The Neurobiology of Trust: the Important Role of Emotions

Jan B. Engelmann & Ernst Fehr

Abstract

There is accumulating evidence suggesting that emotions can have a strong impact on social decision-making. However, the neural mechanisms of emotional influences on choice are less well understood to date. Here, we integrate recent results from two independent, but related, research streams in Social Neuroeconomics and Social Neuroscience, which together identify the neural mechanisms involved in the influences of emotions on social choice. Specifically, research in Social Neuroeconomics has shown that social decisions, such as trust taking, involve commonly ignored emotional considerations in addition to economic considerations related to payouts. Cooperative social interactions are consistently associated with approach emotions and activation within core structures of the reward system, the ventromedial prefrontal Cortex (VMPFC) and the ventral Striatum (VS), reflective of a social reward bonus above and beyond financial gain. Uncooperative social interactions, on the other hand, are consistently associated with aversive emotions and activation within core structures of the avoidance system, the anterior insula and amygdala, reflective of an emotional penalty above and beyond financial loss. These results are paralleled by recent findings in Social Neuroscience that underline the role of emotions in social interactions. Positive social feedback is consistently associated with approach emotions and activation within core structures of the reward system, the VMPFC and the VS. On the other hand, negative social feedback is consistently associated with aversive emotions and activation within core structures of the avoidance system, the anterior insula and the amygdala. Jointly, results from Social Neuroeconomics and Social Neuroscience suggest that social decision-making partially relies on emotional brain systems that signal the magnitude of positive and negative anticipatory emotions about the pro- and anti-social intentions of interaction partners. Therefore, anticipatory emotions associated with social approval and rejection can have important, but often ignored, influences on social choices. These considerations call for the integration of emotions into theories of social decision-making.
The central role of emotions in social decision-making

Increasing evidence indicates that emotions influence decision-making (Cohn, Engelmann, Fehr, & Maréchal, 2015; Engelmann & Hare, in press; Engelmann, Meyer, Fehr, & Ruff, 2015; Harlé & Sanfey, 2007; 2010; Lerner, Li, Valdesolo, & Kassam, 2015; Loewenstein, Weber, Hsee, & Welch, 2001; Phelps, Lempert, & Sokol-Hessner, 2014). This is consistent with recent results in cognitive and affective neuroscience that blur the traditional distinction between emotional and cognitive processes and underline their interactive nature (Pessoa, 2008; Phelps, 2006). Specifically, it has been shown that cognitive processes, such as attention and memory, rely heavily on emotional information to control goal-directed behavior (e.g., Anderson & Phelps, 2001; Hamann, Ely, Grafton, & Kilts, 1999; Lim, Padmala, & Pessoa, 2009; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). Decision-making constitutes one form of goal-directed behavior that is particularly complex and involves multiple component processes. These include, at minimum, forming a perceptual representation and computing the values of the available choice options, planning and executing an action to obtain the chosen outcomes and learning about the outcomes of the decision to improve future choices (Rangel & Hare, 2010; Rangel, Camerer, & Montague, 2008). Moreover, decisions made in a social context involve considerations of others’ well-being and prediction of their actions (Fehr & Camerer, 2007; Sanfey, 2007). These choice processes are subserved by multiple cognitive mechanisms including attention (e.g., Hare, Malmaud, & Rangel, 2011; Lim, O’Doherty, & Rangel, 2011; Rangel, 2010), memory (e.g., Bechara & Martin, 2004; Hinson, Jameson, & Whitney, 2003), learning (e.g., Niv, Edlund, Dayan, & O’Doherty, 2012; Schonberg, Daw, Joel, & O’Doherty, 2007; Schultz, 2002), and, in a social context, perspective taking (Gallagher & Frith, 2003; Saxe, 2006). Given the well-documented interactions between emotions and choice-relevant cognitive processes, in conjunction with the significant overlap in the neural circuitry of these processes (Engelmann & Hare, in press), emotions can be expected to influence choice-related cognitive mechanisms at all stages of the decision process.

Multiple theories of emotion agree that the brain generates behavior, among other things, via two opposing motivational systems, the approach and avoidance system, which mediate behavioral responses to reinforcement and can be linked to different underlying neural circuitry (Alcaro & Panksepp, 2011; Cacioppo & Gardner, 2003; Cloninger, 1987; Davidson, Ekman, Saron, Senulis, & Friesen, 1990; J. A. Gray, 1987; J. R. Gray, 2001; Lang, Bradley, & Cuthbert, 1998; Schneirla, 1959). Approach (appetitive) motivation is responsible for orchestrating behavior that
increases the probability of rewarding outcomes and is intimately related to positive affect. Neuroscientific research has identified neural circuitry within the mesolimbic and mesocortical dopamine system that encodes appetitive value across species (the “reward system”), with central projection sites in ventral striatum (VS) and ventromedial prefrontal cortex (VMPFC) (e.g., Bartra, McGuire, & Kable, 2013; Haber & Knutson, 2009; Levy & Glimcher, 2012; McClure, York, & Montague, 2004). Avoidance (aversive) motivation, on the other hand, is responsible for guiding behavior to avoid punishments and threats and is intimately related to negative affect. Aversive value has been shown to be encoded in a network of regions (the “avoidance system”) that includes anterior insula (Nitschke, Sarinopoulos, Mackiewicz, Schaefer, & Davidson, 2006; Paulus & Stein, 2006; Wager, Phan, Liberzon, & Taylor, 2003), lateral orbitofrontal cortex (O’Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001a; Small, Zatorre, Dagher, Evans, & Jones-Gotman, 2001) and amygdala (O’Doherty, Rolls, Francis, Bowtell, & McGlone, 2001b; Phelps & LeDoux, 2005; Wager et al., 2003). Supporting evidence for the consistent and specific involvement of these regions in processing aversive events is provided by quantitative forward and reverse inference analyses conducted in the context of large-scale automated meta-analyses (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011).

Here we integrate recent evidence from the fields of Social Neuroeconomics and Social Neuroscience to show that social decision-making, with a focus on trust taking, relies in important ways on the emotions and behavioral tendencies elicited by the approach and avoidance neural systems. We first review behavioral and neural evidence demonstrating that a particular form of social anxiety, betrayal aversion, influences trust decisions. We show that the importance of social emotions generalizes to decision-making in the context of other social decision-making tasks, such as the Ultimatum and Prisoners Dilemma games. Drawing upon recent developments in Social Neuroscience that underline the role of social emotions in interpersonal interactions, we argue that the anticipation of both approach- and avoidance-related social emotions elicited by the potential pro- and anti-social intentions of other players can influence social decision-making in important ways and in addition to economic considerations concerning monetary outcomes. Together, recent evidence from Social

---

1 Neurosynth, a tool for conducting large-scale meta-analyses, allows for the quantitative identification of cognitive states from brain activity patterns, commonly referred to as reverse inference. The reverse inference analysis for the term aversive identifies a network of regions, which are specifically activated during aversive events and include the right anterior insula (peak voxel at 42, 30, -2), bilateral amygdala (peak voxel at left: -22, -2, 18; right: 28, -4, -18), right lateral orbitofrontal cortex (peak voxel at 42, 46, -14) and dorsal anterior cingulate cortex (peak voxel at -2, 12, 28).
Neuroscience and Social Neuroeconomics calls for a more integrated theory of social decision-making that incorporates the influence of emotions.

**Trust decisions involve emotional consideration**

The importance of emotions in social decision-making is best illustrated by a concrete example underlining the role of social anxiety in decisions to trust. Trust is an essential component of human relationships that permeates not only interpersonal interactions, but is also an essential building block of economic transactions (Algan & Cahuc, 2013; Knack & Keefer, 1997; Zak & Knack, 2001). Trust has been extensively studied in Behavioral Economics and Social Neuroeconomics (Fehr, 2009; Rilling & Sanfey, 2011). The standard way to experimentally assess trust taking is the Trust Game (Berg, Dickhaut, & McCabe, 1995). In this game, two players interact by sequentially exchanging monetary amounts. First, the investor (player 1) and the trustee (player 2) receive an identical endowment from the experimenter. The investor can transfer any amount of his endowment to the trustee (player 2). If the investor decides to send a positive amount, the trustee receives the tripled amount of what the investor transferred. Next, the trustee decides whether and how much of his current holdings, which consist of her endowment plus the tripled transfer, to share with the investor. These rules are known by both players and provide an opportunity for mutual gain if the investor transfers a positive amount and the trustee is willing to share part of his resources. The amount sent by the investor reflects trust-taking, while the amount returned by the trustee reflects a willingness to reciprocate trust (“reciprocity”).

The trust game thus captures the notion of trust as a behavior that makes an individual (the investor) vulnerable to the actions of another person (the trustee) (Coleman, 1990; Fehr, 2009). The investor’s motive for taking this social risk is that mutual cooperation can increase not only his financial wellbeing (and that of the trustee), but also, as we argue below, that cooperation in itself is rewarding. That the trust game captures complex choice mechanisms, which involve emotions and thereby go beyond simple exchanges of money, has repeatedly been demonstrated. A number of recent behavioral studies have suggested that non-reciprocated trust elicits an emotional reaction associated with betrayal, and therefore carries a cost above and beyond the loss of money (e.g., Bohnet & Zeckhauser, 2004; Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). One of the clearest
demonstrations of the involvement of social emotions in trust taking comes from a series of experiments conducted by Bohnet and Zeckhauser (2004, 2008). The authors investigated betrayal aversion using a modified version of the trust game in which both the investor and the trustee had only two choices; the investor could only decide to trust or not to trust (imagine, for simplicity, that the above described trust game gives the investor only two choices – to invest nothing or to invest the whole endowment) and the trustee could only decide to send back a very low share or a fair share of his holdings in case that the investor trusted her (imagine, again for simplicity, that the trustee can only send back nothing or equally share her holdings). In the experiments of Bohnet and Zeckhauser many investors and many trustees simultaneously participated in the experiment and they were randomly matched with each other. However, the investors did not know whether they were matched with a trustworthy or a greedy trustee. The authors exploit this fact and asked the investors to state the minimum share of trustworthy trustees (i.e., those who share their holdings equally) that need to be present in order to make them willing to trust. The statement of such a minimum share of trustworthy trustees that need to be present basically boils down to making trusting conditional on the existence of a minimum probability of getting trust repaid, a measure that the authors called the “minimum acceptance probability” (MAP). If the actual share of trustworthy trustees was equal to or exceeded the required share (i.e., the MAP), the trusting action was implemented. If the actual share fell below the stated MAP, the non-trusting action was implemented. This procedure is closely related to the Becker-DeGroot-Marschak method (BDM, Becker, DeGroot, & Marschak, 1964) and therefore incentive compatible. To assess the specificity of risk attitude in social compared to non-social settings, participants played two additional control games, a risk game, in which participants stated their MAP to accept a gamble over a sure outcome (with payoffs for the investor that were exactly identical to the payoffs in the binary trust game described above), and a risky dictator game that was identical to the trust game, except that the decision to be trustworthy or not was determined by a random mechanism and not by the 2nd player in the game.

Results from an initial investigation indicate significantly greater MAPs in the trust game compared to the risk game (Bohnet & Zeckhauser, 2004). This result has subsequently been replicated across six different countries (Bohnet et al., 2008). Participants are therefore more reluctant to trade a sure payout for a gamble in a social compared to a non-social setting, indicating that factors beyond mere risk aversion influence trust taking. One such factor was
revealed by a comparison between MAPs in the trust and risky dictator games. The trust game differed from the risky dictator game only with respect to the mechanism that determines the payout for both players - a trustee that can intentionally behave in an untrustworthy manner in the binary trust game or a random mechanism that lacks intentionality in the risky dictator game. A comparison between the MAPs in the two games therefore controls for outcome based social preferences and reveals betrayal aversion. Results show significantly greater MAPs in the trust compared to the risky dictator game, reflective of betrayal aversion. Moreover, results were replicated in six different countries (Bohnet et al., 2008), indicating that betrayal aversion is a robust phenomenon across different cultural contexts.

A recent fMRI study directly investigated the neural correlates of betrayal aversion (Aimone, Houser, & Weber, 2014). Specifically, participants played two types of trust games in the role of the investor while undergoing scanning: a standard binary trust game, in which a trustee decided payout distributions, and a computer-mediated game, where the payout distribution to the two players was determined by a random mechanism equivalent to the risky dictator game of Bohnet and Zeckhauser (2004). Consistent with prior results (Bohnet et al., 2008; Bohnet & Zeckhauser, 2004), Aimone et al. (2014) observed significantly more trust taking in the computer-mediated game compared to the standard game, confirming the presence of betrayal aversion. This effect is reflected at the neural level by increased activity in the anterior insula, both when participants played with a human compared to a computer mediator and when they decided to give compared to withhold trust. Moreover, increasing levels of betrayal aversion were associated with increasing anterior insula activation during trust decisions with human counterparts compared to computer mediators. Together, these results implicate the anterior insula in betrayal aversion. Given the consistent role of the anterior insula in processing negative affect (Kuhnen & Knutson, 2005; Nitschke et al., 2006; Paulus & Stein, 2006), as well as social emotions (Lamm & Singer, 2010), the authors conclude that the heightened insular activity during human interactions is reflective of an emotional warning signal of potential future betrayal of trust.

These results are supported and extended by a series of pharmacological and neuroimaging studies investigating the role of oxytocin (OXT) in trust decisions (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Kosfeld et al., 2005; Mikolajczak et al., 2010). Oxytocin is a uniquely mammalian neuropeptide that is synthesized in the hypothalamus and can act as a neurotransmitter within the central nervous system. OXT’s action on the brain can influence
social behavior, including maternal behavior and pair bonding, as well as social motivation and sexual behavior (for review see Meyer-Lindenberg, Domes, Kirsch, & Heinrichs, 2011; Skuse & Gallagher, 2009) and facilitate social cognition (Domes et al., 2007a; Domes, Heinrichs, Michel, Berger, & Herpertz, 2007b; Guastella et al., 2010; Guastella, Mitchell, & Mathews, 2008; Kirsch et al., 2005). Kosfeld et al. (2005) tested the hypothesis that OXT enhances trust by administering the trust game to two groups of participants, the treatment group that received synthetic oxytocin and the control group that received an inactive placebo. Their results show increased trust taking in the OXT relative to the control group. Importantly, the effect of oxytocin was specific for trust taking, as OXT administration had no effect on transfer rates in the non-social risk game (despite the fact that the choice context was perfectly matched), as well as on trustees’ reciprocity. Given prior results demonstrating that hesitation to trust is in part driven by betrayal aversion (Aimone et al., 2014; Bohnet et al., 2008; Bohnet & Zeckhauser, 2004; Kosfeld et al., 2005), in combination with oxytocin’s role in reducing social anxiety (Kirsch et al., 2005; Labuschagne et al., 2010), results from Kosfeld et al. (2005) suggest that one mechanisms by which OXT mediates trust-taking is by reducing betrayal aversion.

A neuroimaging investigation by Baumgartner et al. (2008) confirmed the role of OXT in trust decisions and revealed the underlying neural circuitry involved in OXT’s effects on trust taking. While undergoing fMRI, two groups of participants (OXT and placebo control) played both the Trust Game and a matched risk game. Importantly, at the halfway point of the experiment participants received negative feedback, revealing that decisions to trust (and take risks) were not returned on half the trials. Behavioral adaptation to feedback was significantly affected by the administration of oxytocin, such that average post-feedback transfers decreased relative to pre-feedback in the placebo group, while an increase in post-feedback average transfers was observed in the OXT group. No differential effect of feedback on choices in the OXT and control group were observed in the risk game, indicating that OXT exerts its behavioral effect in the domain of social risks taken in the Trust Game. Moreover, neuroimaging findings demonstrate increased activity in the amygdala during trust taking in the post-feedback (relative to pre-feedback) phase in the placebo, but not the OXT group. Administration of OXT thus suppressed behavioral and neural adaptation to trust betrayal. Taken together with an extensive literature implicating the amygdala in processing aversive emotions and emotional relevance detection (e.g., Phelps & LeDoux, 2005), as well as studies showing that OXT decreases fear responses by modulating activity in the amygdala (Domes et al., 2007a; Kirsch et al., 2005; Labuschagne et al., 2010), these findings are consistent with the notion that OXT reduces emotional reactions to
trust betrayal by decreasing reactivity of the amygdala. Since such effects are specific to social risk taking, OXT administration likely leads to a reduction of social anxiety.

Taken together, these results suggest that emotions are important mediators of trust decisions. Specifically, betrayal aversion, which is a form of social anxiety, has been postulated to reduce the propensity to trust relative to matched situations, in which financial loss is not associated with intentional betrayal by another person. These results are supported by recent neuroimaging and pharmacological evidence, indicating that betrayal aversion is processed in a core structure of the avoidance system, the anterior insula, while learning about trustworthiness of another is mediated by a central projection site of the anterior insula (Shi & Cassell, 1998), the amygdala.

**Social decisions are influenced by anticipated emotional costs above and beyond potential material losses**

There is a large body of behavioral and neuroimaging evidence that supports the notion that, in addition to betrayal, unreciprocated cooperation and unfairness elicit aversive emotions and influence social decisions. The ultimatum game has been widely employed to investigate emotional reactions in response to unfair treatment. In this game, an endowment (e.g., 10 MU) is given to one player, the proposer, who decides on a division of this money with another player, the responder. The responder can accept the division (e.g., 50/50 split), in which case both players receive their respective amounts, or reject the division, in which case both players receive 0 MU. If subjects care only about their own income, they should accept any amount greater than zero MU. This prediction does not match empirical observations, which have repeatedly shown rejections of 20/80 divisions. Rejections of unfair offers are not reconcilable with models of decision-making that assume self-interest maximization and neglect social motives (and the underlying social emotions) such as fairness concerns, as rejections entail a financial loss for the responder and no direct benefit in one-shot interactions.

There is now considerable evidence suggesting that social emotions, such as anger, constitute a central motivating factor of costly punishment in the ultimatum game. One line of research investigated the role of emotions in costly punishment decisions via self-reports. While this approach has various methodological limitations (e.g., Nisbett & Wilson, 1977), it remains an important means to assess the subjective feeling states associated with social interactions. In
one of the first investigations of the role of emotions in UG decisions (Pillutla & Murnighan, 1996), participants made decisions about whether to accept or reject UG offers in the role of the responder and rated their emotional reactions to each offer. Results indicate that, in a classical UG, in which participants know the exact amount that is being split by the proposer, UG rejections of small offers are associated with feelings of anger and perceptions of unfairness. Interestingly, the UG was played in various conditions, one of which was designed to trigger emotional responses to unfairness. Specifically, under conditions in which the responder received a small monetary amount from the experimenter in case he rejected the offer (outside option), other players’ proposals that equaled this amount were rejected in almost 75% of the cases, and rejection rates were still high (ca. 40%) when the offer was slightly larger than the outside option. The authors argue that such low offers in conditions of low outside options are a signal of bad intentions, as proposers could use their knowledge about the responder’s low alternative income to intentionally reduce their own offer. This notion is supported by the emotional reactions of responders. A large majority of participants who rejected such offers reported feeling angry (72%). Moreover, a subset of participants who accepted unfair offers to avoid financial loss (outside option < offer) still reported feeling angry (21%). Importantly, emotional reactions to unfair offers were associated with actual decisions to reject, as demonstrated in logistic regressions showing that greater levels of anger are predictive of increases in rejection likelihoods.

Results from Tabibnia et al. (2007) show that the unfairness level of UG offers is significantly associated with self-reported contempt. Importantly, their study design employed a modified version of the UG, in which offer size and fairness level were varied independently across trials. This allowed for the dissociation of offer fairness from offer size and thus for investigating the emotional influences on choice above and beyond monetary considerations (see also Crockett, Clark, Tabibnia, Lieberman, & Robbins, 2008; Gradin et al., 2015). For instance, the same offer (e.g., 2MU) could be presented in a fair context (when the proposers’ endowment was 4MU), or in an unfair context (when the proposers’ endowment was 10MU). The significant relationship between level of unfairness and contempt persisted even after controlling for offer amount, indicating that unfairness can elicit aversive emotions. Another experiment showed that, given the opportunity, participants not only readily verbalize their emotional reactions to unfair UG offers, but such verbalization also reduces rejection rates (Xiao & Houser, 2005). A large amount of participants (90%) expressed their negative emotions in response to unfair offers of 20% of
the proposer’s endowment. Moreover, when comparing rejection rates in a standard UG with a modified version that allowed responders to send unconstrained messages, significantly reduced rejection rates of unfair offers were found in the latter version of the game. These results indicate that participants use costly punishment in the standard ultimatum game in which participants have no other way to communicate their emotions. When given the opportunity for alternative ways of expressing their contempt, however, these are preferred over costly punishment.

Emotions also influence decisions in games other than the UG. In one experiment, Rilling et al. (2008) assessed self-reported emotional reactions to unreciprocated cooperation in a PD game. Unreciprocated cooperation (CD) was associated with greater self-reports of aversive emotions that include betrayal, anger, envy, irritation, sadness and disappointment. Finally, punishment of free riders in public goods games is associated with increased negative emotions. Fehr and Gaechter (2002) show that participants expressed significant anger towards free riders in a hypothetical public goods game scenario. Moreover, the intensity of self-reported negative emotions towards free riders increased as the average contributions from other players increased. The punishment pattern in the public goods game – that high contributors tend to punish low contributors and the positive correlation between the punishment of free-riders and the free-riders’ deviation from average contribution – suggests that negative emotions are an important proximate factor behind altruistic punishment.

A further way to assess emotional reactions to unfairness is via psychophysiological measures of autonomic nervous system activation, such as Galvanic Skin Conductance Responses (SCR), which reflect emotional arousal (Figner & Murphy, 2011). Using skin conductance responses as a proxy for emotional arousal, van’t Wout et al. (2006) show significantly enhanced skin conductance responses for unfair compared to fair offers. Importantly, enhanced emotional arousal was specific to interactions with another human player, as it was not observed during an equivalently framed game with a computer proposer. Participants in Civai et al. (2010) played both a classical UG and a third-party UG, in which they made decisions on behalf of another person. Civai and colleagues measured skin conductance responses in these two UG games and probed participants’ emotional responses. The authors show significant negative emotions for unfair offers and significant positive emotions for fair offers in both game types, with the intensity of the emotional reaction being stronger in the classical UG compared to the third party UG. Investigating the relationship between SCR and rejections of the most unfair offers,
the authors show greater SCR amplitudes for classical UG rejections compared to acceptance, but also compared to decisions made in the third-party version of the UG. These results confirm findings from van’t Wout and colleagues (2006), showing that aversive emotional reactions to unfair offers, as indexed by SCR responses during decision-making and emotional self-reports, are related to classical UG decisions. Emotional reactions may be weaker as social distance increases (Strombach et al., 2015), for instance during third-party UG decisions when the unfairness is not directed at the self.

Functional neuroimaging studies lend support to the picture that is emerging from behavioral experiments, namely that negative emotions are important drivers of social decision making. Sanfey et al. (2003) investigated the neural correlates of responder decisions in the UG. Unfair offers elicited greater activation in bilateral anterior insula (AI), dorsolateral prefrontal cortex (DLPFC), and anterior cingulate cortex (ACC). Control and computer offers did not yield responses in these areas, underlining the importance of interactions with another human for insular activation patterns. The involvement of the anterior insula in UG decisions is strengthened by further evidence showing increasing insula activity with increasing unfairness, greater insula activity on trials during which the unfair offer was rejected and an association between rejection likelihood and insula activity. Follow-up studies generally support the involvement of the insula in fairness considerations, with some investigations showing activation of AI during UG decisions (Corradi-Dell’Acqua, Civai, Rumiati, & Fink, 2013; Güroğlu, van den Bos, Rombouts, & Crone, 2010; Güroğlu, van den Bos, van Dijk, Rombouts, & Crone, 2011; Tabibnia, Satpute, & Lieberman, 2008), while another implicates an important projection site of AI, the amygdala, in UG rejections (Gospic et al., 2011). Results demonstrating anger and contempt in response to unfair UG offers (Civai, Corradi-Dell’Acqua, Gamer, & Rumiati, 2010; Pillutla & Murnighan, 1996; Tabibnia et al., 2008) on the one hand, and the involvement of the anterior insula in both UG decisions and negative emotional states on the other (Nitschke et al., 2006; Paulus & Stein, 2006; Wager et al., 2003) provide converging evidence for the hypothesis that insula activation during unfair UG offers may reflect the emotional aspects of UG decisions. This notion is further corroborated by recent results from social neuroscience demonstrating the involvement of anterior insula in aversive emotional reactions to social rejection (Eisenberger, 2012).

---

1 Quantitative reverse inference analysis (neurosynth term aversive, Yarkoni et al., 2011) that identifies the right anterior insula (peak voxel at 42, 30, -2) and its projection sites in the amygdala (peak voxel at left: -22, -2, -18; right: 28, -4, -18) as regions that are selectively activated during aversive events lend further support to this idea.
Social rejection elicits aversive emotions that are processed in circuitry related to physical pain and negative affect

Humans care deeply about how they are evaluated by others. This fact is well captured in the high prevalence of public speaking anxiety (Pull, 2012), as well as research on social stress showing that social evaluation triggers a strong stress response in a great majority of participants (Dawans, Kirschbaum, & Heinrichs, 2011; Kirschbaum, Pirke, & Hellhammer, 1993). These effects of social evaluation are driven in part by reputation concerns (Fehr & Fischbacher, 2003; Izuma, 2012). Moreover, recent research has demonstrated that social approval can elicit positive emotions and is largely processed in core reward circuitry (Izuma, 2012), while social rejection can elicit aversive emotions and is largely processed in neural circuitry implicated in processing the affective component of pain (Eisenberger, 2012).

Although only a relatively recent research endeavor, great progress has been made in identifying the neural circuitry involved in processing social evaluation. Research investigating the behavioral and neural mechanisms underlying rejection have led to the proposal that social and physical pain share common psychological and neural mechanisms (Pain Overlap Theory; Eisenberger, 2012; Eisenberger & Lieberman, 2004; MacDonald & Leary, 2005). Indeed, behavioral experiments have demonstrated similarities in the psychological responses to physical and social pain. Specifically, participants recalling past episodes of social pain (e.g., betrayal by a close person) reported levels of experienced pain that were statistically indistinguishable from participants recalling past episodes of physical pain (Chen, Williams, Fitness, & Newton, 2008; Riva, Wirth, & Williams, 2011). Both recall and induction of physical pain (via the immersion of one hand in cold water) and social pain (via a virtual computer game of social exclusion outlined in detail below) led to equivalent decreases in feelings of self-esteem and control, as well as increases in negative affect and the desire to aggress (Riva et al., 2011). Moreover, physical pain induction significantly heightened feelings of being ignored and excluded, which are typically associated with social pain, further supporting the emotional overlap between social and physical pain.

In parallel, neuroimaging research on the neural circuitry of social rejection has revealed that social rejection is processed within regions implicated in physical pain processing. In an initial study investigating the neural circuitry of social exclusion (Eisenberger, Lieberman, & Williams,
participants played a virtual ball tossing game (the “Cyberball game”, Williams & Jarvis, 2006) with two other human participants. In one condition, virtual players included participants in the game by throwing the ball to participants, while in another, participants were actively excluded by the other two players. Results revealed that explicit social exclusion triggered greater self-reported distress and was associated with enhanced activation in dorsal ACC and anterior insula. Interestingly, both regions have consistently been implicated in processing the affective component of pain. Research on the neural basis of pain emphasizes the presence of distinct neural pathways for processing the somatosensory and affective components of pain (e.g., Bushnell, Čeko, & Low, 2013; Price, 2000; Rainville, Duncan, Price, Carrier, & Bushnell, 1997). Specifically, the brain first needs to represent the sensory aspects of pain, such as its location on the body, as well as its quality and intensity. This information is subsequently employed to assign an aversive emotional value to the painful sensation, which is crucial for motivating protective behavior that terminates the painful stimulation. Imaging studies have shown that distinct, but interacting neural pathways process the sensory and affective components of pain, with primary (S1) and secondary (S2) somatosensory cortices encoding the sensory aspects (e.g., Bushnell et al., 1999; Treede, Kenshalo, Gracely, & Jones, 1999), while anterior insula (AI) and dorsal anterior cingulate cortex (dACC) encode the affective components (e.g., Rainville et al., 1997; Shackman et al., 2011; Wiech et al., 2010). Eisenberger et al. (2003) therefore concluded that regions that process the affective component of pain, such as anterior insula (AI) and dorsal anterior cingulate cortex (dACC), are also intimately involved in processing social pain (Eisenberger, 2012).

While results from this initial study suggest the presence of shared circuitry for physical and social pain, a direct demonstration of overlap in the neural circuitry requires assessment of both physical and social pain in the same participants. Kross et al. (2011) addressed this shortcoming and investigated the overlap of physical and social pain in the same sample of participants. Participants who had recently experienced an unwanted relationship break up viewed photographs of their ex-partner while thinking about being rejected (compared to thinking about a recent positive experience when viewing a photograph depicting a friend), in the social pain task. In the physical pain task, participants experienced noxious (compared to non-noxious) thermal stimulation. Imaging results confirm that neural circuitry encoding physical and social pain overlapped in dACC and AI, confirming previous reports. Conjunction analysis also revealed significant overlap between social and physical pain in somatosensory brain systems, which encode the sensory component of pain, including thalamus and secondary somatosensory
cortex, as well as the opercular insular region and dorsal posterior insula (see also Fisher, Brown, Aron, Strong, & Mashek, 2010). These results suggest a more extensive overlap in the neural circuitry for social and physical pain under conditions of intense social pain, such as rejection from a loved one.

Correlations between self-reported feelings of distress and neural sensitivity to experienced social exclusion in AI and dACC can more closely identify the involvement of neural activation patterns in these regions in the affective component of social pain. Indeed, participants reporting greater distress to social exclusion showed greater social exclusion-related activity in ACC (Eisenberger et al., 2003). More recent studies using the cyberball game further underline the relationship between social rejection-related neural signals in AI and dACC and self-reported feelings of distress. A number of studies have shown a positive correlation between self-reported feelings of social distress and activity in dACC (DeWall et al., 2012; Eisenberger et al., 2003; Masten, Telzer, Fuligni, Lieberman, & Eisenberger, 2012), as well as AI (DeWall et al., 2012; Masten et al., 2012; 2009). These results have been corroborated in a series of studies showing that greater trait sensitivity to social exclusion (e.g., low self-esteem, anxious attachment) has been associated with increased social exclusion related signal in AI and dACC (DeWall et al., 2012; Onoda et al., 2010), while reduced sensitivity to social exclusion (e.g., social support, avoidant attachment) has been associated with decreased social exclusion related signal in these regions (DeWall et al., 2012; Eisenberger, Taylor, Gable, Hilmert, & Lieberman, 2007; Masten et al., 2012). A recent investigation of felt understanding demonstrates an even closer link between social emotions and brain activity in the anterior insula (Morelli, Torre, & Eisenberger, 2014). Specifically, the authors show that trial-by-trial signal changes in anterior insula and dmPFC tracked subjective ratings of feelings of non-understanding. Moreover, participants high in rejection sensitivity showed greater activity in the AI when they received negative feedback indicating non-understanding. Jointly, these results consistently implicate BOLD responses in anterior insula and dACC in the affective responses associated with social exclusion.

One consequence of shared neural circuitry for physical and social pain is that analgesics known to reduce physical pain would be expected to similarly influence social pain. Indeed, De Wall et al. (2010) show that, compared to placebo, daily intake of acetaminophen (a physical pain suppressant commonly known as ibuprofen) for a period of three weeks reduced self-reports of social pain during daily life, as well as neural responsivity to social exclusion in dACC, anterior
insula, and amygdala. Interestingly, the analgesic effects of acetaminophen have been shown to be mediated via the cannabinoid system (Mallet et al., 2008), activation of which has been shown to reduce anxiety (Patel & Hillard, 2006), as well as amygdala reactivity to social threat (Phan et al., 2008). Together, these results implicate the amygdala, a structure with a high density of cannabinoid CB1 receptors (Herkenham et al., 1990; Katona et al., 2001), in social pain, which parallels results from Social Neuroeconomics (e.g., Baumgartner et al., 2008). Furthermore, they suggest an interesting relationship between the cannabinoid system and social pain perception that provides a venue for future research on the neural mechanisms of alleviating social pain and social anxiety.

Taken together, these results provide converging evidence for the notion that aversive social emotions and the affective aspects of physical pain share underlying neural circuitry within dACC and AI. Moreover, results from social neuroscience provide a putative neural mechanism underlying aversive emotions important for social decisions, such as fear of betrayal, as well as anger and contempt due to unfairness and unreciprocated cooperation. The anterior insula in particular emerges as a neural hub important for aversive emotions in social dilemmas and social situations involving unfairness, as it activates during situations signifying trust betrayal (Aimone et al., 2014) and unfairness (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), as well as in the context of social rejection, where AI activity correlates with feelings of social distress and tracks trial-by-trial changes in social emotions (DeWall et al., 2012; Masten et al., 2009; 2012; Morelli et al., 2014).

**Social acceptance elicits approach emotions that are processed in core reward circuitry**

On the flip side, social neuroscience research investigating the neural basis of positive social feedback has demonstrated that different forms of social approval are processed within core reward circuitry in ventral striatum and ventromedial PFC, suggesting that social approval has affective properties that resemble primary rewards. Izuma et al. (2008) conducted one of the first studies investigating the behavioral and neural effects of social approval. Participants performed both a simple gambling task, in which outcomes were monetary rewards, as well as a matched social reward task, in which outcomes were positive evaluations of their personality by strangers. Despite the fact that the currency of reward differed, imaging results indicated that positive outcomes in both the monetary and social domain are processed in overlapping regions
of ventral striatum and follow similar activation patterns. Importantly, this was paralleled by behavioral reports, indicating positive social feedback to be desirable and to increase participants’ subjective happiness.

Follow up studies confirm this initial report of overlapping activation patterns in reward circuitry in response to monetary and social rewards. Receiving positive feedback reflective of first impressions from strangers based on participants’ appearance in a photograph was associated with activation in core reward regions including ventral striatum and VMPFC (Davey, Allen, Harrison, Dwyer, & Yücel, 2010). Moreover, when positive feedback came from strangers that participants regarded highly, greater activity was observed in VMPFC and amygdala. In another study (Spreckelmeyer et al., 2009), overlapping regions in VS also showed enhanced activation during reward anticipation of both monetary and social rewards when comparing brain responses during a standard version of the Monetary Incentive Delay (MID, Knutson, Westdorp, Kaiser, & Hommer, 2000) task to those during a social version of the MID. Moreover, it has repeatedly been shown that activity in VS parametrically tracks both monetary and social rewards (increasing intensity of happy facial expressions) particularly during reward anticipation (Rademacher, Salama, Gründer, & Spreckelmeyer, 2013; Spreckelmeyer et al., 2009).

Another series of studies employed a social feedback task that allowed for the dissociation of social approval from expectancy violations (Somerville, Heatherton, & Kelley, 2006). Participants viewed pictures of unfamiliar faces and communicated their expectation about whether the depicted individual would like them. After the participants’ initial rating, they received feedback about whether the individual accepted or rejected the participant. Social feedback thus varied in terms of congruency with participants’ expectations (confirm vs. contradict), and valence (positive vs. negative). Results revealed a neural dissociation between social feedback valence and expectancy, such that positive social feedback was preferentially processed in VMPFC, while expectancy violations were processed in dorsal ACC. Moreover, anticipation of positive social evaluations from peers was processed in VS, in addition to VMPFC, in one follow-up study (Moor, van Leijenhorst, Rombouts, Crone, & Van der Molen, 2010) and VS and DMPFC in another (Powers, Somerville, Kelley, & Heatherton, 2013). Individuals higher in self-reported rejection sensitivity also showed greater VS and DMPFC responses during anticipation of positive feedback (Powers et al., 2013). Finally, Morelli et al. (2014) show that feeling understood is tracked by signal in the VS. In this study, participants received feedback from strangers about how much they sympathized with a videotaped description of an important and
emotional event from the participants’ lives. Feeling understood (relative to not feeling understood) activated VS and middle insula. Importantly, the authors also demonstrate a close relationship between neural signals and social emotions, as activity in VS and TPJ also parametrically tracked subjective ratings of felt understanding on a trial-by-trial basis (Morelli et al., 2014; Rademacher et al., 2013; Spreckelmeyer et al., 2009).

Taken together, these studies consistently show that (1) activations in response to social rewards overlap with core reward circuitry in VS and VMPFC, (2) disparate social rewards that include social approval, first impressions, the feeling of being understood and happy facial expressions are encoded in VS and VMPFC across different tasks and contexts (Davey et al., 2010; Izuma, Saito, & Sadato, 2008; Moor et al., 2010; Morelli et al., 2014; Powers et al., 2013; Rademacher et al., 2013), and (3) VS and VMPFC often parametrically track social rewards in a manner that parallels their processing of monetary rewards (Morelli et al., 2014). Moreover, positive social evaluations elicit positive emotions (Izuma et al., 2008). Together, the behavioral and neuroimaging evidence suggests that positive social interactions are rewarding at the behavioral and neural level3. This is consistent with the view that positive social outcomes are processed by the brain in a manner that parallels the neural processing of primary rewards and money, which supports theories claiming that core reward circuitry encodes subjective values of disparate rewards on a common scale (Montague & Berns, 2002).

Social decisions are influenced by anticipated emotional benefits above and beyond potential material gains

These considerations directly translate to social interactions that are commonly investigated experimentally using social decision-making tasks, such as the Ultimatum Game (UG), the Prisoner’s Dilemma Game (PD) and the Trust Game (TG). In particular, the results reviewed above predict that cooperation has rewarding properties, by association with positive social feedback and acceptance, above and beyond the monetary outcomes that can be earned from cooperation. Support for this notion would be provided by evidence showing that pro-social actions of interaction partners in social choice tasks elicit (1) approach emotions and (2) activation in core reward circuitry.

3 This notion is further corroborated by quantitative reverse inference analysis (neurosynth term reward, Yarkoni et al., 2011) that identifies a network consisting of bilateral ventral striatum (peak voxel at left: -12, 10, -8; right: 10, 12, -6), ventromedial Prefrontal Cortex (peak voxel at 2, 58, -8) and brain stem (peak voxel at 4, -18, -14) to be selectively activated during rewarding events.
Indeed, Social Neuroeconomic research has repeatedly shown that cooperation in social dilemmas elicits positive emotions and is processed in core reward circuitry. In an initial investigation of cooperative behavior in the context of a social dilemma, Rilling et al. (2002) show that mutual cooperation (CC) in a repeated Prisoner’s Dilemma Game was associated with enhanced activations in VS and VMPFC. Importantly, mutual cooperation showed greater BOLD signal in these regions than trials that led to greater financial outcomes (DC), which is consistent with the view that cooperation was more rewarding than greater financial payoffs. Moreover, greater activity in VS and VMPFC was obtained only when monetary gains were due to social cooperation, but not when the same monetary outcome was accomplished in the nonsocial control condition. Further support for the involvement of VMPFC in cooperation is provided by a study by Decety et al. (2004), in which participants played a computer game that required the completion of a pattern with help from another player. Even in the absence of monetary payouts, the VMPFC showed prolonged enhanced activation when participants faced a cooperative player compared to a competitive one. Together, these results support the notion that cooperation is associated with activity in core reward circuitry.

Given that brain activation patterns do not necessarily reflect the emotional states of the participants during cooperation (Poldrack, 2006; 2011), it is important to assess actual emotional responses of participants to the different game outcomes. Initial evidence for positive emotional reactions to cooperation comes from interviews conducted by Rilling et al. (2002), which indicate that mutual cooperation (CC) was the most personally satisfying outcome. Extending these results, Rilling et al. (2008) assessed emotional reactions to PD outcomes in a questionnaire. Participants reported enhanced positive emotions due to cooperation. Importantly, enhancements of positive emotions that were specific to CC outcomes included trust, camaraderie and happiness (Rilling et al., 2008). The most direct evidence for the strong relationship between cooperation and positive emotions has been provided by a study by Tabibnia et al. (2007), which assessed the relationship between emotions and responder decisions in the UG. The researchers employed a study design that allowed for dissociation of emotional reactions to unfairness from reactions to low offer amounts (see detailed design outline above). Findings revealed that fairness was associated with greater happiness ratings, a result that persisted even after controlling for offer amount. Neurally, fairness (relative to unfair offers) elicited greater activation in VS, VMPFC, and amygdala. Again, this relationship persisted even after controlling for the size of the offer,
indicating that both emotional reactions and brain responses were specific to fairness considerations and not driven by monetary stake size. Finally, Xiao and Houser (2005) show that participants, when given the opportunity to communicate their emotions about UG offers, express exclusively positive emotions when receiving fair, but not unfair, offers.

The results reviewed above indicate that positive social interactions (1) elicit approach emotions associated with reward above and beyond monetary payoffs, and (2) are associated with enhanced brain activity in core regions of reward circuitry, including VS and VMPFC. Taken together with results from Social Neuroscience that attribute rewarding properties to social acceptance, emotional reactions to cooperative actions likely reflect inferences about the pro-social intentions of interaction partners and mimic emotions associated with social acceptance.

**Concluding Remarks**

Results from two independent streams of research in Social Neuroeconomics and Social Neuroscience jointly emphasize the importance of both approach and avoidance emotions in social interactions. The studies reviewed above provide converging evidence – using multiple methods, which include self-reports of subjective emotional state, choice behavior, psychophysiology and neuroimaging – that supports the involvement of emotions in social decisions. On the one hand, unsuccessful social interactions in interactive games such as the Trust Game, Ultimatum Game and Prisoners Dilemma have been shown to elicit aversive emotions related to betrayal, unfairness, and unreciprocated cooperation. On the other hand, successful social interactions in these games are associated with approach emotions. Moreover, our review of an independent research stream in Social Neuroscience identifies a striking overlap in two neural systems, one in core reward circuitry in VS and VMPFC that processes positive social feedback and successful cooperation, the other in a core structure of the avoidance system, the anterior insula, which emerges as a neural hub important for processing aversive emotions in social decision-making tasks. Anticipatory emotions related to the potential pro- and anti-social actions of interaction partners in these tasks therefore mimic emotional reactions to social acceptance and rejection, both neurally and behaviorally. Jointly, results from Social Neuroeconomics and Social Neuroscience suggest that social decision-making partially relies on emotional brain systems that signal the magnitude of positive and negative feelings about the pro- and anti-social intentions of interaction partners above and
beyond the material benefits and costs that result from their actions. Importantly, the anticipatory emotions associated with social approval and rejection can have central, but often ignored, influences on social choices. These considerations call for the integration of emotions into theories of social decision-making.
References


Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing attention on the health aspects of foods...


