Neural responses to sanction threats in two-party economic exchange

Jian Li[a,1], Erte Xiao[b], Daniel Houser[c], and P. Read Montague[a,d,2]

[b]Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213; [c]Interdisciplinary Center for Economic Science, Department of Economics, George Mason University, Fairfax, VA 22030; and [a]Department of Neuroscience and [d]Menninger Department of Psychiatry & Behavioral Sciences, Baylor College of Medicine, Houston, TX 77030

Communicated by Vernon L. Smith, George Mason University, Fairfax, VA, August 11, 2009 (received for review April 22, 2008)

Sanctions are used ubiquitously to enforce obedience to social norms. However, recent field studies and laboratory experiments have demonstrated that cooperation is sometimes reduced when incentives meant to promote prosocial decisions are added to the environment. Although various explanations for this effect have been suggested, the neural foundations of the effect have not been fully explored. Using a modified trust game, we found that trustees reciprocate relatively less when facing sanction threats, and that the presence of sanctions significantly reduces trustee's brain activities involved in social reward valuation [in the ventromedial prefrontal cortex (VMPFC), lateral orbitofrontal cortex, and amygdala] while it simultaneously increases brain activities in the parietal cortex, which has been implicated in rational decision making. Moreover, we found that neural activity in a trustee's VMPFC area predicts her future level of cooperation under both sanction and no-sanction conditions, and that this predictive activity can be dynamically modulated by the presence of a sanction threat.

cooperation | neuroimaging | perception shift | punishment | social norms

Sanctions are ubiquitous in modern human societies (1). The purpose of sanctions is to enforce norm obedience beyond the level that humans might achieve in the absence of punishment (2–4). However several recent field studies and laboratory experiments have established that adding monetary sanctions to an environment can reduce cooperation (5–7). Substantial speculation has arisen surrounding the source of this counterintuitive effect, including the possibility that the presence of sanctions might change individuals' perceptions of the environment, thus crowding out internal motivations for cooperation (5–8). The imposition of sanctions also might be perceived as a signal of distrust (9–11) and might create a hostile atmosphere (12, 13), leading to decreased cooperation.

Previous behavioral experiments have sought to distinguish these competing explanations. For example, a recent study (5) reported data from an experiment aimed at determining the relative importance of intentions and incentives in producing noncooperative behavior. Participants played a one-shot investment experiment in pairs. Investors sent a certain amount to trustees, requested a return on that investment, and, in some treatments, could threaten sanctions to enforce their requests. Decisions by trustees facing threats imposed (or not) by investors were compared with decisions by trustees facing threats imposed (or not) by nature. The main finding was that when not threatened, trustees typically decided to return a positive amount less than the investor requested, but when threatened, that decision was less common. This result is the same whether the sanction is imposed by a human investor or by nature, suggesting that the detrimental effect of sanctions on cooperation might not hinge specifically on trustees' perceptions of investor intentions. One explanation for such effects has been called the "perception shift" hypothesis, where a nonthreatened subject makes decisions directed by social norms and shifts to utility-driven choices in the presence of threats. In this paper, we pursue the neural substrates of such effects using an economic exchange game equipped with the possibility that a player can threaten to sanction his or her partner.

The specific brain areas of interest to the perception shift hypothesis are reasonably well established. The parietal cortex has been shown to activate in self-interested economic decision making, especially expected utility calculations (14–16). Neural networks involved in social rewards also have been heavily researched (17–28). Of particular interest to us is the orbitofrontal cortex (OFC), which is known to be reliably involved in social reward evaluation and decision making processes (15, 17, 19, 28–31). But despite the substantial neuropsychology and psychiatry literature pointing to the importance of the prefrontal cortex and the OFC in social recognition and interaction (19, 21–25, 32, 33), ours are among the first experiments informing the OFC's role in perceiving and evaluating threats of sanctions. In particular, we investigate (i) how activation patterns in the OFC depend on whether one is threatened with sanctions and (ii) whether the activity of the medial area of the OFC, the ventromedial prefrontal cortex (VMPFC), a brain area that appears to be pivotal in human decision making (15, 17, 18, 34–38), also predicts subjects' social exchange decisions.

Our study used event-related functional magnetic resonance imaging (fMRI) and an investment game that has been used previously to reliably elicit detrimental sanction effects (5, 9) [Fig. 1; also see supporting information (SI) Fig. S1]. In this game, 2 mutually anonymous participants are paired together for 10 trials. One player is assigned the role of investor and the other is assigned the role of trustee, and both players are given 10 monetary units (MUs) at the beginning of each trial (Figs. S1 and S2). The subject pairs, as well as the subjects' roles within each pair, remain fixed for the entire 10 rounds. The investor moves first and makes 3 consecutive decisions: (i) the amount of money to send to the trustee (the amount of money was tripled on the way to the trustee), (ii) the amount of money to request back from the trustee, and (iii) whether or not to impose a threat (i.e., a monetary sanction). The sanction is a fixed loss—a 4-MU deduction from the trustee's final earnings should the trustee not send back the requested amount (Fig. S1). We collected blood oxygen level–dependent (BOLD) images from trustees while they made decisions in the investment game. Investor brain activity was not monitored. Because participants played the game in fixed pairs, reputation presumably could accumulate throughout the experiment. But this presents no difficulties for our analysis, because we focus on sanction–no-sanction contrasts

Author contributions: J.L., E.X., D.H., and P.R.M. designed research; J.L., E.X., D.H., and P.R.M. performed research; J.L. and P.R.M. analyzed data; and J.L., E.X., D.H., and P.R.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]Present address: Department of Psychology, New York University, New York, NY 10003.

[2]To whom correspondence should be addressed. E-mail: rmontague@hnl.bcm.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908855106/DCSupplemental.

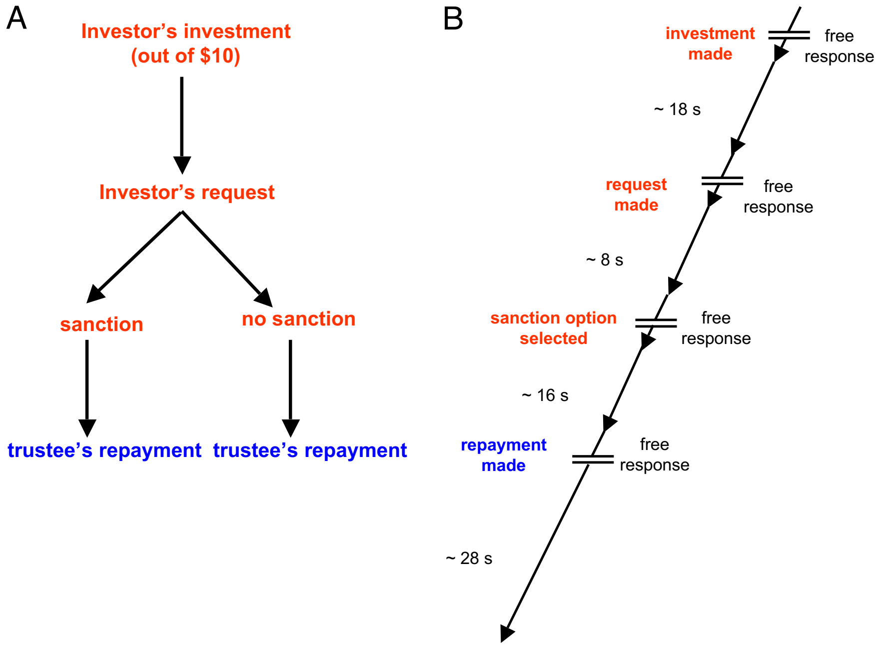

Fig. 1. Experiment task. The task involves 2 subjects sequentially exchanging MUs. Investors' choices are labeled in red; trustees' decisions, in blue. (*A*) The investor makes 3 decisions sequentially: investment amount, back-transfer request, and whether or not to threaten sanctions. Then the trustee makes the back-transfer decision. (*B*) Experiment timing. After each player makes her decision, the results are displayed simultaneously to both subjects. A total of 10 rounds are played, and at the end of each round each player's earnings are revealed to both players (also see Figs. S1 and S2).

across all rounds and subjects, thereby controlling for any reputation effects.

Results

Sanction Decisions and Their Effect on Trustees' Repayment Decisions. On average, investors imposed threats of sanctions 49.3% of the time following a trustee's decision to defect and 46% of the time following a trustee's cooperation. Out of 52 investors, 8 imposed sanctions on every trial, while 11 never imposed a sanction. Overall, an investor's decision to impose a threat was uncorrelated with whether or not a trustee defected in the previous period (P = .78; two-sample χ^2 test); however, an investor was more likely to use sanctions in a given trial if (i) the trustee defected in the previous trial and (ii) a sanction had not been used in that previous trial (χ^2 = 23.38; P = .001). Overall, investors chose the sanction option 46.3% of the time, ranging from a high of 53.7% (round 9) to a low of 37.0% (round 1). Using a mixed-effect analysis including a one-sample t test and logistic regression, we found that the correlation between the use of sanctions and the round number did not survive statistical thresholds (average sigmoid slope, 1.64; P = .053). Three important variables—investor's investment (mean slope, −0.048; P = .52), investor's request (mean slope, −0.013; P = .87), and trustee's repayment (mean slope, −0.03; P = .64)—are not correlated with round numbers.

To assess trustees' behavioral responses to sanction threats, we first plot an "equal split" strategy as a baseline (Fig. 2*B*, dotted line). This strategy could emerge if a trustee were to treat the tripled investment amount as a common good and demand half of it. We compare this to trustees' mean real repayments when threatened and when not threatened with sanctions (Fig. 2*B*, blue and red lines, respectively). Each vertical line in the figures

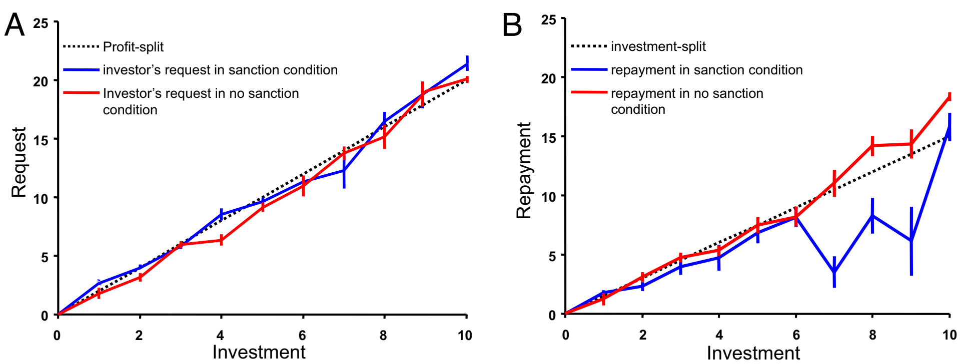

Fig. 2. Summary of players' decisions when sanctions are threatened versus not threatened. Error bars represent SEM. (*A*) The investor's request as a function of the investment amount. The dotted line indicates a request of two-thirds of the tripled investment amount, which implies equal earnings for investor and trustee. The blue and red curves indicate investors' requests under the threat and no-threat of sanctions condition, respectively. (*B*) The trustee's repayment as a function of investor's investment. The dotted line indicates a back-transfer amount of half of the tripled investment. The blue and red curves indicate trustee's back-transfer under the threat and no-threat of sanctions condition, respectively (also see Fig. S3).

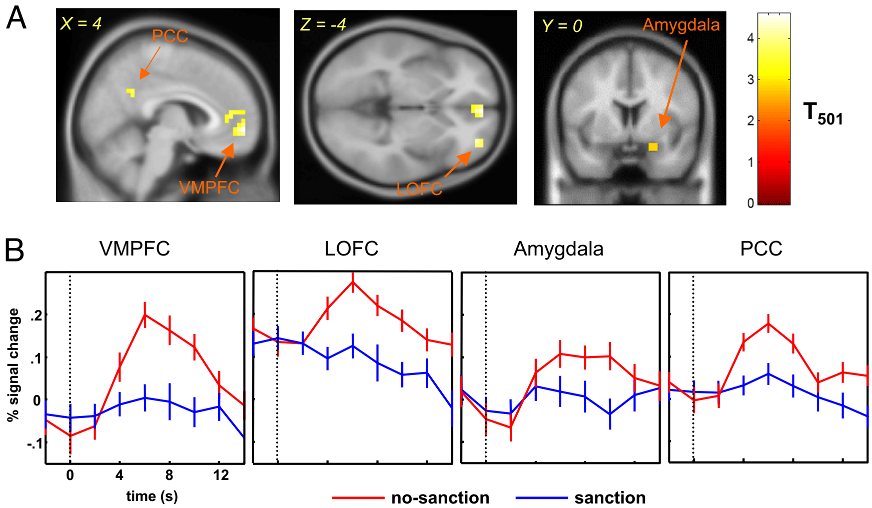

Fig. 3. The trustee's brain regions showing greater activation in the no-sanction condition than in the sanction condition ($P < .001$, uncorrected; cluster size $k > 5$ voxels). (*A*) A random-effects GLM analysis reveals several brain regions significantly more activated by the revelation of no sanction. These regions include the VMPFC (peak activation MNI coordinate [4 56 −4]), right amygdala (peak activation MNI coordinate [24 0 −20]), right LOFC (peak activation MNI coordinate [32 52 −4]), and PCC (peak activation MNI coordinate [4 −24 36]). (*B*) Mean event-related time courses of the 4 brain regions. The dashed line indicates the time onset; error bars are SEM. Bold signals in the VMPFC, LOFC, amygdala, and PCC are all significantly greater when the trustee is in the no-sanction condition (red traces) than when she is in the sanction condition (blue traces).

represents 1 SE of the trustees' mean repayment in both conditions. The trustee's repayment when threatened with sanctions is significantly different between sanction and no-sanction cases ($P < .05$; two-sample t test); see Fig. S3 and Table S1 for details. The difference is greater when the investments are larger (>6). Overall, trustees' average repayments are 6.05 MUs in sanction cases and 12.04 MUs in no-sanction cases (Table S1). Thus, the difference in repayment amounts cannot be explained solely by the possibility that trustees choose to keep 4 extra MUs in the sanction condition as compensation for the cost of the sanction.

Previous research suggests that trustees' repayments also might depend on whether the investor used the sanction to enforce an "unfair" request (5) (defined as a request for 2/3 of the tripled investment amount, which is the amount that equalizes investor and trustee earnings). To investigate unfair requests, we first explored investor behavior by plotting the back-transfer request against the investment decision for both the sanction and no-sanction conditions (Fig. 2*A*, blue and red lines, respectively). The dotted line in that figure indicates a request of 2/3 of the tripled investment. It is apparent that the investors' requests do not differ significantly between the sanction and no-sanction conditions ($P = .9$; t test), nor are the averages significantly different on average from equal-earnings requests ($P_{\text{no-sanction}} = .9$; $P_{\text{sanction}} = .9$). With respect to trustees' decisions, consistent with previous studies (5), we find that sanctions have a detrimental effect on trustees' returns both when the investor's back-transfer request is fair and when it is unfair, and that these detrimental effects are not statistically significantly different. In particular, a fair request results in a mean return equal to 53% of the tripled investment amount, while combining sanctions with a fair request reduces returns to 47% on average. When the request is unfair, the analogous change is from 59% to 47%; this between-condition difference (a 6% vs 12% reduction) is not statistically significant ($P > .15$, two-tailed Wilcoxon test). Previous reports suggest that subjects in repeated games might adopt sophisticated Nash equilibrium strategies (39,40), and we specifically tested those hypotheses (see *SI Text* for more details).

Trustees' Neural Responses to the Revelation of Sanctions. To gain insight into the neural underpinnings of this effect, we used a standard general linear model analysis (GLM) to compare trustees' brain responses in cases where sanctions were and were not threatened by the investor. The sanction–no-sanction contrast did not identify any prefrontal brain activities at $P < .001$ (uncorrected, 5 continuous voxels; see Table S3), but the no-sanction—sanction contrast did reveal differential activation in areas implicated in social reward processing (Fig. 3; Table S2). These brain areas include the VMPFC (peak activity at MNI [4 56 −4]), lateral OFC (LOFC; peak activity at MNI [32 52 −4]), posterior cingulated cortex (PCC peak activity at MNI [4 −24 36]), and right amygdala (peak activity at MNI [24 0 −20]). We conducted a region-of-interest (ROI) analysis to further investigate these results (Fig. 3*B*). In the figure, the vertical dotted line indicates the point at which either the sanction or no-sanction screen was revealed, and the red and blue curves represent brain activities in the no-sanction and sanction conditions, respectively (23–25, 31, 35).

Brain Activity Predicts Trustee's Repayment. We used standard parametric regression analysis to explore whether a trustee's neural activity at the revelation of the sanction screen might predict her subsequent back-transfer decision (which was made about 10 or 15 seconds later). Because the absolute back-transfer from a trustee does not inform a trustee's intention to cooperate, it is sensible to normalize the back-transfer by the maximum amount that the trustee could have sent (i.e, the tripled investment amount). The back-transfer–to–tripled-transfer amount ratio is a useful measure of a trustee's willingness to cooperate. Our analysis revealed a brain area at the superior frontal gyrus (DLPFC) (peak activity at MNI [24 52 20]; $P < .005$, uncorrected) (Fig. 4*A*). The activity of this area is negatively correlated with the back-transfer–to–investment amount ratio. Further ROI analysis demonstrated that as this back-transfer ratio increases, the BOLD signal at the DLPFC area decreases, and it returns to the baseline level when the trustee cooperates fully (Fig. 4*A*, *Bottom*; each vertical bar represents 1 SEM). Positive parametric regression analysis identified several brain areas, including the medial frontal gyrus (38), the inferior frontal

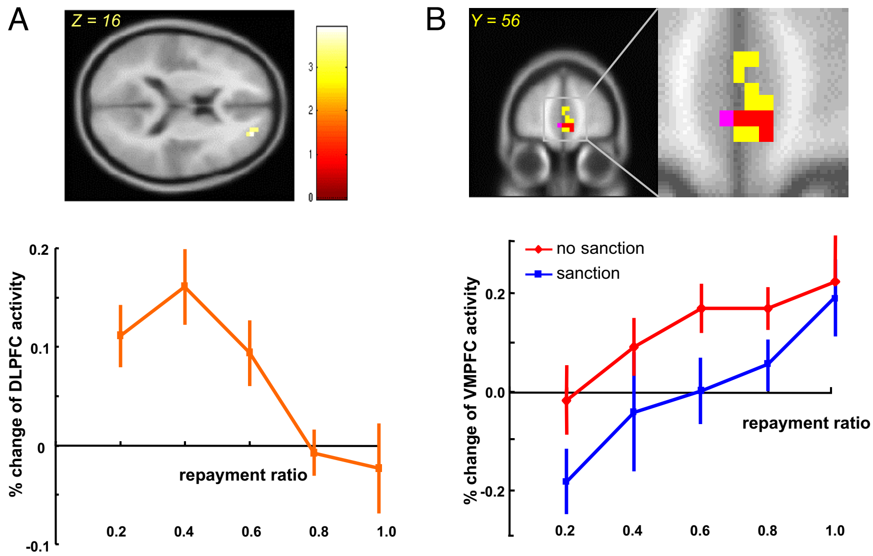

Fig. 4. Trustees' brain regions whose activations are parametrically correlated with trustees' normalized back-transfer (defined as the ratio of the back-transfer and the tripled investment amount). (*A*) Brain activity at dorsal lateral prefrontal cortex (DLPFC; peak activation MNI coordinate [24 52 20]) is negatively correlated with trustees' normalized back-transfers ($P < .001$, uncorrected; cluster size, $k > 5$ voxels). (*B*) A GLM ($P < .005$, uncorrected; cluster size, $k > 5$ voxels) showing that a subset of voxels (peak activation MNI coordinate [-4 56 −4]; purple) in the VMPFC area (yellow, with the overlap in orange) previously identified in Fig. 3*A* strongly and positively predicts trustees' normalized back-transfers. Further ROI analysis indicates that the VMPFC activity is correlated with trustees' normalized back-transfers in both sanction and no-sanction conditions. The slopes of the 2 curves (red and blue) do not differ significantly ($P = .1$, t test) while the intercept of the no-sanction curve (red) is significantly greater than that of the sanction curve (blue; $P < .01$, t test).

cortex, the middle temporal cortex, and the occipital cortex (Fig. 4*B* and Table S4; $P < .005$, uncorrected). Interestingly, one of those brain areas, the area in the VMPFC (peak activity at MNI [-4 56 −4]; Fig. 4*B*, purple) overlaps significantly with the VMPFC region identified in the previous sanction–no-sanction contrast (Fig. 4*B*, yellow; overlapping area indicated in orange).

The ROI analysis (Fig. 4*B*, *Bottom*) demonstrates this unique pattern of VMPFC activation. Although the VMPFC activity correlates with the repayment ratio in general, further separation of the VMPFC BOLD signal into sanction and no-sanction categories reveals a shift of the BOLD signal in both conditions (Fig. 4*B*; sanction in blue, no-sanction in red). Moreover, there is only weak evidence of differing slope coefficients ($P = .1$, two-sample t test); the intercepts are significantly different, however ($P < .01$, t test). It is also interesting to note that when the trustee plans to completely defect in the no-sanction situation, VMPFC activity remains at baseline, but when the trustee plans to defect under the sanction condition, VMPFC activity is well below baseline ($P < .05$, t test). The fact that brain activity at the VMFPC precedes the trustee's actual repayment choice by 10–15 seconds suggests that this brain area might be heavily involved in the trustee's final decision making and might generate a BOLD signal predicting the trustee's repayment ratio. This signal is thus responsive, in that it is susceptible to social cues (i.e., whether or not the trustee is threatened by sanctions), and also acts as a predictive signal, parametrically modulating the trustee's final repayment.

Discussion

Using an iterated version of the trust game with a sanction component, we have demonstrated an aversive effect of sanctions on human cooperation as measured by trustee's repayment in the investment game (5) (Fig. 2*B*). Recent theories that incorporate other preferences (particularly inequality aversion and kindness) shed light on motives for trustees' decisions in standard trust games (6, 41–49) but cannot explain the detrimental effect of punishment on reciprocity. We hypothesized that this effect might be due to a "perception shift" from norm-sensitive choices to utility-based choices.

Differential Brain Activities in the Sanction–No-Sanction Contrast. Our perception shift hypothesis suggests that trustees not threatened with sanctions make their reciprocity decision within a social context and are directed by social norms. Indeed, we found that when a trustee learns that he or she has not been threatened with sanctions, a neural network including the VMPFC, right amygdala, LOFC, and PCC is activated. Activation of these reward-related pathways supports our hypothesis for several reasons. Recent studies have found elevated brain activity in the LOFC area when subjects choose to comply with social norms (50, 52), while the medial part of the OFC (VMPFC) may be involved in preference generation and final decision making (17, 30, 33, 52–54). Although amygdala activation in humans has been associated with negative emotions and fear conditioning, emerging evidence suggests that the amygdala might be equally important to reward processing (22, 52, 53, 55–59). In addition, reciprocal connections between the amygdala and OFC have been studied extensively, and the functional interaction between these two regions is thought to be essential in goal-directed behaviors (53, 54, 56–59).

Differential Brain Activation in the Sanction–No-Sanction Contrast. The sanction–no-sanction contrast did not reveal any differential brain responses in the prefrontal cortex. Instead, we observed bilateral parietal cortex (LIP) activation (Table S3). Parietal activity has been linked to the representation of expected utility in primate research and "rational" choices in both primates and humans (16, 60). Our finding of no differential activation of social or emotional systems under sanction threats seems to cast some doubt on the role of negative "intentions" in affecting behavior in this environment. Instead, this finding provides convergent support for the "cognitive shift" hypothesis that credible threats of sanctions generate a cognitive shift that diminishes social motivations and increases the likelihood of market-oriented earnings maximizing behavior (5–8).

Evidence of the VMPFC as a Neural Integrator. The perception shift hypothesis requires the presence of a neural integrator to evaluate and compare inputs from various neural networks. Such an integrator would be expected to produce a signal that reliably predicts subjects' decisions. The VMPFC is anatomically and functionally well suited to play this role, in that it projects to several brain areas that are heavily involved in reward valuation, preference generation, and decision making (e.g., striatum, amygdala, hippocampus, parietal cortex) and also is known to have intense local connections with the LOFC. In investigating whether VMPFC activation predicts decisions, we indeed found that VMPFC activity is positively correlated with trustees' repayment ratio in both the sanction and no-sanction conditions. The specific brain area, revealed by linear regression analysis using the trustees' repayment ratio as an independent regressor, overlaps with the VMPFC area previously identified using the sanction–no-sanction contrast (61) (Fig. 4*B*). We also performed a ROI analysis of the overlapped region of the VMPFC. A simple linear fit of VMPFC activation on repayment amount in both sanction and no-sanction conditions indicated no statistically significant difference in the estimated slope coefficients between conditions, but a statistically significant difference in intercepts (Figs. 3 and 4*B*).

Our findings regarding the VMPFC echo those of previous studies in which investigators, using a different paradigm, reported data suggesting that activations in a neural network including the VMFPC positively reinforce reciprocal altruism (41). But our study is unique in that it shows that VMPFC activity not only predicts trustee's reciprocal decisions, but also is susceptible to emotionally salient social cues (particularly sanction or no sanction). Taken together, these results may point to a common ground for the neural representation and interaction of monetary and social rewards (18, 38, 58, 59, 61).

Methods

Task Description. Healthy subjects age 18–58 years ($n = 104$; 61 females; mean age, 28.2 ± 0.7 years) participated in the task. Half of the subjects were randomly assigned as investors, and the other half were assigned as trustees. The 52 investors (36 females) ranged in age from 20 to 58 years (mean age, 31.1 ± 1.2 years), and the 52 trustees (25 females) ranged in age from 18 to 35 years (mean age, 25.4 ± 0.4 years). All subjects had normal or corrected vision and had no previous or current neurologic or psychiatric conditions or structural brain abnormalities. All subjects were recruited through advertisements in local newspapers and internal school flyers. Informed consent was obtained using consent from approved by the Baylor College of Medicine's Institutional Review Board.

For testing, the subject lay supine with the head in the scanner bore and observed the rear-projected computer screen via a 45-degree mirror mounted above the face on the head coil. The subject's choices were registered using 2 fMRI-compatible button boxes.

Image Analysis and Statistical Analysis. See SI Text for details.

ACKNOWLEDGMENTS. We thank N. Apple for the experimental design and C. Bracero and J. McGee for the fMRI image collection. This research was supported by grants from the Kane Family Foundation (to P.R.M.), the National Institute of Neurological Disorders and Stroke (NS-045790, to P.R.M.), and the National Institute on Drug Abuse (DA-11723, to P.R.M.).

ECONOMIC SCIENCES

NEUROSCIENCE

1. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
2. Bolton P, Dewatripont M (2005) *Contract Theory* (MIT Press, Cambridge, MA), pp 553–600.
3. Camerer CF (2003) . *Behavioral Game Theory*: *Experiments in Strategic Interaction* (Princeton Univ Press, Princeton, NJ).
4. Andreoni J, Harbaugh W, Vesterlund L (2003) The carrot or the stick: Rewards, punishments, and cooperation. *Am Econ Rev* 93:893–902.
5. Houser D, Xiao E, McCabe K, Smith V (2008) When punishment fails: Research on sanctions, intentions and non-cooperation. *Games Econ Behav* 62:509–532.
6. Gneezy U, Rustichini A (2000) A fine is a price. *J Legal Studies* 29:1–17.
7. Deci EL, Koestner RM, Ryan RA (1999) Meta-analytic review of experiments examining the effect of extrinsic rewards on intrinsic motivation. *Psychol Bull* 125:627–668.
8. Lepper M, Greene D (1978) The hidden cost of reward. *New Perspectives on the Psychology of Human Motivation* (Wiley, New York).
9. Fehr E, Rockenbach B (2003) Detrimental effects of sanctions on human altruism. *Nature* 422:137–140.
10. Dickinson D, Villeval M (2004) Does monitoring decrease work effort? The complementarity between agency and crowding-out theories. *IZA Discussion Papers* 1222.
11. Frey B (1993) Does monitoring increase work effort? The rivalry between trust and loyalty. *Econ Inquiry* 31:663–670.
12. Bewley T (1999) *Why Wages Don't Fall During a Recession* (Harvard Univ Press, Cambridge, MA).
13. Frey B (1998) in *Not Just for the Money*: *An Economic Theory of Personal Motivation* (Beacon Press, Boston, MA).
14. Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron* 44:365–378.
15. Glimcher PW (2002) Decisions, decisions, decisions: Choosing a neurobiological theory of choice. *Neuron* 36:323–332.
16. Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
17. Glimcher PW, Rustichini A (2004) Neuroeconomics: The consilience of brain and decision. *Science* 306:447–452.
18. Kringelbach ML (2005) The human orbitofrontal cortex: Linking reward to hedonic experience. *Nat Rev Neurosci* 6:691–702.
19. Seguin JR (2004) Neurocognitive elements of antisocial behavior: Relevance of an orbitofrontal cortex account. *Brain Cogn* 55:185–197.
20. Camille N, et al. (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304:1167–1170.
21. Adolphs R (2001) The neurobiology of social cognition. *Curr Opin Neurobiol* 11:231–239.
22. Veit R, et al. (2002) Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neurosci Lett* 328:233–236.
23. Adolphs R (2003) Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci* 4:165–178.
24. Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005) Opinion: The neural basis of human moral cognition. *Nat Rev Neurosci* 6:799–809.
25. Pellis SM, et al. (2006) The effects of orbital frontal cortex damage on the modulation of defensive responses by rats in playful and nonplayful social contexts. *Behav Neurosci* 120:72–84.
26. King-Casas B, et al. (2005) Getting to know you: Reputation and trust in a two-person economic exchange. *Science* 308:78–83.
27. King-Casas B, et al. (2008) The rupture and repair of cooperation in borderline personality disorder. *Science* 321:806–810.
28. Fellows LK, Farah MJ (2003) Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain* 126:1830–1837.
29. Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 8:170–177.
30. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310:1680–1683.
31. Volz KG, Schubotz RI, von Cramon DY (2006) Decision-making and the frontal lobes. *Curr Opin Neurol* 19:401–406.
32. Coccaro EF, McCloskey MS, Fitzgerald DA, Phan KL (2007) Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biol Psychiatry* 62:168–178.
33. Schaefer M, Rotte M (2007) Thinking on luxury or pragmatic brand products: Brain responses to different categories of culturally based brands. *Brain Res* 1165:98–104.
34. Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295.
35. Bechara A, Damasio H, Damasio AR, Lee GP (1999) Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J Neurosci* 19:5473–5481.
36. Damasio A (2006) in *Descartes' Error* (Vintage Rand, London), p 352.
37. Kringelbach ML, Rolls ET (2004) The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Prog Neurobiol* 72:41–72.
38. McClure SM, et al. (2004) Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44:379–387.
39. Friedman J (1985) Cooperative equilibria in finite-horizon noncooperative supergames. *J Econ Theor* 35:390–398.
40. Benoit JP, Krishna V (1985) Finitely repeated games. *Econometrica* 53:905–922.
41. Rilling J, et al. (2002) A neural basis for social cooperation. *Neuron* 35:395–405.
42. Falk A, Fischbacher U (2006) A theory of reciprocity. *Games Econ Behav* 54:293–315.
43. Falk A, Fehr E, Fischbacher U (2003) On the nature of fair behavior. *Econ Inquiry* 41:20–26.
44. Cox J (2004) How to identify trust and reciprocity. *Games Econ Behav* 46:260–281.
45. Engelmann D, Strobel M (2004) Inequity aversion, efficiency and maximin preference in simple distribution experiments. *Am Econ Rev* 94:857–869.
46. Rabin M (1993) Incorporating fairness into game theory and economics. *Am Econ Rev* 83:1281–1302.
47. Fehr E, Schmidt K (1999) A theory of fairness, competition and cooperation. *Q J Econ* 114:817–868.
48. Dufwenberg M, Kirchsteiger G (2004) A theory of sequential reciprocity. *Games Econ Behav* 47:268–298.

49. Ellingsen T, Johannesson M (2008) Pride and prejudice: The human side of incentive theory. *Am Econ Rev* 98:990–1008.
50. Montague PR, Lohrenz T (2007) To detect and correct: Norm violations and their enforcement. *Neuron* 56:14–18.
51. Spitzer M, Fischbacher U, Herrnberger B, Gron G, Fehr E (2007) The neural signature of social norm compliance. *Neuron* 56:185–196.
52. Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301:1104–1107.
53. Winstanley CA, Theobald DE, Cardinal RN, Robbins TW (2004) Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J Neurosci* 24:4718–4722.
54. Arana FS, et al. (2003) Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *J Neurosci* 23:9632–9638.
55. Schultz W (2000) Multiple reward signals in the brain. *Nat Rev Neurosci* 1:199–207.
56. Baxter MG, Parker A, Lindner CC, Izquierdo AD, Murray EA (2000) Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J Neurosci* 20:4311–4319.
57. Moll J, et al. (2002) The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci* 22:2730–2736.
58. Holland PC, Gallagher M (2004) Amygdala–frontal interactions and reward expectancy. *Curr Opin Neurobiol* 14:148–155.
59. Schoenbaum G, Chiba AA, Gallagher M (1998) Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci* 1:155–159.
60. McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306:503–507.
61. Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:46–48.