Vol 463 | 25 February 2010 | doi:10.1038/nature08785

Neural evidence for inequality-averse social preferences

Elizabeth Tricomi[1], Antonio Rangel[2,3], Colin F. Camerer[2,3] & John P. O'Doherty[2,3,4]

A popular hypothesis in the social sciences is that humans have social preferences to reduce inequality in outcome distributions because it has a negative impact on their experienced reward[1–3]. Although there is a large body of behavioural and anthropological evidence consistent with the predictions of these theories[1,4–6], there is no direct neural evidence for the existence of inequality-averse preferences. Such evidence would be especially useful because some behaviours that are consistent with a dislike for unequal outcomes could also be explained by concerns for social image[7] or reciprocity[8,9], which do not require a direct aversion towards inequality. Here we use functional MRI to test directly for the existence of inequality-averse social preferences in the human brain. Inequality was created by recruiting pairs of subjects and giving one of them a large monetary endowment. While both subjects evaluated further monetary transfers from the experimenter to themselves and to the other participant, we measured neural responses in the ventral striatum and ventromedial prefrontal cortex, two areas that have been shown to be involved in the valuation of monetary and primary rewards in both social and non-social contexts[10–14]. Consistent with inequality-averse models of social preferences, we find that activity in these areas was more responsive to transfers to others than to self in the 'high-pay' subject, whereas the activity of the 'low-pay' subject showed the opposite pattern. These results provide direct evidence for the validity of this class of models, and also show that the brain's reward circuitry is sensitive to both advantageous and disadvantageous inequality.

A pervasive notion in social science is that human preferences and behaviour are sensitive to inequality considerations[1–3]. This suggestion is based on considerable experimental and field evidence. Small-scale societies often share through norms of informal social insurance[15], and large-scale societies share through governmental welfare transfers and progressive taxation[16]. In corporations, wages are typically secret and vary less than productivity does, as though workers have a strong aversion for earning less than others do[17]. Workers also seem to reciprocate when they feel companies have treated them well[18], but withdraw effort when they feel wronged[19]. These social patterns have been replicated under controlled conditions in many behavioural economics experiments: participants regularly share wealth with strangers[20], punish non-cooperators at a cost to themselves[4,20,21], and reject unfair divisions of a pool of money[22].

Subjective ratings, and preferences inferred from choices, provide evidence that subjects like transfers that reduce inequality[5]. However, it is unknown whether reward structures in the brain respond to self–other monetary gains in ways that reflect a preference for reducing inequality. The missing neural evidence is important because most of the behavioural observations consistent with a dislike for unequal distributions can also be explained by concerns for social image[7] or reciprocity[8,9], rather than an aversion to inequality. Furthermore, the behavioural evidence is mixed about people's dislike of advantageous inequality (that is, a willingness to decrease their own payoff to improve those of people who are worse off). Some studies indicate that the better-off will pay to reduce an outcome gap, but other studies suggest they will pay only to maintain or to increase their relative status[23,24].

We used functional magnetic resonance imaging (fMRI) to test directly for the presence of inequality-averse social preferences in the human brain, for both positive and negative inequality. A stark and salient inequality in overall pay for experimental participation was created by recruiting 20 pairs of unacquainted male participants. They each received $30 base pay, and then drew balls labelled 'rich' or 'poor.' The 'rich' (high-pay) participant received a $50 bonus to the base pay and the 'poor' (low-pay) participant received no bonus. We then scanned the participants as they each rated their subjective valuations for further positive potential monetary transfers from the experimenter to themselves and to the other player (Fig. 1). Each transfer ranged from $0 to $50 and the set of transfers was symmetrical over the two participants. At the end of the experiment the transfers from a randomly chosen trial were paid to the subjects.

Our test focused on the response of the ventral striatum and ventromedial prefrontal cortex (vmPFC), areas known to be involved in the valuation of stimuli at the time of decision making[25,26] and with processing the experienced subjective value of receiving monetary reward and other rewards[11]. We characterized the pattern of stated inequality preferences by regressing each subject's ratings on the transfer amount to themselves and to the other person. Both groups rated transfers to themselves positively, indicating that they valued having higher earnings to themselves, although the value was lower for the high-pay group than for the low-pay group (Fig. 1b; $t_{(38)} = -3.8$, $P = 0.0005$). Consistent with previous studies[5], the low-pay group disliked falling farther behind the high-pay group ('disadvantageous inequality aversion'), because they rated positive transfers to the high-pay participants negatively, even though these transfers had no effect on their own earnings ($t = 3.4$, $P = 0.003$). Conversely, the high-pay group seemed to value transfers that closed the gap between their earnings and those of the low-pay group ('advantageous inequality aversion'), because they rated transfers to the low-pay participants positively ($t = 2.8$, $P = 0.01$). This difference in whether transfers to others are valued negatively (by the low-pay group) or positively (by the high-pay group) resulted in a significant group-by-recipient interaction ($F_{(1,20)} = 9.4$, $P < 0.01$).

In the fMRI data, on trials in which there was a transfer only to oneself, activation in the ventral striatum (coordinates −9,12,−6; $t_{(19)} = 3.6$, $P < 0.05$, small-volume-corrected) and vmPFC correlated significantly with the magnitude of the monetary transfer (coordinates −9,39,−9; $t_{(19)} = 3.75$, $P < 0.05$, small-volume-corrected),

[1]Psychology Department, Rutgers University, Newark, New Jersey 07102, USA. [2]Division of the Humanities and Social Sciences, [3]Computational and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA. [4]School of Psychology and Trinity College Institute of Neuroscience, Trinity College, Dublin 2, Ireland.

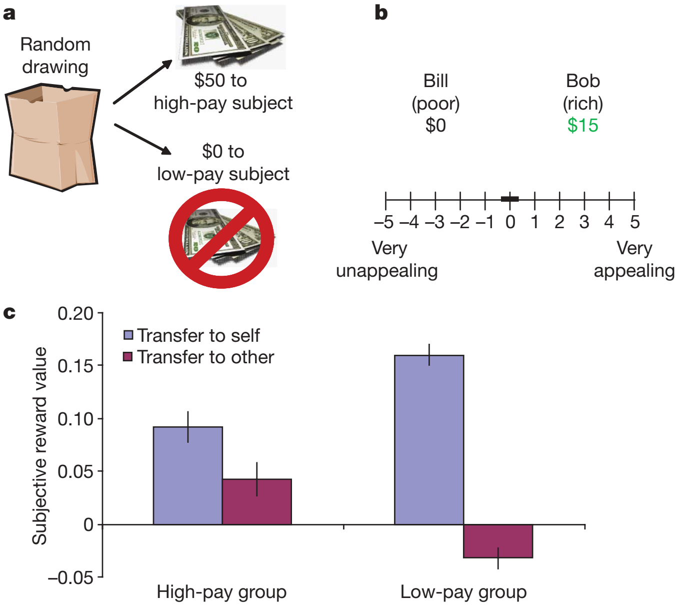

Figure 1 | Effect of inequality manipulation on behaviour. a, Two participants each drew a ball from hat. One, labelled 'rich', led to a bonus payment of $50. The other, labelled 'poor', entailed no bonus. **b**, Each participant was then scanned while they rated further potential monetary transfers to themselves and the other player on a scale from −5 to 5. **c**, Estimated coefficients from a linear regression of the behavioural ratings on the transfers to self and other. The high-pay group rated transfers to themselves less highly than the low-pay group did, while rating transfers to the other player more highly. Error bars represent s.e.m.

confirming previous findings implicating these regions in reward-related processing[11,25,26]. Next we tested for effects of our inequality manipulation on activity in these areas by computing a contrast of the difference in fMRI responses for transfers to self minus transfers to other in each of the two groups. In both regions this contrast was significantly greater for the low-pay group than for the high-pay group (Figs 2a and 3a; see also Supplementary Table 1; $t_{(38)} > 3.3$, $P < 0.05$, small-volume-corrected). This is further illustrated in Figs 2b and 3b, which show that both the ventral striatum and mPFC of low-pay subjects responded more strongly to transfers to self than other, whereas the opposite pattern was observed for high-pay subjects.

We conducted further analyses to address two potential concerns about our results. First, because high-pay and low-pay subjects had different initial wealth levels, the differences in their ratings and neural activity might have been due to 'wealth effects' and not to equity considerations. To address this we ran two behavioural versions of the experiment in which both subjects had the same initial wealth, so they were either both high-pay or both low-pay. As shown in Supplementary Fig. 1, there were no significant differences in the ratings of the two groups, which rules out this alternative explanation.

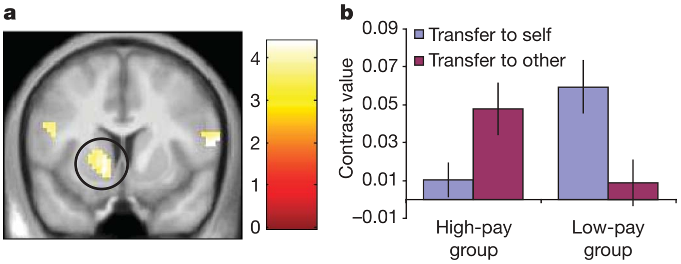

Figure 2 | Effect of inequality manipulation on activity in the ventral striatum. a, In the ventral striatum (circled), the summary statistic, '(transfer to self) minus (transfer to other)', was significantly greater for the low-pay group than the high-pay group ($P < 0.05$, small-volume-corrected). The image is shown at $P < 0.001$, uncorrected. **b**, Parameter estimates for each of the parametric regressors in the general linear model. Error bars represent s.e.m.

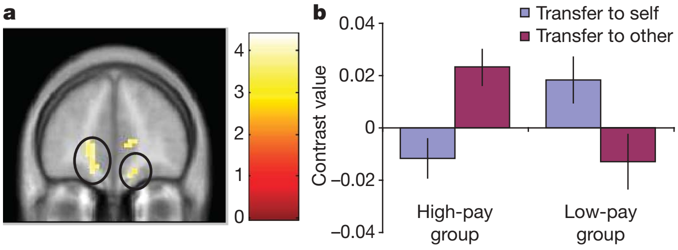

Figure 3 | Effect of inequality manipulation on activity in the vmPFC. a, In the vmPFC, bilaterally (circled regions), activity in the '(transfer to self) minus (transfer to other)' contrast was also significantly greater for the low-pay group than high-pay group ($P < 0.05$, small-volume-corrected). The image is shown at $P < 0.001$, uncorrected. **b**, Parameter estimates for each of the parametric regressors in the general linear model. Error bars represent s.e.m.

Second, because activity in ventral striatum and vmPFC has been associated with prediction errors measuring the degree of surprise in receiving rewards[10], there is a concern that some of the difference in activity between rich and poor might be explained by differences in expectations, which have nothing to do with social preferences. To address this we estimated another model in which neural activity was assumed to be modulated by prediction errors for the transfers to oneself, using an initial expected value of 50 for the high-pay subjects and 0 for the low-pay subjects, but not by the transfers to the other subject (see Methods for details). We then performed a Bayesian model comparison of the two models. The results indicate that the original model provides a better fit to the data in our regions of interest. In the left vmPFC the probability that our valuation model was a better explanation of the BOLD (blood oxygen level dependent) response profile than the prediction error model (exceedance probability; EP) was 0.92, whereas for the right vmPFC the EP was 0.9998 and for the left ventral striatum the EP was 0.76.

Although we found that both behavioural ratings and neural activity in vmPFC and striatum showed significant effects of advantageous and disadvantageous inequality aversion, the pattern of advantageous inequality aversion measured in the brain data was discernibly different from that expressed in the behavioural ratings. For high-pay subjects, neural responses associated with transfers to themselves were lower than those associated with transfers to the other, whereas the low-pay subjects showed the opposite pattern. In contrast, the stated preferences of the high-pay subjects showed a greater valuation for their own transfers than for transfers to the other. This apparent incongruity between stated behavioural ratings and brain data indicates that basic reward structures in the brain may reflect even stronger equity considerations than is necessarily expressed or acted on at the behavioural level. These findings raise the possibility that even when basic reward responses reflect strong equity considerations, in some cases additional factors may intercede to moderate the influence of such equity judgements on behaviour, such as strategizing[15] (under situations in which a competition is perceived between individuals), or the engagement of self-serving biases such as judgements of deservingness or need.

Our results provide direct neurobiological evidence in support of the existence of inequality-averse social preferences in the human brain. Although the objective values of the transfers shown to the high-pay and low-pay subjects was equivalent, the subjective value of the transfers was influenced by who received the initial $50 endowment, and activity in the striatum and vmPFC reflected this influence. This builds on a growing body of research showing that experienced subjective reward signals in the striatum and vmPFC are modulated by a variety of other social factors[13,14,27–30]. Furthermore, given that the initial wealth manipulation was small relative to the subjects' overall income, our results also underscore the strong sensitivity of how the brain's reward circuitry responds to equity considerations. This provides insight into why equity concerns seem to be such a pervasive and fundamental feature of human social exchange.

METHODS SUMMARY

Twenty pairs of healthy, previously unacquainted male participants participated in the experiment. They were each paid a base $30 fee; each then drew a ball from a hat, labelled 'rich' or 'poor'. The 'rich' (high-pay) subject received an immediate payment of $50, whereas the 'poor' (low-pay) player received no bonus payment. They then performed an identical task in consecutive fMRI scanning sessions. In each trial subjects viewed possible monetary transfers from the experimenter to themselves and to the other player, ranging from $0 to $50 (Fig. 1). Participants rated how appealing they found the possible transfers on a scale of -5 (very unappealing) to 5 (very appealing). After both players' scans, a single trial was randomly picked from the set, and the transfers for that trial were paid out.

For each participant, the behavioural ratings were regressed on the transfers to self and other. The resulting coefficients were pooled into a mixed-effects group analysis using two-sided *t*-tests to determine whether the high-pay and low-pay groups differed in their social preferences. We estimated the parameters of a general linear model of the fMRI data that included parametric effects of the transfer amounts on the BOLD signal. A random (between-subject) effects analysis was then used to identify regions that responded differentially to transfers to self versus transfers to the other player in the two groups. Images are displayed with a voxel-wise significance threshold of $P < 0.001$. Results are reported with a corrected significance threshold of $P < 0.05$, based on a small-volume correction within an 8-mm sphere centred on coordinates in the striatum and vmPFC taken from previous studies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 June; accepted 22 December 2009.

1. Fehr, E. & Schmidt, K. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114,** 817–868 (1999).
2. Adams, J. in *Advances in Experimental Social Psychology* (ed. Berkowitz, L.) 267–299 (Academic, 1965).
3. Bolton, D. & Ockenfels, A. ERC: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **82,** 166–193 (2000).
4. Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. Egalitarian motives in humans. *Nature* **446,** 794–796 (2007).
5. Loewenstein, G. F., Thompson, L. & Baserman, M. H. Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* **57,** 426–441 (1989).
6. Henrich, J. *et al.* 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28,** 795–815 (2005).
7. Andreoni, J. & Bernheim, B. Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects. *Econometrica* **77,** 1607–1636 (2009).
8. Falk, A., Fehr, E. & Fischbacher, U. Testing theories of fairness—intentions matter. *Games Econ. Behav.* **62,** 287–303 (2008).
9. Rabin, M. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **83,** 1281–1302 (1993).
10. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
11. O'Doherty, J. P. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* **14,** 769–776 (2004).
12. Fliessbach, K. *et al.* Social comparison affects reward-related brain activity in the human ventral striatum. *Science* **318,** 1305–1308 (2007).
13. Harbaugh, W. T., Mayr, U. & Burghart, D. R. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* **316,** 1622–1625 (2007).
14. Tabibnia, G., Satpute, A. B. & Lieberman, M. D. The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* **19,** 339–347 (2008).
15. Gurven, M. The evolution of contingent cooperation. *Curr. Anthropol.* **47,** 185–192 (2006).
16. Kakwani, N. C. Measurement of tax progressivity: an international comparison. *Econ. J.* **87,** 71–80 (1977).
17. Frank, R. H. Are workers paid their marginal products. *Am. Econ. Rev.* **74,** 549–571 (1984).
18. Akerlof, G. A. & Yellen, J. L. The fair wage-effort hypothesis and unemployment. *Q. J. Econ.* **105,** 255–283 (1990).
19. Krueger, A. B. & Mas, A. Strikes, scabs, and tread separations: labor strife and the production of defective Bridgestone/Firestone tires. *J. Polit. Econ.* **112,** 253–289 (2004).
20. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425,** 785–791 (2003).
21. Fehr, E. & Gachter, S. Altruistic punishment in humans. *Nature* **415,** 137–140 (2002).
22. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the Ultimatum Game. *Science* **300,** 1755–1758 (2003).
23. Frank, R. H. *Choosing the Right Pond: Human Behavior and the Quest for Status* (Oxford Univ. Press, 1985).
24. Herrmann, B., Thoni, C. & Gachter, S. Antisocial punishment across societies. *Science* **319,** 1362–1367 (2008).
25. Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W. & Rangel, A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* **28,** 5623–5630 (2008).
26. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27,** 9984–9988 (2007).
27. Rilling, J. *et al.* A neural basis for social cooperation. *Neuron* **35,** 395–405 (2002).
28. Moll, J. *et al.* Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl Acad. Sci. USA* **103,** 15623–15628 (2006).
29. Singer, T. *et al.* Empathic neural responses are modulated by the perceived fairness of others. *Nature* **439,** 466–469 (2006).
30. Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L. & Camerer, C. F. Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J. Neurosci.* **29,** 2188–2192 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Bushong, N. Tetrault, S. Bray, J. Gläscher and R. Lee for their assistance. This work was supported by grants from the US National Science Foundation to J.O.D. (0617174), from the Human Frontiers of Science Program to C.F.C., and from the Gordon and Betty Moore Foundation to J.O.D., A.R., C.F.C. and the Caltech Brain Imaging Center.

Author Contributions E.T., A.R., C.F.C. and J.O.D. designed the experiment, E.T. conducted the experiment and analysed the data, and E.T., A.R., C.F.C. and J.O.D. discussed the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.O.D. (johnpodoherty@gmail.com).

METHODS

Participants. Forty-two healthy, right-handed participants participated in the experiment; two were excluded for excessive head movement, leaving 40 subjects in the analysis (age 21.9 ± 3.6 years (mean ± s.d.); range 18–34 years). Because of potential gender differences in social behaviour[29,31], the study included males only. Participants were pre-screened to ensure that they were within one standard deviation from the mean on the altruism subscore of the NEO PI-R personality questionnaire[32]. All participants gave informed consent, and the Institutional Review Board at the California Institute of Technology approved the study.

Experimental procedure. Pairs of participants who did not know each other performed the experiment. They were each paid a base $30 fee at the beginning of the experiment, and then each drew a ball from a hat, labelled 'rich' or 'poor'. The 'rich' (high-pay) player received an immediate payment of $50, whereas the 'poor' (low-pay) player received no bonus payment. They then performed an identical task, in consecutive fMRI scanning sessions (with scan order determined by another random draw).

In addition, two other behavioural versions of the experiment were performed with high-pay/high-pay and low-pay/low-pay pairs. The purpose of these experiments was to see whether any behavioural effects observed in the main experiment could be due to windfall gains rather than due to the inequality manipulation.

Stimulus presentation and behavioural data acquisition were implemented in Matlab (The Mathworks Inc.) with the Cogent 2000 toolbox (Wellcome Department of Imaging Neuroscience). For each trial of the experimental task, the participant viewed possible monetary transfers from the experimenter to himself and to the other player (Fig. 1). There were three trial types: transfer to high-pay player only, transfer to low-pay player only, and transfer to both players. Positive transfers varied randomly between $1 and $50. After 2 s, a rating scale appeared on the screen, and the participants rated how appealing they found the transfer on a scale of -5 (very unappealing) to 5 (very appealing) by moving a cursor along the scale with button presses. The rating period lasted 4 s and was followed by a jittered intertrial interval of 1–7 s.

Additional control trials were included in which X's were shown in place of the transfers. For these trials, a second, grey cursor was shown at a random value on the rating scale. Participants were asked to match the main cursor with the grey cursor and enter the response. The trials were randomly interspersed, with 120 experimental trials (40 of each type) and 60 control trials. The trials were broken up into four sessions of about 8 min each. After both players' scans, a single trial was randomly picked from the set, and the transfers for that trial were paid out.

fMRI data acquisition. A 3-T Trio scanner (Siemens) and an eight-channel phased array coil was used to acquire high-resolution T1-weighted structural images (1 mm × 1 mm × 1 mm) for anatomical localization and T2*-weighted echo planar images (45 slices, 3 mm × 3 mm × 3 mm voxels, TR = 2.65 s, TE = 30 ms, flip angle = 80°, FoV = 192 mm × 192 mm, slice gap = 0 mm). Each image was acquired in an oblique orientation of 30° to the anterior-commissure–posterior-commissure (AC–PC) axis, which reduces signal dropout in the vmPFC relative to AC–PC-aligned images[33].

Behavioural data analysis. For each participant, the behavioural ratings were regressed on the transfers to self and other. The resulting parameter estimates were used as summary statistics for between-subject random effects inference; the parameter estimates were pooled into a mixed-effects group analysis with the use of two-sided *t*-tests to determine whether the high-pay and low-pay groups differed in their social preferences. A two-way analysis of variance with experimental group (high-pay, low-pay) as a between-subjects factor and recipient (self, other) as a within-subjects factor was used to examine whether these parameter estimates differed between the experimental conditions. Post-hoc *t*-tests were used to determine whether the parameter estimates differed significantly between the high-pay and low-pay players for transfers to oneself and transfers to the other player.

fMRI preprocessing. We used SPM5 (Wellcome Department of Imaging Neuroscience) for fMRI data analysis. The images were slice-time corrected, realigned to the first volume to correct for subject motion, spatially transformed to match the Montreal Neurological Institute EPI template, and spatially smoothed with a Gaussian kernel (8 mm, full-width at half-maximum). We also applied intensity normalization and high-pass filtering (filter width 128 s) to the imaging data.

fMRI analysis 1. We estimated the parameters of a general linear model for each participant to generate voxel-wise statistical parametric maps of brain activation. For each participant we constructed an fMRI design matrix by modelling the following regressors for each session: 'task' (modelled as a 0-s duration event at the onset of each trial of the experimental task), 'transfer to self' (a parametric modulator of the task regressor indicating the transfer amount to oneself), 'transfer to other' (a parametric modulator of the task regressor indicating the transfer amount to the other player) and 'control' (a 0-s duration event at the onset of each control trial). The regressors were convolved with a canonical haemodynamic response function. Regressors of no interest were also generated by using the realignment parameters from the image preprocessing to further correct for residual subject motion. The parameter estimates from this first-level analysis were then entered into a random (between-subject) effects group analysis, and linear contrasts were used to identify regions that responded differentially to transfers to self versus transfers to the other player, for the high-pay versus low-pay players.

Because of previous studies indicating a role for the ventral striatum and vmPFC in processing reward value, we had an a priori hypothesis that these regions would be involved in our study. A small-volume correction was performed based on the averaged MNI coordinates from previous studies[12–14,27–29,34–38] (right vmPFC: $x = 9$, $y = 45$, $z = -13$; left vmPFC: $x = -8$, $y = 40$, $z = -13$; left striatum: $x = -10$, $y = 9$, $z = 0$), with a corrected significance threshold of $P < 0.05$. For anatomical localization, the statistical maps were rendered on the average of all subjects' structural images. For completeness, Supplementary Tables 1 and 2 list all regions displaying an effect with a voxel-wise significance threshold of $P < 0.001$, uncorrected. However, we cannot make strong conclusions about the regions outside our regions of interest, because we did not have a priori hypotheses regarding them, and no regions survived a whole-brain correction at a threshold of $P < 0.05$.

To quantify the effects we found to be significant, the underlying parameter estimates were plotted for each region of interest, by extracting the average parameter estimates from each activated cluster from each subject separately and plotting the average of those across subjects.

fMRI analysis 2. We estimated an additional related general model in which prediction error measures of the transfer to self were used as the sole parametric modulator. The prediction error measure was calculated from the following equations: $B(1) = 0$ for the low-pay group, $B(1) = 50$ for the high-pay group; and for all $t > 1$: $E_p(t) = V(t) - B(t)$; $B(t+1) = B(t) + \lambda \times E_p(t)$, where E_p is the prediction error, t the trial number, $V(t)$ is the value of the transfer to oneself on trial t, $\lambda = 0.5$ and B is the expected value. The data were modelled separately with λ values ranging from 0.1 to 0.7; $\lambda = 0.5$ was found to provide the best fit for the current data, and this value was used in the subsequent Bayesian model comparison.

Bayesian model comparison. We performed a Bayesian model comparison of the fit of the two general linear models at the random effects level, comparing the model in which we tested for a significant interaction in the responses to transfers of self compared with others between high-pay and low-pay subjects with the alternative prediction error account[39]. This analysis outputs an exceedance probability that assigns a probability to the event that one model accounts better than the other model for neural activity in a given voxel. We report the results of the Bayesian model comparison at the a priori coordinates used for the small-volume corrections, so as not to bias the results by using regions or coordinates from our voxel-wise analysis.

31. Andreoni, J. & Vesterlund, L. Which is the fair sex? Gender differences in altruism. *Q. J. Econ.* **116,** 293–312 (2001).
32. Costa, P. T. Jr & McCrae, R. R. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual* (Psychological Assessment Resources, Inc., 1992).
33. Deichmann, R., Gottfried, J. A., Hutton, C. & Turner, R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* **19,** 430–441 (2003).
34. de Quervain, D. J. *et al.* The neural basis of altruistic punishment. *Science* **305,** 1254–1258 (2004).
35. Nieuwenhuis, S. *et al.* Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage* **25,** 1302–1309 (2005).
36. Breiter, H. C., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30,** 619–639 (2001).
37. O'Doherty, J., Critchley, H., Deichmann, R. & Dolan, R. J. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J. Neurosci.* **23,** 7931–7939 (2003).
38. Lohrenz, T., McCabe, K., Camerer, C. F. & Montague, P. R. Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl Acad. Sci. USA* **104,** 9493–9498 (2007).
39. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46,** 1004–1017 (2009).