The neurobiology of rewards and values in social decision making

RUFF, Christian C., FEHR, Ernst

Abstract

How does our brain choose the best course of action? Choices between material goods are thought to be steered by neural value signals that encode the rewarding properties of the choice options. Social decisions, by contrast, are traditionally thought to rely on neural representations of the self and others. However, recent studies show that many types of social decisions may also involve neural value computations. This suggests a unified mechanism for motivational control of behaviour that may incorporate both social and non-social factors. In this Review, we outline a theoretical framework that may help to identify possible overlaps and differences between the neural processes that guide social and non-social decision making.

Reference


DOI: 10.1038/nrn3776
The neurobiology of rewards and values in social decision making

Christian C. Ruff and Ernst Fehr

Abstract | How does our brain choose the best course of action? Choices between material goods are thought to be steered by neural value signals that encode the rewarding properties of the choice options. Social decisions, by contrast, are traditionally thought to rely on neural representations of the self and others. However, recent studies show that many types of social decisions may also involve neural value computations. This suggests a unified mechanism for motivational control of behaviour that may incorporate both social and non-social factors. In this Review, we outline a theoretical framework that may help to identify possible overlaps and differences between the neural processes that guide social and non-social decision making.

Of all the decisions in your life, which were the most important? Chances are high that many of them had a social component, be it when deciding about others (for example, whether to marry someone), for others (for example, which school to send your child to) or jointly with others (for example, whether to buy a house with your partner). Pathological changes specifically in the social aspects of decision making are pervasive in many psychiatric1 and neurological2 disorders and can have devastating consequences for individual and collective welfare. A good understanding of the neural basis of social decision making is thus an important aim of scientists in many disciplines, from basic neuroscience to medicine, psychology and economics.

The neural mechanisms underlying social and non-social types of choices have traditionally been examined from different theoretical angles. In studies of non-social choices (for example, purchasing decisions), researchers have often focused on neural value computations associated with the rewarding properties of the choice options3,4. On the basis of this research, several brain structures have been proposed as elements of a brain valuation system (BOX 1). The neural value signals in these structures are thought to represent a ‘common currency’ for assessing the motivational relevance of all possible stimuli or actions, which is essential for guiding decision making across varying contexts5,6. By contrast, researchers studying social decisions have traditionally paid less attention to motivational processes, such as reward and value coding; instead, they have focused on identifying neurocognitive processes that may have uniquely evolved to guide social behaviour7. This emphasis on ‘social-specific cognition’ has identified, for example, that different visual areas seem to be functionally specialized for the perception of faces or bodies8 and that the dorsomedial prefrontal cortex (dmPFC) and the temporoparietal junction (TPJ) may be specifically involved in representing the intentions, emotions or actions of other people9–11. The idea of ‘social-specific cognition’ thus implies that the control of social and non-social behaviour should fundamentally differ in terms of neural architecture and information processing demands12.

However, this view is challenged by recent findings that social decisions may also draw on value-related brain processes that strongly resemble those seen during non-social decisions. This apparent overlap in neural representations of motivational relevance for social and non-social choices is often taken to implicitly extend the ‘common currency’ idea to decisions based on social factors. This proposal — that choices in both social and non-social situations are steered by identical neural value computations — therefore questions whether the motivational control of social behaviour requires dedicated neural processes.

In this Review, we propose a framework for studying the possible neurobiological links in the motivational control of social and non-social decision making. We first outline two competing schemas of the value signals that drive both types of decisions; these schemas represent different conceptual combinations of the ideas of a ‘common currency’ and of ‘social-specific cognition’. We then discuss the congruency of these two schemas with empirical results. A major challenge in this context is the sheer diversity of social situations and neural effects that are being investigated under the umbrella terms...
of ‘social decisions’, ‘rewards’ and ‘values’. We therefore group the discussion of empirical findings about social decision making by the type of social situation and the neural value process under consideration. We finish the Review with suggestions for future work that may clarify how neural processes devoted to value coding and reward coding can account for both social and non-social factors in influencing choices.

Social valuation: a conceptual framework

There are two schemas that illustrate how neural value representations underlying social decisions could relate to those driving non-social choices (FIG. 1). These two schemas have been implicitly alluded to in the literature but have rarely been explicitly compared. Although they are not mutually exclusive, the two schemas make contradicting predictions.

The ‘extended common currency schema’ (FIG. 1a) assumes that identical neural processes assign motivational relevance to social and non-social factors. However, these value-related processes may incorporate input from different brain areas that compute cognitive information that is relevant for the social or non-social choice. This schema therefore predicts that both types of choices induce similar activity in the brain value system but that these shared value representations change their functional connectivity with other, domain-specific brain regions in social versus non-social choice contexts.

By contrast, the ‘social-valuation-specific schema’ (FIG. 1b) proposes that social rewards and values are processed in a dedicated neural circuitry that evolved specifically to deal with interactions with others. In this schema, the neurons in this circuitry derive values based on social information by using types of neural value computations similar to those used by the neurons that encode non-social value representations. The two types of neurons may be located in close spatial proximity (perhaps owing to the need for mutual interactions) or in different brain regions. This schema therefore predicts that social and non-social neural value signals are implemented in different spatial patterns of brain activity but nevertheless follow similar computational principles.

Classes of social valuation. Before we discuss empirical support for these two schemas, it is important to note that studies of social decision making address a wide range of contexts and choices that — at first glance — mainly seem to be linked by the fact that more than one
Discounting is often thought to represent a general economic principle. However, the brain processes social and non-social factors differently. Several brain regions are specialized for each type of information. These areas interact with other brain regions to determine the value of social and non-social factors during decision making.

**Neural valuation of other people**

**Experiencing rewards or punishments from specific others.** Humans frequently interact with others, and many aspects of these direct interactions may be rewarding (or punishing). A basic example is the mere perception of an attractive person, which can elicit neural activity in several areas of the brain's reward circuitry, including the orbitofrontal cortex (OFC) and amygdala. Similar effects have been found when participants viewed erotic photos or faces with positive affect. Although this neural activity may reflect low-level visual differences between different pictures, several studies have linked neural activations to some measures of behavioural preference. For example, one study found that the increase in the blood-oxygen-level-dependent (BOLD) signal in the OFC during viewing more versus less attractive faces predicted the amount of money participants were willing to spend to view the more attractive faces. Moreover, choices between viewing different attractive faces or erotic photos follow similar economic principles (for example, delay discounting) as those established for choices between non-social rewards. Thus, differences in the task and stimulus, social and monetary reward outcomes led to comparable BOLD activations in the ventral striatum. Similar activations have also been found in the ventral striatum and ventromedial PFC when participants read that other depicted individuals liked or wanted to meet them. Finally, ventral striatum activity has been observed during outcomes that indicate changes in the social status of the participants in the social interactions. Thus, neural experienced values during social interactions can encode both immediate hedonic aspects and longer-lasting social consequences of basic social gestures (for example, changes in reputation or status).

Conceptually related findings have been reported for social punishments. Social exclusion, as simulated in the context of a game paradigm, led to BOLD signal increases in the anterior cingulate cortex (ACC). These rejection-related BOLD responses have been proposed to reflect processes that are related to (socially induced) pain, as they correlated with social stress-related inflammatory responses, were reduced by administration of...
Empathy with someone’s regions involved in social cognition, such as the TPJ of functional connectivity of these valuation areas with encoding non-social rewards. Moreover, the rewarding and being excluded or rejected) have been linked to neu
example, perceiving others, receiving interest or approval the vmPFC receiving feedback indicating a match in interest activated the ACC, whereas feedback that a depicted partner had rejected the par
responses to physical pain a conventional painkiller and overlapped with BOLD responses to physical pain. Moreover, in a direct comparison of punishments and rewards, receiving social feedback that a depicted partner had rejected the participant’s romantic interest activated the ACC, whereas receiving feedback indicating a match in interest activated the vmPFC.

Thus, several aspects of direct social interactions (for example, perceiving others, receiving interest or approval and being excluded or rejected) have been linked to neural activity in brain structures that are also involved in encoding non-social rewards. Moreover, the rewarding properties of social stimuli can also relate to the strength of functional connectivity of these valuation areas with regions involved in social cognition, such as the TPJ. These results therefore provide support for the extended common currency schema. However, only a few studies have directly compared activity patterns elicited by social outcomes with those due to non-social financial rewards. Although these studies have found areas of overlapping activity, they have also identified activity differences. For example, one study indicated that financial rewards and erotic pictures elicit activity in distinct portions of the OFC, whereas other studies showed that either type of reward uniquely activates voxels in the ventral striatum or the vmPFC. These results may therefore also be consistent with predictions from the social-valuation-specific schema. As few studies have parametrically varied the degree of reward and measured the hedonic and behavioural impact of the rewards, it is difficult to determine whether slightly different activation patterns for social versus non-social rewards and punishments indeed reflect fundamentally different neural computations or simply show differences in the reward value of the different types of feedback.

Decisions involving specific other individuals. All of the studies discussed above investigated neural expected values at the time of the (rewarding or non-rewarding) outcome. Different types of neural value representation (namely, decision values and anticipated values; see BOX 1) might be needed during actual choices that require participants to rely on subjective value estimates in the absence of any explicit reward. Several of the studies mentioned above have provided evidence for such value representations. For example, value-related activity can already be observed in the ventral striatum and vmPFC during the mere anticipation of social gestures or face stimuli with rewarding or punishing properties. Interestingly, studies have shown that the presence of observers during decision making — and thus the possibility of future approval and reputation gains — changed value-related responses in the ventral striatum during risky choices and altruistic choices. The degree of change in ventral striatum activity during social observation correlated with changes in overt behaviour, suggesting that the mere anticipation

<table>
<thead>
<tr>
<th>Decision-making stage (value type)</th>
<th>Target (and reference frame) of valuation</th>
<th>Value of other people (for oneself)</th>
<th>Value of other people’s experiences (for them)</th>
<th>Value of social constellation (for normative social principle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice (decision value and anticipated value)</td>
<td>Deciding whether to marry someone</td>
<td>Choosing a school for your child</td>
<td>Altruistic punishment of norm violations</td>
<td></td>
</tr>
<tr>
<td>Outcome (experienced value)</td>
<td>Being smiled at</td>
<td>Empathy with someone’s pain</td>
<td>Enjoying fair distributions</td>
<td></td>
</tr>
<tr>
<td>Learning (anticipated value)</td>
<td>Learning about someone’s trustworthiness</td>
<td>Learning about someone else’s preferences</td>
<td>Changing your opinion to increase social conformity</td>
<td></td>
</tr>
</tbody>
</table>

a conventional painkiller and overlapped with BOLD responses to physical pain. Moreover, in a direct comparison of punishments and rewards, receiving social feedback that a depicted partner had rejected the participant’s romantic interest activated the ACC, whereas receiving feedback indicating a match in interest activated the vmPFC.

Thus, several aspects of direct social interactions (for example, perceiving others, receiving interest or approval and being excluded or rejected) have been linked to neural activity in brain structures that are also involved in encoding non-social rewards. Moreover, the rewarding properties of social stimuli can also relate to the strength of functional connectivity of these valuation areas with regions involved in social cognition, such as the TPJ. These results therefore provide support for the extended common currency schema. However, only a few studies have directly compared activity patterns elicited by social outcomes with those due to non-social financial rewards. Although these studies have found areas of overlapping activity, they have also identified activity differences. For example, one study indicated that financial rewards and erotic pictures elicit activity in distinct portions of the OFC, whereas other studies showed that either type of reward uniquely activates voxels in the ventral striatum or the vmPFC. These results may therefore also be consistent with predictions from the social-valuation-specific schema. As few studies have parametrically varied the degree of reward and measured the hedonic and behavioural impact of the rewards, it is difficult to determine whether slightly different activation patterns for social versus non-social rewards and punishments indeed reflect fundamentally different neural computations or simply show differences in the reward value of the different types of feedback.

Decisions involving specific other individuals. All of the studies discussed above investigated neural expected values at the time of the (rewarding or non-rewarding) outcome. Different types of neural value representation (namely, decision values and anticipated values; see BOX 1) might be needed during actual choices that require participants to rely on subjective value estimates in the absence of any explicit reward.

Several of the studies mentioned above have provided evidence for such value representations. For example, value-related activity can already be observed in the ventral striatum and vmPFC during the mere anticipation of social gestures or face stimuli with rewarding or punishing properties. Interestingly, studies have shown that the presence of observers during decision making — and thus the possibility of future approval and reputation gains — changed value-related responses in the ventral striatum during risky choices and altruistic choices. The degree of change in ventral striatum activity during social observation correlated with changes in overt behaviour, suggesting that the mere anticipation
of peer feedback may introduce a neural ‘anticipated value’ that changes decision-value computations and, therefore, choice.

A slightly different line of research has examined neural decision values during choices about voluntary giving to specific others. Several studies suggest that humans make such decisions by integrating the monetary costs and social benefits of the decisions in value-related brain areas. Voluntary giving may therefore resemble a social form of purchase decision (the ‘purchased item’ is the specific other’s welfare). Giving money to liked others has indeed been found to elicit a neural ‘experienced value’: sharing a financial reward with a friend elicited more subjective happiness and more ventral striatum activity than sharing the same reward with a confederate or computer.

During decisions to transfer a portion of one’s endowment to a specific charity, BOLD activity in the ventral striatum is stronger than during mandatory donations or during decisions about similar payments to oneself. These BOLD activity increases in the ventral striatum may therefore reflect valuation of anticipated hedonic feelings or social consequences (for example, gains in reputation or social status) of the donation compared with the short-term monetary consequences. In line with this interpretation, a study found that BOLD activity in the OFC co-varied with the freely chosen amount of money given to a specific charity. As the OFC showed functional connectivity with the anterior insula and the posterior superior temporal sulcus during these decisions, it was argued that social valuation computations in the vmPFC may depend on input from specialized regions that compute the desirability of each charity.

The findings of these studies suggest that learning about the trustworthiness of others is associated with activity in the brain’s reward circuitry that mimics activity seen during learning about (non-social) reward values of objects. For instance, information about the trustee’s moral character or experience with a trustee’s behaviour changed both neural activity in the striatum during the outcome phase of each round and subsequent investment behaviour in the game, and these effects changed over the course of the experiment. Specifically, striatal responses tied to trust decisions were expressed during the outcome phase of trials early in the experiment (when opponents’ reputations were not yet known) and then shifted forward to the decision phase of each trial, possibly now indicating learned reward predictions for each opponent. Moreover, these trust-learning effects are strongly modulated by prior expectations (as is the case for non-social reward-learning effects): unexpected reciprocity in any trial induced stronger ventral striatum activity and higher investments in subsequent encounters compared with expected reciprocity, suggesting that this activity may reflect a reward prediction error similar to that observed during non-social forms of reward learning. However, striatal prediction errors and behavioural learning during the outcome of trust decisions can also be suppressed when information about the opponents’ trustworthiness is presented, suggesting that strong priors can overrule the importance of new information gathered during single trials.

The outcomes of the trust game used in these studies were purely financial, and one may therefore wonder whether the learning effects described above are really uniquely social or simply reflect that the other people predict a financial outcome. The latter interpretation is contradicted by studies of non-financial social rewards: faces with positive affect or symbolic gestures that are fully unrelated to monetary outcomes elicit behavioural effects and prediction-error signalling in the ventral striatum similar to those seen for monetary rewards. Thus, stimuli and actions that are uniquely encountered in social interactions can reinforce behaviour through neural mechanisms that are similar to those underlying non-social reinforcement with money. However, social-specific effects were observed for (punishment) learning by trust violations in the context of a pharmacological study that involved oxytocin, a neuropeptide that plays a central part in the ability to form social attachments in mammals and that can enhance trusting behaviour. In the placebo participants taking part in this study, receiving information that an opponent had breached their trust decreased trusting behaviour and increased BOLD levels in the amygdala and the caudate. As these behavioural and neural effects were not seen in a lottery game that matched this experiment in terms of financial investments and outcomes, they were specifically tied to the social dimension of breaches of trust rather than their financial consequences. Crucially, these effects were strongly reduced in participants who had been given oxytocin.

Learning about others. Interactions with a specific person often involve repeated contacts, and information about the other person strongly shapes behaviour in these interactive contexts. A series of studies using the trust game have suggested that learning about the trustworthiness of others is associated with activity in the brain’s reward circuitry that mimics activity seen during learning about (non-social) reward values of objects. For instance, information about the trustee’s moral character or experience with a trustee’s behaviour changed both neural activity in the striatum during the outcome phase of each round and subsequent investment behaviour in the game, and these effects changed over the course of the experiment. Specifically, striatal responses tied to trust decisions were expressed during the outcome phase of trials early in the experiment (when opponents’ reputations were not yet known) and then shifted forward to the decision phase of each trial, possibly now indicating learned reward predictions for each opponent. Moreover, these trust-learning effects are strongly modulated by prior expectations (as is the case for non-social reward-learning effects): unexpected reciprocity in any trial induced stronger ventral striatum activity and higher investments in subsequent encounters compared with expected reciprocity, suggesting that this activity may reflect a reward prediction error similar to that observed during non-social forms of reward learning. However, striatal prediction errors and behavioural learning during the outcome of trust decisions can also be suppressed when information about the opponents’ trustworthiness is presented, suggesting that strong priors can overrule the importance of new information gathered during single trials.

The outcomes of the trust game used in these studies were purely financial, and one may therefore wonder whether the learning effects described above are really uniquely social or simply reflect that the other people predict a financial outcome. The latter interpretation is contradicted by studies of non-financial social rewards: faces with positive affect or symbolic gestures that are fully unrelated to monetary outcomes elicit behavioural effects and prediction-error signalling in the ventral striatum similar to those seen for monetary rewards. Thus, stimuli and actions that are uniquely encountered in social interactions can reinforce behaviour through neural mechanisms that are similar to those underlying non-social reinforcement with money. However, social-specific effects were observed for (punishment) learning by trust violations in the context of a pharmacological study that involved oxytocin, a neuropeptide that plays a central part in the ability to form social attachments in mammals and that can enhance trusting behaviour. In the placebo participants taking part in this study, receiving information that an opponent had breached their trust decreased trusting behaviour and increased BOLD levels in the amygdala and the caudate. As these behavioural and neural effects were not seen in a lottery game that matched this experiment in terms of financial investments and outcomes, they were specifically tied to the social dimension of breaches of trust rather than their financial consequences. Crucially, these effects were strongly reduced in participants who had been given oxytocin. This suggests that oxytocin influences trusting behaviour by blocking social experience-related changes in neural activity in the amygdala and caudate. At a more general level, this finding illustrates how experimental modulations of neural function (for example, pharmacological interventions) can provide evidence that social decision making is causally steered by neural valuation processes.

The findings on social learning by trust violation reviewed above seem to be consistent with the social-valuation-specific schema, as they demonstrate neural and behavioural learning effects that are unique to social information. A similar point is made by several recent demonstrations that, during social interactions, prediction-error computations can be observed in neural structures outside the classic reward system if the updated predictions concern specific features of
Humans and other species have the capability to learn vicariously from the actions and outcomes of others. Several studies have shown that merely observing others receiving rewards and punishments can trigger neural experienced values that are similar to those elicited by personal reward experience. Social interactions and decisions can influence neural activity and reward processing in the vmPFC and ventral striatum. Whether these two prediction-error representations are integrated by functional connectivity between these areas is an open question.

**Vicarious neural valuation**

**Vicarious experience of rewards and punishment.** Humans and other species have the capability to learn by observation. Current debate centers on whether this learning depends on neural mechanisms similar to those for learning based on personal reward experience. Several studies have shown that merely observing others receiving rewards and punishments can trigger neural experienced value representations similar to those elicited by a personal experience of these stimuli. One study showed that the reported pleasure of watching others win monetary rewards was associated with BOLD activity increases in the ventral striatum that overlapped with (but were weaker than) activations elicited when participants won these rewards themselves. In another study, monetary wins for the participants or for a friend elicited ventral striatum activity that was higher than activity evoked by comparable wins for an unrelated partner. Similarly, vicarious punishments, such as watching others receive painful shocks or smelling disgusting odours, can trigger neural responses in the ACC and the insula that are similar to those induced by the personal experience of these shocks or smells. Such vicarious activation may even be induced by very abstract punishments, such as embarrassing social situations. Importantly, these vicarious neural experienced values can predict later choice behavior: activation of the anterior insula (or ventral striatum) in subjects who were viewing other people receiving painful shocks correlated with their subsequent decisions to help (or not to help) the observed person by choosing to endure some of these painful shocks themselves. These latter neural and behavioral effects were strongly modulated by whether the observed person was perceived as a member of the subject’s own social group or a rival group.
of value-related neural processing in the ventral striatum. Here, participants made stock purchase decisions based on non-social information (previous outcomes) and social information about each stock (by passively observing other individuals’ decisions and outcomes). Both types of information had distinct, additive effects on choice and ventral striatum activity. These findings of fully independent social and non-social influences on neural value coding may provide a possible neural basis for puzzling social influences on financial decisions that seem to contradict non-social evidence (for example, irrational herd behaviour during market bubbles; see also REF. 67).

Humans not only decide for themselves but can also make purely vicarious choices for others that have no direct impact on their own outcome. Three recent studies suggested that such vicarious decisions may involve neural valuation processes that are similar to those involved in choices that have consequences for oneself68–70. In one of these studies, BOLD activity in the vmPFC related to the participant’s estimates of how much a pictured other person would contribute to a public good69. No vmPFC activation was observed if the participant estimated how much the depicted person would invest in a personal saving context. A second study showed that vmPFC BOLD signals correlated with the number of monetary bids for specific movie DVDs independently of whether the outcome (and cost) of the bids affected the participant or another depicted person70. The third study showed that when participants alternated between choosing for themselves and for another person, the vmPFC flexibly encoded the decision-value signal relevant for the current choice, independently of whether it was the participant or the other person who received the resulting outcome69 (FIG. 2c). By contrast, a region in the dmPFC always encoded the corresponding value signal for the other, currently choice-irrelevant reference frame69.

These studies thus suggest that decision-value signals in the vmPFC can flexibly encode the value of both self-related and vicarious choices. Moreover, the same OFC region showed distinct functional connectivity with parietal regions during vicarious (compared to self-related) choices69. These results therefore support the common currency schema, both in terms of the overlap of neural activity in value regions and the distinct functional connectivity of these regions during social choice.

Vicarious learning. Observing others may not only influence the current choice but may also help humans and animals to learn the value of stimuli and actions for future behaviour. For instance, observing that others receive painful shocks following a visual stimulus can lead to conditioned fear responses and amygdala activation to this visual stimulus, even though participants never experienced the shocks themselves71. For positive values, several further studies also suggested that vicarious learning may follow computational principles that are similar to those for learning through personal experience62–72.

For example, one recent study induced prediction errors at the outcome stage of altruistic choices by unexpectedly changing the outcome, either for the...
participant or for a charity benefiting from the choice. Both types of outcome changes (participant or charity) led to reward-prediction errors in the ventral striatum\(^2\), thus demonstrating vicarious learning of outcome probabilities for another agent. Another study showed that the dorsal striatum encodes comparable reward prediction errors when participants make a choice and benefit from the rewards or when they simply observe another person’s choices and outcomes\(^3\). These studies thus suggest that both experience-based and vicarious learning of anticipated values have the same neural substrate (which supports the common currency schema). However, two other studies suggest that vicarious social learning also draws on distinct prediction-error computations in different neural structures\(^4,5\). These studies showed both vicarious reward prediction errors in the vmPFC (FIG. 2d) and action-prediction errors in either the dorsolateral PFC (dlPFC)\(^7\) or the dmPFC\(^7\) that encoded violations of expectations about which action the other participant would choose. These signals occurred in addition to BOLD signals in the ventral striatum that reflected the prediction-error signal for the participant’s own expected rewards\(^5\). Thus, in analogy to learning about specific others, vicarious learning may not only involve updating of anticipated values in reward-coding regions but also spatially and functionally distinct prediction errors that nevertheless follow similar computational principles (which supports the social-valuation-specific schema).

**Neural valuation of normative social principles**

Humans base their decisions not only on their self-interest but also take account of normative social principles (for example, one should be fair, one should be polite, and so on) that are often at odds with maximizing personal payout. Behaviour in line with these principles can be observed in a whole range of fully anonymous situations and cultural contexts and is thus unlikely to merely reflect properties of interactions with other specific individuals (discussed above). Exactly what motivates people to display this behaviour is a topic of debate, but it is often assumed that individual behaviour reflecting these principles may have played a crucial part in the evolution of human social behaviour\(^5\). Several studies now propose that one proximate mechanism for maintaining this behaviour may involve neural valuation processes that overlap with those underlying non-social rewards.

**Social principles affect reward experience.** Neural valuation of financial outcomes can be strongly modulated by observing the simultaneous outcomes for other people in the same situation, even though these outcomes should not matter when viewed from the perspective of material self-interest. In one study, two participants simultaneously played a game in which the payoffs for a correct performance were sometimes divided unevenly between both participants (that is, one participant received more than the other for a similar performance). Here, BOLD activity in a given participant’s ventral striatum co-varied with the payoff to this player when the other player did not receive any money, but it reflected relative payoff (compared with the other’s gain) when both players received money\(^7\). Another study found that this relative reward coding did not occur if the observed outcome was not actually paid to another person\(^9\), underlining that the modulation of ventral striatum activity was due to ‘social comparison’ of one’s own outcome to that of others (rather than merely to the numerical discrepancy between two outcomes).

Such comparison-based modulation of neural reward processing may lie at the heart of the phenomenon of inequality aversion\(^8\). Recent findings suggest that the behavioural enjoyment and BOLD activity in the ventral striatum and OFC that are elicited by financial payouts to one of two players strongly reflect differences in the initial capital of both players. Participants with less initial capital showed greater enjoyment ratings and BOLD activity in response to payments to themselves than to comparable payments to other players, whereas participants with a higher starting capital showed the opposite pattern\(^10\) (FIG. 3a). These neural signals may thus reflect experienced values in the ventral striatum and OFC that track changes in the outcome difference between two individuals rather than changes in each individual’s outcome. That differences in such neural value coding may lie at the heart of individual differences in inequality aversion was suggested by a study that used behavioural measures to identify participants with either dislike (‘prosocials’) or indifference (‘individualists’) to inequality. In this study, prosocials, but not individualists, rated unequal payments to two participants as less attractive than equal payments, and they showed a higher BOLD signal in the amygdala during these outcomes\(^11\) (FIG. 3b). Thus, neural activity in the brain’s valuation circuitry in response to financial rewards can directly incorporate information about the outcomes of others. This supports the idea that social principles can influence the very neural representations that encode one’s own financial rewards (as per the extended common currency schema). Such modulatory social influences on neural processing may underlie behavioural findings that salary satisfaction depends more on comparisons with average income in society than on absolute salary magnitude\(^12\).

Another example of modulatory effects on neural expected-value representations in line with social principles concerns the phenomenon of cooperation. It is widely thought that cooperative behaviour has evolved in many species through the beneficial effects it may have on reproductive fitness\(^8\), but it is unclear how it may be instantiated and maintained in individuals. Several neuroimaging studies have found that cooperative behaviour leads to BOLD increases in reward-related brain areas, such as the ventral striatum and OFC\(^13-17\). In one such study, two participants played a repeated prisoner’s dilemma game, where the outcome revealed whether both had chosen the same strategy — and hence had cooperated — or had defected\(^14\). Both the ventral striatum and the OFCs showed higher BOLD activity during outcomes that indicated mutual cooperation. Crucially, these activations exceeded those
Social influences on brain activity during choice behaviour and learning. a | Inequality modulates value-related blood-oxygen-level-dependent (BOLD) signals in the ventromedial prefrontal cortex (vmPFC) and striatum. Participants were randomly given a higher endowment (high-pay group) or lower endowment (low-pay group) than an interaction partner. Identical monetary transfers from the experimenter to both players led to different neural responses, with higher BOLD activity for transfers to the player who had less starting capital. b | Social value orientation is associated with BOLD activation in the amygdala in response to unequal financial payoffs to self and other. Participants with a prosocial value orientation show increased amygdala activity for large differences between payoffs to themselves or the other, whereas participants with an individualistic value orientation show the opposite. Each bar represents the strength of this relation for one participant. c | The risk associated with non-monetary social decisions (that is, the probability in a fictitious moral dilemma that humans will die as a consequence of the participant’s choice) elicits BOLD activity in the anterior insula; this neural measure predicted how strongly participants adapted their choices to avoid this risk. d | BOLD signals in the ventral striatum during face attractiveness ratings change if the ratings of other people disagree, indicating influences of social conformity on ventral striatum responses. Faces originally matched in attractiveness elicit higher (or lower) ratings and ventral striatum BOLD activity after feedback that peers rate these faces as higher (or lower) than oneself. Part a is from REF. 80, Nature Publishing Group. Part b is from REF. 81, Nature Publishing Group. Part c is adapted with permission from REF. 98, Cell Press/Elsevier. Part d is adapted with permission from REF. 102, © 2011 by SAGE Publications.

during non-social control conditions (in which the identical financial outcome or a choice outcome was determined by a computer). These findings were interpreted as indicating an experienced value of cooperation per se that may bias individuals to display cooperative behaviour.

Social principles influence decisions. Social principles not only influence reward experience but can also bias decisions and even lead participants to accept a personal cost for implementing a particular principle. Several studies have shown that these puzzling behaviours may relate to influences on neural decision-value computations. For instance, humans frequently sacrifice some of their own resources to punish violators of social norms, a behavioural tendency that may have evolved to stabilize cooperation and social-group cohesion. Such altruistic punishment is associated with feelings of satisfaction and with neural activity in the ventral striatum and the OFC, even though it resulted in a financial loss for the participant. Other examples for influences of fairness-related social principles on decision making come from fMRI studies of the ultimatum game. In these studies, rejection of unfair outcome distributions elicited activation of the anterior insula, whereas acceptance of fair offers elicited BOLD activity in the ventral striatum, vmPFC and amygdala. These neural responses were not restricted to situations in which the participants’ decisions reflected costly punishments for unfair offers: BOLD activity in the vmPFC and anterior insula were also modulated by fairness if participants decided about the outcome distributions themselves. The (un)fairness of choices resulting in outcome distributions is therefore encoded in neural structures that overlap with those often presumed to encode anticipated values of non-social choices (which is in accordance with the extended common currency schema). However, in close similarity to the effects of vicarious reward experience (discussed above), fairness-related neural responses and the associated behaviour can strongly depend on contextual factors, such as the proposer’s intentions, the relation of a specific offer’s distribution to the average in a series of offers or whether the opponent belongs to the player’s social group or a rival group. This contradicts a direct, automatic mapping of the objective ‘fairness’ of others’ decisions to neural responses in the reward system and raises questions about the origins of these value-related responses. One possible mechanism by which social principles may influence value-coding regions is through functional connectivity with areas involved in social cognition (such as the TPJ and dmPFC), as has been observed for basic social reward experience and for altruistic and vicarious choices.

Valuation-related neural brain structures may also be implicated in abstract social decisions that do not involve any personal gains or losses for the participant. For instance, decisions about the distribution of money between different groups of African children elicited social-related neural activity related to both equity (how fairly is the money split) and efficiency (how much money in total is given). Inequity correlated with BOLD activity in the insula, whereas efficiency correlated with BOLD activity in the striatum, suggesting separate neural coding of both of these putatively decision-relevant social principles in the brain’s reward circuitry. Finally, it seems that even social decisions that do not involve money may involve similar neural processes as those involved in classic forms of value-based decision making. For example, in one study, moral judgements about courses of actions that resulted in varying hypothetical numbers of deaths activated the OFC and the ventral striatum, and activity in both brain structures correlated with
value learning in line with social principles. Although numerous studies have demonstrated that normative social principles can influence neural value computations during choices, it is relatively unknown how these principles influence value-based learning processes. Some insight has been provided by studies on social conformity — the general tendency to align behaviour and attitudes with those of others, which may facilitate the creation and maintenance of social groups and information transmission within a population. These studies have shown that neural mechanisms that underlie the influences of conformity on an individual's valuation of objects are similar to the neural mechanisms involved in non-social reward-based learning. For instance, when the attractiveness ratings of faces or music choices made by other people do not match those of the participant, then BOLD activity increases in the rostral ACC and decreases in the ventral striatum (thus reflecting higher ventral striatum activity during conformity) of the participant. These neural effects of disagreement are more pronounced in participants who subsequently change their ratings and for those face stimuli that the participant subsequently rates differently to conform with others' opinion. This suggests that these neural responses in the rostral ACC and ventral striatum reflect conformity-related prediction-error signals associated with updating the estimated value of the stimuli. Importantly, being confronted with the discrepant opinion of others also influences the neural valuation of the same items during subsequent encounters: face stimuli that were initially rated as more (or less) attractive by a peer group than by the participant yielded higher (or lower) attractiveness ratings and ventral striatum and OFC activity when the participant rated them a second time [Fig. 5c].

These studies thus suggest that conformity-related changes in behaviour are brought about by similar prediction-error learning signals to those involved in non-social forms of learning. However, no studies have directly compared these two types of learning, so it is currently unclear whether these results provide support for truly domain-general value processing (that is, for the extended common currency schema) or for common principles of neural processing that are implemented in different neural structures (that is, for the social-valuation-specific schema).

Is social neural valuation distinct?

Most of the evidence discussed above supports the notion that social behaviour is controlled by reward- and motivation-related neural processes in the ventral striatum, vmPFC, amygdala and insula, which strongly resemble the neural value signals that were originally identified in the context of non-social decision making. These neural processes reflect experienced value coding of social rewards at the time of consumption, computation of the anticipated or decision value associated with a given social choice and prediction-error-like neural signals during social learning. As discussed throughout the article, all of these signals have been observed during different classes of situations that reflect either social interactions with specific others, vicarious decision making and learning or influences of abstract social principles on valuation and behaviour. As the only exception to this, there is so far little evidence that learning-related value signals are involved in the control of behaviour that conforms to social principles, possibly because such behaviour usually does not change rapidly from trial to trial in the context of behavioural experiments.

The findings discussed above are clearly at odds with the notion that all social aspects of our environment — including their motivational and hedonic value — are processed in fully separate, dedicated neural circuitries. Numerous findings indeed suggest that BOLD activity in the ventral striatum, vmPFC, amygdala and/or insula encodes a 'common neural currency' that assigns motivational properties to all stimuli and situations, irrespective of whether they relate to specific others, to abstract social principles or to material objects. This conclusion, however, may be premature, as some aspects of the findings discussed above suggest that there are also differences between social and non-social neural valuation.

For example, is unclear whether the BOLD changes in overlapping areas for social and non-social decisions involve the same neurons. fMRI may not provide adequate spatial resolution to resolve whether different types of decisions recruit distinct sub-areas or neural populations of a given brain area. Single-unit recording studies in non-human primates are beginning to identify different types of neurons in the striatum, OFC and ACC that may selectively encode social versus non-social aspects of rewards. For example, different types of neurons in the striatum have been found to respond to social (images of conspecifics) or non-social (juice) rewards or to be involved in signalling either when a reward was given or when this reward was due to the action of the monkey or a conspecific. Another study found that neurons in the macaque OFC mainly encoded rewards given to the monkey himself, whereas neurons in a subportion of the ACC encoded rewards received by another monkey. These recent studies seem to be consistent with the social-valuation-specific schema and illustrate that invasive recordings (and possibly brain stimulation methods) in non-human primates may offer important evidence for arbitrating between the two competing schemas of social reward processing, provided that all aspects of social decision making reviewed here can be identified and reliably measured in these species (BOX 4).
Box 4 | Evolutionary aspects of social rewards and valuation

The rewarding properties of social behaviour may have evolved to facilitate group cohesion and cooperation and may thus also be observable in non-human primates, our evolutionary cousins. A few studies have provided evidence that social behaviour is closely tied to reward processing in non-human primates. For example, macaques are willing to trade access to food for the possibility of viewing pictures of socially relevant others215, indicating that they value these social stimuli more than they value the food. This valuation of social information seems to be dependent on the integrity of the anterior cingulate gyrus145. Macaques can also learn to represent an opponent’s actions in reward-based competitive interactive games through distinct neural signals in the prefrontal cortex244. Finally, monkeys can show vicarious reinforcement when their own actions lead to rewards for other monkeys147, a behavioural tendency that is amplified by oxytocin administration48 and that apparently involves distinct types of neurons in the anterior cingulate cortex207.

Many other social aspects of reward-based learning and decision making have only been demonstrated behaviourally in non-human primates. For example, brown capuchin monkeys can learn foraging strategies purely vicariously, by observing conspecifics’ ‘rewarded choice’293. There is also considerable debate regarding the degree to which non-human primates display behaviour that is in line with normative social principles, such as altruistic food sharing220,221, altruistic helping233, inequity aversion224,225 and fairness-related behaviour242. Whether these behaviours engage reward- and value-related neural processes as they do in humans remains an interesting question for further research. These comparative studies may require methodological developments to enable parallel studies in monkeys and humans (for example, methods to image cerebral blood flow in chimpanzees195,198, but they may also provide converging evidence on social rewards and values using methods that are not readily available for human studies, such as single-unit recording108,146 and experimental lesions185.

Reverse inference

A scientific strategy of inferring the presence of a specific mental process when observing brain activity that has been correlated in previous studies with this process. This strategy can lead to invalid conclusions when used carelessly (because activity in a brain region may often be triggered by several different mental processes) and should be applied with caution.

Repetition suppression

Phenomenon that repetition of the same stimulus elicits reduced activity in neurons specialized for this stimulus, possibly reflecting adaptation.

Multivariate pattern analyses

Neuroimaging analyses that do not examine activity changes in each voxel in isolation but rather identify patterns of activity changes across spatially distributed voxels.

Adaptive coding

Neural computations that are not fixed but adapt to the environment. For example, the same range of neural activity can encode different value ranges in different settings.

If different types of neurons underlie valuation of social versus non-social information, then even neural activity at the level of BOLD signals may show crucial dissociations in a given brain area. Such dissociations are less apparent in the literature, as relatively few studies have directly compared social and non-social rewards or decisions in the same participants; many conclusions about possible neural overlaps have therefore been based on reverse inference107 or on comparisons between the results of different studies. When social and non-social decisions are directly compared, it can be extremely difficult to match different classes of stimuli so as to provide similar subjective values, which is a prerequisite for an objective test of whether an identical scale assigns value in both situations. These concerns notwithstanding, a growing number of studies have reported apparent functional specialization or domain-specific activity within regions of the reward circuitry for social versus non-social aspects of decision-relevant information124,125,137,138. This may be most apparent for tasks that require value learning: here, similar prediction-error computations are reflected in activity of different areas of the neural valuation circuitry, depending on whether these computations involved information about social context or information about non-social rewards. Moreover, some pharmacological studies also suggest that there are differences in the precise neuronal underpinnings of social and non-social value coding. For example, the effects of oxytocin on social behaviour include neural modulations in value-related regions without concomitant dopamine release195, whereas BOLD responses in these same regions for non-social decisions are clearly affected by pharmacological modulation of the dopamine system199. All of this suggests that social and non-social decision making may rely on parallel neural computations that follow similar principles but that are nevertheless instantiated in adjacent or overlapping neural circuits that are specialized for processing one type of information (in accordance with the social-valuation-specific schema). Evidence for such distinct neural populations within a given cortical region may perhaps be obtained using methods such as high-resolution fMRI, repetition suppression paradigms111 or multivariate pattern analyses112.

Finally, even shared neural processes in some regions may nevertheless depend on input from, and interactions with, distinct brain networks that encode either social or non-social aspects of the environment. In this case, the main difference in neural processing during social and non-social decisions may not lie in local value computations but rather in the remote neural regions that may provide the information on which these computations are based. Consistent with this view, several types of social decisions reviewed above have been associated with responses in regions outside the classic valuation circuitry (such as the dlPFC111, TP111,113 and dmPFC114,115) that may in principle provide input that is relevant for the construction of uniquely social values. This possibility has been formally tested with connectivity analyses38,45,57,70,74,113,115, which have revealed that different types of social decision making and learning involve functional coupling between BOLD responses in ‘classic’ valuation regions and BOLD responses in regions outside the reward circuitry (for example, in the TP or dmPFC). However, the specificity of these effects in social contexts remains to be established by direct comparisons of how social versus non-social decisions may change patterns of connectivity, and how the strength of this connectivity relates to behavioural performance. These tests may also be useful for studying the mechanisms by which social value-related brain responses change with characteristics of the target of valuation59,60,64,65 or choice situation92,93.

Outlook

We hope to have illustrated many interesting parallels that are emerging in the neuroscientific study of social and individual decision making, most of which relate to reward coding and neural value computations. These findings in crease our understanding of how social aspects of our environment can be integrated with non-social information in order to control our behaviour. Most importantly, however, these parallels lead to interesting questions that may guide the neurobiological study of social decision making in coming years. In closing, we outline some of these questions.

The apparent overlap in neural processing for social and non-social decisions suggests that some phenomena discovered in both domains may have a common neural origin. For example, some modulatory effects of social comparisons on reward-related striatal BOLD activity77,78,80,81 show conceptual similarities to findings on adaptive coding of reward value in the non-social domain116, perhaps indicating that social comparisons may provide just another ‘frame of reference’ for neural
REVIEWS


2. Damasio, A. R., Tranel, D. & Damasio, H. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomously to social stimuli. Behav. Brain Res. 41, 81–94 (1990).


This study shows that monetary gains and observed social gestures elicit largely overlapping patterns of BOLD activity in the striatum, therefore suggesting that both types of outcomes trigger related value computations.


Reputation for reciprocity engages the brain reward to violations of trust during social interactions. This study demonstrates that oxytocin has specific Perceptions of moral character modulate the neural related signals.

Increased functional connectivity between the 30 from regions involved in social cognition. during charitable decision making incorporate input the pursuit of social reward and the avoidance of social punishment. 

Rushworth, M. F. S. Associative learning of social exchanges. 


This study demonstrates, with various behavioural and neural measures, that social gestures have courses.


This study shows that social versus non-social rewards are processed by distinct types of neurons in the striatum of the macaque. 


125. Morishima, Y.; Schunk, D.; Bruhin, A.; Ruff, C. C.; & Fehr, E. Linking brain structure and activation in temptation to explain the neurobiology of human altruism. Neuron 75, 73–79 (2012). This study shows that the individual tendency for altruistic giving is systematically related to both the structure and the functional response profile of the TPJ.


