Neural responses to sanction threats in two-party economic exchange

J a L [a,1], E X a [b], Da H [c], a d P. R ad M a [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)

Sa c a d b c b d c ca . H , c d d ad ab a a d - a d a c a d c d c a ca d c a add d . A . a a a c a b . . d, a da c a b d. U a d d . a , d a c ca a ac . a c a , a d a c a c ca d c ' b a ac d c a a d a a [da a c (VMPFC), a a b a c , a d a . da a] - a c a b a c a a c , c a b ca d a a d c a . M , d a a ac a ' VMPFC a a d c c a d b a c a d - a c c d , a d a d c a c a c ca b d a ca d a d b c a a c a .

cooperation | neuroimaging | perception shift | punishment | social norms

S (1). (2, 4). H (5, 7). - (5, 8). (9, 11) (12, 13), P . F , (5) . P . I D ( ) ( ) . O . I

(14, 16). N (17, 28). O . (OFC), (15, 17, 19, 28, 31). B . OFC (19, 21, 25, 32, 33), OFC' I , ( ) OFC ( ) OFC, (MPFC), (15, 17, 18, 34, 38), O (MRI) (5, 9) F . 1; (I) F . 1 . I , 2 10 . O 10 (M) (F . 1a 2). 10 3 : ( ) ( ), ( ) , ( ) (. .,). 4-M (F . 1). (BOLD) . I . B . B

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.

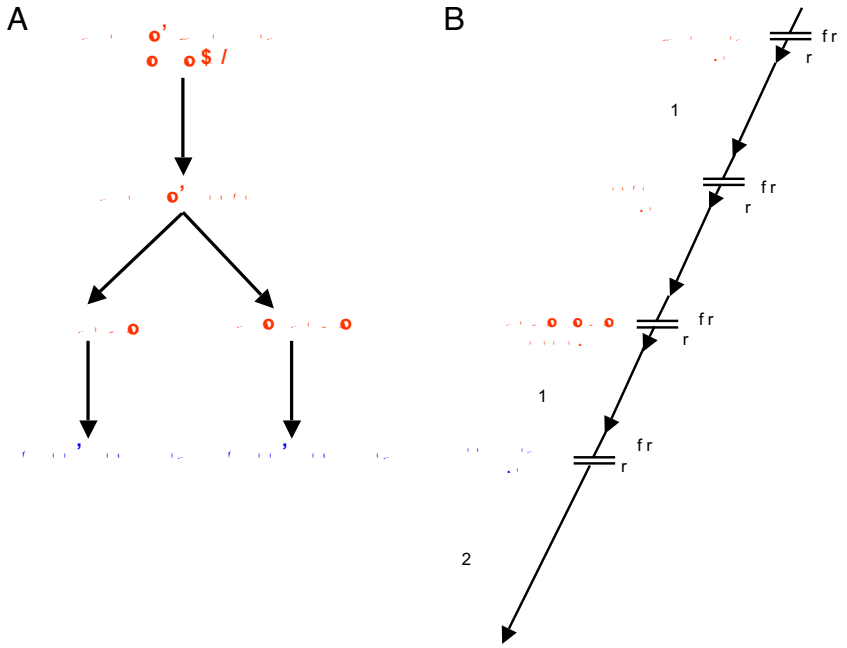
The authors declare no conflict of interest.

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.

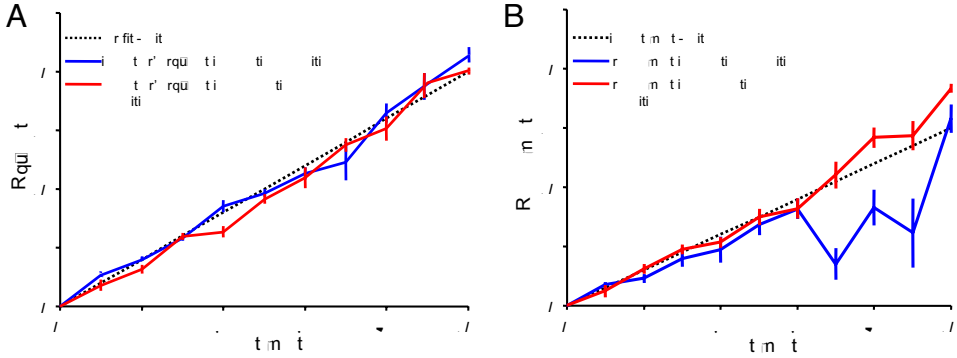


F. . 1. Experiment task. The task involves 2 subjects sequentially exchanging MUs. Investors' choices are labeled in red; trustees' decisions, in blue. (*A*) The investor makes 3 decisions sequentially: investment amount, back-transfer request, and whether or not to threaten sanctions. Then the trustee 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).

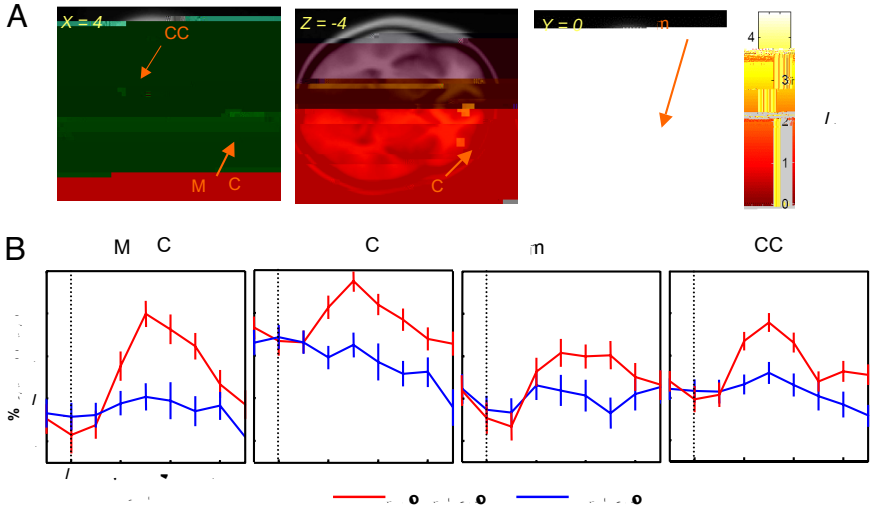
Results

Sanction Decisions and Their Effect on Trustees' Repayment Decisions.

O , 49.3% 46% . O 52 , 8 11 O ; (*P* ¤ .78; - *χ*²); , ( ) ( ) (*χ*² ¤ 23.38; *P* ¤ .001). O , 46.3% , 53.7% (9) 37.0% (1). - - t , 1.64; *P* ¤ .053). (, ¤ 0.048; *P* ¤ .52), (, ¤ 0.013; *P* ¤ .87), (; ¤ 0.03; *P* ¤ .64) (F . 2*B*,). (F . 2*B*,). □



F. . 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).



F . 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).

1 P

(*P* ▫ .05; t); F . 3 1 (▫ 6). O , 6.05 M 12:04 M (1). 4 . M

P

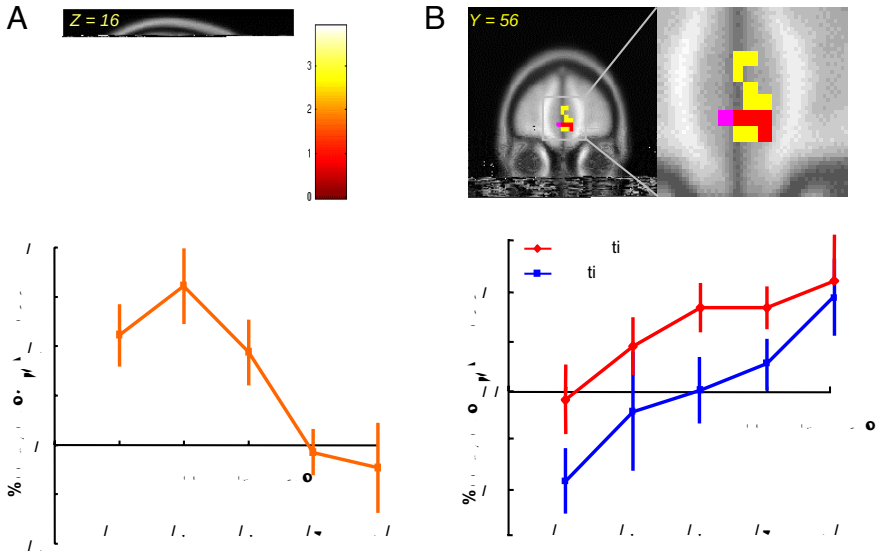
" (5) (2/3). (F . 2*A*,). 2/3 . I (*P* ▫ .9; *t*), ($P_{no\text{-}sanction}$ ▫ .9; $P_{sanction}$ ▫ .9). (5), . I 53% 47% 59% 47%; (▫ 6% 12%). P (*P* ▫ .15, N (39,40), (*SI Text*).

Trustees' Neural Responses to the Revelation of Sanctions.

(GLM) *P* ▫ .001 (, 5 ; 3), (F . 3; 2). MPFC (MNI 4 56 ▫ 4), OFC (LOFC; MNI 32 52 ▫ 4), (PCC MNI 4 ▫ 24 36), (MNI 24 0 ▫ 20). (ROI) (F . 3*B*). 1 (23 25, 31, 35).

Brain Activity Predicts Trustee's Repayment.

10 15). B (. ,). O (DLPFC) (MNI 24 52 20 ; *P* ▫ .005,) (F . 4*A*). . P ROI BOLD DLPFC (F . 4*A*, *Bottom*; 1 M). P (38),



F. .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).

(F . 4*B* 4; P ▫ .005,). I MPFC (MNI -4 56 ▫4 ; F . 4*B*,) MPFC (F . 4*B*, ;). ROI (F . 4*B*, *Bottom*) MPFC . A MPFC MPFC BOLD BOLD (F . 4*B*;). M (P ▫ .1, t); (P ▫ .01, t). I MPFC MPFC (P ▫ .05, t). MFPC 10 15 BOLD (.),

Discussion

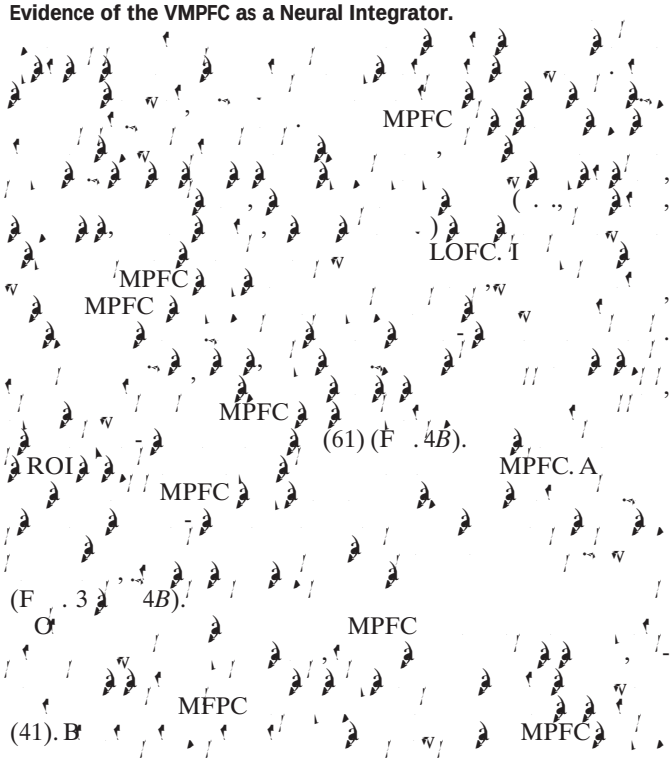
(5) (F . 2*B*). R ( ) (6, 41, 49) .

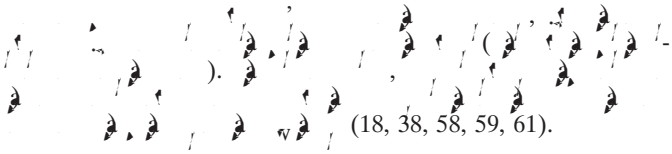
”

Differential Brain Activities in the Sanction–No-Sanction Contrast. O . I MPFC, LOFC, PCC . A . R LOFC (50, 52), OFC (MPFC) (17, 30, 33, 52, 54). A (22, 52, 53, 55, 59). I OFC (53, 54, 56, 59).

Differential Brain Activation in the Sanction–No-Sanction Contrast. . . I . (LIP) (3). P (16, 60). O . I , ” (5, 8).

Evidence of the VMPFC as a Neural Integrator.





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 face on 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.

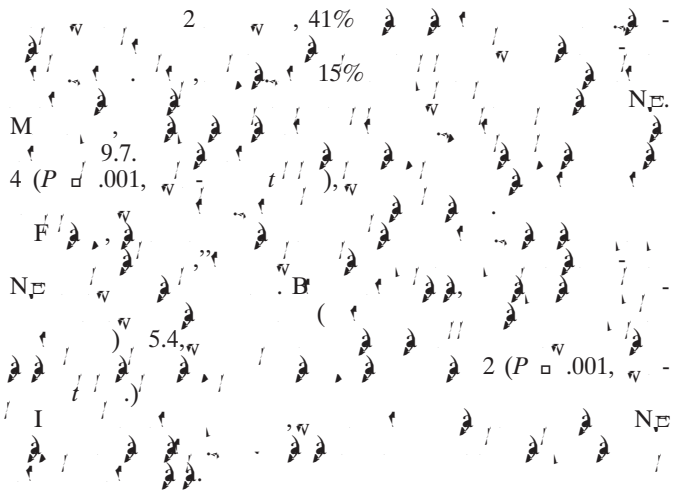
Image Analysis and Statistical Analysis. See SI Text for details.

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.).

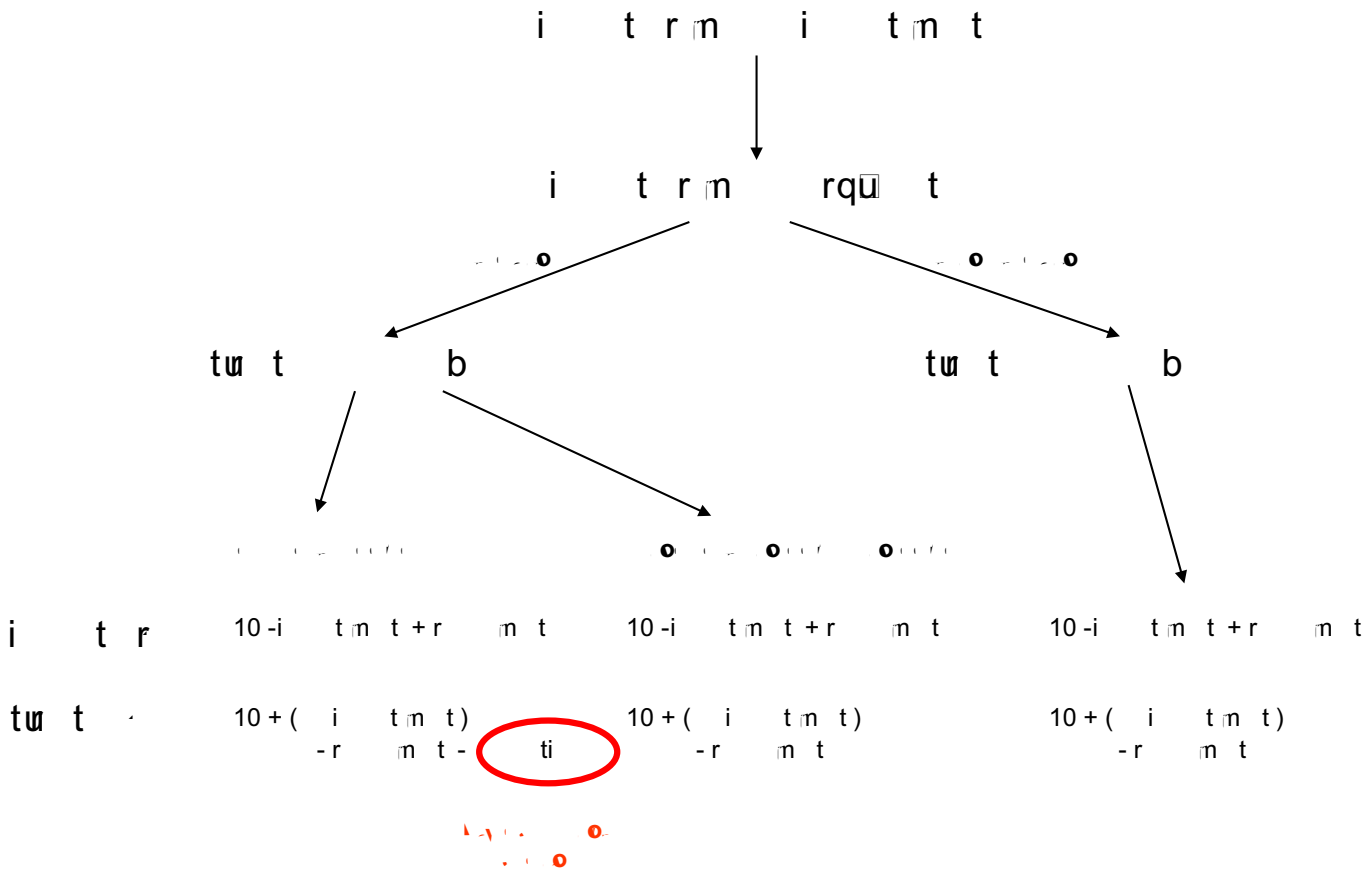
1. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
2. Bolton P, Dewatripont M (2005) *Contract Theory* (MIT Press, Cambridge, MA), pp 553–600.
3. Camerer CF (2003) . *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ Press, Princeton, NJ).
4. 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.

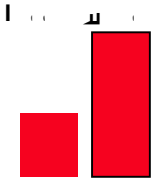
S . I a
L a . 10.1073/ a .0908855106

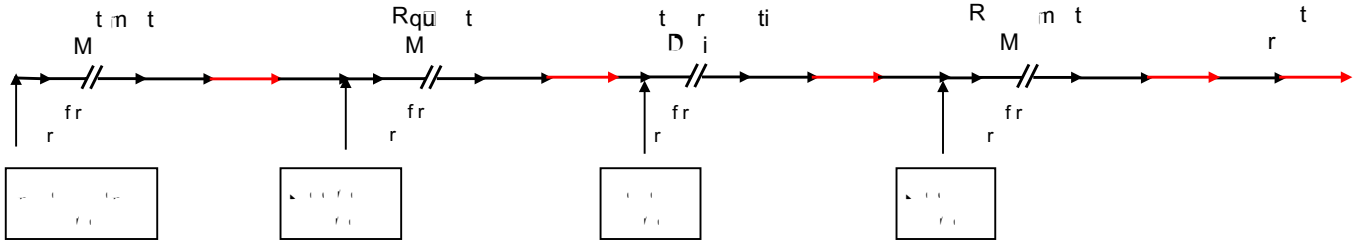


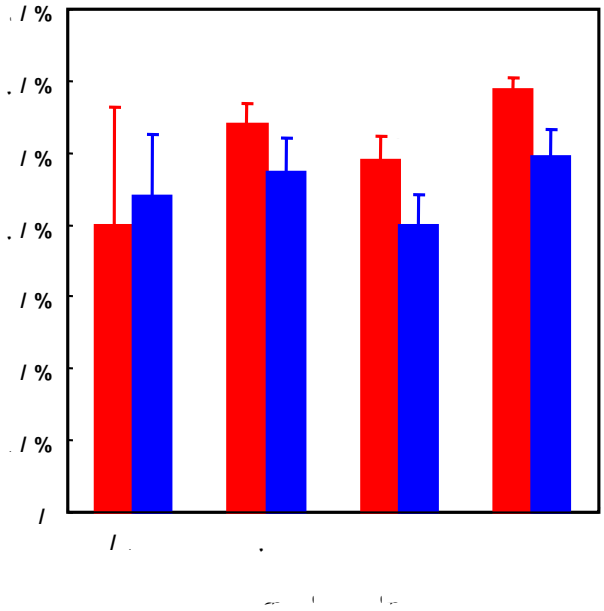
Exploring the "Once Commodity, Always Commodity" Hypothesis. P

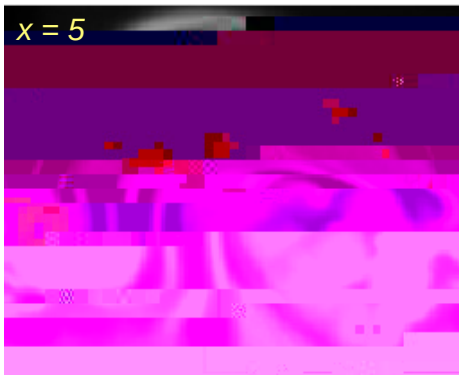


F. . S1. The 2-player investment game. Two players are paired anonymously. Both the investor and the trustee are endowed with 10 points at the beginning of each round of the experiment (10 rounds total). The investor first decides how many points to invest, how many to request back and, whether or not to threaten punishment. The trustee observes these 3 pieces of information and then decides how many points to send back to the investor. If the trustee returns less than the investor requested, and if the investor chose the threat option, then a penalty of 4 points is deducted from the trustee's final earnings. If the threat was not chosen, then the trustee's and the investor's earnings depend only on the amounts sent and returned, respectively, as described above.




M
M
M




x = 5

Tab S1. A ą b a a d a a d

	Sanction	No sanction	Significance
Investment	4.89	7.09	*
Request	10.6	13.89	-
Request/(3 ▫ investment)	0.72	0.64	*
Repayment	6.05	12.04	-
Repayment/(3 ▫ investment)	0.46	0.55	*
Repayment/request	0.67	0.89	*
Investor's payoff	11.58	14.95	*
Trustee's payoff	17.01	19.22	-

*Statistical significance

Tab S2. B a d a ac a d - a c a c c d

Region of activation	Peak MNI coordinates			Voxels	*z*-value
	X	*Y*	*Z*		
Medial frontal gyrus (R)	4	56	▫4	83	4.45
Superior temporal gyrus (R)	48	16	▫16	52	4.52
Superior temporal gyrus (L)	▫48	16	▫12	31	3.76
Lateral frontal gyrus (R)	32	52	▫4	15	4.03
Superior frontal gyrus (R)	20	40	48	35	3.78
Superior frontal gyrus (L)	▫28	40	36	24	3.26
Occipital lobe (R)	12	▫92	12	12	3.07
Occipital lobe (L)	▫16	▫88	▫16	19	3.58
Precuneus (R)	4	▫52	32	12	3.49
Posterior cingulate cortex	4	▫24	36	11	3.41
Inferior frontal gyrus (R)	52	24	4	5	2.78
Amgdala (R)	24	0	▫20	7	2.7

Regions with ≥5 significant voxels were identified at P ▫ .005 (uncorrected).

Tab S3. B a d a ac a d a c - a c c d

Region of activation	Peak MNI coordinates			Voxels	*z*-value
	X	*Y*	*Z*		
Parietal lobe (L)	−24	−60	52	72	3.99
Parietal lobe (R)	28	−48	40	81	4.13
Inferior temporal gyrus	−44	−68	−4	67	4.1
Temporal lobe	28	−68	20	27	3.29
Precentral gyrus (R)	44	−4	36	68	3.97
Precentral gyrus (L)	−44	−8	36	80	3.79
Fusiform gyrus (R)	36	−48	−16	18	3.63
Medial frontal gyrus	−8	−24	68	17	3.3
Midbrain	4	−12	−12	59	4.17
Cerebellum	24	−48	−36	44	4.19

Regions with ≥5 significant voxels were identified at P < .005 (uncorrected).