Published in final edited form as:
Neuron. 2010 August 12; 67(3): 511–521. doi:10.1016/j.neuron.2010.06.030.

When giving is good: Ventromedial prefrontal cortex activation for others' intentions

Jeffrey C. Cooper, **Tamar A. Kreps**, **Taylor Wiebe**, **Tristana Pirkl**, and **Brian Knutson**
Department of Psychology, Stanford University

Summary

In social decision-making, people care both about others' outcomes and their intentions to help or harm. How the brain integrates representations of others' intentions with their outcomes, however, is unknown. In this study, participants inferred others' decisions in an economic game during functional magnetic resonance imaging. When the game was described in terms of donations, ventromedial prefrontal cortex (VMPFC) activation increased for inferring generous play and decreased for inferring selfish play. When the game was described in terms of individual savings, however, VMPFC activation did not distinguish between strategies. Distinct medial prefrontal regions also encoded consistency with situational norms. A separate network, including right temporoparietal junction and parahippocampal gyrus, was more activated for inferential errors in the donation than in the savings condition. These results for the first time demonstrate that neural responses to others' generosity or selfishness depend not only on their actions but also on their perceived intentions.

Introduction

People often have to evaluate others' decisions – for instance, to decide whether a car seller's offered price is fair, or to arbitrate between an employer and employee in a wage conflict. In these evaluations, people generally care about outcomes, such as how much money is at stake or how much each party earns, but also about intentions, such as whether the seller is honest or the employer is negotiating fairly. Participants in economic games, for example, will sacrifice their own monetary payoffs to punish selfish players or reward generous players (de Quervain et al., 2004; Fehr and Gachter, 2002; Fischbacher et al., 2001). These evaluations are commonly analyzed with reciprocity-based theories of social decision-making (Falk and Fischbacher, 2005; Frank, 1988; Sobel, 2005). Reciprocity-based theories propose formal models of preferences about others' outcomes, or "social preferences," in which people prefer rewards for others with helpful intentions and punishments for others with harmful intentions.

These theories generally assume that others' intentions are judged by observing past actions, such as how the seller has treated other buyers or how the employer has negotiated before. Judgments based on others' actions, however, can be biased by a range of individual and situational factors, such as the observer's personality, the stereotypes he or she holds, or what other information is provided (Kelley, 1973; Marston, 1976). If participants in similar

Correspondence concerning this article should be addressed to Jeffrey C. Cooper, now at Trinity College Institute of Neuroscience, Lloyd Institute, Trinity College, Dublin 2, Ireland. jeff.cooper@tcd.ie.

economic games, for example, are given different intentions for the other players (that they are compelled by the experimenter, or by chance), then the same selfish or generous actions are judged less harshly or kindly (Garrett and Libby, 1973; Singer et al., 2004).

Neural correlates of social preferences are likely to reflect both others' objective outcomes and their perceived intentions. The ventromedial prefrontal cortex (VMPFC) is one good candidate among neural structures that might represent social preferences. The medial PFC is important for a wide range of social cognitive tasks and social behaviors, and different regions seem to encode different components of social cognitive processing (Amodio and Frith, 2006). In particular, VMPFC activation correlates both with preferences for tangible personal outcomes (like money or food) and for social outcomes (like viewing attractive others or discovering that another person likes you) across a wide range of incentives and tasks (Chib et al., 2009; Davey et al., 2009; Knutson et al., 2005; Knutson and Wimmer, 2007; O'Doherty et al., 2003; Plassmann et al., 2007; Somerville et al., 2006). The VMPFC also plays a critical role in social cognition and empathy more broadly (Adolphs, 2009; Shamay-Tsoory et al., 2009), which suggests that it may be engaged when people consider others' outcomes as well as their own (Lieberman, 2007).

To test whether VMPFC activation reflects both others' outcomes and their intentions, we examined participants in an event-related functional magnetic resonance imaging (FMRI) study, as well as a separate behavioral study, while they observed other people playing a repeated public goods game (Ledyard, 1995) framed with one of two different descriptions (Figure 1). In both versions of the observed game, each player on each trial decided how much of a $10 endowment to contribute to a group investment, which was then doubled and split equally between players. The game presents a tension between contributing, which benefits the group, and not contributing, which benefits the individual player. In the "Donation" condition, the game was described in terms of donations and the group consequences of generous or selfish play. This description was designed to evoke emotional responses to players' actions, and to highlight players' helpful or harmful intentions toward others (Fehr and Gachter, 2002; Frank, 1988). In the "Savings" condition, the game was described in terms of personal savings and the individual consequences of risky or prudent play. This kind of description reduces personal contributions in the public goods game and related social dilemmas (Andreoni, 1995; Liberman et al., 2004), and was designed to minimize emotional judgments about players' intentions toward others.

In both conditions, participants on each trial inferred how much they expected the group of players to contribute (in the Donation condition) or save (in the Savings condition). Next, they saw each player's actual contribution amount. In the Donation condition, "High" (compared to "Low") inferences corresponded to high contributions and hence larger monetary outcomes for the group; in the Savings condition, "High" inferences corresponded to high individual savings and hence larger monetary outcomes for some individual players but not the group. The objective monetary outcomes were identical across conditions, and participants had no personal monetary stake in the game in either condition. The effect of the task descriptions on players' perceived intentions was measured by examining changes in liking for players (in the FMRI study) and interpersonal perceptions of players (in the behavioral study). If the different descriptions suggested different intentions behind the same observed actions, then participants should report increased liking for generous players and decreased liking for selfish players in the Donation condition, but not in the Savings condition.

Although we examined activation across the whole brain, our key hypotheses related to activation in the VMPFC. Evaluating others' decisions might recruit the VMPFC in three possible ways. First, the VMPFC might represent only the value of personal outcomes. In

this case, its activation should not distinguish between any observed outcomes, since participants had no personal stake in the game. Second, the VMPFC might "simulate" the objective value of others' outcomes regardless of how they are described. In this case, since the observed monetary payoffs were identical between conditions, any VMPFC activation should be also identical between conditions. Finally, VMPFC activation might reflect preferences about others' outcomes that incorporate their perceived intentions. In this case, its activation should represent the value of players' contributions in the Donation condition (when those contributions affected players' perceived intentions to help or harm others), but not in the Savings condition (when they did not).

Results

Note

For consistency, results are described in terms of contributions. Contributions match the numbers that participants in the Donation condition saw, but are reversed from what participants in the Savings condition saw (for example, an $8 contribution was seen as a $2 savings). When describing inferences, however, we retain the original framing, to match the words that all participants saw. Results associated with High (vs. Low) inferences in the Donation condition are thus associated with contributions of $20 or more, while results associated with High (vs. Low) inferences in the Savings condition are associated with savings of $20 or more.

Behavior (FMRI study)

Performance was measured as the percentage of correct inferences, averaged over blocks of 15 trials within conditions. Correct inferences increased over time, and did not differ between conditions (Table 1). Performance was above chance for all blocks; the worst performance was 59.63% correct in the first block of the Savings condition ($t(17) = 2.57$, $p = 0.02$). Polynomial contrasts indicated only a significant linear effect ($F(1, 36) = 10.28$, $p = 0.003$). There was no main effect of condition ($F(1, 36) = 0.004$, $p = 0.95$) or interaction between condition and time ($F(3, 108) = 0.09$, $p = 0.97$). Performance reached an identical plateau in both conditions of about two-thirds correct. For comparison, a participant with perfect knowledge of all players' strategies could have been correct on 76% of trials, due to the probabilistic nature of the task.

Reaction time (averaged over blocks of 15 trials) declined over time, but also did not differ between conditions (Table 1). Polynomial contrasts indicated both linear ($F(1, 36) = 8.65$, $p = 0.006$) and quadratic ($F(1, 36) = 5.47$, $p = 0.025$) effects, such that the speeding of reaction time declined over blocks. There was no main effect of condition ($F(1, 36) = 2.17$, $p = 0.15$) or interaction between condition and time ($F(3,108) = 1.69$, $p = 0.17$), indicating that participants spent similar amounts of time making inferences between conditions.

To assess explicit learning, players' actual average contributions were used to predict participants' post-task estimates, using a mixed linear model (MLM) with actual contribution, condition, and their interaction as predictors. Perfect learning would correspond to an average estimate (i.e., model intercept) of $5 and an actual contribution slope of $1. Participants made highly accurate estimates of average contributions. Actual contribution significantly predicted estimated contribution ($F(1,228) = 458.79$, $p < 0.001$), and the model intercept of $4.90 did not differ significantly from $5.00 (95% confidence interval [CI]: $4.68 to $5.09). The actual contribution slope was $0.88, significantly less than $1 (95% CI: $0.80 to $0.96), indicating that participants tended to overestimate low contributions and underestimate high contributions. Condition had no main effect or interaction (main effect: $F(1,228) = 0.03$, $p = 0.85$; interaction: $F(1,228) = 1.37$, $p = 0.24$).

Participants therefore made similarly accurate estimates of numerical contributions in both conditions.

Participants inferred High slightly more often than Low, but this bias did not differ between conditions (mean High in Donation = 56.33%, *SEM* = 1.67%; mean High in Savings = 55.00%, *SEM* = 2.10%; $t(36) = 0.50$, $p = 0.62$). Reaction times were also faster for High than Low inferences by about 70 ms (High $M = 1018.34$, *SEM* = 35.08; Low $M = 1084.60$, *SEM* = 37.51; $F(1, 36) = 10.00$, $p = 0.003$), but this advantage did not differ by condition ($F(1, 36) = 0.94$, $p = 0.34$). High and Low inferences and reaction times were thus comparable across conditions.

Participants were asked how much they themselves would have contributed if they had played (framed either as a question about donating or saving). Participants' own hypothetical contributions were significantly higher in the Donation condition ($M = \$5.85$, *SEM* = 0.43) than in the Savings condition ($M = \$3.56$, *SEM* = 0.53; $t(36) = 3.40$, $p = 0.002$), suggesting participants viewed contributions more favorably in the Donation condition.

Next, to test how observed contributions and the task descriptions influenced players' perceived intentions, changes in liking for players were predicted using an MLM with estimated contribution, condition, and their interaction as predictors (including initial liking and the quadratic and random effects of estimated contribution as covariates of no interest). We hypothesized that estimated contribution would increase liking in the Donation condition, but that this effect would be reduced in the Savings condition.

Both hypotheses were supported (Figure 2). In the Donation condition, high contributors were liked and low contributors were disliked ($F(1,37.05) = 34.07$, $p < 0.001$). This effect was symmetrical for high and low contributors, such that average liking across players did not differ from zero ($F(1,37.75) = 1.27$, $p = 0.27$). The effect was qualified by a significant interaction with condition ($F(1,36.60) = 10.82$, $p = 0.002$), reflecting a reduced effect of estimated contribution on liking in the Savings condition. There was no main effect of condition ($F(1,37.93) = 2.53$, $p = 0.12$), indicating that participants did not differ between conditions in their average liking across players.

Behavioral study

Participants in the behavioral study performed the same task as FMRI participants, again in either the Donation or Savings condition. Accuracy and reaction time were similar to the FMRI study (see Supplemental Table 1 online), and participants were again highly accurate in their explicit learning in both conditions (actual contribution effect on estimated contribution: $F(1,504) = 711.62$, $p < 0.001$; interaction with condition: $F(1,504) = 0.29$, $p = 0.59$). As well, participants' own hypothetical contributions were again higher in the Donation ($M = \$5.83$, *SEM* = 0.33) than in the Savings condition ($M = \$2.81$, *SEM* = 0.28; $t(82) = 6.90$, $p < 0.001$). The effect of players' estimated contribution on liking for players was also replicated. Participants liked high contributors and disliked low contributors in the Donation condition (main effect of estimated contribution: $F(1,86.68) = 76.82$, $p < 0.001$), but this effect was significantly reduced in the Savings condition (interaction: $F(1,84.98) = 39.76$, $p < 0.001$). There was again no main effect of condition on liking ($F(1,81.58) = 0.68$, $p = 0.41$).

To examine how interpersonal perceptions might be connected to liking, we also asked participants to judge players' interpersonal traits before and after the task - specifically, their dominance and their friendliness. One interpretation of the task condition's effect on liking might be that participants in the Savings condition saw players' contributions as purely individual decisions, unconnected to a social group. Differences in VMPFC activation

between conditions might then be due to differences in "how social" the situation was perceived to be (Harris et al., 2007). If the Savings condition changed whether players' contributions were perceived to have social meaning at all, then this condition should also decrease whether contributions affected any interpersonal traits.

Instead, however, the effect of task condition was selective for perceived intentions to help or harm (see Supplemental Figure 1 online). Specifically, judgments of dominance were unaffected by condition, such that low contributors were judged to be dominant in both conditions ($F(1,87.63) = 7.40$, $p = 0.01$). The interaction of this effect with condition was not significant ($F(1,85.69) = 0.40$, $p = 0.53$). By contrast, judgments of friendliness showed an identical pattern to liking; high contributors were judged to be friendly and low contributors were judged to be unfriendly in the Donation condition (main effect of estimated contribution: $F(1,86.58) = 82.38$, $p < 0.001$), but this effect was significantly reduced in the Savings condition (interaction: $F(1,84.47) = 21.00$, $p < 0.001$). There were no main effects of condition for either trait (dominance: $F(1,78.44) = 0.10$, $p = 0.75$, friendliness: $F(1,79.44) = 1.58$, $p = 0.21$).

This pattern suggests that contributions in both Donation and Savings conditions had social meaning; participants in both conditions saw low contributions as indicative of dominance (i.e., placing individual goals before others'). However, in the Donation condition only, those contributions also influenced perceptions of friendliness, an interpersonal dimension identified with the intention to help or harm others (Fiske et al., 2007).

Brain activation during inference phase

If VMPFC activation integrated others' objective monetary outcomes with their intentions to help or harm, the difference in activation between High and Low inferred contributions should be larger in the Donation condition (when contributions influenced player likability and perceptions of friendliness) than in the Savings condition (when they did not). To test the interaction of inferred contribution and condition, contrast images for making High vs. Low inferences were calculated within participants. These contrast images were then compared between Donation and Savings condition participants in an independent-sample *t*-test.

As predicted, VMPFC activation distinguished between High and Low inferences in the Donation but not the Savings condition (Figure 3). The comparison revealed a cluster in VMPFC (x/y/z = 0/42/−8 mm, peak $Z = 3.75$, extent = 58 voxels, $p = 0.046$ corrected), as well as several other regions including rostromedial prefrontal cortex (RMPFC), right middle temporal gyrus, and medial precuneus. The reverse interaction activated only one cluster in medial parietal cortex (Table 2). VMPFC activation timecourses suggested that the interaction was driven by increased activation for High inferences and decreased activation for Low inferences in the Donation condition, with little difference between High and Low inferences in the Savings condition. (See also Supplemental Table 2 online for activations within conditions.)

One important experimental control involved subjective certainty about inferences, which might also modulate prefrontal activation (Doya, 2008; Rushworth and Behrens, 2008). A reinforcement learning model was fit to each participant's behavior to estimate inferential certainty on each trial. After including regressors for certainty and estimated contribution sum, a smaller cluster of VMPFC was still activated for the interaction between inference and condition (see Supplemental Table 2 online), suggesting that VMPFC activation was not due to differences in subjective certainty between conditions.

We also examined the main effects of inference type (High or Low) across conditions. High inferences corresponded to opposite monetary outcomes between conditions (high donations or high savings), as did Low, but both kinds of inferences also shared several features; for example, High inferences always corresponded to larger numbers, and were always more consistent with the situational norms. To test the main effects, within-participant High vs. Low contrast images for both conditions were averaged in a one-sample *t*-test.

Only one area was more active for making High vs. Low inferences in both conditions, a cluster in right rostromedial PFC (Table 2; Figure 4). Several areas, however, were more active for Low vs. High inferences in both conditions. These included anterior cingulate (ACC) overlapping dorsal MPFC (x/y/z = −2/18/44 mm, peak $Z = 4.04$, extent = 124 voxels, $p < 0.001$ corrected), right dorsolateral prefrontal cortex (DLPFC), anterior insula, and occipital cortex. Several areas of medial and lateral frontal cortex, then, encoded the difference between High and Low inferences identically across conditions, even though those inferences corresponded to opposite monetary outcomes in different conditions.

Brain activation during feedback

Since players' contributions influenced liking in the Donation but not Savings condition, participants must have updated their beliefs about participants differently between conditions in response to observing those contributions. The difference between conditions in how players updated their beliefs might be reflected by differential neural activation to observing the contributions on the trial-by-trial level. Brain areas that were more engaged for learning about contributions in the Donation than in the Savings condition might be involved not just in learning numerical amounts, but specifically in learning or updating beliefs about players' likability or their intentions to help and harm.

To examine whether neural responses to learning about players' contributions differed between conditions, we examined activation correlated with inferential errors, which were estimated by a reinforcement learning model that accurately predicted participants' actual inferences (see Supplemental Experimental Procedures online for model details). The model estimated errors in a participant's inferred contribution for every player on every trial (positive for higher-than-expected contributions and negative for lower-than-expected contributions in both conditions). Imaging regressors then correlated these inferential errors with trial-by-trial activation when feedback was displayed. Within-participant contrast images that averaged across error regressors for all players were constructed; as before, these contrast images were compared in an independent-sample *t*-test between Donation and Savings conditions. Greater contrast values for the Donation than the Savings condition would indicate (on average) more activation for higher-than-expected contributions and less activation for lower-than-expected contributions.

Several regions in fact responded to inferential errors more in the Donation condition than in the Savings condition (Table 3 and Figure 5; see also Supplemental Table 3 online for activations within condition). Inferential error activation was greater for the Donation condition in the right parahippocampal gyrus, left DLPFC, right temporoparietal junction, and cuneus. Inferential error activation was greater for the Savings condition only in left middle frontal gyrus. These clusters met the exploratory cluster-size threshold, but none was large enough to meet the whole-brain corrected threshold.

Averaged across both conditions, inferential errors positively correlated with activation in right parietal cortex, such that higher-than-expected contributions increased activation in this region in both the Donation and Savings conditions (Table 3). Inferential errors in both conditions correlated negatively with clusters in posterior cingulate bordering on the parietal

cortex and in occipital cortex. These clusters also met only the exploratory cluster-size threshold.

Discussion

When evaluating others' decisions, people consider both their outcomes as well as their intentions. To determine how others' outcomes and intentions were integrated neurally, the current study examined individuals in two conditions of a novel social observation task in separate FMRI and behavioral experiments. All participants made inferences about the outcomes of players in a public goods game in which the participant had no personal stake, and during which the players used strategies ranging from generous to selfish. Participants in the Donation condition saw the game in terms of donations that helped or harmed other people, while participants in the Savings condition saw the same game in terms of savings that individuals maximized in a series of risky market investments. The Donation condition was designed to evoke emotional judgments of players' intentions to help or harm others, while the Savings condition was designed to disengage those judgments.

In the VMPFC, a key structure for evaluating personal outcomes, activation for others' outcomes was significantly affected by judgments of their intentions. In the Donation condition, VMPFC activation increased for high contribution inferences, which helped the group, and decreased for low contribution inferences, which harmed the group. In the Savings condition, however, VMPFC activation did not significantly vary when participants inferred high versus low contributions.

These findings are consistent with the idea that VMPFC activation reflects an integrated evaluation that can guide decisions. In this study, though, these evaluations were solely about others' outcomes. If VMPFC activation only represented personal outcomes, this region should not have responded in either condition, since participants had no monetary stake in the game and knew they would not interact with the players. If, by contrast, the VMPFC only simulated others' objective outcomes during observation, its activation should not have distinguished between conditions, as the observed monetary outcomes in the Savings and Donation conditions were identical. Neither of these accounts matches the current findings.

Instead, a social preference account suggests that participants preferred high contributions to low contributions in the Donation condition, but did not distinguish between them in the Savings condition. Why would preferences for the same outcomes differ between conditions? The clearest possibility is that different descriptions of the public goods game evoked different emotional judgments of players' intentions. Reciprocity-based theories of social preferences suggest that others' intentions to help or harm others play a key role in determining a personal response to their outcomes. Typically, those intentions are judged from behavior. For instance, a player who pursues a "nice" strategy (i.e., donates to the group) is seen as more likable than one who pursues a "nasty" strategy (i.e., withholds donations), and hence rewards for the nice player are preferred.

This judgment process, however, is not fixed; in this study, judgments differed across conditions. High and low contributions only suggested the intention to help or harm others in the Donation condition, as confirmed by changes in liking and ratings of friendliness in the Donation but not in the Savings condition. One possibility is that the framing manipulation changed the basis for moral evaluation. Low contributors were always perceived to put the individual before the group (as confirmed by ratings of dominance in both conditions). In the Donation condition, when pro-group norms were promoted, these norm violations were seen as antisocial and unlikable. In the Savings condition, when pro-

individual norms were promoted, low contributions were no longer a violation or an offense. Antisocial actions that seem justified do not generate the same level of outrage as the same actions evaluated as spiteful or competitive.

One caution about these conclusions is that the VMPFC is involved in social cognitive processing beyond simple evaluation of outcome preferences (Adolphs, 2009; Amodio and Frith, 2006). Although the Donation and Savings conditions were designed to be matched on social cognitive demands as closely as possible, differences in features of the social context beyond perceived intentions may also have contributed to differences in VMPFC activation.

Other regions of the medial PFC encoded different representations of players' actions that might correspond with different kinds of judgments. Rostromedial PFC was more active for inferring High in both conditions (at the exploratory cluster threshold), even though High inferences corresponded to different monetary outcomes (contributing or saving) across conditions. High inferences, however, were always more consistent with situational norms, as well as the participants' own hypothetical donations. This region has been linked to "mentalizing," the process of considering others' mental states and intentions (Amodio and Frith, 2006; Mitchell et al., 2005; Walter et al., 2004). In particular, this region is more active when considering intentions with clearer explanations, or when judging others who are more similar to the self (Harris et al., 2005; Mitchell et al., 2006). Activation in this region may thus reflect a situational norm for High inferences across conditions; this norm would provide a clearer reason for High donations/savings than for Low and may have led participants to feel more similar to those following the norm.

By contrast, Low inferences in both conditions activated a network including ACC, DLPFC, and insula. The ACC and DLPFC especially are involved in response conflicts like overriding a prepotent response, as in the Stroop or oddball tasks (Amodio and Frith, 2006; Barch et al., 2001; Carter et al., 1998), while the insula has been linked to detecting and processing uncertainty (Platt and Huettel, 2008). Activation of this network suggests that there was a prepotent response towards High inferences, an idea supported by the choice bias and slower reaction times in Low inferences. This interpretation is again consistent with a situational norm across conditions towards High and away from Low inferences, regardless of the monetary outcomes. Low outcomes may have seemed less likely or desirable due to the condition's described norms, and hence inferring Low may have required overriding the "default" prediction about players' behavior.

Taken together, these results suggest that in more dorsal MPFC (RMPFC and ACC), neural representations of players' intentions were relatively less sensitive to their objective monetary outcomes, and more sensitive to whether their behavior was consistent or inconsistent with the situational norms. Inferring behavior consistent with the condition's norm activated mentalizing regions, while inferring inconsistent behavior activated regions linked to response conflict, even when that norm objectively reversed between donating and saving.

Neural responses to learning about players' actual contributions also varied between conditions. A reinforcement learning model was fit to participants' inferences to estimate their trial-by-trial learning about how much each player tended to contribute. Similar models have a long tradition in social psychological accounts of impression formation (Anderson, 1971; Kashima and Kerekes, 1994), parallel to but distinct from their use in studying reward learning (Sutton and Barto, 1998). FMRI studies of learning about others have found that error terms in these models correlate with activation in several brain regions including the striatum, medial PFC, and right temporoparietal junction (TPJ; Behrens et al., 2008; King-Casas et al., 2005).

Inferential errors were associated with greater activation (positive and negative) in the right TPJ for participants in the Donation condition. Right TPJ activation has been associated with judging others' intentions in a variety of other social cognitive tasks (Castelli et al., 2000; Saxe, 2006). Activation in this region suggests that in the Donation condition, participants saw contributions as more informative about players' intentions to help or harm, consistent with the greater effect of contributions on liking in this condition. Inferential errors in the Donation condition were also associated with greater activation in the right parahippocampal gyrus and left DLPFC, which have been linked to explicit memory encoding (Gabrieli, 1998; Wagner et al., 1998). Donation participants learned more about players' intentions from the same numerical feedback; that learning may have changed existing cognitive representations about the players, such as beliefs about their personality traits. This interpretation is consistent with social psychological models using reinforcement learning algorithms (Kashima and Kerekes, 1994), in which inferential errors play a similar role to reward prediction errors in studies of incentive learning – that is, improving existing beliefs about others' traits based on feedback.

We did not detect between-condition differences in the striatum; while reward prediction errors have been linked especially to striatal activation (McClure et al., 2003), inferential errors (without a personal reward at stake) may instead be associated with activation in regions that support social learning and memory like the TPJ and medial temporal lobe. Another possibility is that the study lacked sufficient power to detect a between-condition difference; in the within-condition results (see Supplemental Table 3 online), putamen activation correlated with inferential errors in the Donation but not Savings condition, providing speculative evidence for this possibility. An important qualifier on all of the inferential error conclusions is that these clusters were activated only at the exploratory cluster threshold; future research will be needed to determine how robustly these models account for brain activation and behavior during learning about others.

These findings extend a growing line of research on how social contextual factors can modulate neural representations of others' outcomes. Others' outcomes, such as donations to charity, can activate reward-sensitive regions like the ventral striatum, even when observers have no personal stake (Harbaugh et al., 2007). These activations, though, can depend on emotional judgments of those others. Individuals watching others receive electric shocks, for example, had reduced activation in pain-sensitive regions such as the insula and ACC if those others had played unfairly in a prior economic game (Singer et al., 2006). In another study, when individuals read about others' misfortunes, they had greater activation in the ventral striatum when they envied those people than when they did not (Takahashi et al., 2009). Contextual modulation can also account for reactions to others' decisions and rewards. In one study, in which participants played economic games with fictional partners given likable, neutral, or unlikable backstories, the ventral caudate was activated only in response to cooperative decisions if the partner was unlikable or neutral, but was activated for both cooperation and non-cooperation if the partner was likable (Delgado et al., 2005). In another study, individuals watching another player win in a gambling game had greater ventral striatal activation when that player had previously expressed likable (compared to unlikable) personal traits in a taped interview. Further, VMPFC activation in response to those wins was modulated by subjective similarity to the player (but not by liking of him or her; Mobbs et al., 2009).

The current findings suggest that observing others' outcomes can activate neural structures that are also recruited by personal outcomes, such as the VMPFC, and highlight again that this activity depends upon the social context. Earlier studies, however, manipulated this context by manipulating the others' actual behavior (e.g., changing whether they had done likable things or played fairly). For the first time, this study demonstrates that VMPFC

activation is affected by others' intentions independent of their actions. Generous and selfish players made the same contributions in both conditions. Participants only judged their intentions as helpful or harmful, however, when contributing was described in terms of its consequences for the group – and only then did VMPFC activation vary.

The effect of the task descriptions on liking has implications for reciprocity-based models of social preference. These models typically assume individuals automatically judge the "niceness" or "nastiness" of other players, echoing evolutionary accounts of altruism that rely on knowing others' past reputations (Axelrod and Hamilton, 1981; Nowak and Sigmund, 1998). The current findings support these accounts, since merely observing others' contributions in a public goods game can drive formation of strong preferences. Social liking and disliking can persist well beyond observation of a single act and can influence unrelated decisions (Byrne, 1971), and thus the current results imply that deciding to contribute can have a long-term impact on one's reputation. At the same time, the results emphasize that the judgment of others' niceness or nastiness is not fixed by their behavior, but depends on how it is described. The offered price for a car might seem high when the seller's high profits are emphasized, but the seller might highlight the need to pay his or her staff; similarly, an arbitrator might view a wage offer differently when it is described as an "institutional savings measure" instead of a "pay cut." In the current study, a high contributor might choose to be described as a "high donator," while a low contributor might choose the "high saver" description. These descriptions influence both observers' judgments and their neural responses to observing contributions, or failures to contribute.

In summary, when individuals observed others in an economic game described in terms of donations to the group, they liked high contributors and disliked low contributors even when they had no personal stake in the game. In this Donation condition, VMPFC activation increased when inferring generous play and decreased when inferring selfish play. When participants observed the same game in a Savings condition that described play in terms of individual savings, neither VMPFC activation nor liking changed during inference. Regardless of the objective outcomes in each condition, rostromedial PFC activation increased for inferring behavior consistent with the condition's norm, while ACC, DLPFC, and insula activation increased for inconsistent behavior. In addition, inferential errors for observed contributions recruited brain regions linked to social cognition and memory (including the TPJ, DLPFC, and parahippocampal gyrus) in the Donation more than the Savings condition. These findings are consistent with the idea that in the Donation condition, individuals perceived contributions as more informative of others' intentions to help or harm, and that those intentions were integrated with the value of others' outcomes in the VMPFC. This region may thus play a key role in representing preferences about others' outcomes, above and beyond one's own. Those preferences, though, depend crucially on the perceived intentions behind others' actions.

Experimental Procedures

Participants

In the FMRI study, 38 individuals participated for cash after recruitment online, 20 in the Donation condition (10 women and 10 men) and 18 in the Savings condition (8 women and 10 men). Two additional participants, both in the Savings condition, were not analyzed due to self-reported drowsiness. FMRI participants were right-handed, fluent English speakers, ethnically representative of the Stanford community, and between the ages of 18 and 46 (M = 21.34, SEM = 1.01). FMRI participants also had no metal or medical device implants, no history of neurological or cardiovascular disorder, and no current psychiatric diagnosis or psychotropic prescriptions.

In the behavioral study, 84 individuals participated for cash after recruitment online, 42 in the Donation condition (25 women and 17 men) and 42 in the Savings condition (22 women and 20 men). Behavioral participants were fluent English speakers, ethnically representative of the Stanford community, and between the ages of 18 and 40 (M = 20.48, SEM = 0.35). All participants gave informed consent for a protocol approved by the Institutional Review Board of the Stanford University School of Medicine.

Materials and setting

Two sets of target faces were used in the FMRI study, one each in the Donation and Savings conditions; two faces overlapped between sets. Target faces were drawn from the Productive Aging Laboratory Face Database (Minear and Park, 2004). Two sets of target faces drawn from the same database were also used in the behavioral study, one all-male and one all-female. Target gender was counterbalanced across conditions, but was not analyzed in this study. All target faces were European-American, between 18 and 25, and had neutral expressions. All photos were of the full face in color on a gray background, cropped to 120 × 140 pixels. No rating differences were found between sets prior to the study. Scanning was conducted at the Richard M. Lucas Center for Imaging (Stanford, CA). Stimuli were projected using E-Prime 1.1 (Psychology Software Tools, Inc.; Pittsburgh, PA).

FMRI study experimental design and task

Before task—After informed consent, each participant was told he or she was going to make predictions about the outcomes of an economic game from an earlier experiment. Before further instructions, participants were left alone to fill out judgment questionnaires for each player, each with that player's face and the liking scale. Liking was measured with the two-item version of the Interpersonal Judgment Scale (Byrne, 1971). Item 1 asked "How much do you think you would like this person?" while Item 2 asked "How much would you like to work with this person in an experiment?" Both items were anchored at 1 by *dislike very much*, at 5 by *neither like nor dislike*, and at 9 by *like very much*.

The observed public goods game—With the experimenter, participants then read a series of instructions describing the observed game, a six-person repeated public goods game (Ledyard, 1995). On each round of the game, four of the six players are selected to play while the other two sit out. Those four are each given $10 and asked to decide how much to contribute to a common investment (from $0 to $10 in whole dollars). None of the players know who else has been selected to play on that round or how much they contribute. The experimenter then doubles the common investment and splits it into four equal shares. Each player then receives one share, plus any amount she did not contribute, into her bank account. High contributions thus improve group outcomes, because more money is doubled and split, but low contributions improve individual outcomes, providing an incentive for each individual to not contribute. After the description, participants played four practice rounds of the public goods game as players and reviewed two further examples to ensure they understood the payoffs. Players were tested identically in both conditions, to make certain the understanding of payoffs did not differ between conditions.

The task—Participants then read a set of instructions describing their inference task (Figure 1). Each trial had three phases. First (the *face phase*, 8 s), the players on that round appeared. Faces appeared individually from left to right every 2 s and remained on screen for the rest of the trial. Second (the *inference phase*, 3–5 s), the words "High" and "Low" appeared on screen. The participant inferred High if she believed the four players together contributed $20 or more to the common investment on that round, and Low otherwise. Inferences were made using a button box held in the right hand. The side of the screen (left or right) and associated button for each option was counterbalanced randomly between

trials. Participants had 3 s to infer, followed by a jitter delay of either 0, 1, or 2 s (equal numbers of trials for each delay, ordered randomly). Finally (the *feedback phase*, 5 s), each player's actual contribution appeared, as did the total amount and the actual group outcome (the word "High" or "Low", in green for correct or red for incorrect). The feedback phase was followed by a fixation cross of 2–4 s to make each trial's length 20 s.

Participants were told the rounds they saw were not presented in their original order from the earlier experiment, but were randomly ordered to prevent them from tracking sequences of contributions. The instructions asked participants to learn each player's average contribution level and use that level to make inferences. Participants then played three practice trials with cartoon faces and asked any questions before entering the scanner.

The task included 60 trials. Each possible combination of players appeared an equal number of times; each player therefore appeared 40 times. The primary manipulation was average contribution level. Each player's contributions were pre-designed so that the six players ranged from highly generous to highly selfish. Each player made contributions within a three-dollar range (12 trials at each of three values) for 90% of their appearances (36 of 40 trials); the remaining 10% of contributions for each player were random amounts outside of the player's range to increase plausibility. The ranges were $10 to $8, $9 to $7, $7 to $5, $5 to $3, $3 to $1, and $2 to $0. For example, the highest contributor gave $10, $9, and $8 12 times each, and a random amount between $0 and $7 4 times. The order of all contributions was randomized for every participant. About half of each participant's trials were High (mean High trials in Donation = 52.50%, *SEM* = 0.70%; mean High trials in Savings = 53.43%, *SEM* = 0.71%). Face photos were randomly assigned to players across participants.

Donation and Savings conditions—In the Donation condition, instructions emphasized group outcomes for the public goods game and described high contributions as positive. For example, contributions were called "donations" throughout; participants were also told the game is called "the Public Goods Game," that "donating to the common investment improved how every other player did on that round," and that the game is used "to study charitable donation behavior and how people invest in public goods like parks and schools." As well, all inferences and feedback during the instructions and task were shown in terms of contributions (i.e., the amount a player gave to the common investment.)

The Savings condition differed from the Donation condition in two ways. First, the instructions emphasized individual outcomes, and highlighted both the risk of contributing and the safety of not contributing. Participants were told the game was called "the Stock Market Game," that "investing is risky," that "the optimal decision is to save $10," and that the game is used "to study risk-taking behavior in markets and compulsive gambling." Second, inferences and feedback during the instructions and the task were shown not in terms of contributions, but rather in terms of savings (i.e., the amount a player kept for herself). For example, an $8 contribution was shown as $8 in the Donation condition, but as $2 in the Savings condition. The terms "saving" and "not saving" substituted for "not donating" and "donating" throughout the instructions. During the task, Savings participants inferred group savings, instead of contributions, on each trial; participants inferred High if they believed the four players together saved $20 or more, and Low otherwise. Feedback was then shown as savings amounts (instead of contributions) for that trial. All other details were identical to the Donation condition.

After inference task—Participants completed identical player judgment questionnaires in a waiting room. The final packet also asked participants to estimate the average amount each player contributed (in the Donation condition) or saved (in the Savings condition) in a single round. It also asked how much participants would have contributed themselves (in the

Donation condition) or saved themselves (in the Savings condition) on average in a single round if they had been a player. Afterwards, they were thanked and fully debriefed about the origins of the contributions and the aims of the study. Participation took about 90 min, and each participant was paid $40 in cash.

Imaging

Participants were scanned with a General Electric 1.5 T Signa scanner using the standard head coil, with a bite bar and padding to minimize head motion. Functional images covered the whole brain with 24 contiguous 4-mm thick axial slices (TR = 2 s, TE = 40 ms, flip = 90°, 3.75 × 3.75-mm in-plane voxel size, 64 × 64 matrix), collected using a T2*-sensitive spiral in/out pulse sequence that minimizes dropout in ventral frontal regions (Glover and Law, 2001; Preston et al., 2004). Each participant's functional run consisted of 606 images; the first 3 were then discarded to account for magnetic equilibration. Shimming was performed immediately before the functional run with custom software. An in-plane structural image was acquired before the shim (24 contiguous 4-mm thick axial slices; TR = 14 ms; TR = 400 ms, 0.94 × 0.94-mm in-plane resolution, 256 × 256 matrix), and a high-resolution structural was acquired after the functional run (3-D acquisition; T1-weighted SPGR sequence; 0.86 × 0.86 × 1.5-mm voxel size; 256 × 256 × 116 matrix).

Behavioral study experimental design and task

Behavioral participants performed a nearly identical task. The observed public goods game and contribution levels, inference task (including number of trials), and framing manipulation between Donation and Savings conditions were all identical to the FMRI study. Task timing was identical, except that the faces appeared all at once instead of over 8 s, and the inference phase was self-paced. Participation took about 60 min, and each participant was paid $12 in cash.

The key difference between FMRI and behavioral studies was in the player judgment questionnaires before and after the task. Participants again rated liking with the two-item Interpersonal Judgment Scale, but they also rated players on eight interpersonal adjectives spanning the interpersonal circumplex (e.g., assertive, antisocial; Knutson, 1996; Wiggins, 1979). These ratings were combined to create pre- and post-task ratings of dominance and friendliness that were then analyzed identically to liking (see Supplemental Experimental Procedures for details).

Statistical analysis

Inferences were analyzed using repeated-measures ANOVA and post hoc *t*-tests in SPSS 17.0 (SPSS, Inc.; Chicago). Reaction time was log-transformed before testing to correct for its skewed distribution. Liking, estimated contributions, and interpersonal ratings were analyzed with mixed linear models (MLM) using the MIXED command, treating players as the lower-level unit within participants. Models were estimated using maximum likelihood and diagonal covariance. All predictors were centered on the experiment mean and examined for approximate normality. The two liking items were averaged together for all analyses ($\alpha = 0.79$). Liking and interpersonal ratings were analyzed as changes from before to after the task, with initial ratings included in the model as a covariate. All tests were two-tailed.

Imaging data was analyzed with SPM5 (Wellcome Department of Imaging Neuroscience; London), using standard spatial preprocessing (slice timing correction, realignment, normalization, and spatial smoothing; see Supplemental Experimental Procedures for details). Two different models of experimental effects were used (see Supplemental Experimental Procedures for details). The "standard model" examined experimental events

across conditions, with separate regressors for High and Low inferences crossed with subsequently correct or incorrect inferences. The "reinforcement learning model" used the output of a reinforcement learning algorithm to create regressors for each participant's trial-by-trial estimate of the average contribution level for each player on each trial, as well as player-specific inferential errors on each trial. Inferential errors were calculated as the difference between actual contribution and estimated contribution level (i.e., increasing with actual contribution, decreasing with estimated contribution level). All regressors of interest were convolved with the SPM5 canonical hemodynamic response function. Six regressors modeling residual head motion (x, y, z, pitch, roll, and yaw) and a constant term were also included.

Models were estimated using restricted maximum likelihood and an AR(1) model for temporal autocorrelation. A high-pass filter (cutoff 90 s) removed low-frequency noise. Beta-weight images for each regressor were combined to form appropriate contrast images for each within-participant comparison (e.g., High vs. Low). Between-condition comparisons (e.g., Donation vs. Savings) were then made with independent-sample *t*-tests on within-participant contrasts. Peak activations are reported in MNI coordinates, as in SPM5.

Activations were thresholded voxelwise at $p < 0.001$. Family-wise error correction for multiple comparisons across the whole brain at $p < 0.05$ was achieved by using a cluster-size threshold estimated for each contrast using Gaussian random field theory (as standard in SPM5; Worsley et al., 1996). Cluster-size thresholds for the contrasts reported ranged from 41 to 65 voxels (noted in tables). As this correction tends to be conservative, we also report all activations above the exploratory cluster-size threshold of 10 voxels (Lieberman and Cunningham, 2009).

Highlights

- Response to others in identical economic games depends on game description.
- Liking, VMPFC distinguish generous from selfish play in "group donation" game only.
- Distinct MPFC regions encode consistency with norms regardless of outcome.
- Right TPJ and MTL are more activated by learning errors in donation game.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge suggestions and support from Samuel M. McClure, Arthur Aron, James Gross, Anthony Wagner, Kacey Ballard, Gregory Samanez-Larkin, and the SPAN lab. This work was funded by the National Science Foundation (0748915 to B.K.) and the National Institute of Mental Health (5T32MH020006-10 to J.C.C.).

References

Adolphs R. The social brain: neural basis of social knowledge. Ann Rev Psychol. 2009; 60:693–716. [PubMed: 18771388]

Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci. 2006; 7:268–277. [PubMed: 16552413]

Anderson NH. Integration theory and attitude change. Psychol Rev. 1971; 78:171–206.

Andreoni J. Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. Q J Econ. 1995; 110:1–21.

Axelrod R, Hamilton WD. The evolution of cooperation. Science. 1981; 211:1390–1396. [PubMed: 7466396]

Barch DM, Braver TS, Akbudak E, Conturo TE, Ollinger J, Snyder A. Anterior cingulate cortex and response conflict: Effects of response modality and processing domain. Cereb Cortex. 2001; 11:837–848. [PubMed: 11532889]

Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MF. Associative learning of social value. Nature. 2008; 456:245–249. [PubMed: 19005555]

Byrne, D. The attraction paradigm. New York: Academic Press; 1971.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll DC, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science. 1998; 280:747–749. [PubMed: 9563953]

Castelli F, Happe F, Frith U, Frith C. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage. 2000; 12:314–325. [PubMed: 10944414]

Chib VS, Rangel A, Shimojo S, O'Doherty JP. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J Neurosci. 2009; 29:12315–12320. [PubMed: 19793990]

Davey CG, Allen NB, Harrison BJ, Dwyer DB, Yucel M. Being liked activates primary reward and midline self-related brain regions. Hum Brain Mapp. 2009

de Quervain DJ, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E. The neural basis of altruistic punishment. Science. 2004; 305:1254–1258. [PubMed: 15333831]

Doya K. Modulators of decision making. Nat Neurosci. 2008; 11:410–416. [PubMed: 18368048]

Falk, A.; Fischbacher, U. Modeling fairness and reciprocity. In: Gintis, H.; Bowles, S.; Boyd, R.; Fehr, E., editors. Moral sentiments and material interests: The foundation of cooperation in economic life. Cambridge: MIT Press; 2005.

Fehr E, Gachter S. Altruistic punishment in humans. Nature. 2002; 415:137–140. [PubMed: 11805825]

Fischbacher U, Gachter S, Fehr E. Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett. 2001; 71:397–404.

Fiske ST, Cuddy AJC, Glick P. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci. 2007; 11:77–83. [PubMed: 17188552]

Frank, RH. Passions within reason: the strategic role of the emotions. New York: Norton; 1988.

Gabrieli JDE. Cognitive neuroscience of human memory. Ann Rev Psychol. 1998; 49:87–115. [PubMed: 9496622]

Garrett J, Libby WL. Role of intentionality in mediating responses to inequity in dyad. J Pers Soc Psychol. 1973; 28:21–27.

Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med. 2001; 46:515–522. [PubMed: 11550244]

Harbaugh WT, Mayr U, Burghart DR. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science. 2007; 316:1622–1625. [PubMed: 17569866]

Harris LT, Mcclure S, van den Bos W, Cohen J, Fiske ST. Regions of the MPFC differentially tuned to social and nonsocial affective evaluation. Cognitive, Affective & Behavioral Neuroscience. 2007; 7:309–316.

Harris LT, Todorov A, Fiske ST. Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. Neuroimage. 2005; 28:763–769. [PubMed: 16046148]

Kashima Y, Kerekes ARZ. A distributed-memory model of averaging phenoman in person impression-formation. J Exp Soc Psychol. 1994; 30:407–455.

Kelley HH. Processes of causal attribution. Am Psychol. 1973; 28:107–128.

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR. Getting to know you: Reputation and trust in a two-person economic exchange. Science. 2005; 308:78–83. [PubMed: 15802598]

Knutson B. Facial expressions of emotion influence interpersonal trait inferences. J Nonverbal Beh. 1996; 20:165–182.

Knutson B, Taylor J, Kaufman MT, Peterson R, Glover G. Distributed neural representation of expected value. J Neurosci. 2005; 25:4806–4812. [PubMed: 15888656]

Knutson, B.; Wimmer, GE. Reward: Neural circuitry for social valuation. In: Harmon-Jones, E.; Winkielman, P., editors. Social neuroscience: Integrating biological and psychological explanations of social behavior. New York: The Guilford Press; 2007.

Ledyard, JO. Public goods: a survey of experimental research. In: Kagel, JH.; Roth, AE., editors. The Handbook of Experimental Economics. Princeton, NJ: Princeton University Press; 1995. p. 111-194.

Liberman V, Samuels SM, Ross L. The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. Pers Soc Psychol Bull. 2004; 30:1175–1185. [PubMed: 15359020]

Lieberman MD. Social cognitive neuroscience: A review of core processes. Ann Rev Psychol. 2007; 58:259–289. [PubMed: 17002553]

Lieberman MD, Cunningham WA. Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc Cogn Affect Neurosci. 2009; 4:423–428. [PubMed: 20035017]

Marston BJ. Trait attribution to a described action as a function of changes in salient information. J Res Pers. 1976; 10:245–255.

McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 2003; 38:339–346. [PubMed: 12718866]

Minear M, Park DC. A lifespan database of adult facial stimuli. Behav Res Methods Instrum Comput. 2004; 36:630–633. [PubMed: 15641408]

Mitchell JP, Macrae CN, Banaji M. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron. 2006; 50:655–663. [PubMed: 16701214]

Mitchell JP, Neil Macrae C, Banaji MR. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. Neuroimage. 2005; 26:251–257. [PubMed: 15862225]

Mobbs D, Yu R, Meyer M, Passamonti L, Seymour B, Calder AJ, Schweizer S, Frith CD, Dalgleish T. A key role for similarity in vicarious reward. Science. 2009; 324:900. [PubMed: 19443777]

Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature. 1998; 393:573–577. [PubMed: 9634232]

O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia. 2003; 41:147–155. [PubMed: 12459213]

Plassmann H, O'Doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci. 2007; 27:9984–9988. [PubMed: 17855612]

Platt M, Huettel S. Risky business: the neuroeconomics of decision making under uncertainty. Nat Neurosci. 2008; 11:398–403. [PubMed: 18368046]

Preston AR, Thomason ME, Ochsner KN, Cooper JC, Glover GH. Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5 and 3 T. Neuroimage. 2004; 21:291–301. [PubMed: 14741667]

Rushworth MF, Behrens T. Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci. 2008; 11:389–397. [PubMed: 18368045]

Saxe R. Uniquely human social cognition. Curr Opin Neurobiol. 2006; 16:235–239. [PubMed: 16546372]

Shamay-Tsoory SG, Aharon-Peretz J, Perry D. Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain. 2009; 132:617–627. [PubMed: 18971202]

Singer T, Kiebel SJ, Winston JS, Dolan R, Frith CD. Brain responses to the acquired moral status of faces. Neuron. 2004; 41:653–662. [PubMed: 14980212]

Singer T, Seymour B, O'doherty J, Stephan K, Dolan R, Frith CD. Empathic neural responses are modulated by the perceived fairness of others. Nature. 2006; 439:466–469. [PubMed: 16421576]

Sobel J. Interdependent preferences and reciprocity. J Econ Lit. 2005; 43:392–436.

Somerville LH, Heatherton TF, Kelley WM. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. Nat Neurosci. 2006; 9:1007–1008. [PubMed: 16819523]

Sutton, RS.; Barto, AG. Reinforcement learning: An introduction. Cambridge, MA: The MIT Press; 1998.

Takahashi H, Kato M, Matsuura M, Mobbs D, Suhara T, Okubo Y. When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. Science. 2009; 323:937–939. [PubMed: 19213918]

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 1998; 281:1188–1191. [PubMed: 9712582]

Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. J Cogn Neurosci. 2004; 16:1854–1863. [PubMed: 15701234]

Wiggins JS. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. J Pers Soc Psychol. 1979; 37:395–412.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant voxels in images of cerebral activation. Hum Brain Mapp. 1996; 4:58–73. [PubMed: 20408186]

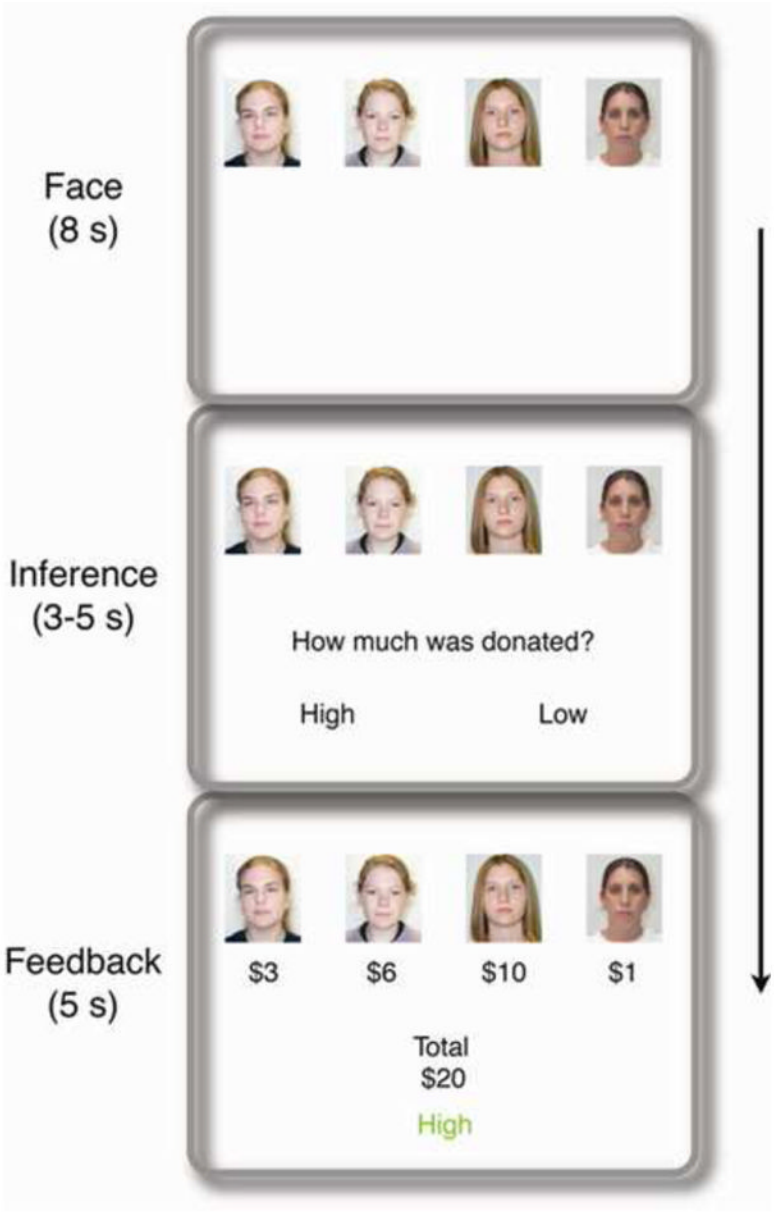

Figure 1. Timeline for a single trial

Participants observed and made inferences about other players in a repeated public goods game. Participants did not play themselves and had no personal monetary stake in the game. In the observed game, each player on each trial decided how much of a $10 endowment to contribute to a group investment that was then doubled and split equally between players on that trial. On each trial, participants first saw the four players for that round (*face phase*, 8 s), then inferred whether their contributions would total $20 or more ("High") or less than $20 ("Low"; *inference phase*, 3–5 s). Each player's actual contribution was then displayed under her face, along with the total contribution and the correct outcome (*feedback phase*, 5 s). A Donation-condition trial is shown; the Savings condition was identical except all inferences and feedback were in terms of savings ($10 - the contribution amount) rather than contributions.

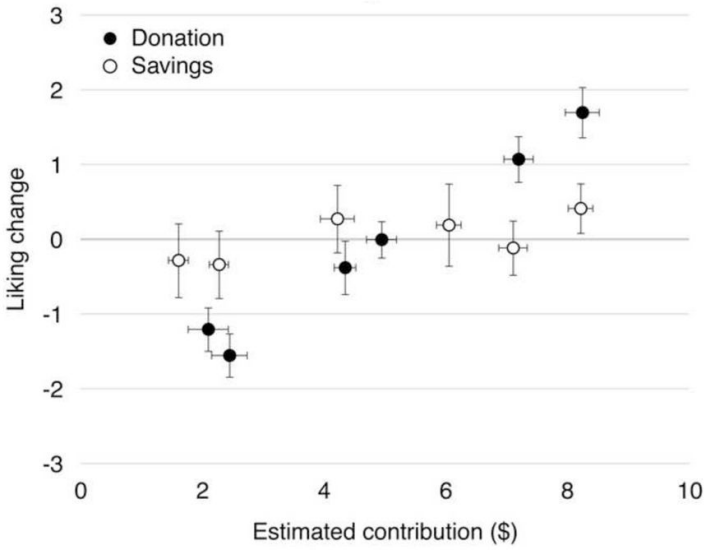

Figure 2. Estimated contribution predicts liking in the Donation but not in the Savings condition

Points represent liking change from before to after the task for each player, plotted against the estimated average contribution for that player. Participants in the Savings condition saw savings amounts ($10 – contributions); contributions are displayed here for clarity. Error bars are standard errors across participants. See also Supplemental Figure 1 online for changes in interpersonal ratings.

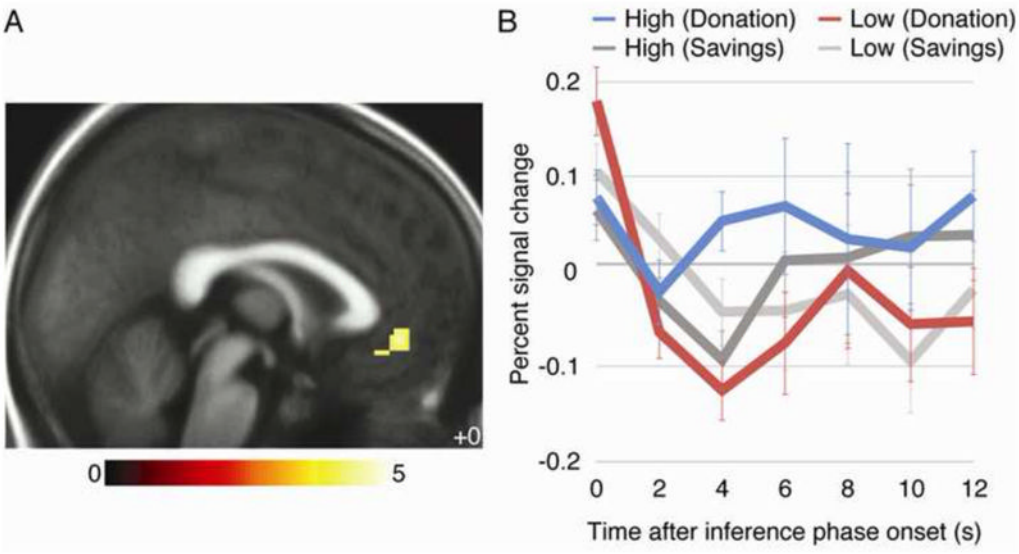

Figure 3. Ventromedial prefrontal cortex is selectively activated for High inferences in the Donation condition

A) Activation for High vs. Low inferences, greater for Donation than Savings condition. Color bar indicates *t*-statistic. Activations thresholded voxelwise at $p < 0.001$ with a 10-voxel extent minimum for display. B) Timecourse of activation from 8-mm spherical region of interest centered on VMPFC functional peak, beginning at inference-phase onset. Error bars are standard errors across participants.

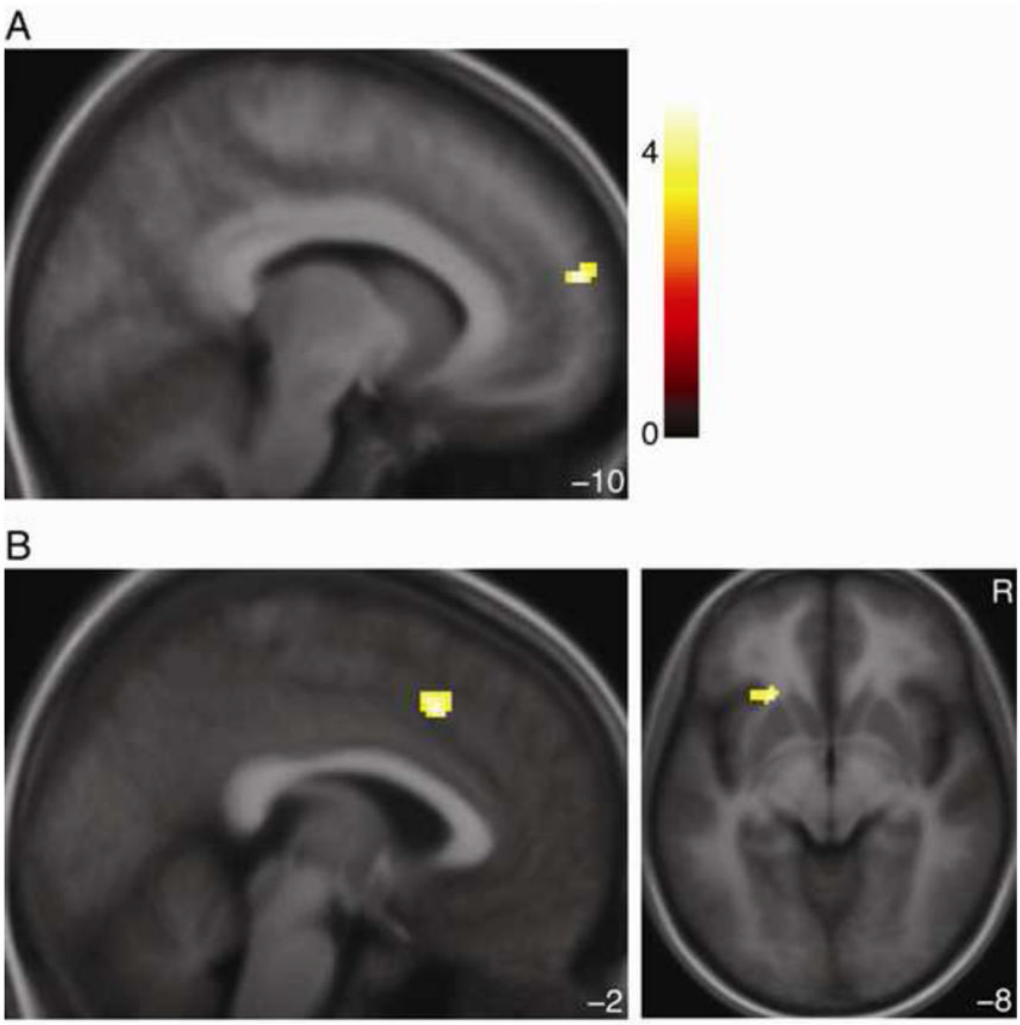

Figure 4. Rostromedial and dorsomedial prefrontal cortex distinguish High and Low inferences across conditions

A) Activation for High vs. Low inferences in both conditions. B) Activation for Low vs. High inferences in both conditions. R indicates right. Color bar indicates *t*-statistic for both panels. Activations thresholded voxelwise at $p < 0.001$ with a 10-voxel extent minimum for display.

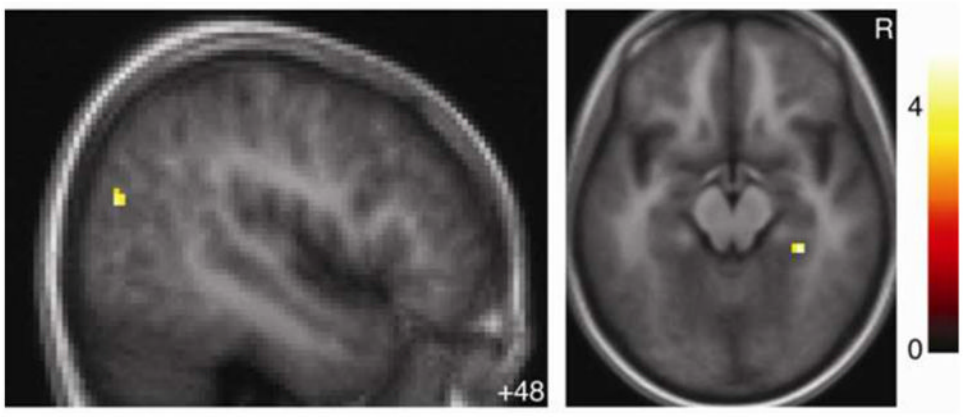

Figure 5. Activation for inferential errors between conditions

Regions where activation for inferential errors averaged across all players was greater for the Donation than the Savings condition. R indicates right. Color bar indicates *t*-statistic. Activations thresholded voxelwise at $p < 0.001$ with a 10-voxel extent minimum for display.

Table 1

Performance and reaction time

Block of trials	% correct (*SEM*)		Reaction time, ms (*SEM*)	
	Donation	**Savings**	**Donation**	**Savings**
1st	59.67 (2.34)	59.63 (3.75)	1095.38 (48.93)	1161.97 (67.82)
2nd	62.67 (2.96)	63.70 (2.10)	970.01 (42.16)	1051.15 (61.37)
3rd	68.67 (1.82)	67.04 (2.32)	987.84 (48.92)	1026.48 (69.32)
4th	68.00 (3.08)	68.15 (2.32)	930.07 (44.31)	1105.13 (70.67)

Note. $n = 38$ (20 in Donation condition, 18 in Savings condition). Blocks are 15 trials long. Standard errors of the mean (*SEM*) are calculated within block and condition. See also Supplemental Table 1 online.

Table 2

Activation during inference phase (standard model)

Region	Peak Z-score	X	Y	Z	Cluster size (vox)
High > Low (Donation > Savings)					
Middle temporal gyrus	4.41	62	−24	−14	40
Medial precuneus	4.07	14	−54	60	17
Ventromedial PFC	3.75	0	42	−8	58*
Rostromedial PFC	3.70	4	62	0	12
Thalamus	3.54	8	−6	8	12
High > Low (Savings > Donation)					
Medial parietal cortex	4.18	−8	−28	66	13
High > Low (Both conditions)					
Rostromedial PFC	4.03	−10	62	18	29
Low > High (Both conditions)					
Dorsolateral PFC	4.17	24	22	46	14
Dorsomedial PFC	4.04	−2	18	44	124*
Anterior cingulate	3.69	8	26	36	*a*
Anterior insula/putamen	3.99	−24	22	−8	34
Cuneus	3.63	12	−78	14	17
Cuneus	3.51	16	−68	16	12
Dorsolateral PFC	3.41	40	24	34	13

Note.

[a] indicates subpeaks within a cluster. PFC = prefrontal cortex.

* cluster size p < 0.05 corrected for multiple comparisons across the whole brain.

Activations in table were thresholded voxelwise at $p < 0.001$ and with a cluster size ≥ 10 voxels (whole-brain corrected cluster-size threshold = 57 voxels). T-statistics were converted to Z-scores for reporting. Coordinates are reported in MNI/ICBM152 coordinates, as in SPM5. Resampled voxel size was 2 × 2 × 2 mm. See also Supplemental Table 2 online for activation within conditions.

Table 3

Activation correlated with inferential errors during feedback phase (reinforcement learning model)

Region	Peak Z-score	X	Y	Z	Cluster size (vox)
Donation > Savings					
Parahippocampal gyrus	4.25	32	−38	−16	15
Dorsolateral PFC	3.92	−40	6	30	11
Cuneus	3.74	4	−68	18	28
Temporoparietal junction	3.72	48	−74	20	13
Dorsolateral PFC	3.58	−20	16	54	15
Savings > Donation					
Middle frontal gyrus	4.28	−30	22	18	16
Both conditions (positive correlations)					
Lateral parietal cortex	4.15	44	−24	48	21
Both conditions (negative correlations)					
Posterior cingulate	4.04	34	−64	12	12
Cuneus	3.80	18	−80	24	13

Note. PFC = prefrontal cortex. Activations in table were thresholded voxelwise at $p < 0.001$ and with a cluster size ≥ 10 voxels (whole-brain corrected cluster-size threshold = 65 voxels). T-statistics were converted to Z-scores for reporting. Coordinates are reported in MNI/ICBM152 coordinates, as in SPM5. Resampled voxel size was 2 × 2 × 2 mm. See also Supplemental Table 3 online for activation within conditions.