Neuroscience and Biobehavioral Reviews 33 (2009) 305–313

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

# Rules of social exchange: Game theory, individual differences and psychopathology

Julia Wischniewski [a], Sabine Windmann [b], Georg Juckel [a], Martin Brüne [a,*]

[a] Department of Psychiatry, University of Bochum, LWL University-Hospital, Alexandrinenstr. 1, D-44791 Bochum, Germany
[b] Institute of Psychology, Johann-Wolfgang-Goethe University Frankfurt, Germany

ARTICLE INFO

*Article history:*
Received 16 June 2008
Received in revised form 22 September 2008
Accepted 23 September 2008

*Keywords:*
Neuroeconomics
Social exchange
Game theory
Altruistic punishment
Individual differences
Psychopathology

ABSTRACT

Human social interaction is rarely guided by pure reason. Instead, in situation in which humans have the option to cooperate, to defect, or to punish non-cooperative behavior of another person, they quite uniformly tend to reciprocate "good" deeds, reject unfair proposals, and try to enforce obedience to social rules and norms in non-cooperative individuals ("free-riders"), even if the punishment incurs costs to the punisher. Abundant research using various game theoretical approaches has examined these apparently irrational human behaviors. This article reviews the evolutionary rationale of how such behavior could have been favored by selection. It explores the cognitive mechanisms required to compute possible scenarios of cooperation, defection, and the detection of cheating. Moreover, the article summarizes recent research developments into individual differences in behavior, which suggest that temperament and character as well as between- and within-sex differences in hormonal status influence behavior in social exchange. Finally, we present an overview over studies that have addressed the question of how neuropsychiatric disorders may alter performance in game theoretical paradigms, and propose how empirical approaches into this fascinating field can advance our understanding of human nature.


**Contents**



## 1. Introduction

As humans, we are proud of our rationality. We even consider our ability to reason as one of our highest achievements, separating us from the rest of the animal kingdom and enabling us to behave in a logical and comprehensive manner. Standard models of *homo economicus* suggest that human behavior is universally based on deliberate and controlled thinking that is free from biases, and strives to maximize personal benefit (i.e., subjective utility), regardless of social and emotional context. This view has recently been challenged, following observations that human behavior is all but logical when it comes to the distribution of resources between individuals, groups or nations, while our introspective access to these processes is limited (Fehr and Fischbacher, 2004a). In reality, our behavior in situations involving give or take is guided by momentary states such as affection, empathy, or anger, as well as by gene–environment interactions influencing personality traits, and gender. But what exactly is it that let someone cooperate in one situation and defect in another? What makes us accept or reject an offer, and what guides our perception of fairness or unfairness?

Recent research in the evolutionary neurosciences has begun to unveil the factors involved in complex decision-making in

* Corresponding author. Tel.: +49 234 5077 155; fax: +49 234 5077 234.
*E-mail address:* Martin.Bruene@rub.de (M. Brüne).

0149-7634/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2008.09.008

situations of social exchange. It has become increasingly clear that humans have evolved cognitive and emotional motives that guide their behavior towards cooperation, defection, and even sanctioning of unfair behavior (Trivers, 1971; Axelrod and Hamilton, 1981; Cosmides, 1989; Fehr and Fischbacher, 2004b; Nowak, 2006; Wilson, 2006). Similarly, although not explicitly evolutionarily informed, Relational Models Theory suggests that humans, across cultures, deal with communal sharing, social ranking, imbalances of equality and market pricing in very similar ways (Fiske, 1992). Empirical evidence comes from behavioral observation and brain imaging studies during performance of tasks involving decisions about the distribution of (virtual) goods (e.g., Sanfey et al., 2003; de Quervain et al., 2004). However, as most of these studies report group effects, there is a paucity of research into individual differences in behavior, even though it is implicitly clear that character and temperament as well as situational contingencies influence an individual's attitude towards cooperation or non-cooperative alternatives.

Human social groups comprise a spectrum of individual behavioral "morphs" occupying discrete ecological niches and differing in attitudes towards exploitation of resources and contribution to the welfare of the community. Such differences may be reflected in variation in personality traits including extremes of variation akin to psychopathological conditions (Mealey, 1995; Troisi, 2005), and are highly relevant for understanding the nature of human cooperation.

Moreover, it is well conceivable that there are scenarios in which almost everyone would cooperate, as well as scenarios in which perhaps everybody would defect. In other words, contextual information is vital for an individual's benefit-cost evaluation in a given situation. For example, people who themselves have abundant resources at hand are arguably more likely to share with others in need. By contrast, in situations in which an individual feels threatened, the likelihood of cooperating with a stranger is probably weakened. Like individual differences, such contextual factors influencing decision-making during social exchange have largely been disregarded in experimental research. Similarly, the possibility that decision-making can be influenced in one way or the other by an individual's incapacity to make use of contextual information – as is the case in autism or schizophrenia – has hardly been explored empirically.

In this article, we seek to review the evolutionary background of human social behavior as far as it pertains to the exchange of goods or resources between unrelated parties. We then describe the major game theoretical approaches and the neurobiology of altruism, defection and punishment, with special emphasis on the underlying cognitive and emotional mechanisms necessary for successful understanding of complex social interaction. Finally, we discuss the role individual differences including psychopathological conditions may have in illuminating the importance of contextual information processing for social exchange, as well as ways to experimentally manipulate conditions to examine within-subject variation in behavior.

## 2. Evolutionary background of social exchange

Cooperation between genetically unrelated individuals is a highly positively selected and perhaps quite unique trait in humans (Fehr and Rockenbach, 2004). Although some studies suggest that chimpanzees, our closest extant primate relatives, also show some forms of altruistic behavior towards non-kin (Warneken and Tomasello, 2006; Warneken et al., 2007) as well as signs of inequity aversion (Brosnan and de Waal, 2003), findings are not unequivocal in this respect (Jensen et al., 2007). Cooperation seems to be much more strongly selected in humans compared to other ape species such that universally accepted rules of social exchange evolved as "the decisive organizing principle of human society" (Nowak, 2006).

Trivers (1971) was the first to lay the theoretical groundwork for the understanding of altruistic behavior between genetically unrelated individuals within the modern evolutionary synthesis of "inclusive fitness theory" (Hamilton, 1964). Based on his theoretical outline, it can be predicted that different forms of cooperation between non-kin individuals can be distinguished according to the directness of reciprocity. Direct reciprocity implies that there are repeated interactions between the same two individuals (or groups of individuals), and that both have resources that are attractive to one another (Trivers, 1971) such that both parties receive direct benefit from exchange (Hammerstein and Leimar, 2006), referred to as "conditional cooperation". Accordingly, people tend to cooperate if their counterpart behaves in the same way. Conversely, defection by one party is seen as a legitimate reason for the other party to retaliate (Fehr and Fischbacher, 2004a,b). The efficacy of such reciprocity can be experienced by both parties within relatively short periods of time.

In everyday life, however, other forms of altruism exist that are not likely to be reciprocated directly—for example, when someone helps a stranger who has lost his way. Such behavior is usually referred to as *indirect reciprocity*, where the benefit may lie in improved reputation or gain in social status (Nowak, 2006)—a potential pay-off that may lie in the more distant future. Nevertheless, such behavior can be frequently observed, because humans are concerned about the impressions other people get of them; usually, helpful deeds are approved by significant others or the community, and hence may be suitable to help raise one's social status (Nowak and Sigmund, 1998). In evolutionary terms indirect reciprocity may also serve as an "honest signal": Individuals who are willing to take costs without the (direct) prospect of getting anything in return may in fact be showing that they can afford giving away "surplus" resources.

Above and beyond direct and indirect reciprocity, however, cooperative behavior can sometimes appear to be entirely altruistic, without the prospects of ever getting anything in return. This form of "strong" altruism may be less prevalent than reciprocal forms of cooperation. Until recently, its mere existence has caused problems for evolutionary theorists, particularly those who have denied that group selection may have played a role in human evolution (e.g., Dawkins, 1976). The core problem is that, although strong altruism may benefit the social group as a whole, any mutation that brings about individuals with more egoistic tendencies would proliferate, since selfish individuals could easily exploit the altruism of others and increase their own reproductive fitness at the expense of the group. Eventually, this should lead to a selective advantage of selfish genes within the genepool and, conversely, to extinction of strong altruism.

For groups to survive it has therefore been crucial to keep the number of "free-riders" low by means of vigilantly monitoring the behaviors of group members for their willingness to cooperate or defect (Mealey, 1995), and sanction non-cooperative behavior (Boyd and Richerson, 1992, 2002). In support of this assumption, several studies have shown that humans possess the ability to detect cheating behavior almost effortlessly compared to processing more abstract information that conveys essentially the same meaning (Cosmides, 1989; Tooby and Cosmides, 1992; Gigerenzer and Hug, 1992). Moreover, all known human societies have established social rules and standards that enforce cooperative behavior within the group, especially obedience to norms of fairness and equity. In fact, most people tend to feel uncomfortable when witnessing somebody being cheated upon by another person and experience satisfaction from observing or administering

punishment to norm-violators (de Quervain et al., 2004; Singer et al., 2006). Accordingly, and in line with inclusive fitness theory (Hamilton, 1964; Williams, 1966), groups of highly cooperative individuals can be assumed to have the highest average individual fitness, which declines with the number of defectors in a particular population (Nowak, 2006). Sustaining mutual cooperation at a high level may have promoted new levels of societal organization with increasing specialization and diversity both biologically and culturally (Tooby and Devore, 1987)—a notion that is perfectly in accordance with Trivers' (1971) conceptualization of reciprocal altruism.

To understand the environmental contingencies in which individuals cooperate or defect, it is necessary to experimentally manipulate conditions that may provoke cooperation, defection or punishment of "free-riders". However, translating the evolutionary scenarios of social exchange into empirical settings has proven difficult, which is understandable, given the complexity of real-life interactions between individuals, groups or larger societal organizations. Accordingly, experimental reductionism seems inevitable. In addition to the groundbreaking studies into the cognitive mechanisms of cheating detection using variations of Wason's selection task (1966) by modifying the channel of how information is provided, but not the amount or abstract complexity of information (Cosmides, 1989; Tooby and Cosmides, 1992), more recent research has focused on evolutionary game theoretical scenarios to examine behavior in (virtual) social interactions more directly (Axelrod and Hamilton, 1981).

## 3. Game theoretical approaches

Narrowly defined, the subject of game theory is the distribution of resources between two or more parties—individuals, groups, or nations (Nowak, 2006). In a broader sense, however, game theory is pertinent to virtually every dynamic interaction between sentient beings (Wilson, 2006). Within this framework, several games have been developed to examine subjects' behavior in cooperative scenarios, which differ in complexity according to the number of participants and repetitions of social exchange.

The Public Goods Game is played by an optional number of players who receive a defined amount of money or tokens at the beginning of the exchange scenario. Participants are asked to simultaneously invest their money in a common pool (the public good) without knowing the allowance of the other players. The experimenter multiplies the whole sum by a factor that is larger than one and smaller than the number of players, and returns an equal share of that money to each of the players. This means that the players all profit equally from the public goods, irrespective of how much they have invested before. This scenario is therefore suitable to examine the extent to which players are tempted to choose a "free-rider" strategy. If someone keeps his/her own money or tokens while letting the others make contributions, his/her return will exceed those of the other players (Fehr and Fischbacher, 2004a,b; Brandt et al., 2006; Hauert et al., 2006). Accordingly, it has been observed that, if played repetitively, individual investments in Public Goods Games decline from initially high levels to lower levels (Ledyard, 1995), unless sanctions are in place by which non-cooperative players can be punished.

The effect of punishing non-cooperators has been demonstrated in a one-shot Public Goods Game played by four players (Fehr and Gächter, 2002). After making their decisions how much to contribute to the public good, players were informed about the investment of the other players, followed by the option to punish non-cooperative individuals. Predictably, under such conditions investments increased sharply and remained stable across trials. Punishment was mostly exerted on defectors and executed by cooperators. The degree of punishment was largely determined by the magnitude of deviation of defectors' investment from the group's average investment. This behavior supports the assumption that the observation of "free-riding" induces negative emotions in cooperators, which in turn increases the likelihood of sanctioning the defector's behavior, even if the interaction is only a singular event such that players cannot expect to ever meet again (Fehr and Gächter, 2002).

The experimental setting of the Public Goods Game comes closest to "real-life" scenarios of the distribution of goods within an established social group that collects taxes, donations, fees, or other investments, and/or pays off returns equally to its members regardless of individual contribution. Accordingly, one would expect cultural differences in how investment and extraction of goods is socially appreciated. However, two studies into this matter have revealed somewhat contradictory results. Whereas Okada and Riedl (1999) reported no differences in contribution rates between groups with different cultural backgrounds, another study found that Japanese participants were more likely to sanction non-cooperators relative to US Americans (Cason et al., 2002), a finding that could reflect the greater emphasis on group cohesion and cooperation in Japanese society compared to the US population. Surprisingly, Public Goods Games have also revealed that "antisocial" punishment may be provoked if punishing occurs anonymously (Herrmann et al., 2008). In other words, depending on cultural background, i.e., whether societal structures are more collectivistic or individualistic, individuals may tend to punish altruistic behavior (hence the term "antisocial" punishment), particularly if the rules of law can easily been undermined. Conversely, strong norms of civic cooperation constrain antisocial punishment in altruistic societies, and hence, promote the punishment of "free-riding".

A special form of the Public Goods game is the Trust Game played between two individuals. One player, called the investor, is endowed with a particular amount of money, some proportion of which he/she can pass on to the other player (the trustee). The proportion is multiplied by the experimenter, usually by a factor of three. The other player then decides how much of the resulting sum of money he/she wants to give back to the investor. While the concept of *homo economicus* suggests that neither player will share any money with the other player, the actual observation is that the investor does indeed send a significant share of his/her endowment to the other player, and that most trustees reciprocate.

Similarly, the Prisoner's Dilemma Game (Zeeman, 1980) represents another modification of the Public Goods Game with just two players participating. The players have to decide simultaneously whether or not they wish to cooperate or defect. If both players cooperate, they receive a certain amount of money units (MU), for example, 10 MU. If both players defect, they get a lower amount, for example, 5 MU. In the case that one player defects and the other cooperates, the latter one receives only 1 MU and the other one 15 MU. Theoretically, cooperation is not the best strategy in this scenario, because without knowing the strategy chosen by the other player, non-cooperation is associated with a higher expectancy value. Interestingly, most people nevertheless cooperate in the Prisoner's Dilemma at a considerably high rate (Rilling et al., 2004). One reason for such "trustful" behavior in both the Trust Game and the Prisoner's Dilemma Game could be the implicitly accepted social norm of conditional cooperation. People tend to cooperate with the expectation that their counterpart is willing to cooperate in return. In contrast, most people would consider it appropriate to defect when the partner defects, too (a shorthand for this kind of interaction is "tit-for-tat; Axelrod and Hamilton, 1981).

In a similar scenario, called the Ultimatum Game, two players have to split up a sum of money (e.g., 10 MU). Player A is told to make a proposal how (s)he would like to distribute the money. In contrast to the Prisoner's Dilemma Game, Player B has the option to either accept or decline the offer. If B agrees, the sum will be split according to Player A's proposal, but if B rejects, both receive nothing at all (Falk and Fischbacher, 2000). The outcome of a rejection relative to acceptance of any offered amount larger than zero is therefore costly not only for player A, but also for player B, which is why this game paradigm can be considered to tap into a very simple form of altruistic punishment (even though in the strict sense, it is not altruistic, because no third-party is involved). A wealth of studies has shown that, independent of the amount that has to be divided, average offers on the first trial hover around 40% of the total sum. Smaller offers (around 20%) have at least a 50% probability of being rejected (Güth et al., 1982; Falk and Fischbacher, 2000; Camerer, 2003b; Sanfey et al., 2003).

From a strictly rational point of view, player B starts with no money and ought to be satisfied with every amount (s)he receives, hence should never reject an offer that is larger than zero. In reality, however, the perception of being treated in an unjust manner usually stirs up negative emotions and leads individuals to reject unfair offers, even in single shot games (Sanfey et al., 2003), because humans seem to have a natural aversion against perceived inequity (Fehr and Schmidt, 1999). Put differently, unfair offers may induce a conflict between two types of reactions; the rational (cognitive) motivation to accept, and the irrational (emotional) motive to retaliate at one's own expense (Sanfey et al., 2003). Solving this motivational conflict requires a considerable amount of computational resources reflected in the activation of an extended neural network (see below).

Since player A in the Ultimatum Game knows in advance that his or her offer can be rejected by Player B, an interesting question is how people would behave if they did not have to fear any kind of direct or indirect punishment. In the Dictator Game, the responder does not have the opportunity to reject the offer, but is forced to accept (Falk and Fischbacher, 2000). Hence, from the point of view of the proposer, the Dictator Game is a more direct measure of altruism than the Ultimatum Game (Camerer, 2003a,b,c). Research using this paradigm has revealed a rather disillusioning but not unexpected image of people's magnanimity. In the Dictator Game offers by player A usually revolve around only 15% of the whole sum, albeit with considerable interindividual differences (Charness and Gneezy, 2003).

It is therefore obvious that individuals' decisions involving altruistically motivated behavior are highly context-dependent, including expectations of sanctions. It is now interesting to take a closer look at how much people in the position of a potential punisher actually perform to reinforce reciprocity and fairness. This perspective has recently been carved out using a paradigm referred to as "altruistic punishment" (AP) of non-cooperative behavior. AP involves "a selfless personal cost to the punisher that is never likely to be recovered" (Seymour et al., 2007). The crucial observation is that humans – at least in unstressed conditions – are willing to spend time and money on punishing uncooperative behavior, even when they just witness an unfair interaction between others without being personally involved (de Quervain et al., 2004; Fehr and Fischbacher, 2004b). Roughly 60% of non-participating observers of a Dictator Game between two unrelated players selflessly sanction dictators who are proposing less than 50% of the whole sum, even at their own monetary cost. The responder's expectations of punishment increase with the decline of the transferred money, and so do the sanctions of the third-party player. Likewise, studies that have implemented the possibility of AP in the Prisoner's Dilemma game have shown that defectors are frequently sanctioned by an (observing) third-party in approximately 50% of cases, whereas punishment rates drop to about 20% if the partner of the non-cooperative player has been observed to defect as well (Fehr and Fischbacher, 2004a).

In sum, experimental evidence suggests that a delicate balance exists between cooperative and non-cooperative strategies of social exchange. Moreover, individuals seem to have a clear motivation to punish non-cooperative behavior within their social in-group—quite exactly what Trivers (1971) proposed would be expected for the establishment of reciprocal altruism in long-lived social animals like humans.

This complexity of human social behavior was probably one of the crucial driving forces of brain evolution, simply because computational resources have had to be vast in order to oversee the manifold behavioral options of coalition formation, reciprocality or "free-riding" of group members (Brothers, 1990; Dunbar, 2003). In other words, reproductive success critically depended upon the ability to detect cheaters, to form trustful relationships, but also to conceal one's own intentions to defect.

## 4. Evolved cognitive and emotional mechanisms of social exchange and their neuronal correlates

Altruistic behavior and enforcement of social rules and norms require sophisticated cognitive and emotional abilities, which are represented in extended neural networks involving phylogenetically old and new structures (Wilson, 2006). The extraordinary gregariousness of humans has given rise to the hypothesis that human brains are essentially social by design (Trivers, 1971; Brothers, 1990; Dunbar, 2003). This notion is not at odds with the observation that between-group competition may be intense; on the contrary, social intelligence embraces the option for deception, cheating and cooperation, depending on environmental circumstances, resource availability and in-group–out-group distinction. At the neurocognitive level, theory of mind, episodic memory, reward prediction, the ability to tolerate reward-delay, as well as a set of culturally formed moral principles and social conventions contribute to the decision of whether or not to cooperate (Axelrod and Hamilton, 1981; Boyd, 2006; Brüne and Brüne-Cohrs, 2006). Some of these faculties may be distinguishable as domain-specific modules (Fodor, 1983; Tooby and Cosmides, 1992), but are nonetheless functionally intimately intertwined and converge in complex decision-making in social exchange situations.

One primary region involved in social cognition and emotion regulation is the prefrontal cortex (PFC). Numerous brain lesion studies have revealed that people with damage to the PFC have extreme difficulties in maintaining reciprocal social relationships. The oft-cited case of Phineas Gage, for example, illustrates how a personality can change from a rather friendly and relaxed character to an unreliable and moody person with severe violations of social rules and norms despite preserved intellectual abilities. The problem of such patients is not that they have forgotten how to behave or what would be morally appropriate; their theoretical knowledge is still accessible to them. The problem lies within the dysfunctional integration of decision-making, and consideration of future consequences of current behavior (Damasio, 1994; Fellows and Farah, 2005). Similarly, patients with degenerative brain diseases including Parkinson's disease (McNamara et al., 2007), frontotemporal dementia (Gregory et al., 2002; Lough et al., 2006), focal brain damage to the medial PFC (MPFC) (Koenigs and Tranel, 2007), autistic disorders or schizophrenia (Cutting and Murphy, 1990; Agay et al., 2008) also are impaired in appreciating social norms, and hence frequently violate rules of social exchange, even in situations where cooperation would be the most beneficial option. The reason for the similarities in

behavioral performance between such diverse clinical disorders is that the structures affected by the disease process are key components of the social cognition network (Sanfey et al., 2003; Rilling et al., 2004).

For example, McNamara et al. (2007) found that patients with Parkinson's disease adopted a "Machiavellian" attitude (i.e., showed more exploitive and non-cooperative behavior) associated with deficits in prefrontal functioning.

Within the PFC, two sub-regions seem to be specifically important for evaluation and execution of social exchange. The first is the dorsolateral prefrontal cortex (DLPFC), a structure that has been shown to play an important role in evaluating fair versus unfair offers, probably with some hemispheric differences. When functional "lesions" were induced in the right DLPFC using low-frequency repetitive transcranial magnetic stimulation (rTMS) acceptance rates of unfair offers were significantly higher relative to stimulation of the left DLPFC (Knoch et al., 2006). Moreover, the response latencies were similar to fair and unfair offers after right DLPFC stimulation, whereas left DLPFC stimulation was associated with differentially slower response latency to unfair offers. Surprisingly, however, retrospective fairness-judgments were unaffected by the stimulation. These results support the hypothesis that the right DLPFC is involved in overriding selfish-interest motives, like taking as much money as one can, in favor of fairness or equity-motives (Fehr and Schmidt, 1999; van't Wout et al., 2005). This latter tendency can no longer be followed after disruption of the right-hemispheric DLPFC, rendering selfish motives dominant.

Imaging studies suggest that the DLPFC shows a rather constant activation during processing of unfair offers, irrespective of the degree of unfairness, consistent with its involvement in goal-maintenance, working-memory, and executive control processes (Bechara et al., 1998; Miller and Cohen, 2001; Sanfey et al., 2003). Again, these findings suggest a conflict between two possible reaction types: The tendency to accept an offer, which is cognitively motivated by the wish to maximize imminent financial gains, as opposed to a socio-emotional motivation such as inequity aversion, personal reputation and other direct, indirect, or strongly altruistic motives.

The second sub-region that is crucially involved in situations of social exchange is the medial prefrontal cortex (MPFC). This region contributes to social decision-making by predicting and monitoring behavioral outcomes (reward and punishment), and by regulating emotions and behavioral impulses accordingly. The MPFC can be understood to play an integrative role in an emotional network combining inputs from different sensory and mnemonic modalities (Paus, 2001; Kringelbach, 2005). For instance, Koenigs and Tranel (2007) found that patients with damage to the ventromedial PFC (VMPFC) consistently show lower acceptance rates for unfair offers in the Ultimatum Game than healthy controls and a group of differently brain-injured people. It is possible that this difference is related to patients' insensitivity to reward in combination with their reduced ability to control negative emotions (i.e., anger towards the proposer).

Moreover, decision-making in complex social interactions requires interpreting intentions and developing a theory of others' minds, and this ability is also critically mediated by MPFC function (Frith and Frith, 2003; Amodio and Frith, 2006). Abundant research has shown that people with autistic spectrum disorders, for instance, have profound deficits in tasks that involve theory of mind (Baron-Cohen, 1995), and that impaired theory of mind in autism is associated with medial prefrontal dysfunction (but also with a defective mirror neuron system; Williams et al., 2001). Consequently, in studies using game theoretical approaches, low offers in the Ultimatum Game were much more likely to be accepted by children with autism and healthy younger children (who have not yet developed a theory of mind). In the initial round of the game, autistic children who acted as proposers displayed a balanced preference for even offers or extremely unfair offers, suggesting a lack of understanding of fairness norms which may require a theory of mind (Sally and Hill, 2006).

Similar to autistic children, patients with schizophrenia are known to have difficulties in theory of mind (Frith, 2004). Unlike autistic individuals, however, schizophrenic patients have been reported to make overly fair offers in the Ultimatum Game compared to a control group (Agay et al., 2008). Another difference between schizophrenic patients and normal individuals in that study was that schizophrenic patients did not adjust their proposals to the reaction they received from the responder in the previous trial. Instead, schizophrenic subjects raised their offer following trials in which recipients had actually accepted. Interestingly, when schizophrenic patients were recipients in the Ultimatum Game themselves, their behavior did not differ from that of control participants. These results could support the assumption that schizophrenic patients have a deficit in strategic thinking (Sullivan and Allen, 1999), possibly associated with theory of mind deficits (Mazza et al., 2003).

Other studies have used functional imaging to investigate the role of the MPFC in socioeconomic decision-making. One fMRI study using a guessing-task (coin tossing: head or tail) comprising four conditions (with and without the opportunity to cooperate, paired with and without financial reward) showed activation in the MPFC, temporal pole and temporo-parietal junctions in the cooperation condition relative to the non-cooperation condition (Elliott et al., 2006). The reward-condition led to further activation in the VMPFC, overlapping with some of the cooperation-related activation, suggesting that cooperation *per se* embodies a rewarding element. Similarly, an fMRI study using versions of the Ultimatum Game and Prisoner's Dilemma revealed activations in the paracingulate gyrus and the posterior superior temporal sulcus during cooperation relative to defection (Rilling et al., 2004).

Consistent with findings from lesion studies (Koenigs and Tranel, 2007), Sanfey et al. (2003) found in an fMRI study in healthy subjects that unfair offers made in the Ultimatum Game by a human player compared to a computer elicit activity in structures pertaining to the phylogenetically older limbic system, namely the bilateral anterior insula and the anterior cingulate cortex (ACC). The anterior insula showed a significantly heightened activation for the most unfair offers (like 8:1 or 9:1), and the intensity of the insula activation correlated with the number of rejections of unfair offers. In addition, unfair offers were associated with increasing activation of the ACC. This activation pattern highlights the role of the ACC as an important interface involved in the integration of conflicting information (Botvinick et al., 1999).

The findings of specific neuronal correlates to unfair offers are supported by a study by van't Wout et al. (2006) who measured skin conductance response (SCR) of responders in the Ultimatum Game. The SCR is regarded as a measure of autonomic arousal as an indicator for emotional involvement (Bechara et al., 1997, 1998, 2000a,b). Unfair offers by human proposers (unlike trials using a computer as proposer) were rejected at a rate that increased steadily with the degree of unfairness. The SCR showed a corresponding pattern that correlated with the unfairness of the offers. The emotional response to unfair offers was much higher than to fair offers and led to a significantly higher rate of rejections, but only when offers were made by a human proposer.

Besides the VMPFC, the DLPFC, the bilateral anterior insula, and the ACC, other neuronal structures involved in social exchange comprise striatal areas including the caudate nucleus, the nucleus accumbens, and the thalamus. These regions were found to be

active in a PET-study examining the neuronal structures involved in altruistic punishment (de Quervain et al., 2004). The caudate nucleus, which is crucially involved in reward processing and linking reward to behavior (Knutson et al., 2001; Knutson and Cooper, 2005), was activated in conditions where the counterpart was punished for his intentional cheating. In this condition the strength of the caudate activation was positively correlated with the investment in punishment. In contrast, in situations where the desire to punish could not be satisfied, caudate activation was below-average. Furthermore, there was higher thalamus activation in conditions in which subjects afterwards verbalized a strong desire to punish, in addition to activations observed in the VMPFC and the medial orbitofrontal cortex (OFC), which lends further support to the assumption of a tight interaction between phylogenetically older and younger brain structures in complex socioeconomic decision-making (Wilson, 2006). A major conclusion from this study is that people apparently gain satisfaction from sanctioning norm-violations in a monetary exchange game, even if the punishing act incurs costs. Alternatively, Knutson (2004) has suggested that caudate activation mirrors the anticipation of satisfaction rather than satisfaction itself. This interpretation is supported by the observation of increased striatal activity by the anticipation of monetary gain (Knutson et al., 2001).

In summary, it can be pointed out that an extended neural network contributes to the evaluation of costs and benefits of social exchange and to socioeconomic decision-making, including cooperation and altruistic punishment. These brain regions encompass cortical midline structures including the ACC, VMPFC, medial OFC, but also the DLPFC, insula, the caudate nucleus and the thalamus. Structural or functional lesions to these regions and activation patterns found in healthy people during functional brain imaging strikingly match those brain regions that have been found necessary to perform tasks involving theory of mind and reward prediction tasks (Sanfey et al., 2003; Rilling et al., 2004; Elliott et al., 2006). Interestingly, the cortical brain areas are the ones that enlarged relatively recently in primate evolution, and mature ontogenetically late in humans as reflected in the offset of myelination and synaptic pruning (Rakic et al., 1986; Pfefferbaum et al., 1994; Fuster, 1997; Giedd, 2004; Segalowitz and Davies, 2004).

All studies mentioned in the previous paragraphs have focused on group effects or average performance across individuals. However, in naturalistic settings one would expect differential decision-making depending on contextual information and differences according to individual predispositions. In other words, individual differences in behavior in social exchange situations depend on the availability of contextual information or the ability to process contextual information using evolved mechanisms such as "theory of mind" and autobiographic information, including acquired social rules and norms.

## 5. Individual differences

In contrast to the wealth of research into human behavior using game-theoretical approaches focusing on average group effects, there is a paucity of studies into individual differences in performance on social exchange and reward tasks. This can be considered a much neglected issue in neuroeconomics, perhaps even in the behavioral sciences in general. Individual differences can concern attitudes towards cooperation and rule-obedience, as well as the propensity to choose a free-riding strategy. Accordingly, one would expect measurable differences in actual performance. For example, it is conceivable that individual differences in character and temperament may influence behavior in game theoretical models. Such behavioral differences may, in part, be mediated by genetic differences, which, depending of gene–environment interactions, possibly predispose an individual to behave more selfishly or altruistically. For example, antisocial behavior has been found to be linked to genetic variation of the MAOA and serotonin transporter coding genes, however, to manifest only if associated with adverse childhood experiences (Caspi et al., 2002).

Two studies have shown the relevance of gene–environment interactions for behavior in social exchange games. Knafo et al. (2007) investigated individual differences in allocating behavior in the Dictator Game with relation to the length of the vasopressin 1a receptor RS3 (AVRP1), a receptor that has been shown to play an important role in affiliative behavior in mammals (Hammock and Young, 2005). The study found that participants with shorter versions of the AVRP1a offered significantly less money to recipients than participants with a longer version of the allele. The other study compared behavior of monozygotic and dizygotic twins born in the US or Sweden, and found that monozygotic twins behaved more similarly than dizygotic twins in a trust game. The heritability of the cooperative trait was estimated to be around 20% in the Swedish sample and 10% in the US sample, again suggesting that both genetic and cultural factors play a role (Cesarini et al., 2008). In a similar vein, Wallace et al. (2007) reported from their study with mono- and dizygotic twins playing the Ultimatum Game that around 40% of the variation in the responders' rejection behavior could be explained by additive genetic effects, indicating a sizeable effect of genetics on performance in game theoretical conceptions, with shared environment explaining less of the variance than non-shared environmental effects.

If the tendency to act cooperatively had a genetic basis, individual differences in social exchange games should predictably be associated with personality traits. A few studies do indeed support this notion. In a recent fMRI study, Spitzer et al. (2007) focused on Machiavellian attitudes of the proposer in a paradigm combining the Dictator Game and the Ultimatum Game, which included a punishment option that could be imposed by the recipient if treated unfair by the proposer. fMRI data revealed stronger activation in the DLPFC and the orbitolateral prefrontal cortices bilaterally when the proposer saw him/herself confronted with a punishment threat, which correlated with an increase in the transfer offer on the behavioral side (from 10 to 40 MUs on average). Also, heightened bilateral caudate activation was found in the social punishment condition indicating an arousal that was associated with the expected (but still uncertain) punishment reduction after the amount of offered MU had been increased. In line with Knutson et al.'s hypothesis (2001) that caudate activity may be associated with anticipated but yet uncertain monetary gains or punishment in a monetary reward-delay task, subjects with pronounced Machiavellian personality characteristics (high Machs) – a tendency to deceive and manipulate others for personal profit – showed higher activation in brain regions associated with evaluation of punishment threats (left anterior OFC), and in those areas of the brain associated with representation of emotional states (insula) in the punishment condition relative to the control condition. Impending punishment led to an increase of transferred money, whereas the actual transfer level was negatively correlated with Machiavellianism in the non-punishment condition (Spitzer et al., 2007). These findings suggest that high Machs seem to pretend increased cooperation when threatened with punishment, whereas they behaved selfishly when no punishment was to be expected.

Another fMRI study into the association of impulsivity and a reward or loss condition revealed that individuals with impulsive personality disorder showed significantly less activation in the prefrontal cortex during a reward task than normal controls (Völlm

et al., 2007). There was also a negative correlation between impulsivity scores on the Barrat Impulsivity Scale (BIS; Barratt, 1985) with responses in the OFC during both the reward and the loss task. ACC activation in the loss condition was found only in the patient group, which could indicate that financial loss requires greater information processing capacity in patients relative to controls.

Amodio et al. (2007) showed in an fMRI and event-related potential study that self-reported liberalists, relative to conservatives, revealed significantly stronger conflict-related ACC activity in a Go/No-Go task where response inhibition was required. The authors interpreted these findings as supportive of the assumption that individual differences in basic neurocognitive processes such as self-regulation, conflict-monitoring, and decision-making may cause different tendencies in political attitude.

Scheres and Sanfey (2006) examined how individual differences in basic psychological processes influence economic decision-making in the Ultimatum Game and the Dictator Game. They measured personality differences with two subscales of the Behavioral Activation Scale (BAS; Carver and White, 1994), i.e., reward responsiveness and drive. Higher scores on the two scales correlated with higher offers in the Ultimatum Game but with lower offers in the Dictator Game. Moreover, higher discrepancies between offers in the Ultimatum Game and the Dictator Game were associated with higher scores on the reward responsiveness scale. The authors argue that higher scores on reward responsiveness and drive lead to pursuing different strategies, on the one hand maximizing the likelihood of reward when depending on a game partner and on the other maximizing the sum when there's no possibility of being sanctioned.

Further individual differences can be predicted on the basis of sex differences in social behavior. Women, compared to men, are often more cooperative, particularly in same sex versus opposite sex scenarios, because women are selected to cooperate with other women, whereas men typically compete with one another (Trivers, 1971). In line with this assumption, studies have shown that cooperative behavior is supported by higher oxytocin levels in the brain (Kirsch et al., 2005; Kosfeld et al., 2005).

A recent study using a Public Goods Game among students that were made to believe to be competing with students of another university confirmed this prediction (Van Vugt et al., 2007). However, another study showed that average offers in the Ultimatum Game were unaffected by the proposer's gender, although as recipients, women were confronted with lower offers than men (Solnick, 2001; Eckel and Grossman, 2001). Solnick (2001) reported that unfair offers proposed by women were the most likely to be rejected, and that the highest rejection rate was found in women to women interactions. Conversely, Eckel and Grossman (2001) – using a repeated Ultimatum Game – reported high rejection rates for offers made by men, whereas those from women to women were the least likely to be rejected.

Another study hypothesized that sexual attractiveness could influence the acceptance of offers in the Ultimatum Game (van den Bergh and Dewitte, 2006). Indeed, higher acceptance rates of unfair offers were found when male players were confronted with photos of highly attractive women or lingerie models before playing the game, compared to a condition in which they watched photographs of landscapes prior to the game. Men with a low digit ratio (2D:4D, i.e., ratio of the length of the index finger to the length of the ring finger), which reflects an estimate of high prenatal testosterone exposure (Manning, 2002), were particularly inclined to accept unfair offers in the game when previously confronted with sexually arousing pictures.

In a similar vein, the actual level of testosterone seemed to have an influence on the acceptance of low offers in the Ultimatum Game. Men who rejected low offers were found to have a significantly higher testosterone level than men who accepted (Burnham, 2007). It is possible that low offers are more likely to cause aggression in men with higher levels of testosterone, which in turn influences their cognitive evaluation in a way that makes them perceive offers as more unfair, compared with individuals with lower testosterone levels.

## 6. Discussion

We have reviewed the literature on game theoretical approaches into the question of how humans solve problems associated with social exchange. Cooperation and sanctioning of uncooperative behavior is governed by cognitive and emotional mechanisms that evolved in humans in response to the need for mutual cooperation in complex social groups. We identified theory of mind, reward prediction, and appreciation of social norms as necessary (though perhaps not sufficient) mechanisms involved in social exchange.

Empirical evidence suggests that the evaluation of social exchange scenarios is maintained by an extended neural network. This neural network comprises frontal brain regions including the DLPFC, VMPFC, ACC and insula, which are activated when people evaluate fairness and trustworthiness of others in social interactions (Koenigs and Tranel, 2007; Sanfey et al., 2003; Knoch et al., 2006). Moreover, the ACC, the temporal pole, the temporo-parietal junction and the posterior superior temporal sulcus are activated in evaluating cooperation and defection (Rilling et al., 2004; Elliott et al., 2006), where frontal regions are more active in recognizing deceptive behavior compared with cooperation (Lissek et al., 2008).

If individuals detect that another intends to defect, they are often willing to invest some of their own resources to reinforce cooperation, even if they are not the primary target of defection, a behavior known as altruistic punishment. Punishing others to reinforce cooperation seems to be associated with the experience of reward. Imaging studies demonstrated that the thalamus and the caudate nucleus are activated during altruistic punishment (de Quervain et al., 2004) or during anticipation of others being punished for non-cooperative behavior (Knutson, 2004).

Until recently, most studies into behavior using game theoretical models have focused on group effects, while neglecting individual differences. A few studies suggest, however, that differences in temperament and character including traits such as Machiavellianism, impulsivity, reward responsiveness, and motivation influence an individual's behavior both in terms of willingness to cooperate and propensity to defect (Spitzer et al., 2007; Scheres and Sanfey, 2006). Moreover, recent research has revealed that individuals differ with regards to tolerance of others' non-cooperative behavior (Drebel et al., 2008). Interestingly, high-status individuals ("winners") are apparently more tolerant towards defection, whereas low-status individuals ("losers") more often tend to punish, perhaps because the latter are more inclined to display resentment when (subjectively) treated unfair.

In addition to temperament and character differences, several studies could show that men and women differ in behavior in social exchange situations (Eckel and Grossman, 2001; Solnick, 2001). Such differences can partly be explained by differences in sex hormones and bonding hormones (Kosfeld et al., 2005; Kirsch et al., 2005; Burnham, 2007). We believe that these differences at the proximate level reflect evolved sex differences in behavior. Intrasexual competition in males, for example, may increase the likelihood to defect in social exchange situations where the recipient is another male who is not part of the proposer's male-to-male alliance. Conversely, a male would probably not dare to

defect if this threatened his coalition with other males. A different situation would emerge if proposer and recipient are of opposite sex, especially if features such as sexual attractiveness are involved (van den Bergh and Dewitte, 2006). In future studies, these interindividual contingencies with regards to behavior in social exchange situations need to be carved out in more detail.

Moreover, we consider it a fruitful approach to examine the impact of gene–environment interaction on behavioral performance in social exchange scenarios. We predict that some genetic polymorphisms, especially if associated with adverse early rearing conditions, would produce measurable behavioral deviations from the average performance of individuals in game theoretical scenarios. For example, Caspi et al. (2002) could demonstrate that polymorphic variations of the serotonin transporter coding gene predispose to antisocial behavior only if carriers of these variants have experienced childhood maltreatment such as abuse or emotional neglect. In a way, such a gene–environment interaction could increase the likelihood for defection in social interaction, because individuals brought up under such conditions are perhaps "imprinted" to adopt opportunistic and exploitative interpersonal behavioral strategies (Belsky et al., 1991).

Another source of evidence of individual differences in social exchange could be advanced by studying cooperation and defection as well as punishing behavior in patients with neuropsychiatric disorders who are impaired in their abilities to evaluate contextual information (Agay et al., 2008; Koenigs and Tranel, 2007). This line of research should include conditions that bridge the invisible line between "normalcy" and "pathology" such as psychopathy (Hare, 2006), thus fostering a dimensional view on abnormal psychology.

Finally, further research ought to focus on within-subject variation in behavior including contextual factors such as mood, empathy, or current resource availability. Low mood or aggression, antipathy and resource scarcity almost certainly constrain one's willingness to cooperate with others, probably depending on the degree of genetic relatedness (kinship selection; Hamilton, 1964; Williams, 1966), as well as perception of the moral and social attitude of a game partner (Delgado et al., 2005). These factors could be experimentally manipulated by various means including rTMS, or emotional priming (Harlé and Sanfey, 2007). It is well conceivable that, depending on environmental contingencies, everybody can be biased in his or her decision to refuse cooperation or to punish others for their (perceived) misbehavior.

Understanding the biological underpinnings of individual differences in social exchange could be essential for theoretical conceptualizations of gene-culture co-evolution. This research may ultimately help to better understand the difficult question how different forms of altruism could have become selectively advantageous, and provide clues as to how we can promote and cultivate principles of cooperation, altruism, and self-control through education and legislation. After all, it is these kinds of social interactions we need to rely on to be able to address the problems of a global world whose population continues to grow but whose resources are limited (Hardin, 1968).

## References


Agay, N., Kron, S., Carmel, Z., Mendlovic, S., Levkovitz, Y., 2008. Ultimatum bargaining behavior of people affected by schizophrenia. Psychiat. Res. 157, 39–46.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.

Amodio, D.M., Jost, J.T., Master, S.L., Yee, C.M., 2007. Neurocognitive correlates of liberalism and conservatism. Nat. Neurosci. 10, 1246–1247.

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 211, 1390–1396.

Baron-Cohen, S., 1995. Mindblindness: An Essay on Autism and Theory of Mind. Bradford/MIT Press, Cambridge, MA.

Barratt, E.S., 1985. Impulsiveness subtraits: arousal and information processing. In: Spence, J.T., Izard, C.E. (Eds.), Motivation, Emotion, and Personality. Elsevier, North Holland, pp. 137–146.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R., 1997. Deciding advantageously before knowing the advantageous strategy. Science 275, 1293–1295.

Bechara, A., Damasio, H., Tranel, D., Anderson, S.W., 1998. Dissociation of working memory from decision making within the human prefrontal cortex. J. Neurosci. 18, 428–437.

Bechara, A., Damasio, H., Damasio, A.R., 2000a. Emotion, decision making and the orbitofrontal cortex. Cereb. Cortex 10, 295–307.

Bechara, A., Tranel, D., Damasio, H., 2000b. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain 123, 2189–2202.

Belsky, J., Steinberg, L., Draper, P., 1991. Childhood experience, interpersonal development, and reproductive strategy: and evolutionary theory of socialization. Child. Dev. 62, 647–670.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402, 179–181.

Boyd, R., 2006. The puzzle of human sociality. Science 314, 1555–1556.

Boyd, R., Richerson, P.J., 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethol. Sociobiol. 13, 171–195.

Boyd, R., Richerson, P.J., 2002. Group beneficial norms can spread rapidly in a structured population. J. Theor. Biol. 215, 287–296.

Brandt, H., Hauert, C., Sigmund, K., 2006. Punishing and abstaining for public goods. Proc. Natl. Acad. Sci. U.S.A. 103, 495–497.

Brosnan, S.F., de Waal, F.B.M., 2003. Monkeys reject unequal pay. Nature 425, 297–299.

Brothers, L., 1990. The neural basis of primate social communication. Motiv. Emot. 14, 81–91.

Brüne, M., Brüne-Cohrs, U., 2006. Theory of mind-evolution, ontogeny, brain mechanisms and psychopathology. Neurosci. Biobehav. Rev. 30, 437–455.

Burnham, T.C., 2007. High-testosterone men reject low ultimatum game offers. Proc. Biol. Sci. 274, 2327–2330.

Camerer, C.F., 2003a. Behavioural studies of strategic thinking in games. Trends Cogn. Sci. 7, 225–231.

Camerer, C.F., 2003b. Psychology and economics. Strategizing in the brain. Science 300, 1673–1675.

Camerer, C.F., 2003c. Behavioral Game Theory: Experiments in Strategic Interaction. University Press, Princeton.

Carver, C., White, T., 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J. Pers. Soc. Psychol. 67, 319–333.

Cason, T.N., Saijo, T., Yamato, T., 2002. Voluntary participation and spite in Public Good provision experiments: an international comparison. Exp. Eco. 5, 133–153.

Caspi, A., McClay, J., Moffitt, T.E., Mill, J., Martin, J., Craig, I.W., Taylor, A., Poulton, R., 2002. Role of genotype in the cycle of violence in maltreated children. Science 297, 851–854.

Cesarini, D., Dawes, C.T., Fowler, J.H., Johannesson, M., Lichtenstein, P., Wallace, B., 2008. Heritability of cooperative behavior in the trust game. Proc. Natl. Acad. Sci. U.S.A. 105, 3721–3726.

Charness, G., Gneezy, U., 2003. What's in a name? Anonymity and social distance in Dictator and Ultimatum Games. University of California, Santa Barbara, Department of Economics, working paper series.

Cosmides, L., 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition 31, 187–276.

Cutting, J., Murphy, D., 1990. Impaired ability of schizophrenics, relative to manics or depressives, to appreciate social knowledge about their culture. Br. J. Psychiatry 157, 355–358.

Damasio, A., 1994. Descartes Error: Emotion, Reason, and the Human Brain. Avon Books.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press.

de Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E., 2004. The neural basis of altruistic punishment. Science 305, 1254–1258.

Delgado, M.R., Frank, R.H., Phelps, E.A., 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. 8, 1611–1618.

Drebel, A., Rand, D.G., Fudenberg, D., Nowak, M.A., 2008. Winners don't punish. Nature 452, 348–351.

Dunbar, R., 2003. The social brain: Mind, language, and society in evolutionary perspective. Annu. Rev. Anthropol. 32, 163–181.

Eckel, C., Grossman, P., 2001. Chivalry and solidarity in ultimatum games. Econ. Inq. 39, 171–188.

Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., Deakin, J.W., 2006. Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. Soc. Neurosci. 1, 385–395.

Falk, A., Fischbacher, U., 2000. A theory of reciprocity. Institute for Empirical Economic Research. University of Zurich, working paper No. 6.

Fehr, E., Fischbacher, U., 2004a. Social norms and human cooperation. Trends Cogn. Sci. 8, 185–190.

Fehr, E., Fischbacher, U., 2004b. Third-party punishment and social norms. Evol. Hum. Behav. 25, 63–87.

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140.

Fehr, E., Rockenbach, B., 2004. Human altruism: economic, neural, and evolutionary perspectives. Curr. Opin. Neurobiol. 14, 784–790.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114, 817–868.
Fellows, L.K., Farah, M.J., 2005. Dissociable elements of human foresight: a role for the ventromedial frontal lobes in framing the future, but not in discounting future rewards. Neuropsychologia 43, 1214–1221.
Fiske, A.P., 1992. The four elementary forms of sociality: framework for a unified theory of social relations. Psychol. Rev. 99, 689–723.
Fodor, J., 1983. The Modularity of Mind. MIT Press, Cambridge, MA.
Frith, C.D., 2004. Schizophrenia and theory of mind. Psychol. Med. 34, 385–390.
Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 358, 459–473.
Fuster, J.M., 1997. The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe. Lippincott Williams, Wilkins.
Giedd, J.N., 2004. Structural magnetic resonance imaging of the adolescent brain. Ann. N. Y. Acad. Sci. 1021, 77–85.
Gigerenzer, G., Hug, K., 1992. Domain-specific reasoning: social contracts, cheating, and perspective change. Cognition 43, 127–171.
Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., Hodges, J.R., 2002. Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. Brain 125, 752–764.
Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. 3, 367–388.
Hamilton, W.D., 1964. The genetical evolution of social behaviour. I. J. Theor. Biol. 7, 1–16.
Hammerstein, P., Leimar, O., 2006. Cooperating for direct fitness benefits. J. Evol. Biol. 19, 1400–1402 discussion 1426–1436.
Hammock, E.A.D., Young, L.J., 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308, 1630–1634.
Hardin, G., 1968. The tragedy of the commons. Science 162, 1243–1248.
Hare, R.D., 2006. Psychopathy: a clinical and forensic overview. Psychiatr. Clin. North Am. 29, 709–724.
Harlé, K.M., Sanfey, A.G., 2007. Incidental sadness biases social economic decisions in the Ultimatum Game. Emotion 7, 876–881.
Hauert, C., Holmes, M., Doebeli, M., 2006. Evolutionary games and population dynamics: maintenance of cooperation in public goods games. Proc. R. Soc. B 273, 2565–2570.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Jensen, K., Call, J., Tomasello, M., 2007. Chimpanzees are rational maximizers in an ultimatum game. Science 318, 107–109.
Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., Gruppe, H., Mattay, V.S., Gallhofer, B., Meyer-Lindenberg, A., 2005. Oxytocin modulates neural circuitry for social cognition and fear in humans. J. Neurosci. 25, 11489–11493.
Knafo, A., Israel, S., Darvasi, A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., Feldman, E., Lerer, E., Laiba, E., Raz, Y., Nemanov, L., Gritsenko, I., Dina, C., Agam, G., Dean, B., Bornstein, G., Ebstein, R.P., 2007. Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. Genes Brain Behav. 7, 266–275.
Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E., 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science 314, 829–832.
Knutson, B., 2004. Behavior. Sweet revenge? Science 305, 1246–1247.
Knutson, B., Cooper, J.C., 2005. Functional magnetic resonance imaging of reward prediction. Curr. Opin. Neurol. 18, 411–417.
Knutson, B., Adams, C.M., Fong, G.W., Hommer, D., 2001. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J. Neurosci. 21, RC159.
Koenigs, M., Tranel, D., 2007. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. J. Neurosci. 27, 951–956.
Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E., 2005. Oxytocin increases trust in humans. Nature 435, 673–676.
Kringelbach, M.L., 2005. The human orbitofrontal cortex: linking reward to hedonic experience. Nat. Rev. Neurosci. 6, 691–702.
Ledyard, J., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–199.
Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., Juckel, G., Brüne, M., 2008. Cooperation and deception recruit different subsets of the theory-of-mind network. PLoS ONE 3, e2023.
Lough, S., Kipps, C.M., Treise, C., Watson, P., Blair, J.R., Hodges, J.R., 2006. Social reasoning, emotion and empathy in frontotemporal dementia. Neuropsychologia 44, 950–958.
Manning, J.T., 2002. Digit Ratio: A Pointer to Fertility, Behaviour, and Health. Rutgers University Press, New Brunswick, NJ.
Mazza, M., Risio, A.D., Tozzini, C., Roncone, R., Casacchia, M., 2003. Machiavellianism and theory of mind in people affected by schizophrenia. Brain Cogn. 51, 262–269.
McNamara, P., Durso, R., Harris, E., 2007. "Machiavellianism" and frontal dysfunction: evidence from Parkinson's disease. Cognit. Neuropsychiatry 12, 285–300.
Mealey, L., 1995. The sociobiology of sociopathy: an integrated evolutionary model. Behav. Brain Sci. 18, 523–599.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.
Nowak, M.A., 2006. Five rules for the evolution of cooperation. Science 314, 1560–1563.
Nowak, M.A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.
Okada, A., Riedl, A., 1999. When culture does not matter: experimental evidence from coalition formation Ultimatum Games in Austria and Japan. University of Amsterdam (CREED), working paper.
Paus, T., 2001. Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat. Rev. Neurosci. 2, 417–424.
Pfefferbaum, A., Mathalon, D.H., Sullivan, E.V., Rawles, J.M., Zipursky, R.B., Lim, K.O., 1994. A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. Arch. Neurol. 51, 874–887.
Rakic, P., Bourgeois, J.P., Eckenhoff, M.F., Zecevic, N., Goldman-Rakic, P.S., 1986. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. Science 232, 232–235.
Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. NeuroImage 22, 1694–1703.
Sally, D., Hill, E., 2006. The development of interpersonal strategy: autism, theory-of-mind, cooperation and fairness. J. Econ. Psychol. 27, 73–97.
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the Ultimatum Game. Science 300, 1755–1758.
Scheres, A., Sanfey, A.G., 2006. Individual differences in decision making: drive and reward responsiveness affect strategic bargaining in economic games. Behav. Brain Funct. 2, 35.
Segalowitz, S.J., Davies, P.L., 2004. Charting the maturation of the frontal lobe: an electrophysiological strategy. Brain Cogn. 55, 116–133.
Seymour, B., Singer, T., Dolan, R., 2007. The neurobiology of punishment. Nat. Rev. Neurosci. 8, 300–311.
Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D., 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439, 466–469.
Solnick, S., 2001. Gender differences in the ultimatum game. Econ. Inq. 39, 189–200.
Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., Fehr, E., 2007. The neural signature of social norm compliance. Neuron 56, 185–196.
Sullivan, R.J., Allen, J.S., 1999. Social deficits associated with schizophrenia defined in terms of interpersonal Machiavellianism. Acta Psychiatr. Scand. 99, 148–154.
Tooby, J., Cosmides, L., 1992. The psychological foundation of culture. In: Barkow, J., Cosmides, L., Tooby, J. (Eds.), The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York.
Tooby, J., Devore, I., 1987. The reconstruction of hominid behavioral evolution through strategic modelling. In: Kinzey, W.G. (Ed.), The Evolution of Human Behavior: Primate Models. Suny Press, Albany, NY, pp. 183–237.
Trivers, R.L., 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35–57.
Troisi, A., 2005. The concept of alternative strategies and its relevance to psychiatry and clinical psychology. Neurosci. Biobehav. Rev. 29, 159–168.
den Bergh, B.V., Dewitte, S., 2006. Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. Proc. Biol. Sci. 273, 2091–2095.
Van Vugt, M., Cremer, D.D., Janssen, D.P., 2007. Gender differences in cooperation and competition: the male-warrior hypothesis. Psychol. Sci. 18, 19–23.
van't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A., 2005. Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. Neuroreport 16, 1849–1852.
van't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A., 2006. Affective state and decision-making in the Ultimatum Game. Exp. Brain Res. 169, 564–568.
Völlm, B., Richardson, P., McKie, S., Elliott, R., Dolan, M., Deakin, B., 2007. Neuronal correlates of reward and loss in Cluster B personality disorders: a functional magnetic resonance imaging study. Psychiatry Res. 156, 151–167.
Wallace, B., Cesarini, D., Lichtenstein, P., Johannesson, M., 2007. Heritability of ultimatum game responder behavior. Proc. Natl. Acad. Sci. U.S.A. 104, 15631–15634.
Warneken, F., Tomasello, M., 2006. Altruistic helping in human infants and young chimpanzees. Science 311, 1301–1303.
Warneken, F., Hare, B., Melis, A.P., Hanus, D., Tomasello, M., 2007. Spontaneous altruism by chimpanzees and young children. PLoS Biol. 5, e184.
Wason, P.C., Shapiro, D., 1966. Reasoning. In: Foss, B.M. (Ed.), New Horizons in Psychology. Penguin, Harmondsworth.
Williams, G.C., 1966. Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thoughts. Princeton University Press, Princeton, NJ.
Williams, J.H., Whiten, A., Suddendorf, T., Perrett, D.I., 2001. Imitation, mirror neurons and autism. Neurosci. Biobehav. Rev. 25, 287–295.
Wilson, D.R., 2006. The evolutionary neuroscience of human reciprocal sociality: a basic outline for economists. J. Socio-Econ. 35, 626–633.
Zeeman, E.C., 1980. Population Dynamics from Game Theory. Global Theory of Dynamical Systems. Springer, Berlin/Heidelberg.