

# Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism

Yosuke Morishima,<sup>1,\*</sup> Daniel Schunk,<sup>1</sup> Adrian Bruhin,<sup>1</sup> Christian C. Ruff,<sup>1</sup> and Ernst Fehr<sup>1,\*</sup>

<sup>1</sup>Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland

\*Correspondence: [yosuke.morishima@econ.uzh.ch](mailto:yosuke.morishima@econ.uzh.ch) (Y.M.), [ernst.fehr@econ.uzh.ch](mailto:ernst.fehr@econ.uzh.ch) (E.F.)

DOI 10.1016/j.neuron.2012.05.021

## SUMMARY

Human altruism shaped our evolutionary history and pervades social and political life. There are, however, enormous individual differences in altruism. Some people are almost completely selfish, while others display strong altruism, and the factors behind this heterogeneity are only poorly understood. We examine the neuroanatomical basis of these differences with voxel-based morphometry and show that gray matter (GM) volume in the right temporoparietal junction (TPJ) is strongly associated with both individuals' altruism and the individual-specific conditions under which this brain region is recruited during altruistic decision making. Thus, individual differences in GM volume in TPJ not only translate into individual differences in the general propensity to behave altruistically, but they also create a link between brain structure and brain function by indicating the conditions under which individuals are likely to recruit this region when they face a conflict between altruistic and selfish acts.

## INTRODUCTION

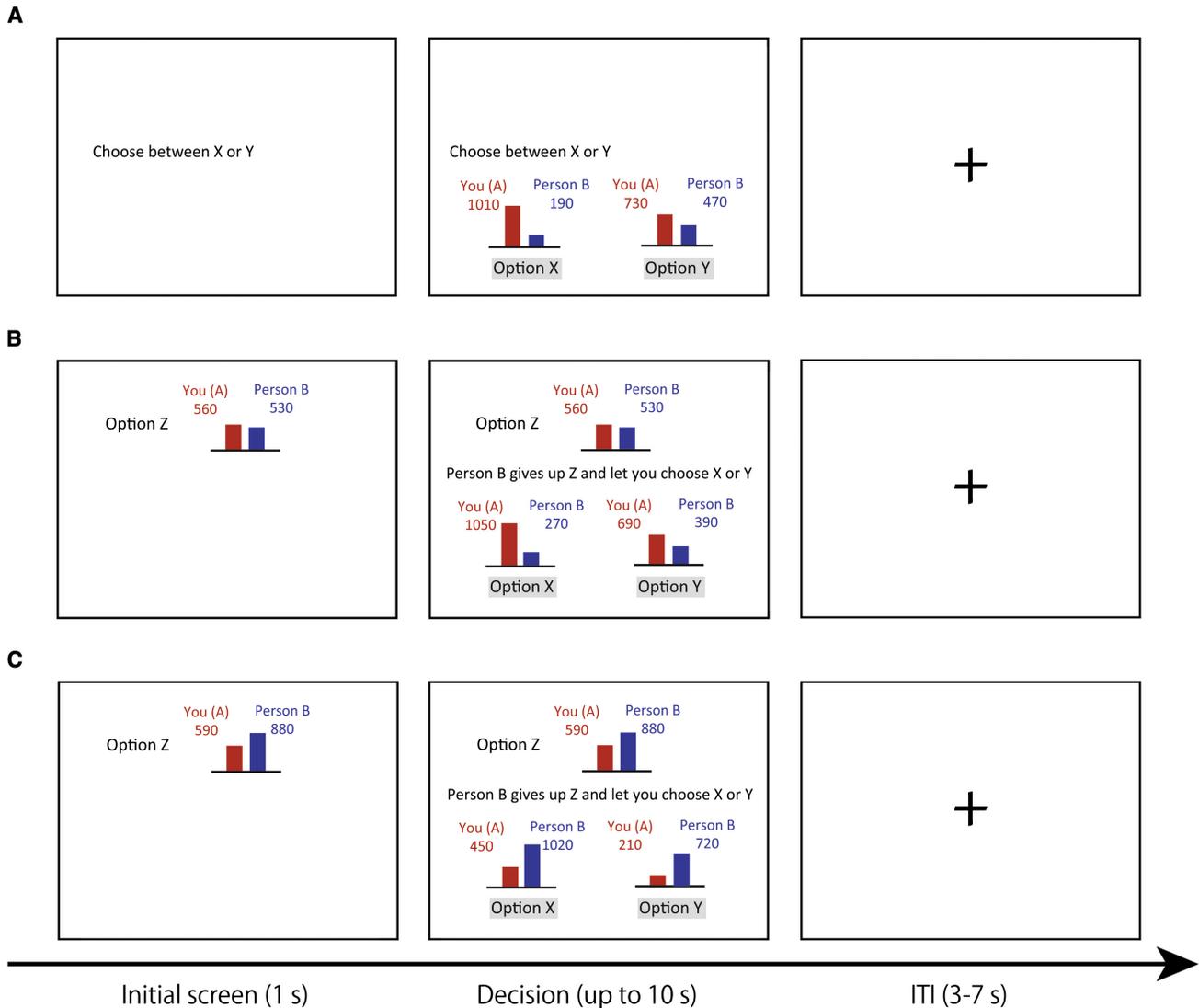
Altruistic acts involve costs for the actor and benefits for another individual. Altruism in most animal species is directed toward genetically related individuals (Hamilton, 1964). In contrast, human altruism goes far beyond helping kin. A significant number of people help strangers and reciprocate favors even when they do not know their interaction partners and will never meet them again (Camerer, 2003; Henrich et al., 2005). Human history has repeatedly shown that some people are even willing to risk their lives in order to contribute to some of the most important public goods—democracy and liberty. However, there is also enormous individual heterogeneity in human altruism, and the sources of individual variation are still very poorly understood. Typically, more traditional variables such as individuals' gender, income, wealth, or education have shown little predictive power (Camerer, 2003; Henrich et al., 2005), and neurobiological variables have only rarely been used as predictors of individual differences in altruism (de Quervain et al., 2004;

Harbaugh et al., 2007; Hare et al., 2010; Moll et al., 2006; Tricomi et al., 2010). Recent applications of brain morphometry indicate that individual differences in brain structure can be useful in understanding individual differences in traits and skills (Kanai and Rees, 2011). We therefore conjectured that variables reflecting relatively stable neuroanatomical individual differences—such as gray matter (GM) volume—may help predict individual differences in altruism.

In humans, altruism is likely to be related to perspective taking, i.e., the ability to take other individuals' perspectives into account. In fact, developmental data suggest that preschoolers who have already acquired theory of mind skills behave more prosocially (Takagishi et al., 2010), and experiments with adults indicate that subjects with better skills in reading others' mental states show more altruistic behavior (Underwood and Moore, 1982). One brain region that has been repeatedly and reliably found to be implicated in tasks requiring the ability to represent and understand others' perspectives is the temporoparietal junction (TPJ) (Decety and Lamm, 2007; Frith and Frith, 2007; Ruby and Decety, 2001; Saxe and Kanwisher, 2003; Young et al., 2010). We therefore hypothesized that GM volume in the TPJ may provide a neuroanatomical basis for individual differences in human altruism.

Research on human social preferences provides behavioral (Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Fehr and Schmidt, 1999) and neural (Tricomi et al., 2010) evidence that other-regarding behaviors and motives depend on the initial payoff allocation between the subject and the subject's partner. In particular, if subjects have a lower initial payoff than their partner (“disadvantageous initial inequality”), they are much less willing to behave altruistically toward the partner compared to a situation with advantageous initial inequality (i.e., when the subject has a higher initial payoff than the partner). In fact, some individuals even reduce the partner's payoff if possible if the latter has a higher initial payoff. In view of the radically different propensities for behaving altruistically in the domain of advantageous and disadvantageous inequality, it may be possible that the neuroanatomical basis for human altruism is not identical across these domains.

In the present study, subjects had to allocate money between themselves and anonymous partners (Figure 1; task description in Experimental Procedures) in a series of binary choice problems. In each trial, subjects faced a binary choice in which they could increase or decrease the partner's monetary payoff. The subjects' cost of changing the partners' payoff varied

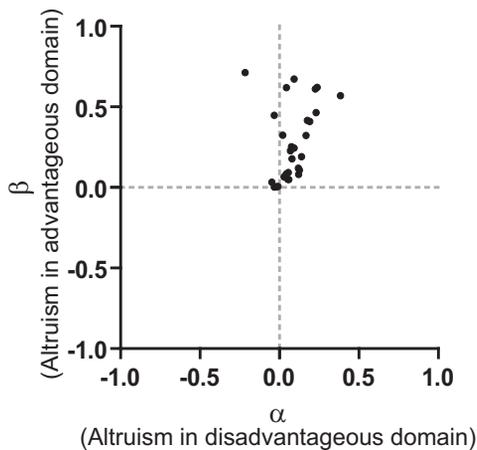


**Figure 1. Behavioral Paradigm**

We used dictator (A) and reciprocity (B and C) games to estimate subjects' social preference parameters. Subjects in the dictator game faced many decision problems in which they had to choose one of two payoff allocations ("options") that assigned money to the subject (person A) and an anonymous partner (person B). For example, the subject in the decision problem presented in (A) could make an altruistic choice (i.e., option Y) that increases B's payoff at a cost to subject A. The subjects were asked to make a decision by pressing the relevant button within 10 s, otherwise the screen moved to the intertrial interval. The actual decision time, after the decision screen appeared, was 2.5 s on average. In the reciprocity games, person B could make an altruistic or a selfish decision by choosing option Z or by letting person A make the choice between X and Y. We asked the subject (player A) how she/he would behave if the other participant (player B) gave up option Z and then allowed player A to choose option X or Y. In the positive reciprocity trials (B), the amount player A earns in both option X and Y is larger than the amount she would earn in option Z, implying that B's decision to give up option Z is an altruistic act toward player A. Thus, if player A has a preference for positive reciprocity, she will reward B's behavior with an altruistic choice, i.e., A will choose the option between X and Y that gives B a higher payoff. In the negative reciprocity trials (C), the amount that player A earns in both options X and Y is less than what she earns in option Z, implying that B's decision to give up option Z is a selfish act toward player A. Thus, if player A has a preference for negative reciprocity, she will sanction B's behavior by choosing the option between X and Y that gives a lower payoff to B. As in the dictator game, subjects were asked to make a decision within 10 s, otherwise the screen moved to the intertrial interval. The actual decision time in a reciprocity game after the decision screen appeared was 2.7 s on average.

across trials, so that the observed-across-trial behaviors enable us to estimate parameters that reflect each subject's preference for altruistic behavior (see [Supplemental Information](#) available online). We then used voxel-based morphometry (VBM) to examine the correlation between brain structure—in terms of relative gray matter volume—and subjects' behavioral prefer-

ences for altruism. We conjectured that gray matter volume in the TPJ might reflect subjects' preferences for altruism and that this fact, if true, could help us understand the link between brain structure and brain activation in TPJ—measured by functional magnetic resonance imaging (fMRI)—during the behavioral task.



**Figure 2. Individual Heterogeneity in Altruism**

Scatter plot of individual preference parameters for altruism in disadvantageous ( $\alpha$ ) and advantageous ( $\beta$ ) situations. Both parameters vary strongly across subjects, and there is no significant correlation between them ( $r = 0.29$ ,  $p = 0.11$ ).

## RESULTS

### Heterogeneous Preferences for Altruistic Behaviors

Our study is based on behavioral experiments ( $n = 30$ ) and a mathematical model of social preferences that enabled us to simultaneously estimate a preference parameter  $\alpha$  for each individual, which measures the subject's preferences for altruistic acts in the domain of disadvantageous inequality, and a parameter  $\beta$ , which measures preferences for altruism in the domain of advantageous inequality. A positive value of  $\alpha$  means that the subject has a preference for increasing the partner's material payoff in the domain of disadvantageous inequality, while a negative value of  $\alpha$  means that the subject prefers reducing the partner's material payoff in this situation; a similar interpretation applies to the  $\beta$  parameter, except that it informs us about the subject's preference in the domain of advantageous inequality. On average,  $\alpha$  (mean 0.085,  $t(28) = 4.06$ ,  $p = 0.004$ ) and  $\beta$  (mean 0.275,  $t(28) = 6.39$ ,  $p < 0.0001$ ) are significantly positive, and there is considerable individual variation (Figure 2). Both parameters are positively correlated, albeit the correlation falls just short of statistical significance ( $r = 0.29$ ,  $p = 0.11$ ). Interestingly, altruism in the domain of advantageous inequality ( $\beta$ ) is significantly higher than altruism in the domain of disadvantageous inequality ( $\alpha$ ,  $t(28) = 4.52$ ,  $p = 0.0001$ ). This indicates that participants are more willing to behave altruistically if altruistic acts decrease inequality (in the advantageous situation) rather than increase inequality (in the disadvantageous situation), suggesting that fairness concerns affect the motivation for altruistic acts.

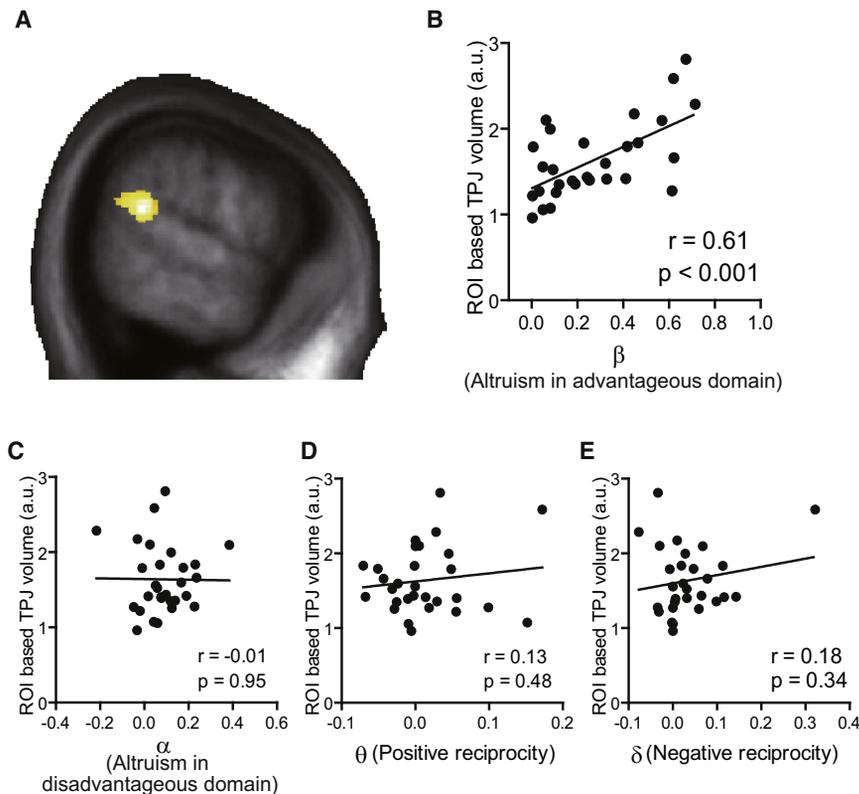
### Gray Matter Volume in the Right TPJ Predicts the General Propensity for Altruistic Acts

To identify possible neurobiological determinants of preferences for altruistic behavior, we used VBM analyses to identify brain regions where local GM volume is significantly correlated with the preference parameters  $\alpha$  and  $\beta$ . We find that GM volume in

the right TPJ displays a strong positive correlation with  $\beta$ , our preference measure of altruism in the domain of advantageous inequality ( $t = 5.94$ ,  $p < 0.05$ , voxelwise whole-brain family-wise error [FWE] corrected) (Figure 3A), while we observe no correlation with preferences for altruism in the domain of disadvantageous inequality  $\alpha$  ( $p > 0.05$ , uncorrected). Moreover, a whole-brain analysis shows that no other brain region is (whole-brain FWE corrected) significantly correlated with  $\beta$  (Table S2). If we define a region of interest (ROI) based on previous imaging studies implicating the right TPJ (centered at  $[x, y, z] = [54, -54, 24]$ , see Experimental Procedures) and compute the correlation between GM volume and  $\beta$ , we also obtain a high and significant correlation (Figure 3B,  $r = 0.61$ ,  $p < 0.001$ ), while preferences for altruism in the domain of disadvantageous inequality  $\alpha$  are uncorrelated with GM volume (Figure 3C,  $r = -0.01$ ,  $p = 0.95$ ). These results suggest a specific role of the TPJ in altruistic behaviors in the domain of advantageous inequality.

In addition to measuring the baseline levels of altruistic preferences in the domain of advantageous and disadvantageous inequality, our behavioral experiments also enable us to measure preferences for positive and negative reciprocity (Supplemental Information). Based on models of reciprocity developed in economics (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Rabin, 1993), we define positive reciprocity as the motive to respond in a kind manner to acts that are perceived as kind. In contrast, negative reciprocity is defined as the motive to respond in a hostile manner to acts that are perceived as hostile. According to this notion of reciprocity, individuals who are motivated by reciprocity are willing to behave reciprocally even if the reciprocal act is associated with a net cost for the acting party, i.e., even if there are no future material benefits that outweigh the cost of the reciprocal action. Thus, positive reciprocity means that a subject responds altruistically (i.e., increases the partner's payoff at his own cost) to an action of the partner that is perceived as kind relative to a neutral action; negative reciprocity means that a subject decreases a partner's payoff at his own cost in response to an action that is perceived to be hostile relative to a neutral action. We embed the notion of intention-based reciprocity into our model of social preferences that is based on Charness and Rabin (2002) and Fehr and Schmidt (1999). In our extended model, we measure an individual's preferences for positive reciprocity with parameter  $\theta$ , while parameter  $\delta$  represents preferences for negative reciprocity.

Interestingly, neither  $\theta$  nor  $\delta$  is significantly correlated with TPJ GM volume (Figures 3D and 3E) or with any other brain region (Table S2), which further supports the specificity of our finding for baseline altruism in the domain of advantageous inequality. We also conducted a multiple regression analysis to examine the robustness of the association between TPJ GM volume and  $\beta$  while controlling for all other preference parameters ( $\alpha$ ,  $\delta$ ,  $\theta$ ), as well as for age, gender, political attitude, and autistic traits. Again,  $\beta$  is highly significant ( $p = 0.004$ , Table S3), while no other preference parameters are significantly correlated (all  $p > 0.5$ ) with TPJ GM volume. Even if we search for individual voxels within the ROI that show a correlation between GM and the preference parameters ( $\alpha$ ,  $\delta$ ,  $\theta$ ), no correlations emerge. GM volume is uncorrelated with preferences for altruism in



**Figure 3. Gray Matter Volume in the Right TPJ Predicts Altruism in Advantageous Situations**

(A) Statistical parametric map for the correlation between subjects' altruism preferences in the domain of advantageous inequality ( $\beta$ ) and gray matter (GM) volume in the right TPJ (peak:  $x = 63$ ,  $y = -42$ ,  $z = 21$ ,  $t = 5.94$ ,  $Z$  score = 4.61,  $p = 0.049$ , voxelwise whole-brain FWE corrected for multiple comparison). For visualization purposes, voxels that survive at  $p < 0.001$  uncorrected are depicted. (B) A strong positive correlation is observed if we define an a priori ROI in the right TPJ based on previous imaging studies that implicate this region and compute the correlation between GM volume in the ROI and subjects' altruism preferences in the domain of advantageous inequality ( $\beta$ ). (C) Preferences for altruism in the domain of disadvantageous inequality  $\alpha$  are not correlated with GM volume in TPJ. (D and E) Preferences for positive reciprocity  $\theta$  (D) or for negative reciprocity  $\delta$  (E) are also not correlated with GM volume in TPJ. The  $p$  values in (B)–(E) are the values for the bivariate correlations of GM volume in the a priori ROI and the preference parameter reported in these graphs.

the domain of disadvantageous inequality  $\alpha$  ( $p = 0.551$ , small volume [SV] FWE corrected) or with preferences for positive reciprocity  $\theta$  ( $p = 0.581$ , SV FWE corrected) or negative reciprocity  $\delta$  ( $p = 0.629$ , SV FWE corrected).

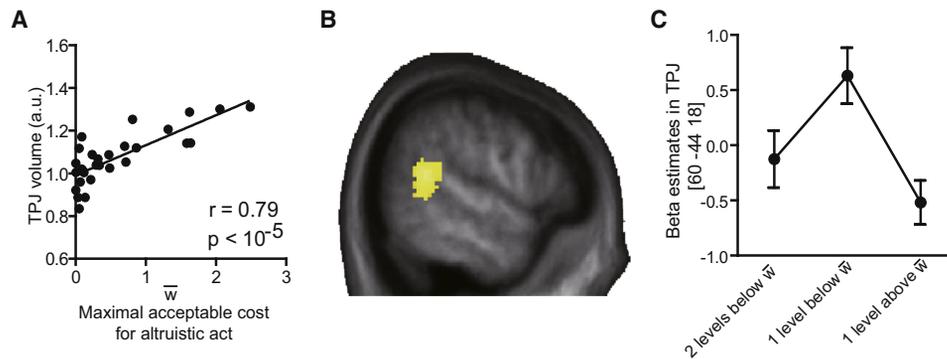
Finally, note that all our results are robust to the exclusion of the participant with extreme values of  $\beta$  and  $\alpha$  (top left data point in Figure 2). When we repeat the analyses without the data from this participant, our main findings remain the same: using the independent ROI specified above,  $\beta$  correlates significantly ( $r = 0.57$ ,  $p = 0.0013$ ) with TPJ GM volume, while all other parameters do not ( $p > 0.10$ ).

### Functional TPJ Activation Reflects the Individual-Specific Willingness to Pay for an Altruistic Act

These findings suggest that GM volume in TPJ may be a crucial neuroanatomical basis for subjects' baseline willingness to behave altruistically because the preference parameter  $\beta$  determines a subject's generosity in the domain of advantageous inequality. This parameter determines, in particular, the maximal cost (denoted by  $\bar{w}$ ) a subject is willing to bear to increase the partner's payoff by a given amount (say by one unit). The higher  $\beta$ , the higher the subject's maximum willingness to pay  $\bar{w}$  to increase the partner's payoff by one unit (see Figure S2). Therefore, subjects with a high  $\beta$  are generally willing to consider behaving altruistically for a much larger range of costly altruistic actions than those with a low value of  $\beta$ . In other words, if the costs of an altruistic act are relatively high, a subject with a relatively high value of  $\beta$  is still willing to consider behaving altruistically, while a subject with a low value of  $\beta$  will behave

selfishly in this situation. This means that  $\bar{w}$  represents a subject-specific cutoff value such that if the actual cost of the altruistic act is below  $\bar{w}$ , the subject will consider making an altruistic choice, while the subject behaves selfishly if the actual cost is above  $\bar{w}$ .

This insight about the role of  $\beta$  (and the implied role of  $\bar{w}$ ), together with the known functional role of the TPJ in perspective-taking tasks (Decety and Lamm, 2007; Frith and Frith, 2007; Saxe and Kanwisher, 2003; Young et al., 2010), can help us establish a link between GM volume in the TPJ and functional activations in TPJ during decision making in our task (in which subjects faced many different cost levels across trials). A high value of  $\beta$  implies a high maximum willingness to pay  $\bar{w}$ , meaning that the correlation between GM volume in right TPJ and  $\beta$  should translate into a correlation between GM volume and  $\bar{w}$  (see Figure 4A). In addition, taking the other individual's perspective seems particularly necessary in those cost situations in which a subject is in principle willing to behave altruistically (i.e., when the actual cost is below  $\bar{w}$ ) but in which self-interest provides a strong obstacle for altruistic acts because the cost is close to  $\bar{w}$ . In contrast, less perspective taking seems necessary in those situations in which the participant will behave selfishly anyway (i.e., when the actual cost is above  $\bar{w}$ ) or in which self-interest is no strong obstacle to behaving altruistically because the costs of altruistic acts is far below  $\bar{w}$ . We can thus predict an inverted U-shaped TPJ activation (in the domain of advantageous inequality) as a function of an individual's  $\bar{w}$ , with a peak at the cost level that is just below the maximally acceptable cost  $\bar{w}$ .



**Figure 4. Functional Activation in Right TPJ Reflects the Individual-Specific Willingness to Pay for an Altruistic Act**

(A) Relationship between individuals' GM volume in TPJ and their maximum willingness to pay  $\bar{w}$  for an altruistic act. Higher GM volume in TPJ is associated with a higher willingness to pay. (B) Statistical parametric map of voxels in the TPJ that exhibit maximal activity for altruistic decisions with a cost just below the individual willingness-to-pay  $\bar{w}$  and an inverted U-shaped activation pattern (MNI coordinates of peak: 60, -44, 18;  $p = 0.003$ , FWE corrected for the cluster displayed in Figure 3A). For visualization purposes, voxels that survive  $p < 0.001$  uncorrected are depicted. (C) Average BOLD signal estimates (and SEM) in the right TPJ as a function of the individual-specific willingness to pay for an altruistic act. We find an inverted U-shaped activity profile around  $\bar{w}$ .

Figures 4B and 4C show that functional TPJ activity (peak coordinate  $[x, y, z] = [60, -44, 18]$ ,  $t$  value = 4.12,  $p = 0.003$  FWE corrected for the volume of the cluster shown in Figure 3A) indeed follows such an activity profile, with the strongest activation for those situations in which the cost of an altruistic act is just below an individual's  $\bar{w}$ . Our results thus indicate that GM volume in TPJ is associated with both subjects' baseline altruism as measured by  $\beta$  and subject-specific functional activity profiles in the TPJ. In other words, GM volume in TPJ correlates with the general propensity to behave altruistically in the domain of advantageous inequality (Figures 3A and 3B), which in turn determines the individual-specific cutoff value of the maximum willingness to pay  $\bar{w}$  (Figure 4A). The subject-specific value of  $\bar{w}$  then determines the cost level for altruistic acts at which the peak of functional brain activation in TPJ occurs (Figures 4B and 4C), which concludes the link between brain structure (as measured by GM volume in right TPJ), individual behavioral tendencies, and patterns of functional brain activity in right TPJ.

## DISCUSSION

The present study demonstrates a link between neuroanatomical brain structure and human altruism: GM volume in the right TPJ, an area that has been shown to be implicated in perspective-taking tasks, is strongly associated with individuals' behavioral altruism in situations of advantageous inequality. These data also provide a plausible biological account of the stability of altruistic preferences. Previous research has documented that individuals' propensity for altruism is relatively stable across time, but these studies did not provide any biological basis for this temporal stability (Benz and Meier, 2008; Van Lange, 1999). The present study shows that anatomical structure, which does not change over short periods of time, can account for the strong heterogeneity in individuals' preferences for altruistic acts.

Furthermore, the link between GM volume in TPJ and subjects' preferences for altruism also provides insights into

the individual-specific conditions under which brain activity in TPJ is recruited when subjects face a tradeoff between economic self-interest and other people's interests. We hypothesized, in particular, that functional brain activation in the right TPJ is highest when the cost of the altruistic act is just below an individual's maximum willingness to pay for the altruistic act, while activation in right TPJ is low when the actual costs exceed or are far below the individual's maximum willingness to pay. Our functional findings support this hypothesis. Taken together, our data thus suggest that the right TPJ is important at the structural-anatomical level for subjects' baseline propensity to behave altruistically, while the concrete extent of an individual's functional TPJ activation is dependent on the context, i.e., on the relationship between the individual's maximum willingness to pay for an altruistic act and the cost of the altruistic act.

Previous functional imaging studies have shown that the right posterior superior temporal cortex (pSTC) is activated during perspective-taking tasks and charitable donation tasks. Hare et al. have shown, for example, that higher activation in this region during decisions on charitable donations reflects the correlation between the subjects' ratings of charities' deservingness and the subjects' actual donation to the charities (Hare et al., 2010). Tankersley et al. have shown that the right pSTC is more activated if subjects passively observe the outcome of an event that triggers money transfers to a charity compared to when they themselves make decisions that have positive monetary consequences for the charity; in addition, this pSTC activation also predicts questionnaire measures of subjects' altruism (Tankersley et al., 2007). These studies, however, do not examine how individual differences in (task-independent) brain structure are related to subjects' behaviorally expressed preferences for altruism; therefore, they do not establish a link between individual differences in brain structure and the individual-specific conditions for the functional activation of TPJ in the altruism task.

In addition to the TPJ, previous imaging studies have shown involvement of other brain structures such as the ventromedial

prefrontal cortex (vmPFC) and ventral or dorsal striatum in altruistic behavior (de Quervain et al., 2004; Krajbich et al., 2009; Krueger et al., 2007; Moll et al., 2006; Tricomi et al., 2010). However, in contrast to the TPJ, these latter areas are routinely found to be involved in nonsocial types of decision making such as reward-seeking behavior, intertemporal decision making, risk taking, and purchasing behavior (Kable and Glimcher, 2007; Kepecs et al., 2008; Knutson et al., 2007; Kuhnen and Knutson, 2005; Padoa-Schioppa and Assad, 2006; Plassmann et al., 2007; Rangel and Hare, 2010; Samejima et al., 2005). Activity in the vmPFC and ventral striatum thus seems to relate to domain-general processes important for many different types of decisions. We thus did not predict these brain areas to be as specific for altruistic decisions as the TPJ with its well-documented role in social cognitive processes such as perspective taking (Decety and Lamm, 2007; Frith and Frith, 2007; Ruby and Decety, 2001; Saxe and Kanwisher, 2003; Young et al., 2010).

Beyond the understanding of the role of the right TPJ in altruistic behavior, our results demonstrate a biological link between inter- and within-individual behavioral variability. Brain structure—in terms of GM volume in a particular brain region—accounts for interindividual variability in subjects' baseline behavioral properties. In addition, the same brain structure also accounts for within-individual variations in behavior dependent on the specific context (which, in our case, is given by the cost of the altruistic act). It is worthwhile to point out that we established this link between inter- and within-individual variability using the estimation of a mathematical model of preferences that captures both the between-subject differences in preferences and the within-subject responses to cost variations. A similar research strategy might also be productively applied to bridge the gap between brain structure and brain function in other behavioral domains.

## EXPERIMENTAL PROCEDURES

### Subjects

Thirty normal healthy adults (17 females; 19–37 years; mean 23.36 years) participated in this study. All subjects gave written informed consent. The study was approved by the ethics committee of the Canton of Zurich. One subject was excluded due to very inconsistent behavior, making the estimation of preference parameters impossible for this subject.

### Behavioral Task and Model

We implemented two types of games, dictator games and reciprocity games. Subjects in the dictator game (player A) were asked to choose one option from two possible allocations of money, option X and option Y (Figure 1A). The reciprocity games allow us to measure preferences for positive and negative reciprocity (Figures 1B and 1C; for details of the task, see Supplemental Experimental Procedures). We applied a model of social preferences in order to estimate each individual's preferences for altruistic acts. Formally, the model can be represented by the following equation:

$$U_A(\Pi_A, \Pi_B) = (1 - \beta r - \alpha s - \theta q + \delta v)\Pi_A + (\beta r + \alpha s + \theta q - \delta v)\Pi_B$$

where  $U_A$  denotes player A's utility,  $\Pi_A$  represents player A's monetary payoff, and  $\Pi_B$  denotes player B's monetary payoff.  $\beta$  and  $\alpha$  are parameters that measure the preference for altruistic acts in the domain of advantageous and disadvantageous situations, respectively. A positive value of  $\theta$  means that the subject has a preference for positive reciprocity, while a positive value of  $\delta$  represents a preference for negative reciprocity. The symbols  $r$ ,  $s$ ,

$q$ , and  $v$  are binary variables that take on the value 1 or 0, depending on the situation in which players A and B are. In particular, the following holds for  $r$ ,  $s$ ,  $q$ , and  $v$ :

- $r = 1$  if  $\Pi_A > \Pi_B$ , and  $r = 0$  otherwise (advantageous inequality);
- $s = 1$  if  $\Pi_A < \Pi_B$ , and  $s = 0$  otherwise (disadvantageous inequality);
- $q = 1$  if player B behaved altruistically toward A and  $q = 0$  otherwise (positive reciprocity);
- $v = 1$  if player B behaved selfishly toward A and  $v = 0$  otherwise (negative reciprocity).

Details of the behavioral model are described in the Supplemental Experimental Procedures.

### Image Acquisition and MRI Data Analysis

We used the Philips Intera whole-body MR Scanner (Philips Medical Systems) at the SNS laboratory of the University of Zurich, equipped with an 8-channel Philips SENSitivity Encoded (SENSE) head coil. High-resolution structural T1-weighted 3D-TFE (3D-turbo fast echo) images (TR = 7.5 s; TE = 3.5 ms; FA = 8 deg; FOV 250 × 250 mm; voxel size 1.04 × 1.04 × 0.6 mm; 301 sagittal slices) were acquired for each participant. The functional images sensitive to blood-oxygen level-dependent (BOLD) contrasts were acquired by T2\*-weighted echo-planar imaging (TR = 1.45 s; TE = 30 ms; inplane resolution of 3 mm in 64 × 64 matrix; 28 slices; slice thickness of 3 mm; 1.5 mm interslice gap). We used SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) for MRI data preprocessing and analysis. Details of the MRI data analysis are described in the Supplemental Experimental Procedures.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, five tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2012.05.021.

## ACKNOWLEDGMENTS

We thank C. Burke, T. Hare, G. Hein, S. Leiberg, and P. Tobler for useful comments on the manuscript, and K.E. Stephan for advice on MRI data analysis. This work was supported by the Swiss National Center of Competence in the Affective Science and the Neurochoice Project of Systems X (E.F.), JSPS (Y.M.), and Naito Foundation (Y.M.).

Accepted: May 16, 2012

Published: July 11, 2012

## REFERENCES

- Benz, M., and Meier, S. (2008). Do people behave in experiments as in the field? evidence from donations. *Exp. Econ.* 11, 268–281.
- Bolton, G.E., and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193.
- Camerer, C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, N.J.: Princeton University Press).
- Charness, G., and Rabin, M. (2002). Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869.
- de Quervain, D.J.F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258.
- Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593.
- Dufwenberg, M., and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games Econ. Behav.* 47, 268–298.
- Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315.

- Fehr, E., and Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* *114*, 817–868.
- Frith, C.D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* *17*, R724–R732.
- Hamilton, W.D. (1964). The genetical evolution of social behaviour. II. *J. Theor. Biol.* *7*, 17–52.
- Harbaugh, W.T., Mayr, U., and Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* *316*, 1622–1625.
- Hare, T.A., Camerer, C.F., Knoepfle, D.T., and Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* *30*, 583–590.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). “Economic man” in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* *28*, 795–815.
- Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* *10*, 1625–1633.
- Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* *12*, 231–242.
- Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* *455*, 227–231.
- Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. *Neuron* *53*, 147–156.
- Krajbich, I., Adolphs, R., Tranel, D., Denburg, N.L., and Camerer, C.F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J. Neurosci.* *29*, 2188–2192.
- Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., and Grafman, J. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. USA* *104*, 20084–20089.
- Kuhnen, C.M., and Knutson, B. (2005). The neural basis of financial risk taking. *Neuron* *47*, 763–770.
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. USA* *103*, 15623–15628.
- Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* *441*, 223–226.
- Plassmann, H., O’Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* *27*, 9984–9988.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* *83*, 1281–1302.
- Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* *20*, 262–270.
- Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* *4*, 546–550.
- Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* *310*, 1337–1340.
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* *19*, 1835–1842.
- Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., and Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *J. Exp. Child Psychol.* *105*, 130–137.
- Tankersley, D., Stowe, C.J., and Huettel, S.A. (2007). Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* *10*, 150–151.
- Tricomi, E., Rangel, A., Camerer, C.F., and O’Doherty, J.P. (2010). Neural evidence for inequality-averse social preferences. *Nature* *463*, 1089–1091.
- Underwood, B., and Moore, B. (1982). Perspective-Taking and Altruism. *Psychol. Bull.* *91*, 143–173.
- Van Lange, P.A.M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *J. Pers. Soc. Psychol.* *77*, 337.
- Young, L., Dodell-Feder, D., and Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* *48*, 2658–2664.