

4-2012

Neural basis of egalitarian behavior

Christopher T. Dawes
New York University

Peter John Loewen
University of Toronto

Darren Schreiber
University of California at San Diego

Alan N. Simmons
Veterans Affairs San Diego Healthcare System

Taru Flagan
University of California at San Diego

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/poliscitheses>

 Part of the [Political Science Commons](#)

Dawes, Christopher T.; Loewen, Peter John; Schreiber, Darren; Simmons, Alan N.; Flagan, Taru; McElreath, Richard; Bokemper, Scott E.; Fowler, James H.; and Paulus, Martin P., "Neural basis of egalitarian behavior" (2012). *Political Science Department -- Theses, Dissertations, and Student Scholarship*. 17.
<http://digitalcommons.unl.edu/poliscitheses/17>

This Article is brought to you for free and open access by the Political Science, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Political Science Department -- Theses, Dissertations, and Student Scholarship by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Christopher T. Dawes, Peter John Loewen, Darren Schreiber, Alan N. Simmons, Taru Flagan, Richard McElreath, Scott E. Bokemper, James H. Fowler, and Martin P. Paulus

Neural basis of egalitarian behavior

Christopher T. Dawes^{a,1}, Peter John Loewen^b, Darren Schreiber^c, Alan N. Simmons^{d,e}, Taru Flagan^{d,f}, Richard McElreath^g, Scott E. Bokemper^h, James H. Fowler^{c,i}, and Martin P. Paulus^{d,e,f}

^aDepartment of Politics, New York University, New York, NY 10012; ^bDepartment of Political Science, University of Toronto, Toronto, ON, Canada M5S 3G3; ^cDepartment of Political Science, ^dDepartment of Psychiatry, ^eLaboratory of Biological Dynamics and Theoretical Medicine, and ^fSchool of Medicine, University of California at San Diego, La Jolla, CA 92103; ^gPsychiatry Service, Veterans Affairs San Diego Healthcare System, San Diego, CA 92161; ^hDepartment of Anthropology, University of California, Davis, CA 95616; and ⁱDepartment of Political Science, University of Nebraska, Lincoln, NE 68588

Edited by Terrence J. Sejnowski, The Salk Institute for Biological Studies, La Jolla, CA, and approved March 2, 2012 (received for review November 18, 2011)

Individuals are willing to sacrifice their own resources to promote equality in groups. These costly choices promote equality and are associated with behavior that supports cooperation in humans, but little is known about the brain processes involved. We use functional MRI to study egalitarian preferences based on behavior observed in the “random income game.” In this game, subjects decide whether to pay a cost to alter group members’ randomly allocated incomes. We specifically examine whether egalitarian behavior is associated with neural activity in the ventromedial prefrontal cortex and the insular cortex, two regions that have been shown to be related to social preferences. Consistent with previous studies, we find significant activation in both regions; however, only the insular cortex activations are significantly associated with measures of revealed and expressed egalitarian preferences elicited outside the scanner. These results are consistent with the notion that brain mechanisms involved in experiencing the emotional states of others underlie egalitarian behavior in humans.

behavioral economics | egalitarianism

An intriguing aspect of human behavior that has long puzzled scholars (1) is that individuals are willing to sacrifice their own resources to promote equality in groups. For example, when dividing resources between oneself and others, people make fair divisions, at a cost to themselves, when the interaction is anonymous and the opportunity for reciprocity does not exist (2). Individuals reject unequal divisions offered by others, even if rejection means neither party receives anything (3); they reject payment for a task after having observed another receiving a higher payment for the same task, even when they had accepted the lower payment before observing others’ payoffs (4); and they voluntarily pay a personal cost to increase the resources of the poorest members of their groups and decrease the resources of the richest members, even when no reputational benefits or reciprocity can be expected (5).

Recent neuroscience studies using functional MRI (fMRI) have begun to identify the neural mechanisms underlying other-oriented behavior. For example, reward-related mechanisms play an important role when individuals engage in costly punishment of people who choose not to contribute to a mutual effort (6), during both mandatory and volitional giving to others (7), and when contemplating more equal divisions of resources (8). However, the focus of these experiments has been on dyadic instead of group interactions. Moreover, although these studies identify brain areas active during valuation, they do not identify which activations can be used to predict which individuals will actually engage in egalitarian behavior outside of a scanner.

To better understand what neural mechanisms might underlie egalitarian behavior in groups, we use a procedure called the “random-income” game. In this game, participants in a group are arbitrarily assigned a level of income and the group is assigned to one of three specific levels of income distribution by a computer (see *SI Methods*). Past analyses of the random-income game have shown that participants are willing to pay to take from the rich and to give to the poor, even in circumstances where the participants have no control of their initial income level and the group’s initial

income distribution. This willingness to pay is driven in part by emotional responses to unequal outcomes (5). Importantly, those with the greatest sensitivity to inequality in the random-income game are also those most likely to engage in costly punishment of noncontributors in a separate public-goods game (9), a behavior that has been shown to promote cooperation in humans (10).

We conducted an fMRI experiment on 20 subjects (10 male, 10 female) to identify brain regions involved in different stages of a random-income game paradigm (Fig. 1). In this experiment, each scanned subject was placed in a group with three subjects outside the scanner in each round, and all subjects made decisions in 20 rounds, giving each scanner subject 60 opportunities to decide whether to send positive or negative tokens to group members of different incomes. Consistent with other studies conducted outside the scanner (5), the behavioral results show that participants were increasingly willing to pay to take money from those with the highest incomes and to give money to those with the lowest (Fig. 1).

Results

Based on results of other studies involving social inequality (11, 12), we focused on activations within the ventromedial prefrontal cortex (vmPFC), which includes the medial orbitofrontal cortex (13). Consistent with this earlier work (8), we find that activation in this area is associated with decision-making in the random-income game. However, we did not find a significant relationship between activation and measures of expressed and revealed egalitarian preferences elicited outside the scanner (Fig. 2).

In addition to the vmPFC, we focused on the insular cortex as a region of interest (ROI) because it has been shown to be an important neural substrate in a diverse set of experiments that involve the relationship between the individual and others. For example, unfair treatment by others in the ultimatum game appears to trigger insular activity (3), with higher activations corresponding to more frequent rejection of the unfair offers (12). Inequality aversion within different contextual situations has also been associated with activation in the insula in other ultimatum-game scenarios (14). Others have reported that individuals with personality disorders exhibit dysfunction in the insula that is associated with impaired trust and social cooperation (15). These and other results suggest that the insular cortex in general is critical for the perception of internal states (16), and that the anterior insula is particularly important for awareness of our own feelings as they relate to others (17), which may be especially relevant for empathy in decision-making situations (18). Insula activation has also been linked to volitional prosocial behavior induced by em-

Author contributions: C.T.D., P.J.L., D.S., A.N.S., T.F., R.M., J.H.F., and M.P.P. designed research; C.T.D., P.J.L., D.S., A.N.S., T.F., J.H.F., and M.P.P. performed research; C.T.D., P.J.L., D.S., A.N.S., J.H.F., and M.P.P. analyzed data; and C.T.D., P.J.L., D.S., A.N.S., S.E.B., J.H.F., and M.P.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: cdawes@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1118653109/-DCSupplemental.

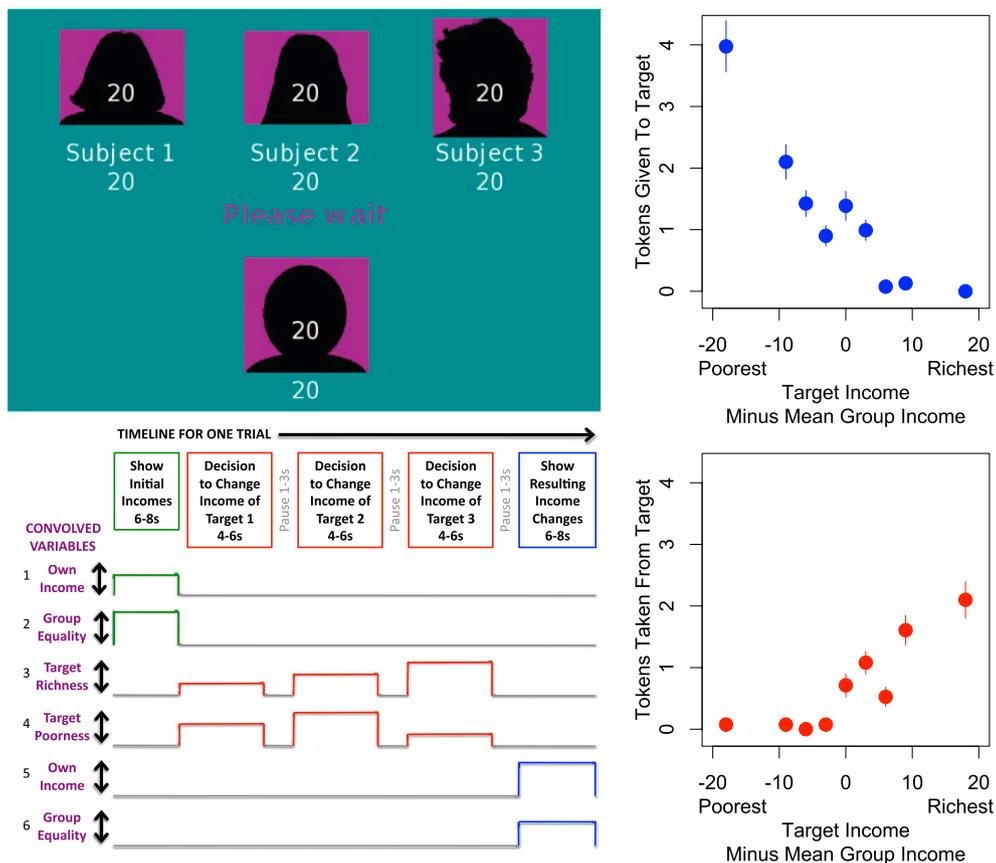


Fig. 1. Sample initial screenshot (*Upper Left*) shows own income at bottom and income of each other group member at top. During decision phases, income beneath silhouette changes to reflect incomes that will result from the subject's actions. The timeline for each trial (*Lower Left*) included an initial phase where randomly drawn incomes for each group member were shown, three decision phases in which the subject sees a target group member's income and then chooses whether or not to give or take tokens from the target, and an outcome phase where resulting incomes for each group member are shown. To identify ROIs, we regressed voxel activations on the convolution of six time-varying covariates with heights depending on initial own income, initial group equality, target "richness" (target income minus mean group income among those who earn more than average), target "poorness" (mean group income minus target income among those who earn less than average), change in one's own income, and change in group equality. Consistent with previously published work (1), scanner subjects tend to give tokens to low earners (*Upper Right*) and take away tokens from high earners (*Lower Right*). Vertical lines indicate SEMs.

pathway toward other's emotional or physical pain (19). Thus, we expected that this brain region would be a likely candidate for involvement in egalitarian behaviors.

Our neuroimaging results support the role of the insular cortex in egalitarian behavior. Specifically, we identified a region within the transition area of lateral inferior frontal gyrus and anterior insula (Fig. 2), where activations correlated significantly with changes in group equality when resulting incomes were shown at the end of each trial. These activations were also significantly associated with two additional measures of egalitarianism. First, the activations correlated with a widely used index of self-reported egalitarianism (Huber regression, $P = 0.009$) (Fig. 2, *Top, Center*). Second, the activations were significantly associated with egalitarian behavior measured by a series of dictator games (*SI Methods*) played after the scan session (Huber regression, $P = 0.05$) (Fig. 2, *Top, Right*).

Discussion

To summarize, this experiment shows that some parts of the brain are more active during egalitarian outcomes, and these activations are correlated with egalitarian behavior inside the scanner. However, a more crucial result is that the activations are also correlated with behavior outside the scanner, including self-reported preferences for egalitarian outcomes and game behavior that reveals how willing subjects are to use their own resources to obtain egalitarian outcomes within their groups. Taken together, the evidence suggests that the anterior insular cortex plays a crit-

ical role in egalitarian behavior in humans. This conclusion is consistent with a broader view of the insular cortex as a neural substrate (17) that processes the relationship of the individual with respect to his or her environment (19). The predominately left-lateralized activation may point toward the possibility of a positive valence or energy-preserving mode- (17) related processing during egalitarian behavior (i.e., individuals may see the group as a greater good that is worth preserving). The fact that the insula is directly involved in physiological, food, and pain-related processing supports the general notion that prosocial behavior, which is important for survival of both the individual and the group/species, is implemented on a fundamental physiological level similar to breathing, heartbeat, hunger, and pain.

Adam Smith (1) contended that moral sentiments like egalitarianism derived from a "fellow-feeling" that would increase with our level of sympathy for others, predicting not merely aversion to inequity, but also our propensity to engage in egalitarian behaviors. The evidence here supports such an interpretation. Although individuals may experience internal rewards when punishing antisocial behavior (6) and may have preferences for social equality (8), our results suggest that it is the brain mechanisms involved in experiencing the emotional and social states of self and others (17–21) that appear to be driving egalitarian behaviors.

Our results have important implications for theories of the evolution of prosocial behavior that suggest culturally transmitted "leveling mechanisms"—for example, food sharing and monogamy—

round and are then provided an opportunity to give up to 10 “negative tokens” or “positive tokens” to other players. Each negative token costs 1 MU and decreases the payoff of a targeted individual by 3 MUs; positive tokens also cost 1 MU, but increase the targeted individual’s payoff by 3 MUs. Groups are randomized after each round to prevent reputation from influencing decisions; interactions between players are strictly anonymous and subjects know this. Furthermore, by allowing participants more than one behavioral alternative, the experiment eliminates possible demand effects (24).

Subjects also completed a self-report survey that included a battery of items, including a standard set of questions that measure egalitarian preferences. At the conclusion of the study, subjects completed five rounds of the modified dictator game (25).

Subjects. All subjects were undergraduate social science majors at the University of California at San Diego and provided informed consent. Based on a pilot study, we chose individuals who bought more positive or negative tokens in relatively unequal conditions of the random-income game to complete an fMRI screening questionnaire. As a result, the subjects selected tend to be more egalitarian than those in the overall pilot sample. Although this selection process potentially makes the sample less representative overall, it increases the number of observations of egalitarian behavior and therefore improves the power of statistical tests. Subjects who were deemed eligible based on the screening questionnaire were invited to participate in the neuroimaging phase of the experiment. The final fMRI sample consisted of 10 males and 10 females that were all very similar in age.

Image Acquisition. fMRI data were collected at the University of California at San Diego on a 3T GE CXX4 Magnet with an eight-channel brain array coil to axially acquire T2*-weighted echo-planar images (EPI) (field-of-view 230 mm, 64 × 64 matrix; 30 2.6-mm thick slices; 1.4 mm gap; TR = 2,000 ms, TE = 32 ms, flip angle = 90°). The basic structural and functional image processing were conducted with the Analysis of Functional NeuroImages (AFNI) software package (26). All EPI images were aligned to the high-resolution anatomical images and resampled to a voxel size of 4 × 4 × 4 mm (from the original 3.7 × 3.7 × 4 mm). Data were temporally smoothed, spatially blurred with a 6-mm FWHM spatial filter, and normalized to Talairach space (via AFNI’s auto Talairach program, followed by visual inspection of each structural image).

Data Analysis. A regression model was constructed for a contrast against the hypothesized voxel activation, based on a blood-oxygen level-dependent hemodynamic response function with 4- to 6-s peaks (Fig. 1). The model is composed of six time-varying convolved regressors with heights that vary based on certain aspects of the trial (Fig. 1). The parameters of the model were varied by round based on the different characteristics of each respective trial of the random-income game. The coefficient for each regressor was established using a standard general linear model.

The regression model consists of three parts that are designed to measure activation during each phase of the trial:

Introduction phase regressors. An “own income” regressor was computed as the initial income assigned to the subject by the computer, and a “group inequality” regressor was calculated as the initial SD of all four incomes in the group.

Decision phase regressors. A “target richness” regressor was defined as the target’s income minus the group average income for the trial. Negative values for this regressor (where the target was poorer than average) were assigned a 0. Similarly a “target poorness” variable was defined as the group average income for the trial minus the target’s income. Negative values (where the target was richer than average) were assigned a 0. These two regressors index the degree to which, during decision-making, the target is either doing better or worse than the group as a whole.

Outcome phase regressors. These regressors were used to indicate the change in individual and group incomes during the trial resulting from token purchases by all four group members. An “own income change” regressor measured the

subject’s final income minus his or her initial income, and a “group equality change” regressor measured the final SD of group incomes minus the initial SD.

A statistical model was used to relate changes in EPI intensity to differences in task characteristics (27). All slices of the EPI scans were temporally aligned following registration to ensure that different relationships with the regressors were not a result of the acquisition of different slices at different times during the repetition interval. EPIs were coregistered using a 3D-coregistration algorithm (28) that has been developed to minimize the amount of image translation and rotation relative to all other images. Six motion parameters were obtained for each subject. Three motion regressors (roll, pitch, and yaw), a linear trend, and constant were included in the model. Based on the fit to the regressors of interest, percent signal changes were calculated on the spatial blurred (6-mm Gaussian blur) normalized brains. Specifically, we used six convolved regressors, two for each phase of the task (Fig. 1). A threshold adjustment method based on Monte-Carlo simulations was applied to prevent identification of false-positive areas of activation (29). Based on these simulations, it was determined that a voxel-wise a priori probability of 0.01 ($t = 2.552$) would result in a corrected cluster-wise activation probability of 0.01 (one-sided) if a minimum volume 320 μL and a connectivity radius of 4.0 mm were to be considered.

Following the individual analysis, we determined how differences in behavior across subjects were related to brain activation differences that are indexed by coefficients on one of the six regressors. Specifically, we considered the behavior during the 20 rounds × 3 group members (60 total) possible decisions and computed two measures of egalitarian behavior: (i) a “take from rich” measure equal to the Pearson correlation between “target richness” and the amount of the change in the target income resulting from the subject’s purchases of positive and negative tokens; (ii) a “give to poor” measure equal to the Pearson correlation between “target poorness” and the amount of the change in the target income resulting from the subject’s purchases of positive and negative tokens. To identify ROIs, these measures were used in a multiple regression analysis with the behavior as the independent measures and the first level analysis as the dependent measure. These results were analyzed in R (30) for group effects and correlations. Using the “take from rich” measure, we identified four ROIs using “anatomically constrained functional regions of interest” (31). For the insular cortex we constructed a probability mask. The probability mask was modeled after a well-documented atlas of various neurological structures. Briefly, to extract a mask for the insular cortex, we used Individual Brain Atlases using Statistical Parametric Mapping software (32), a toolbox for segmenting structural MR images. All programs in this toolbox are developed in MATLAB (33), based on a widely used neuroimaging software package, SPM (Wellcome Trust Centre for Neuroimaging, London, United Kingdom). This package uses the nonlinear registration and gray matter segmentation processes performed through SPM5 subroutines. The group insula mask was obtained by averaging across the individual insular masks and requiring that for the common insula mask each voxel had to cover the insula gray matter in at least 50% of all subjects.

Second-Level Analysis. All second-level analyses were conducted using the statistical programming language R (30) and with SPSS software, version 10. Specifically, a mixed-model analysis was conducted with the R procedure *lme*, which is part of the *nlme* library. The fixed effects were decision type, the random effects were subjects. Moreover, we conducted voxel-wise multiple linear regression analyses with performance during the task as independent measures, and the percent signal change during the different decision-making phases as the dependent measure using the *lm* procedure in R. Finally, we conducted a between-subjects analysis of activations in the identified ROIs (as measured by coefficients on the regressors) and measures of self-reported and behavioral egalitarianism outside the scanner using the *lrm* procedure in R.

ACKNOWLEDGMENTS. J.H.F. thanks the University of California at San Diego Committee on Research for generous research support.

- Smith A (1759) *The Theory of Moral Sentiments* (Printed for A. Millar and A. Kincaid and J. Bell, London).
- Henrich JP, et al. (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford Univ Press, Oxford, New York).
- Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the Ultimatum Game. *Science* 300:1755–1758.
- Brosnan SF, De Waal FB (2003) Monkeys reject unequal pay. *Nature* 425:297–299.
- Dawes CT, Fowler JH, Johnson T, McElreath R, Smirnov O (2007) Egalitarian motives in humans. *Nature* 446:794–796.
- de Quervain DJ, et al. (2004) The neural basis of altruistic punishment. *Science* 305:1254–1258.
- Harbaugh WT, Mayr U, Burghart DR (2007) Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316:1622–1625.
- Tricomi E, Rangel A, Camerer CF, O’Doherty JP (2010) Neural evidence for inequality-averse social preferences. *Nature* 463:1089–1091.
- Johnson T, et al. (2009) The role of egalitarian motives in altruistic punishment. *Econ Lett* 102:192.
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
- Fliessbach K, et al. (2007) Social comparison affects reward-related brain activity in the human ventral striatum. *Science* 318:1305–1308.
- Tabibnia G, Satpute AB, Lieberman D (2008) The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol Sci* 19:339–347.

13. Lacerda AL, Hardan AY, Yorbik O, Keshavan MS (2003) Measurement of the orbitofrontal cortex: A validation study of a new method. *Neuroimage* 19:665–673.
14. Wright ND, Symmonds M, Fleming SM, Dolan RJ (2011) Neural segregation of objective and contextual aspects of fairness. *J Neurosci* 31:5244–5252.
15. King-Casas B, et al. (2008) The rupture and repair of cooperation in borderline personality disorder. *Science* 321:806–810.
16. Craig AD (2002) How do you feel? Interoception: The sense of the physiological condition of the body. *Nat Rev Neurosci* 3:655–666.
17. Craig AD (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 10:59–70.
18. Singer T, Critchley HD, Preuschoff K (2009) A common role of insula in feelings, empathy and uncertainty. *Trends Cogn Sci* 13:334–340.
19. Masten CL, Morelli SA, Eisenberger NI (2011) An fMRI investigation of empathy for ‘social pain’ and subsequent prosocial behavior. *Neuroimage* 55:381–388.
20. Dewall CN, et al. (2010) Acetaminophen reduces social pain: Behavioral and neural evidence. *Psychol Sci* 21:931–937.
21. Immordino-Yang MH, McColl A, Damasio H, Damasio A (2009) Neural correlates of admiration and compassion. *Proc Natl Acad Sci USA* 106:8021–8026.
22. Bowles S (2006) Group competition, reproductive leveling, and the evolution of human altruism. *Science* 314:1569–1572.
23. Boehm C (1999) *Hierarchy in the Forest* (Harvard, Cambridge).
24. Orne MT (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am Psychol* 17:776.
25. Andreoni J, Miller J (2002) Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70:737–753.
26. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
27. Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233.
28. Eddy WF, Fitzgerald M, Noll DC (1996) Improved image registration by using Fourier interpolation. *Magn Reson Med* 36:923–931.
29. Forman SD, et al. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
30. R Development Core Team (2011) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna). Available at <http://cran.r-project.org/>.
31. Johnstone T, et al. (2005) Stability of amygdala BOLD response to fearful faces over multiple scan sessions. *Neuroimage* 25:1112–1123.
32. Alemán-Gómez Y, Melie-García L, Valdés-Hernandez, P (2006) IBASPM: Toolbox for automatic parcellation of brain structures. Available on CD-Rom in *Neuroimage* 31(Suppl 1). Available at <http://www.thomaskoenig.ch/Lester/ibaspm.htm>.
33. MATLAB version 7.10.0. (The MathWorks Inc., Natick, MA). Available at <http://www.mathworks.com>.

Supporting Information

Dawes et al. 10.1073/pnas.1118653109

SI Methods

The supplementary information includes: (i) details of behavioral and attitudinal studies conducted inside and outside the scanner, and (ii) complete instructions for both studies.

Study Details. Pilot study. In the “random-income” game player allocations are randomly drawn from the empirical distribution. Our study had different distributions of allocations based on treatments (T1–T3). In the three treatments allocations are shown to each subject. Each subject was given an opportunity to increase or decrease the income of other subjects. Each subject could buy up to 10 “positive” tokens and up to 10 “negative” tokens for each of the other players. Each negative token costs the subject one token and reduces the target’s income by three tokens. Each positive token costs the subject one token and increases the target’s income by three tokens. One by one, each target’s income was highlighted and the subject was asked if he or she wanted to buy tokens. After subjects make their decisions, they saw the results (unless in an intersession group, see below) and then proceeded to the next treatment.

The distribution of allocations were symmetric and varied in the degree of inequality as follows: T1 = (20, 20, 20, 20), T2 = (11, 17, 23, 29), and T3 = (2, 14, 26, 38). In particular, T1 is the perfect equality condition, and served a contrast for the other conditions that exhibit increasing degrees of inequality. Subjects were randomly assigned incomes such that they were the low, middle-low, middle-high, or high earner at different points during the study.

We ran two types of random income games. In the first type, subjects were matched with three subjects from the same session (S) and in the second type subjects were matched with three subjects from other sessions (O). The purpose of the latter-type groups was to be able to match subject responses to subjects who are chosen for the second functional MRI (fMRI) phase of the study. Each subject participated in both types of groups. Subjects were paid for their decisions in S rounds of the random-income treatments following the conclusion of the session because we had complete information about all player choices at that time. For O rounds, they were paid immediately following the fMRI portion of the study.

The order of group types was randomly determined to prevent the group type and time of payment from influencing subject choices. At the conclusion of the punishment and reward stage for each treatment in which the subject has been placed in a same type S treatment, subjects saw their fellow players’ decisions and the total income they earned for that treatment. At the conclusion of the punishment and reward stage for each treatment in which the subject was placed in a type O treatment, subjects received a message stating that their decisions for that treatment will be matched with a player in another session and they will be paid at a later time. Subjects did not know at the time of their initial decision which type of round they were in or how many type S and O rounds they would play in total.

In total, each subject played 3 rounds of type S and 18 rounds of type O random-income games. We ran six sessions of the random-income game with a total of 140 subjects.

Subject selection for fMRI study. After the conclusion of the pilot study we chose subjects from the pilot who bought more positive or negative tokens in relatively unequal conditions to complete an fMRI screening questionnaire. Those who were deemed eligible were invited to participate in the neuroimaging phase of the experiment. The final fMRI sample consisted of 10 males and 10 females that were all very similar in age.

Study payment. Pilot subjects were paid a show-up fee of \$10. These subjects were paid for all of their S rounds at the conclusion of

each session and were paid at the conclusion of the study for the O rounds. The fMRI subjects were paid a show-up fee of \$50 and were then paid for their actions in the game at the conclusion of the study.

Dictator game. At the time of final payment subjects were asked to play five rounds of the modified dictator game (1). The dictator game involves two “players.” The first player determines a split of some endowment between herself and the other player. In our case, subjects were asked to split 10 lottery tickets between themselves and an anonymous study recipient. The second player simply receives the number of tickets the first player has allocated to her and thus has no strategic role in the game. If the first individual were only concerned with her own well-being, she would keep all of the tickets and pass nothing. Therefore, any positive allocation to the second player is viewed as a revealed-preference measure of altruism.

The modified dictator game (2) enables researchers to distinguish between egalitarian and altruist preference types. Egalitarians prefer to equalize the payoffs to both players in each of the dictator games, whereas altruists simply maximize the amount they give away. In the modified dictator game subjects are asked to play several dictator games in which the price of passing some of the endowment to the other player varies. In our experiment there were five conditions: each ticket kept would yield one chance in lottery for the player originally given the endowment and each ticket passed would yield one chance for the other player (1:1); each ticket kept would yield one chance and each passed would yield two chances (1:2); each ticket kept would yield one chance and each passed would yield three chances (1:3); each ticket kept would yield two chances and each passed would yield one chance (2:1); each ticket kept would yield three chances and each passed would yield one chance (3:1). Egalitarians prefer to equalize the payoffs to both players in each of the dictator games, therefore they should pass 5 tickets in the 1:1 condition, 3.33 tickets in the 1:2 condition, 2.5 tickets in the 1:3 condition, 6.67 tickets in the 2:1 condition, and 7.5 tickets in the 3:1 condition. Tickets were not divisible but the degree to which subjects exhibited egalitarian preferences were determined by how closely they matched this ideal type. Specifically, our measure of egalitarian behavior is equal to the negative of the sum of the square difference in each game between the number of tickets given away and the number an egalitarian would give away.

Subjects were informed that their participation made them eligible to win \$100. They were told that a winner would be drawn randomly and that the number of times an individual’s name was entered in the draw depended on allocations in the dictator game.

Self-reported egalitarianism. We administered to subjects six questions that have been asked repeatedly in the National Election Studies since 1984 (3) to measure self-reported egalitarianism. Each question included response options of Strong Agree, Agree Somewhat, Neither, Disagree Somewhat, Strongly Disagree. These questions included:

- Q1: Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed.
- Q2: We have gone too far in pushing equal rights in this country.
- Q3: One of the big problems in this country is that we don’t give everyone an equal chance.
- Q4: This country would be better off if we worried less about how equal people are.
- Q5: It is not really that big a problem if some people have more of a chance in life than others.

Q6: If people were treated more equally in this country we would have many fewer problems.

A factor analysis showed that these questions all load on the same factor with a Cronbach's $\alpha = 0.88$. The factor from this analysis is treated as the egalitarianism index, with positive values indicating greater self-assessment of egalitarianism. This measure is significantly correlated with the dictator game measure of egalitarianism (Pearson's correlation 0.69, 95% confidence interval 0.23–0.85).

Study Instructions. Random-income game subject instructions. You are now taking part in an economic experiment, which has been financed by various foundations for research. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore very important that you read these instructions carefully.

The instructions, which we have distributed to you, are solely for your private information. It is prohibited to communicate with the other participants during the experiment. Should you have any questions please ask us.

During the experiment we will not speak of Dollars but rather of tokens. During the experiment your entire earnings will be calculated in tokens. At the end of the experiment the total amount of tokens you have earned will be converted to Dollars at the following rate:

$$1 \text{ token} = 5 \text{ cents.}$$

The experiment is divided into periods. In total, the experiment has 21 periods. In each period the participants are divided into groups of four. That is, your group has three other participants in it aside from you. Group composition will change in each period. In each of the 21 periods your group of four is composed of different people. Therefore, in each of the 21 periods you will form a group with three different people. For some of the 21 periods your group will be made up of participants from other sessions.

First Stage

At the beginning of each period the computer will randomly choose a number of tokens to give to each person in your group. The income of each group member from the project is calculated in the same way. An income screen will show you how many tokens you and each member of your group have earned at the first stage.

After the first stage concludes, we will begin the second stage. Do you have any questions?

Second stage

You will see how much the individual group members have earned in the first stage. Please note that who is in your group is randomly determined in each period; the individuals in your group are likely to change over the course of the experiment.

You will now have the opportunity to change or leave unchanged the income of each of the other group members. You can either decrease their income by allocating negative tokens or you can increase their income by allocating positive tokens. The other group members can also reduce or increase your income, if they so wish.

You must decide how many negative or positive tokens to give to each of the other three group members and then move the slider to either the left to give negative tokens or to the right to give positive tokens. If you do not want to change the income of a particular group member, simply do not move the slider. Once you have made a decision regarding a member of your group, hit the Allocate button to move to the next group member.

If you distribute negative tokens, you must pay a cost for each of the negative tokens you allocate. Negative tokens are integers between 0 and 10. The more negative tokens you allocate, the higher your costs. The following formula indicates how much it costs to allocate negative tokens:

$$\text{Cost of negative tokens} = \text{Sum of allocated negative tokens.}$$

Each negative token therefore costs you 1 token. If, for instance, you allocate 2 negative tokens to one group member, you must pay a cost of 2 tokens. If you allocate 9 negative tokens to another member, it will cost you an additional 9 tokens; if you allocate 0 negative tokens to the last group member, there will be no additional cost. Therefore, you have allocated a total of 11 negative tokens and your total costs are 11 tokens (2+9+0).

Each negative token you allocate to a group member reduces his or her income by 3 tokens. If you allocate 0 tokens to a certain group member, you do not change the income of this group member. If, however, you allocate one negative token to a group member, you reduce his or her income by 3 tokens. If you allocate 2 negative tokens to a group member, you reduce his or her income by 6 tokens.

If you distribute positive tokens, you must pay a cost for each of the positive tokens you allocate. Positive tokens are integers between 0 and 10. The more positive tokens you allocate, the higher your costs. The following formula indicates how much it costs to allocate positive tokens:

$$\text{Cost of positive tokens} = \text{Sum of positive tokens.}$$

Each positive token therefore costs you 1 token. If, for instance, you allocate 2 positive tokens to one group member, you must pay a cost of 2 tokens. If you allocate 9 positive tokens to another member, it will cost you an additional 9 tokens; if you allocate 0 positive tokens to the last group member, there will be no additional cost. Therefore, you have allocated a total of 11 positive tokens and your total costs are 11 tokens (2+9+0).

Each positive token you allocate to a group member increases his or her income by 3 tokens. If you allocate 0 positive tokens to a certain group member, you do not change the income of this group member. If, however, you allocate one positive token to a group member, you increase his or her income by 3 tokens. If you allocate 2 positive tokens to a group member, you increase his or her income by 6 tokens.

Whether or by how much the income at the end of the period is in total increased or decreased depends on the total of the received positive or negative tokens. If somebody receives a total of 3 negative tokens (from all other group members in this period), then his or her income would be decreased by 9 tokens. If somebody receives a total of 4 negative tokens, his or her income is reduced by 12 tokens. Likewise, if somebody receives a total of 3 positive tokens (from all other group members in this period), then his or her income would be increased by 9 tokens. If somebody receives a total of 4 positive tokens, his or her income is increased by 12 tokens. Your total income from the two stages is therefore calculated as follows:

$$\begin{aligned} \text{Total income (in tokens) at the end of the period} &= \text{period income} \\ &= \text{income from first stage (1)} \\ &\quad - 3 * (\text{sum of received negative tokens}) (2) \\ &\quad + 3 * (\text{sum of received positive tokens}) (3) \\ &\quad - \text{costs of your negative tokens allocated to others} (4) \\ &\quad - \text{costs of your positive tokens allocated to others} (5) \\ &\text{if } (1) + (2) + (3) + (4) + (5) \geq 0; \\ &= 0 - \text{costs of your distributed tokens} \\ &\text{if } (1) + (2) < 0 \end{aligned}$$

Please note that your income in tokens at the end of the period can be negative, if the costs of your distributed (negative or positive) tokens exceed your income in tokens minus the cost of received negative tokens plus the cost of received positive tokens. You can, however, avoid such losses with certainty through your

