showed that the dorsal striatum is activated by any action that individuals have perceived as being related to a reward.

[10] According to O'Doherty (2004a), the orbitofrontal cortex also encodes the reward stimulus. It is activated by reward or punishment (reinforcement), whereas repeated stimuli lead to a complete extinction of the cortical response if no reward or punishment follows (i.e. habituation).

public good financing receive funds, or gives freely to this association.

An attitude of *cooperation* appears to work in the same way. Thus, in the case of a repeated TG, Delgado *et al.* (2005b) shows that, for player 1, the decision to share (i.e. to trust player 2) rather than to keep the full amount of money increases activity in the ventral striatum and the anterior insula, i.e. regions associated with the reward circuit. Symmetrically, Rilling *et al.* (2002, 2004) and Delgado *et al.* (2005b) highlight the fact that the ventral striatum and the anterior cingulate cortex are activated in player 2's brain when he *intends* to cooperate, even before seeing the amount sent by player 1. The reward associated with loyal behavior is thus intermingled with that associated with being considered trustworthy (there is an externality relation between the act and the emotion, i.e. the person acting is not the one feeling the emotion).

A context of *reciprocity* can also produce a reflexive pleasure. Thus, King Casas *et al.* (2005) showed during a repeated TG that player 2's caudate nucleus is activated when he adopts loyal reciprocal behavior, after player 1 has shown his trust. The resulting pleasure is triggered by the trust shown by player 1 as much as by player 2's intention to behave loyally and to honour the trust vested in him. Furthermore, the magnitude of this activation is positively correlated to the amount player 2 returns. In terms of punitive reciprocity, De Quervain *et al.* (2004) noted that the pain felt by a subject when he realises he is considered disloyal motivates a form of revenge (altruistic punishment) that brings pleasure to the one meting it out, as suggested by the activation of the dorsal striatum, lasting the time of this reaction. Apparently, the more severe the punishment, the more activity in this area.

But the simple existence of a *social relationship* can also activate the reward circuit reflexively. Gains obtained following a successful social interaction with other individuals are a larger source of pleasure for subjects than gains obtained in the form of additional money in a non-social context. Thus, Rilling *et al.* (2002, 2004) show that in an iterative prisoners' dilemma, assuming constant monetary gains, the striatum is activated more when the partner is human rather than a computer.

Even indirectly reflexive acts can become a source of emotion for protagonists via an empathy mechanism[11]. Singer *et al.* (2004) thus establish the fact that the reward circuit is activated in individuals who are shown the face of previously cooperative persons. This mechanism is even visible with regard to strangers

[11] Empathy consists of sharing others' emotions simply by watching or imagining them in a particular state of being (see Preston Waal 2002 for a neurscientific model of empathy, as well as Singer *et al.*, 2004b).

(Jackson *et al.*, 2005; Morrison *et al.*, 2004). *Seeing* individuals be punished that had previously showed loyal behavior increases the activity of the punishment circuit (the insula and the anterior cingulate cortex) of observers (Singer *et al.*, 2006). Conversely, individuals (males) that witness unfair punishment of other individuals show an activation of the reward circuit (striatum).

Some acts can induce a feeling of pleasure in others. Thus, the very feeling of being treated fairly or equitably, or even being considered trustworthy, produces pleasure, as does watching cooperative or loyal persons, even without interacting with them. On the contrary, being a victim of unfair behavior activates the physical disgust region (Sanfey *et al.*, 2003). A common question is whether a fair proposal in the UG is also the one that generates the highest material gain. It thus becomes difficult to dissociate the emotional experience due to a proposal's equity (or lack thereof) from that related to monetary gain. Yet Tabibnia *et al.* (2008) implemented a research protocol that enables these two dimensions to be dissociated. They thus supply information suggesting that a loyal proposal generates pleasure that activates the reward circuit (the ventral striatum, amygdale, ventromedial prefrontal cortex and orbitofrontal cortex). At the same time, when a proposal deemed disloyal is accepted (which implies "swallowing one's pride"), the renouncement of punishment activates neural mechanisms that logically control the negative feeling and regulate emotions (decreased activity of the anterior insula, combined with a more intense activity of the ventromedial prefrontal cortex). In this case, the reward circuit is not activated.

To conclude, emotional factors are essential parameters in the decision-making process in the case of punishment and reprisals inflicted for disloyal or inequitable behavior. Emotional experience has a significant influence on whether a proposal is accepted or refused in the ultimatum game, whether as a consequence of the game situation (Espinoza *et al.*, 2004) or as a driver of the game (Harlé and Sanfey, 2007). Thus, Petit (2009) showed that individuals filled with indignation or anger reject substantially more inequitable proposals than subjects that feel positive emotions. This also appears to be confirmed by the fact that a high level of testosterone is correlated with punishment of inequitable proposals (Burnham, 2007) due to the high "self-esteem" of dominating individuals[12]. Lastly, the neuroimagery work of Sanfey *et al.* (2003) concludes that a disloyal proposal is associated with negative emotional responses (i.e. activation of the anterior insula), whereas De Quervain *et al.* (2004) show that a partner's disloyal attitude is a source of suffering for the individual that is betrayed.

[12] This hormone is linked to behavior aimed at domination.

Prosocial behavior thus results from the individual's immediate emotional experience. More particularly, charity, reciprocity, altruistic punishment and positive social interaction generate gratification that is identical, from the endocrinal and neural standpoint, to economic gains. These findings appear to support the hypothesis that these attitudes have egocentric determinants. This may be a major reason behind the frequency of such behavior.

5. Discussion

The emotion triggered by altruism, empathy, reciprocity, equity, charity or altruistic punishment appears to modify how economists have classified and understood these attitudes. These behaviors are described as *prosocial* in that they are allocentric (i.e. oriented towards others) and are part of a social relationship (Behrens *et al.*, 2009). However, as they are associated with the neural mechanisms of the reward and punishment circuits, they have egocentric determinants. We can thus postulate that, while such behavior is undoubtedly motivated by concern for others or by social norms, it is nevertheless centred on the individual that derives pleasure or displeasure. Depending on the context, the cognitive processes (learning, interpretation) and the individual's personality (reactiveness), a prosocial attitude would thus have egocentric determinants resulting from emotional experience.

This approach enables the concept of "self-interest" to be expanded upon and surpassed. The determinant of a decision/action would no longer be strict economic self-interest[13] but rather the sum of emotional experiences (gratification, pain avoidance, etc.), felt viscerally or calling on deliberative processes.

Indeed, neuroscience has indicated that emotions generate information that influences decision-making, and the absence of emotion in subjects following serious cranial injury may even result in an incapacity to make decisions. Recent progress highlights leading indicators of decisions by identifying changes in neural activity a few seconds before a decision is made. For instance, Knutson and Greer (2008) point out that the emotion triggered by anticipation of a monetary reward, and identified by neural activity, could affect or predict the subject's choice. Activation of the nucleus accumbens, which reflects a certain degree of excitation for the subject, predicts a purchasing decision. Likewise, as explained

[13] A. Hirschman (1980) already showed that the equivalence between "self-interest" and "economic self-interest" resulted from a gradual restriction of meaning.

above, Rilling *et al.* (2002) and Delgado (2005b) show that a player's behavior can be anticipated based on the activity of the ventral striatum and the anterior cingulate cortex.

As the search for emotional reward can thus be a driver in the decision to act, it may be a major reason for prosocial behavior. These discoveries open the way to a better understanding of the deliberative decision-making system by factoring in the true role played by emotions. Integrated into a theoretical framework derived from the economic tradition, these discoveries would mean that when individuals are faced with a choice, they compare reward anticipations for each option and seek to optimise their decision by choosing what offers the highest anticipated reward. According to O'Doherty (2004) and Sanfey *et al.* (2006), the individual's dilemma is one of exploration/exploitation[14], i.e. to arbitrate between exploring new sources of reward and exploiting maximum satisfaction from a given source of pleasure. In the latter case, the emotional influence would be integrated into a calculation of the probability of realising the expected utility (Mellers and McGraw, 2001; Schultz, 2006). *Homo* becomes *sapiens* but is still equipped with perfect rationality enabling the value of emotions to be calculated, anticipated pleasure (or displeasure) to be assessed, and action to be taken based on a cost/reward calculation (for instance, between the pleasure derived from punishing disloyal behavior and that resulting from receiving a small monetary amount).

By revealing the reward aspect of prosocial behavior, neuroscience has shown that these situations are not just sources of "costs" for subjects. Now, we know more about how the deliberative process is carried out in the case of "problematic" decisions, in which the subject's convictions, beliefs, preferences and affects are in conflict (e.g. in the case of choosing between the monetary loss of making a gift and the emotional reward related to the warm glow effect of giving). The resulting emotion could thus be integrated into a logical framework. This appears to be the case, for instance, when activity in the prefrontal cortex – involved in assessing a decision's costs/benefits – is stronger for a costly punishment than for a punishment that implies no costs. Likewise, subjects that show higher activity in the striatum in the case of a punishment implying no costs will also be those that agree to pay a higher price in the case of a costly punishment.

In a UG, Sanfey *et al.* (2003) show that the insula's activity enables us to predict whether a proposal will be accepted or rejected. Intense activity reflects a high probability of rejection, as the insula reflects the state of emotional disgust of the subject that has received an unfair proposal. Likewise, activity in the prefrontal dorsolateral

[14] These terms were initially developed by J. March (1991).

cortex – involved in the execution of tasks, pursuit of a goal, memory – enables us to predict, to a certain extent, that an unfair proposal will be accepted, as reason outweighs the negative emotion triggered by the proposal. Going even further, Van't Wout *et al.* (2005) disrupted the dorsolateral prefrontal cortex using transcranial magnetic stimulation, and observed an increase in the probability of an unfair proposal being rejected. This is because a disruption in the prefrontal cortex cancels its moderating effect, and the disgust triggered by the unfair proposal is no longer counterbalanced by the prospect of a monetary gain.

Along the same lines, but with a different intention, the work of Etkin *et al.* (2006) makes it possible to distinguish between the detection of an emotional conflict and its resolution. As in Sanfey *et al.* (2003) and Van't Wout *et al.* (2008), the activity of the dorsolateral prefrontal cortex and the amygdale signals the existence of a cognitive and emotional conflict. Etkin *et al.* (2006) also highlight the fact that the resolution of these conflicts is associated with activation of the rostral anterior cingulate cortex (rostral ACC). The higher its activity, the stronger the conflict. Yet this activation of the rostral ACC tends to decrease the activity of the amygdale, reflecting a reduction in the internal emotional and cognitive conflict.

Do these cerebral mechanisms signify that emotions are an additional variable for optimising decision-making within the process of formal rationality?[15]

Since leading indicators cannot be predicted, we note only a slight increase in our level of understanding of decision-making mechanisms. Depending on the individual, the warm glow effect is more or less strong; the disgust triggered by an unfair proposal is not the same for everyone, and emotional and cognitive conflicts are largely shaped by the subject's life experiences. Thus, the *determinants of emotions* remain unpredictable, for they are determined by the environment, genes, education, and so forth. The highly subjective nature of emotional experiences appears to explain the wide variety of individual behaviors when faced with the same choices. Even approaches based on variants of the expected utility theory integrating anticipated pleasure show that the same objective result can be the source of very divergent emotional experiences (Mellers and McGraw, 2001). The pleasure or displeasure experienced shows the egocentric character of prosocial behavior but does not enable us to consider emotional experiences as a simple preference in a bundle of goods.

[15] Generally speaking, decision theory contrasts rational decisions based on calculation and "emotionally guided" decisions that go beyond reason or employ pseudo-ratinality to justify behavior triggered by emotions.

Furthermore, there are *sociocognitive biases* that disrupt decision-making. The deliberative mechanism is not infallible and subjects can be victims of it. Thus, Delgado *et al.* (2005b) carried out an experiment in which they manipulated the content of information provided to subjects in a TG. They discovered that when subjects have a positive preconception on their partner, the activity of the caudate nucleus is undifferentiated according to the partner's generosity or selfishness. Subjects appear not to take advantage of the information contained in player 2's responses, as if the individual profiles at their disposal were sufficient. In addition, there is dissonance between the partner's perception (evaluated at the start and end of the experiment) and the sharing decision. Subjects are more clear-minded in their statements (i.e. saying what they think of their partners) than in their choices. They continue to share with individuals whose profile suggests they are reliable and loyal, whereas their statements indicate they do not consider them to be more reliable than others. It would thus appear that moral preconceptions bias perceptions and have a neural effect. In an entirely different field, Bourgeois-Gironde and Guille (in this issue) cast light on neuroscientific work notably involving the monetary illusion, while Bourgeois-Gironde and Aharon (also in this issue) insist on the means at the disposal of politicians to encourage individuals to break with their behavioral biases. In a different way, some recent cerebral imagery research studying the impact of "framing" postulates that emotional experience dominates reasoning (De Martino *et al.*, 2006) because the framing of a problem has a different emotional effect: "keeping 40%" or "losing 60%" of a given amount of money is not transcribed in the same way emotionally, and leads to different behaviors.

Up to now, neuroeconomics research has focused more on deliberative decision-making, and this orientation reflects traditional economic analysis. Yet it would appear that there are indeed *two decision-making systems.* The types of decisions examined up to now are part of the deliberative, controlled system, which is different from the automatic, routine and purely affective system, to reuse the dichotomy laid out by Sloman (2002) and Kahneman (2003). In fact, these two systems coexist. The former arbitrates between gains and losses according to their relative importance and activates areas of the brain such as the dorsolateral and anterior prefrontal cortex and the posterior parietal cortex. The latter activates areas such as the amygdale, the insula cortex, the orbitofrontal cortex, the ACC and the nucleus accumbens. This second system takes greater account of emotion and triggers, in particular, routine behavior, the use of behavioral heuristics and reflexive responses.

Non-conscious sensations and expected emotional gratifications intervene and guide behavior. Damasio (1994)

hypothesised the existence of somatic markers. The orbitofrontal cortex generates unconscious physiological signals that precede and influence the cognitive evaluation and reasoning process when a decision is being made. By introducing a distinction between immediate emotions at the time of the decision and expected emotions according to the result of the decision, Loewenstein *et al.* (2001) underscored that models of expected utility integrate expected emotions but not immediate emotions. Backing proposals put forth by Simon (1967), Loewenstein and his team even conclude that in situations involving a divergence between emotional experience and cognitive evaluation (i.e. formal rationality), behavior is guided by emotions rather than reason. In this perspective, it would be necessary to take into account the influence of "visceral" factors (Loewenstein, 2000) in order to understand economic behaviors, including intertemporal choices.

The research work carried out by Sanfey *et al.* (2006) is probably the best reflection of the current state of knowledge in this field, as it proposes a balanced interpretation. This study reveals that when faced with a disloyal proposal in a UG, two regions of the brain are activated in particular: the anterior insula and the dorsolateral prefrontal cortex. The former region is correlated to emotional processes and the latter to deliberative processes, and as a result, the relative degree of activation explains whether the proposal is accepted or rejected. These results back the idea of an interaction between the two systems, which also appears to be consistent with the results obtained by De Quervain *et al.* (2004) and King Cassas *et al.*. (2005). Sanfey and Chang (2008) propose a description of the neural contours of these two systems, but their interaction and the conditions for their competition/cooperation have yet to be clarified.

The teachings of neuroscience show the complexity of cerebral processes when faced with an internal or external stimulus, a fact that illustrates the need for caution when attempting to explain individual decisions. In light of all the factors explained in this article, we would propose that prosocial behavior results from an individual's emotional experience, which, as it is both distinct from and linked to calculative deliberative mechanisms and normative behaviors (due to the subject belonging to a social group), activates the neural circuits for reward (immediate pleasure or the search for pleasure) and punishment (immediate displeasure or the avoidance of displeasure). Prosocial behavior could thus be considered a product of several non-mutually exclusive determinants that are distinguished by their being extrinsic (for material incentives and social norms) or intrinsic (motivation, emotion).

From an evolutionary standpoint, we can therefore suppose that the neural bases of prosocial behavior were selected in a process of biological, anatomic and social co-evolution. In increasingly

complex and highly organised human groups, the constant combination of antagonistic and cooperative behaviors was the foundation for survival (Morin, 1973). In this perspective, individual fulfilment, which cannot be dissociated from social fulfilment, would imply such a combination – variable and imbalanced, depending on the individuals and the context – that results in pleasure and gratification, or pain and frustration. Cooperation would also be part of *human social and biological nature* as much as selfish and antagonistic behaviors. In this respect, the simplistic portrait of *homo economicus* must be completed with the "invisible hand" of emotions in order to explain economic decisions.

Bibliography

Adams J.S. (1963), "Toward an understanding of inequity", *Journal of Abnormal & Social Psychology*, LXVII, pp. 422-436.

Andreoni J. (1988), "Why free ride? Strategies and learning in Public Goods experiments", *Journal of Public Economics*, XXXVII, pp. 291-304.

Andreoni J. (1990), "Impure altruism and donations to public goods: a theory of 'warm glow' giving", *Economic Journal*, 100, pp. 464-477.

Balleine B., M. Delgado and O. Hikosaka (2007), "The role of the dorsal striatum in reward and decision-making", *The Journal of Neuroscience*, 27(31), August 1, pp. 8161-8165.

Behrens T., L. Hunt and M. Rushworth (2009), "The computation of social behaviour", *Science*, 324, pp. 1160-1164.

Bénamou R. and J. Tirole (2006), "Incentives and prosocial behaviour", *AER*, 96(5), pp. 1652-1678.

Berg J., J. Dickhaut and K. McCabe (1995), "Trust, reciprocity and social history", *Games & Economic Behavior*, X, pp. 122-142.

Binmore K. (1998), *Game theory and the social contract: Just playing*, Cambridge, MIT Press.

Binmore K., J. Gale and L. Samuelson (1995), "Learning to be imperfect: The Ultimatum Game", *Games & Economic Behavior*, 8, pp. 56-90.

Blount S. (1995), "When social outcomes aren't fair: The effect of causal attributions in preferences", *Organizational Behavior and Human Decision Processes*, 63(2), pp.131-144.

Bolton G. et A. Ockenfels (2000), "ERC – A theory of equity, reciprocity and competition", *AER*, 100, pp. 166-193.

Burnham T. (2007), "High-testosterone men reject low ultimatum game offers", *Biological Sciences*, 74 (1623), September 22, pp. 2327–2330.

Camerer C. (2003), *Behavioral game theory: Experiments in strategic interaction*, Roundtable Series in Behavioral Economics, Princeton, New York, Russell Sage Foundation, Princeton University Press.

Camerer C., G. Loewenstein and M. Rabin, eds. (2003), *Advances in Behavioral Economics*, Princeton, Princeton University Press.

Camerer C. and R. Thaler (1985), "Ultimatums, dictators and manners", *Journal of Economic Perspectives*, 9, pp. 209-219.

Carpenter J.P. and J.-C. Cardenas (2008). "Behavioral development economics: Lessons from field labs in the developing world", *Journal of Development Studies*, 44, pp. 337−364.

Casson M. (1997), *Information and organization. A new perspective on the theory of the firm*, Oxford, Clarenton Press.

Charness G. and M. Rabin (2002), "Understanding social preferences with simple tests", *Quarterly Journal of Economics*, 117, pp. 817-869.

Cordonnier L. (1997), *Coopération et réciprocité*, Paris, PUF.

Cox J. (2004), "How to identify trust and reciprocity", *Games & Economic Behavior*, 46 (February), pp. 260-281.

Damasio A. (1994), *Descartes' error: Emotion, reason and the human brain,* New York, Putnam.

De Martino B., D. Kumaran, B. Seymour and R. Dolan (2006), "Frames, biases, and rational decision-making in the human brain", *Science***,** 313 (5787), 4 August, pp. 684-687.

De Quervain D., U. Fischbacher, U. Treyer, V. Schellhammer, M. Schnyder, U. Buck and E. Fehr (2004), "The neural basis of altruistic punishment", *Science*, 305, August 27, pp. 1254-1258.

Delgado M. (2007), "Reward-related responses in the human striatum", *Annals New York Academy of Sciences*, 1104, pp.70-88.

Delgado M., M. Miller, S. Inati and E. Phelps (2005a), "An fMRI study of reward-related probability learning" *Neuroimage*, 24, pp. 862–873.

Delgado M., R. Frank and E. Phelps (2005b), "Perceptions of moral character modulate the neural systems of reward during the trust game", *Nature Neuroscience*, 8(11), pp. 1611–1618.

Decety J., P. Jackson, J. Sommerville, T. Chaminade and A. Meltzoff (2004), "The neural bases of cooperation and competition: an fMRI investigation", *Neuroimage,* 23(2), pp. 744-751.

Deutsch M. (1975), "Equity, equality, and need: what determines which value will be used as the basis for distributive justice?", *Journal of Social Issues*, 31(3), pp.137-149.

Dufwenberg M. and G. Kirchsteiger (2004), "A theory of sequential reciprocity", *Games & Economic Behavior*, 47, 268–298.

Elster J. (1996), "Rationality and the emotions", *Economic Journal*, 106(438), pp. 1386-1397.

Elster J. (1998), “Emotions and economic theory”, *Journal of Economic Literature*, 36(1), pp. 47-74.

Espinoza F., A. Fedokhinin and J. Srivastava (2006), “Anger in Ultimatum bargaining: Emotion outcomes lead to irrational decisions” (extended abstract), *Advances in Consumer Research*, 33(1), pp. 264-265.

Etkin A., T. Egner, D. Peraza, E. Kandel and J. Hirsch (2006), “Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdale”, *Neuron*, 51, pp. 871–82.

Falk A. and U. Fischbacher (2006), “A theory of reciprocity”, *Games & Economic Behavior*, 54, pp. 293–315.

Fehr E. and U. Fischbacher (2003), “The nature of human altruism”, *Nature*, 425, October 23, pp. 785-791.

Fehr E. and S. Gächter (2000), “Fairness and retaliation. The economics of reciprocity”, *Journal of Economic Perspectives*, 14, pp. 159-181.

Fehr E. and K. Schmidt (1999), “Theory of fairness, competition and cooperation”, *Quarterly Journal of Economics*, CXIV, pp. 817-868.

Fehr E. and K. Schmidt (2005), The economics of fairness, reciprocity and altruism – Experimental evidence and new theories, Discussion paper 2005-20, Department of Economics, University of Munich, http://epub.ub.uni-muenchen.de.

Fowler J. (2005), “Altruistic punishment and the origin of cooperation”, PNAS, 102(19), May 10, pp. 7067-7069

Frank R. (1988), *Passions within reason*, New York, Norton.

Fudenberg D. and E. Maskin (1986), “Folk theorem for repeated games with discounting or with incomplete information”, *Econometrica,* 54 (1986), 533-554.

Gächter S., B. Herrmann and C. Thoni (2008a), “Antisocial punishment across societies”, *Science***,** 319, March 7, pp. 1362-1367.

Gächter S., E. Renner and M. Sefton (2008b), “The long-run benefits of punishment”, *Science, 322*, December 5, p. 1510.

Granovetter M. (1985), “Economic action and social structure: The problem of embeddedness”, *American Journal of Sociology,* 91(3), pp. 481-510.

Güth W., R. Schmittberger and B. Schwarze (1982), “An experimental analysis of ultimatum bargaining”, *Journal of Economic Behavior and Organization*, 3, pp. 367-388.

Harbaugh W., U. Mayr and D. Burghart (2007), “Neural responses to taxation and voluntary giving reveal motives for charitable donations”, *Science*, 316, pp. 1622–1625.

Harlé K. and A. Sanfey (2007), “Incidental sadness biaises social economic decision in the Ultimatum Game”, *Emotion*, 7(4), pp. 876-881.

Haruno M., T. Kuroda, K. Doya, K. Toyama, M. Kimura, K. Samejima, H. Imamizu and M. Kawato (2004), "A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task", *Journal of Neurosciences*, 24, pp. 1660–1665.

Henrich J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis and R. McElreath (2001), "In search of homo economicus: behavioral experiments in 15 small-scale societies", *AER*, 91, pp. 73–78.

Henrich J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis and R. McElreath (2004), "Overview and synthesis", in, J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr et H. Gintis (eds), *Foundations of human sociality*, Oxford, Oxford University Press, pp. 8–54.

Hirshleifer J. (1987), "On the emotions as guarantors of threats and promises", in *Dupre J. (ed.), The last on the best*, Cambridge, Mass., MIT Press.

Homans G. (1974), *Social behavior: its elementary forms*, New York, Harcourt Brace Jovanovich.

Jackson P., A. Meltzoff and J. Decety (2005), "How do we perceive the pain of others: A window into the neural processes involved in empathy". *NeuroImage***.** 24, pp. 771-779.

Kahneman D. (2003), "A perspective on judgment and choice: mapping bounded rationality", *American Psychologist*, 58, pp. 697–720

King-Casas B., D. Tomlin, C. Anen, C. Camerer, S. Quartz and P.R. Montague (2005), "Getting to know you: reputation and trust in a two person economic exchange", *Science*, 308 (5718), pp. 78–83.

Knutson B. and S. Greer (2008), "Anticipatory affect: neural correlates and consequences for choice", *Philosophical Transactions of Royal Society*, 363, pp. 3771–3786.

Kreps D. (1990), "Corporate culture and economic theory", in *Alt J. et Shepsle K. (eds), Perspectives on Positive Political Economy*, Cambridge, Cambridge University Press.

Kreps D. and Wilson R. (1982), "Reputation and imperfect information"", *Journal of Economic Theory*, 27(2), pp. 253-279.

Kuhnen C. and B. Knutson (2005), "The neural basis of financial risk-taking", *Neuron*, 47, pp. 763-770.

Ledyard J. (1995), "Public Goods: A survey of experimental research", in: J. Kagel et A. Roth (eds.): *Handbook of Experimental Economics,* Princeton, Princeton University Press.

Levine D. (1998), "Modelling altruism and spitefulness in experiments", *Review in Economic Dynamics*, 1, pp. 593-622.

Loewenstein G. (2000), "Emotions in economic theory and economic behaviour", *AER, Papers and proceedings*, 90(2), May, pp. 426-432.

Loewenstein G., Hsee C., Weber E. and Welch N. (2001), "Risk as feelings", *Psychological Bulletin*, 127(2), pp. 267-286.

March J. (1991), "Exploration and exploitation in organizational learning", *Organization Science*, 2(1), February, pp. 71-87.

Masclet D., C. Noussair, S. Tucker and M.-C. Villeval (2003), "Monetary and non-monetary punishment in the voluntary contributions mechanism", *AER*, 93(1), pp. 366-380.

McCabe K., M. Rigdon and L. Vernon Smith (2003), "Positive reciprocity and intentions in trust games," *Journal of Economic Behavior and Organization*, 52, pp. 267-276.

Mellers B. and McGraw P. (2001), "Anticipated emotions as guides to choice", *Current directions in psychological science*, 10(6), December, pp. 210-214.

Menard C. (1994), "Behaviour rationnel et coopération: le dilemme organisationnel", *Cahiers d'Economie Politique*, 24-25, pp. 185-207.

Moll J., F. Krueger, R. Zahn, M. Pardini, R. de Oliveira-Souza and J. Grafman (2006), "Human fronto–mesolimbic networks guide decisions about charitable donation", PNAS, 109(42), pp. 15623–15628.

Morin E. (1973), *Le paradigme perdu : la nature humaine*, Paris, Seuil.

Morrison I., D. Lloyd., G. di Pellegrino and N. Roberts (2004), "Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue?", *Cognitive, Affective & Behavioral Neuroscience*, 4, pp. 270-278.

Nikiforakis N. (2008), "Punishment and counter-punishment in public good games: Can we really govern ourselves?" *Journal of Public Economics*, 92 (1-2), pp. 91-112.

Nooteboom B. (2002), *Trust*, Cheltenham et Northampton, Edward Elgar.

O'Doherty J.P. (2004a), "Reward representations and reward-related learning in the human brain: insights from neuroimaging", *Current Opinion in Neurobiology*, 14, pp. 769–776.

O'Doherty J.P., Dayan P., Schultz J., Deichmann R., Friston K. and Dolan R.J. (2004b), "Dissociable roles of ventral and dorsal striatum in instrumental conditioning", *Science*, 304, pp. 452-454.

Ollat H. and S. Pirot (2004), "Cortex orbitofrontal, comportement et emotion", *Neuropsychiatrie: tendances et débats*, 25, pp. 25-33.

Özgür G., B. Irlenbusch and B. Rockenbach (2006), "The competitive advantage of sanctioning institutions", *Science*, 312 (5770), 7 April, pp. 108 – 111.

Petit E. (2009), *Emotions et décision économique dans le jeu de l'ultimatum*, Cahiers du GREThA, 3.

Preston S. and F. de Waal (2002), "Empathy: Its ultimate and proximate bases", *Behavioral and Brain Sciences*, 25, pp. 1-72.

Rabin M. (1993), "Incorporating fairness into game theory and economics", *AER*, 83(5), pp. 1281-1302.

Rilling J., D. Gutman, T. Zeh, G. Pagnoni, G. Berns and C. Kilts (2002), "A neural basis for social cooperation", *Neuron*, 35(2), pp. 395–405.

Rilling J., A. Sanfey, J. Aronson, L. Nystrom and J. Cohen (2004), "Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways", *Neuroreport*, 15(16), pp. 2539–2543

Robin S. and B. Ruffieux (1999), "L'économiste au fond du puits: l'expérimentation de la confiance", in C. Thuderoz, V.Magnematin et D. Harrisson (eds), *La confiance*, Paris, Montreal, Gaëtan Morin Editeur, pp. 75-98.

Roth A., V. Prasnikar, M. Okuno-Fujiwara and S. Zamir, (1991), "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study", *AER*, LXXXI, pp. 1068-95.

Sanfey A., J. Rilling, J. Aronson, L. Nystrom and J. Cohen (2003), "The neural basis of economic decision-making in the Ultimatum Game", *Science*, 300, pp. 1755-1758.

Sanfey A., G. Loewenstein, S. McClure and J. Cohen (2006), "Neurosciences: cross-currents in research on decision-making", *Trends in Cognitive Sciences*, 10(3), March, pp. 108-116.

Sanfey A. and L. Chang (2008), "Multiple systems in decision making", *Annals New York Academy of Sciences*, 1128, pp. 53-62.

Schultz W., P. Dayan and P. Montague, (1997), "A neural substrate of prediction and reward", *Science*, 275, pp. 1593-1599.

Schultz W., 2000, "Multiple reward signals in the brain", *Nature Reviews Neuroscience*, 1(3), pp. 199-207.

Simon H. (1967), "Motivational and emotional controls of cognition", *Psychological Review*, 74(1), pp. 29-39.

Singer T., S. Kiebel, J. Winston, R. Dolan and C. Frith (2004a), "Brain responses to the acquired moral status of faces", *Neuron*, 41, pp. 653–662.

Singer T., B. Seymour, J. O'Doherty, H. Kaube, R. Dolan and C. Frith (2004b), "Empathy for pain involves the affective but not sensory components of pain", *Science*, 303 (5661), pp. 1157-1162.

Singer T., B. Seymour, J. O'Doherty, K. Stephan, R. Dolan and C. Frith (2006), "Empathic neural responses are modulated by the perceived fairness of others", *Nature*, 439, pp. 466–469.

Sloman S. (2002), "Two systems of reasoning", in T. Gilovich et D. Griffin, (eds), *Heuristics and biases: The psychology of Intuitive judgment*, Cambridge University Press, pp. 379–396.

Stigler (1981), “Economics or ethics?”, in S. McMurrin (ed), *Tanner lectures on human values*, Cambridge, Cambridge University Press.

Tabibnia G. and M. Lieberman (2007), “Fairness and cooperation are rewarding: evidence from social cognitive neuroscience”, *Annals New York Academy of Sciences*, 1118, pp. 90-101

Tabibnia G., A. Satpute and M. Lieberman (2008), “The sunny side of fairness. Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry)”, *Psychological Science*, 12(4), pp. 339-347.

Thaler R. (2000), “From homo oeconomicus to homo sapiens”, *Journal of Economic Perspectives*, 14(1), Winter, pp. 133-141.

Van't Wout M., R. Kahn, A. Sanfey and A. Aleman (2005), “Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making”, *NeuroReport*, 16, pp. 1849–1852.

Weick K. (1979), *The social psychology of organizing*, New York, Random House.

Weick K. (1995), *Sensemaking in organizations*, Thousand Oaks, Sage.

Williamson O. (1993), “Calculativeness, trust and economic organisation”, *Journal of Law and Economics*, XXXVI, April, pp. 453-502.

Wintrobe R. and A. Breton (1986), “Organizational structure and productivity”, *AER*, 76(3), June, pp. 530-538.