The neurobiology of social decision-making
James K Rilling1,2,3,4, Brooks King-Casas5,6,7 and Alan G Sanfey8

Humans live in highly complex social environments and some of our most important decisions are made in the context of social interactions. Research that probes the neural basis of decision-making in the context of social interactions combines behavioral paradigms from game theory with a variety of methods from neuroscience. The neural correlates of decision making in reciprocal exchange and bargaining games have been probed with functional neuroimaging, transcranial magnetic stimulation, and pharmacological manipulations. These studies have begun to elucidate a set of brain regions and neurotransmitter systems involved in decision-making in social interactions.

Reciprocal exchange
One specific focus of game theory is to model reciprocal exchange, in which an individual provides something of value to a social partner with the expectation that the recipient will reciprocate in the future. Although greed and fear of exploitation threaten the stability of reciprocal exchange, society as a whole is more productive when reciprocity is thriving [2]. Typically, reciprocal exchange is studied via the Trust, and closely related Prisoner’s Dilemma, games. In the Trust Game (TG), a player (the Investor) must decide how much of an endowment to invest with a partner (the Trustee). Once transferred, this money is multiplied by some factor (usually tripled or quadrupled), with the Trustee then having the opportunity to return all, some, or none of the amount back to the Investor. If the Trustee honors trust, and returns money to the Investor, both players can end up with a higher monetary payoff than was originally obtained. However, if the Trustee abuses trust and keeps the entire amount, the Investor ends up with a loss. The well-studied Prisoner’s Dilemma (PD) game is similar to the TG, except that both players simultaneously choose whether or not to trust each other without knowledge of their partner’s choice.

There is considerable evidence to suggest that mesencephalic dopamine projections to the caudate nucleus are involved in decision-making in the context of reciprocal exchange, as in the TG and PDG. Single-cell recording studies in monkeys have demonstrated that midbrain...
dopamine neurons track reward prediction errors [3]. Unexpected rewards increase the firing rate of midbrain dopamine cells, whereas the omission of expected rewards decreases their firing rate. These reward prediction errors provide a signal with which animals and humans can learn about the reward value of stimuli in their environment, thereby shaping decision-making [4]. Midbrain dopamine cells project to both ventral and dorsal striatum, including the caudate nucleus, and activation in the human caudate nucleus is modulated as a function of trial and error learning with feedback, suggesting that the caudate nucleus may allow an organism to learn contingencies between its own responses and either rewarding or punishing outcomes [5]. Several neuroimaging studies have demonstrated that the human caudate tracks a social partner’s decision to reciprocate or not reciprocate cooperation in the TG or PDG [6**,7**,8**,9] (Figure 1). Specifically, reciprocated cooperation activates the caudate nucleus, and unreciprocated cooperation can deactivate this area, in line with the reward prediction error described above. Additionally, caudate activation is associated with increased cooperation or reciprocity in subsequent rounds [7**,8**]. These findings suggest that the caudate may register social prediction errors that then guide decisions about reciprocity. Interestingly, these prediction error signals from partner feedback can be blunted or even absent when people base their decisions on factors other than those gleaned through direct experience, such as the prior moral reputations of social partners [6**]. That is, moral reputations lead to a discounting of feedback information from social partners, demonstrating both top–down and bottom–up influences on the neural basis of social cooperation.

In the PDG, cooperation that is not reciprocated is associated with robust activation of the anterior insula in the cooperator [10**]. Non-reciprocators, along with others who accept benefits without paying expected costs, are known as ‘free-riders’. Many experiments have demonstrated a widespread aversion to ‘free riders’, as evidenced by people’s willingness to punish them at a
personal cost [11,12] and, more recently, by activation in brain reward areas when people successfully punish free riders or observe them receiving punishment [13**,14]. Activation in the anterior insula in response to unreciprocated cooperation may be a neural correlate of the aversive response to free riding, as the anterior insula is responsive to both painful physical stimuli, as well as a variety of negative social interactions, including social exclusion [15], unfair treatment [16**], and observing a loved one in pain [17]. The anterior insula is involved in mapping physiological states of the body, including pain, touch and visceral sensations of autonomic arousal [18–21]. The right anterior insula, in particular, is thought to be a cortical station for interception that may play a role in decision-making by instantiating valenced subjective feeling states [22]. Finally, recent fMRI data implicate right anterior insula in aversive conditioning [23], as well as risk prediction and risk prediction errors [24]. Collectively, these findings suggest that the anterior insula may mark negative social interactions as risky and/or aversive so that individuals learn to avoid such interactions in the future. Indeed, functional connectivity between the anterior insula and the lateral orbitofrontal cortex (OFC) predicts subsequent defection by a player in future interactions with the same non-reciprocating partner. This finding is consistent with evidence that lateral OFC is involved in the evaluation of punishing stimuli that may lead to behavioral changes [25].

Of course, reciprocal altruism cannot be initiated or sustained without trust. How does the human brain allow us to trust others, overcoming our fear of betrayal and aversion to risk and uncertainty? In a TG, intranasal oxytocin (OT) infusion was shown to increase initial monetary transfers by investors [26**]. Recent research suggests that neuropeptides like OT cross the blood-brain barrier after intranasal administration, implying that these effects are due to OT’s action in the brain [27]. Intranasal OT has recently been combined with fMRI to assess its impact on the neural response to threatening faces and scenes [28*,29], and on the neural response during decision-making in a trust game [30**]. In both studies, compared with placebo, OT was associated with reduced amygdala activation, and in the trust game, with increased behavioral expressions of trust. These results are consistent with evidence that OT reduces stress and anxiety [31,32], and that OT decreases activity of central amygdala neurons in the rat [33]. This study suggests one potential mechanism by which OT could increase trust; by reducing anxiety about potential non-reciprocation.

Finally, recent studies have examined the neural correlates of altruistic behavior outside of reciprocal exchange, in tasks where players must decide whether to donate money to charitable organizations. In one study, the decision to voluntarily donate real money to actual charitable organizations was associated with activation in the ventral striatum [34*]. In another, subjects showing stronger ventral striatum activation to mandatory charity donations were more likely to voluntarily give to the charity when given a choice, whereas subjects showing stronger ventral striatum activation to payments to themselves were less likely to voluntarily donate to the charity [35*].

### Bargaining

Bargaining games are another common focus of game theory, with the family of Dictator and Ultimatum games often used to examine responses to equality and inequality. In the Dictator Game (DG), one player (the Proposer) decides how much of an endowment to award to the second player (the Responder). Allocations in this game measure pure altruism, in that the Proposer (usually) sacrifices some personal gain to share the endowment with their partner. The Ultimatum Game (UG) is a variant which examines strategic thinking in the context of two-player bargaining. In the UG, the Proposer and Responder are also asked to divide a sum of money, with the Proposer specifying how this sum should be divided between the two. However, in this case the Responder has the option of accepting or rejecting the offer. If the offer is accepted, the sum is divided as proposed. However, if it is rejected, neither player receives anything. In either event the game is over, that is, there are no subsequent rounds in which to reach agreement. The decision to reject an unfair offer is considered a form of altruistic punishment because the Responder chooses to receive no money rather than the amount offered by the Proposer, presumably to punish the Proposer for making a miserly offer.

Akin to reciprocated cooperation in trust games, receiving a fair offer in the UG is associated with activation in brain regions implicated in reward processing, including the ventral striatum, even after controlling for monetary payoff [36*]. Receiving an unfair UG offer, on the other hand, is associated with both negative emotions and activation of the anterior insula [16**] (Figure 2). The anterior insula response scales to the magnitude of unfairness, and is also stronger when the subject is playing with another human than when engaged with a computer partner. Additionally, anterior insula activation predicts the player’s decision to either accept or reject the unfair offer, with rejections associated with significantly higher activation than acceptances [16**,36*] (Figure 2). Separate measures of emotional arousal support the hypothesis that feedback from the viscera, registered in the anterior insula, influence decision-making in the UG game. An UG study measuring skin-conductance responses, used as an autonomic index of affective state, found higher skin conductance activity for unfair offers, and as with insular activation, this measure discriminated between acceptances and rejections of these offers [37].
Unfair UG offers are also associated with activation in dorsolateral prefrontal cortex (DLPFC) [16] (Figure 2). DLPFC has long been recognized as critically involved in cognitive control, including cognitive control over emotions [38]. Recent studies suggest its specific involvement in overriding prepotent emotional biases, such as when delaying gratification [39] or making utilitarian decisions in the context of moral dilemmas [40]. In the UG, subjects are more likely to accept unfair offers when DLPFC activation exceeds anterior insula activation [16]. This has led to the hypothesis that Ultimatum Game decisions, and difficult social decisions more generally, may involve competition between emotional processing and higher-level controlled or deliberative processing that bias decision-making in opposite ways [41]. Emotional processes are driven by subcortical, limbic and paralimbic structures, whereas the deliberative processes rely on anterior and dorsolateral regions of prefrontal cortex, as well as areas of posterior parietal cortex. Further evidence in support of a role for DLPFC in social decision-making is provided by studies showing that application of repetitive transcranial magnetic stimulation (rTMS) to DLPFC influences decision-making in the Ultimatum Game [42,43].

Both ethnographic [44] and experimental evidence have shown that some people will only uphold fairness norms under threat of punishment [12]. Thus, sensitivity to the threat of punishment is an important motive for norm-abiding behavior in some people. In a recent fMRI study [45], subjects were imaged while playing two different games. In one game, which resembles a DG, player A received a monetary endowment that they could distribute freely between themselves and another player (player B). In this game, player B is a passive recipient of player A’s monetary transfer. In another game, player A again distributed a monetary endowment between themselves and player B, however player B could now choose to pay money to financially punish player A after having been informed of player A’s decision (for example, if player A proposed an unfair distribution). Player A transferred substantially more money to player B in the punishment compared with the non-punishment condition. Those subjects who showed the largest change in monetary transfer from the non-punishment to the punishment condition also showed the greatest increase in activation of the lateral orbitofrontal cortex across conditions. Lateral OFC motivates behavioral changes in response to punishing stimuli [25]. This study suggests that lateral OFC motivates norm-abiding decision-making in response to the threat of punishment.

**Psychopathology**

While the methodological and theoretical approaches of decision neuroscience have begun to uncover the neural computations underlying multiagent interactions and complex social phenomena [46], they also provide an...
opportunity to identify aberrant neural substrates underlying social pathologies. Much recent interest in developmental and adult psychiatry has focused on social features of mental illnesses ranging from autism and schizophrenia to borderline and antisocial personality disorders. The social abnormalities associated with these and other psychiatric illnesses confer tremendous burden on the afflicted, and initial studies of such pathologies have provided detailed accounts of biased processing of social information (e.g. facial affect recognition and eye gaze tasks) and abnormalities in processing social fictions (e.g. ‘theory-of-mind’ tasks).

More recently the approaches detailed above have been used to study social pathologies in active social interactions. Building on previous work which identified neural responses along the bank of cingulate cortex corresponding to decisions made either by oneself or one’s partner [47], a recent study used an active economic exchange to examine agent-specific neural activations in a cohort of autism spectrum disordered adolescents [48]. This study identified a deficit in middle cingulate cortex during the submission of subjects’ own decisions that scaled parametrically with social deficits in autism. In another study using a similar task, the iterated PDG, subjects scoring higher on a measure of psychopathic personality exhibited decreased activation in OFC when choosing to cooperate and decreased activation in DLPFC/rACC when choosing to defect, a pattern of activation hypothesized to reflect prepotent emotional biases for defecting [49].

**Conclusion**

Many of our most important decisions are made in the context of social interactions. Recently, cognitive neuroscientists have begun to investigate the neural correlates of social decisions using tasks derived from a branch of experimental economics known as game theory. The caudate nucleus, and likely the mesolimbic dopamine system that projects to it, registers social prediction errors that guide decisions about reciprocity. Visceral feedback in response to negative social interactions, represented in the anterior insula, influences decision-making in both the PD and UG games. However, these bottom–up biases can sometimes be overridden by top–down cognitive control mediated by the prefrontal cortex. Decisions to trust others are enhanced by oxytocin, which may dampen the fear of betrayal by suppressing amygdala activity. Activation in the ventral striatum seems to motivate decisions to voluntarily donate money to charitable organizations, whereas activation in lateral OFC motivates norm-abiding decision-making in response to the threat of punishment. Finally, social behavioral disorders are associated with abnormalities in the neural substrates supporting social decision-making. Future studies that combine game theory and cognitive neuroscience will continue to advance our understanding of the neural bases of social decision-making (Figure 3).
of special interest
• of outstanding interest


7. Shows that reward prediction error signals during reciprocal exchange can be blunted by prior reputation of a social partner.


11. Used fMRI to image the brains of two interacting partners simultaneously during an iterated trust game. Activation of the caudate nucleus was related both to the reciprocity revealed by a partner’s decision and also predicted one’s own behavioral response to the social partner.


17. Shows that receiving an unfair offer in the ultimatum game was associated with activation in the anterior insula and dorsolateral prefrontal cortex. Activation in the anterior insula scaled to the degree of unfairness of the offer and was stronger for unfair offers from putative human vs. computer partners. When activation in the anterior insula was stronger than activation in DLPCF, subjects were more likely to reject than accept unfair offers, whereas subjects were more likely to accept unfair offers when DLPCF activation exceeded anterior insula activation.


20. Shows that receiving an unfair offer in the ultimatum game was associated with activation in the anterior insula and dorsolateral prefrontal cortex. Activation in the anterior insula scaled to the degree of unfairness of the offer and was stronger for unfair offers from putative human vs. computer partners. When activation in the anterior insula was stronger than activation in DLPCF, subjects were more likely to reject than accept unfair offers, whereas subjects were more likely to accept unfair offers when DLPCF activation exceeded anterior insula activation.


28. Compared with placebo, OT decreased activation in the amygdala to fearful/threatening faces and scenes.


32. McCarthy MM, McDonald CH, Brooks PJ, Goldman D: An anxiolytic action of oxytocin is enhanced by estrogen in the mouse. *Physiol Behav* 1996, **60**:1209-1215.


