A Neurocomputational Model of Altruistic Choice and Its Implications

Highlights

- A drift diffusion model of altruism explains choice, RT, and neural response
- Striatum, TPJ, and vmPFC encode distinct quantities required by the model
- The model predicts RT and time pressure effects on generosity without self-control
- Reward signals confirm model prediction that generosity is sometimes a mistake

Authors

Cendri A. Hutcherson, Benjamin Bushong, Antonio Rangel

Correspondence

chutcher@hss.caltech.edu

In Brief

Hutcherson et al. show that a computational model of altruism accounts for behavioral and neural effects attributed to self-control and/or the value of generosity, without requiring either. The model suggests that some generosity represents choice errors, not genuine social preferences.
A Neurocomputational Model of Altruistic Choice and Its Implications

Cendri A. Hutcherson,1,4,* Benjamin Bushong,1,3 and Antonio Rangel1,4
1Division of the Humanities and Social Sciences
2Computational and Neural Systems
California Institute of Technology, Pasadena, CA 91125, USA
3Department of Economics, Harvard University, Cambridge, MA 02138, USA
4Department of Psychology, University of Toronto, Toronto, ON M1C 1A4, Canada
*Correspondence: chantch@hss.caltech.edu
http://dx.doi.org/10.1016/j.neuron.2015.06.031

SUMMARY

We propose a neurocomputational model of altruistic choice and test it using behavioral and fMRI data from a task in which subjects make choices between real monetary prizes for themselves and another. We show that a multi-attribute drift-diffusion model, in which choice results from accumulation of a relative value signal that linearly weights payoffs for self and other, captures key patterns of choice, reaction time, and neural response in ventral striatum, temporoparietal junction, and ventromedial prefrontal cortex. The model generates several novel insights into the nature of altruism. It explains when and why generous choices are slower or faster than selfish choices, and why they produce greater response in TPJ and vmPFC, without invoking competition between automatic and deliberative processes or reward value for generosity. It also predicts that when one’s own payoffs are valued more than others’, some generous acts may reflect mistakes rather than genuinely pro-social preferences.

INTRODUCTION

Altruism involves helping others at a cost to the self, not only when such behavior is supported by strategic considerations like reciprocity or cooperation (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Nowak and Sigmund, 1998) but even in the absence of expectation for future benefit (e.g., fully anonymous, one-time generosity; Batson, 2011; Fehr and Fischbacher, 2003). A major goal of neuroeconomics is to develop neurocomputational models of altruistic choice, specifying which variables are computed, how they interact to make a decision, and how are they implemented by different brain circuits. Such models have proven useful in domains such as perceptual decision making (Gold and Shadlen, 2007; Heekeren et al., 2008), simple economic choice (Basten et al., 2010; Hunt et al., 2012; Rangel and Clithero, 2013), self-control (Hare et al., 2009; Kable and Glimcher, 2007; Peters and Büchel, 2011; van den Bos and McClure, 2013), and social learning (Behrens et al., 2008; Boorman et al., 2013). We propose a neurocomputational model of simple altruistic choice and test it using behavioral and fMRI data from a modified Dictator Game in which subjects make choices between pairs of real monetary prizes for themselves ($Self$) and another ($Other$). These choices involve a tradeoff between what is best for the self and what is best for the other and thus require people to choose to act selfishly or generously.

Our model assumes that choices are made by assigning an overall value to each option, computed as the weighted linear sum of two specific attributes: monetary prizes for self and other. This type of simple value calculation captures a wide range of behavioral patterns in altruistic choice (Charness and Rabin, 2002; Eckel and Grossman, 1996; Engel, 2011; Fehr and Fischbacher, 2002; Fehr and Fischbacher, 2003). Our model also assumes that the overall value signal is computed with noise and that choices are made using a multi-attribute version of the Drift-Diffusion Model (DDM; Ratcliff and McKoon, 2008; Smith and Ratcliff, 2004). In this algorithm, a noisy relative value signal is integrated at each moment in time and a choice is made when sufficient evidence has accumulated in favor of one of the options. This type of algorithm has been shown to provide accurate descriptions of both choice and reaction time (RT) data (Busemeyer and Townsend, 1993; Hunt et al., 2012; Krajbich et al., 2010; Milosavljevic et al., 2010; Rodriguez et al., 2014; Smith and Ratcliff, 2004), as well as neural response patterns associated with computing and comparing values (Basten et al., 2010; Hare et al., 2011; Hunt et al., 2012) in many non-social domains.

The model suggests neural implementation of two specific quantities. First, values for the attributes $Self$ and $Other$ must be computed independently. Second, an overall value signal must be constructed from the independent attributes. We hypothesized that areas like the temporoparietal junction, precuneus, or medial prefrontal cortex may compute quantities related to the value of these attributes. Prior research strongly implicates these regions in social behavior (Bruneau et al., 2012; Carter and Huettel, 2013; de Vignemont and Singer, 2006; Decety and Jackson, 2006; Hare et al., 2010; Jackson et al., 2005; Moll et al., 2006; Saxe and Powell, 2006; Singer, 2006; Waytz et al., 2012; Zaki and Mitchell, 2011), although their precise computational roles remain poorly understood. Inspired by a large body of work on the neuroeconomics of non-social...
choice (Basten et al., 2010; Hare et al., 2009; Kable and Glimcher, 2007; Lim et al., 2013; McClure et al., 2004; Tom et al., 2007), we additionally hypothesized that the integration of specific attribute signals would occur in ventromedial prefrontal cortex (vmPFC). We explore these hypotheses with our fMRI dataset.

We also highlight three ways in which the development of a computational model of altruistic choice can be used to generate novel insights into the nature of altruistic choice. First, we compare the model’s predictions about RT and neural response for generous versus selfish choices. We find that, for the best-fitting parameters, the model predicts longer RT and higher blood-oxygen-level-dependent (BOLD) response in decision-related regions for generous choices and that the predicted effect sizes match the observed data. Second, we use simulations to identify how model parameters influence altruistic behavior and find that several of these variables (including the relative importance of benefits to self and other and the decision boundaries of the DDM) predict observed individual differences in generosity. Third, we show that the model predicts that generous decisions are sometimes unintended mistakes resulting from the noisy choice process and exploit an aspect of our experimental design to test this using fMRI data.

RESULTS

We collected whole-brain BOLD responses in male subjects while they made 180 real decisions about different allocations of money between themselves and a real-but-anonymous partner. Each trial consisted of a choice phase and an outcome phase (Figure 1A). During the choice phase, the subject saw a proposal consisting of monetary prizes for himself ($Self$) and for another person ($Other$) and had to decide whether to accept or reject it in favor of a constant default prize of $50 for each. On each trial the subject saw one of the nine proposal types depicted in Figures 1B and 2C–2D, with ±$1–$4 random jitter added to avoid habituation. All proposals included one payment below and one payment above the default, creating a choice between generous behavior (benefitting the other at a cost to oneself) and selfish behavior (benefitting oneself at a cost to
another). Subjects indicated their decision using a four-point scale (1 = Strong No, 2 = No, 3 = Yes, 4 = Strong Yes), allowing us to measure both the choice and the value assigned to the proposal. Right-left orientation of the scale varied randomly from scan to scan to reduce motor-related confounds in neural response. Every decision was followed by an outcome phase, during which the decision made by the subject was implemented with 60% probability and reversed with 40% probability. Subjects were told about the 40% probability of choice reversal and that their partner knew their choices might be reversed but were encouraged to simply choose the option they most preferred, since their choice made it more likely to occur (see Supplemental Experimental Procedures for instructions). At the end of the experiment, one trial was randomly selected and its outcome implemented. As shown below, the reversal mechanism allows us to test the extent to which different choices may be decision mistakes, while not changing incentives to pick the best option.

Average Choices Are Relatively Selfish
Subjects made generous choices—maximizing their partner’s payoff ($Other) at a cost to their own ($Self)—in 21% ± 18% (mean ± SD) of trials, sacrificing $3.73 ± $4.84 per trial and giving $8.31 ± $6.86. This level of giving is comparable to other studies of anonymous altruism (Engel, 2011) but also suggests that subjects in general behaved relatively selfishly. There was considerable individual variation in generosity, ranging from 0%–61% generous choices and $0–$22.37 given to the partner. This variation is useful for exploring individual differences, as we do below.

Computational Model
The model is a multi-attribute extension of the standard DDM (Ratcliff and McKoon, 2008; Smith and Ratcliff, 2004). On every trial the choice is based on a dynamically evolving stochastic relative decision value (RDV) signal that provides an estimate of the desirability of the proposed prize ($Self, $Other) relative to the default prize ($50, $50). The signal starts at zero, remaining there for an amount of non-decision time capturing processing and motor delays, given by the parameter NDT. Afterward, it accumulates stochastically at time $t$ according to the difference equation

$$RDV_t = \frac{RDV_{t-1}}{1} + w_{self}(\frac{$Self - $50}{C0}) + w_{other}(\frac{$Other - $50}{C0}) + \epsilon_t,$$

where $Self$ and $Other$ are the proposed prizes for self and other, $w_{self}$ and $w_{other}$ are constant weights, and $\epsilon_t$ denotes white Gaussian noise that is identically and independently distributed with standard deviation $\sigma$. A choice is made the first time the RDV crosses one of two pre-specified barriers. The proposal is accepted if the positive barrier is crossed first and rejected if the negative barrier is crossed first. RT equals the sum of the NDT and crossing time $t$. Building on previous work with time-limited decisions (Churchland et al., 2008; Cisek et al., 2009; Milosavljevic et al., 2010), we allow for the possibility of collapsing barriers, although the model includes fixed barriers as a special case. The upper barrier is described by the equation

$$B_t = be^{-dt},$$

where $b > 0$ is a parameter denoting the initial height of the barrier, $d \geq 0$ is a parameter denoting its exponential rate of decay, and $t$ is measured from the end of the non-decision period. The...

Figure 2. Model Fits to Behavior
(A) Model-predicted versus observed average generosity across subjects. Dashed 45° line represents a perfect match.
(B) Model-predicted versus observed overall response time (RT).
(C and D) Within-subject acceptance likelihood (C; mean ± SEM) and RT (D; mean ± SEM) for each of the 9 proposal types. Observed behavior, gray bars. Predict behavior, blue circles.
Table 1. Parameters of the Best-Fitting DDM for Each Subject

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{\text{Self}}$</td>
<td>0.006</td>
<td>0.002</td>
<td>0</td>
<td>0.0105</td>
</tr>
<tr>
<td>$w_{\text{Other}}$</td>
<td>0.001</td>
<td>0.0026</td>
<td>-0.003</td>
<td>0.009</td>
</tr>
<tr>
<td>NDT</td>
<td>868 ms</td>
<td>241 ms</td>
<td>300 ms</td>
<td>1,300 ms</td>
</tr>
<tr>
<td>$b$</td>
<td>0.23</td>
<td>0.065</td>
<td>0.08</td>
<td>0.32</td>
</tr>
<tr>
<td>$d$</td>
<td>0.00046</td>
<td>0.00022</td>
<td>0</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 1 indicates that the fit for choice behavior (though not RT) was poorer when the proposal involved a sacrifice for the subject (i.e., $S_{\text{Self}}$ amounts below the default). To investigate this issue, we fit a variant of the model that allows the parameters to depend on whether $S_{\text{Self}}$ is more or less than $S_{\text{Other}}$. This alternative model is motivated by previous behavioral work showing that the value placed on $S_{\text{Self}}$ and $S_{\text{Other}}$ can depend on whether the self is coming out ahead or behind (Charness and Rabin, 2002; Engelmann and Strobel, 2004; Fehr and Schmidt, 1999). As detailed in the Supplemental Information, this analysis improves the fit to observed choice behavior when $S_{\text{Self}} < S_{\text{Other}}$, an effect that derives from a higher weight $w_{\text{Self}}$, a lower weight $w_{\text{Other}}$, and a higher threshold parameter. However, because there are no qualitative differences in the analyses reported below when using the more complex model, for simplicity the rest of the analyses utilize the simpler version.

Estimated Model Parameters in the Full Dataset

Having demonstrated that the model accurately predicts out-of-sample choices and RTs, we next examine the best-fitting parameter values using the full dataset (see Table 1). Several results are worth highlighting. First, the average NDT (868 ms) is larger than that usually found for DDMs (Milosavljevic et al., 2010; Ratcliff and McKoon, 2008). We attribute this to the additional time subjects may have needed to determine the payoffs on each trial and translate that into a graded response. Second, both $w_{\text{Self}}$ and $w_{\text{Other}}$ are significantly larger than zero on average (both $p < 0.003$). The model suggests that an integrated value signal is used to make choices. We provide neural evidence of such a signal by estimating a general linear model to identify regions in which BOLD responses correlate positively with the value assigned to proposals at the time of decision, measured by the four-point response scale (1 = Strong-No to 4 = Strong-Yes). Several regions satisfy this property, including a region of vmPFC (p < 0.05, whole-brain corrected [WBC]; Figure 3A; Table S3) that encodes stimulus values at the time of decision in a wide range of tasks (Clithero and Rangel, 2014; Kable and Glimcher, 2007).

The Model Accurately Predicts Out-of-Sample Choice and RT

We used a maximum likelihood method based on simulated likelihood functions to estimate the best-fitting parameters of the model in a randomly selected half of the data. We used these parameters to test the fit between model predictions and observed data on the other half, separately for each subject (see Experimental Procedures for details). Model predictions capture interindividual differences in mean donations (mean Pearson’s $r_{49} = 0.94$, $p < 0.0001$) and RTs ($r_{49} = 0.96$, $p < 0.0001$) quite well (Figures 2A and 2B). The model also captures intra-individual differences in acceptance rates (mean $r = 0.88$, one-sample $t_{49} = 45.05$, $p < 0.0001$; Figure 2C) and in RTs (mean $r = 0.53$, one-sample $t_{49} = 11.79$, $p < 0.0001$; Figure 2D) across different trial types.

Although this suggests that the model described above fits well, Figure 2C indicates that the fit for choice behavior (though not RT) was poorer when the proposal involved a sacrifice for the subject (i.e., $S_{\text{Self}}$ amounts below the default). To investigate this issue, we fit a variant of the model that allows the parameters to depend on whether $S_{\text{Self}}$ is more or less than $S_{\text{Other}}$. This alternative model is motivated by previous behavioral work showing that the value placed on $S_{\text{Self}}$ and $S_{\text{Other}}$ can depend on whether the self is coming out ahead or behind (Charness and Rabin, 2002; Engelmann and Strobel, 2004; Fehr and Schmidt, 1999). As detailed in the Supplemental Information, this analysis improves the fit to observed choice behavior when $S_{\text{Self}} < S_{\text{Other}}$, an effect that derives from a higher weight $w_{\text{Self}}$, a lower weight $w_{\text{Other}}$, and a higher threshold parameter. However, because there are no qualitative differences in the analyses reported below when using the more complex model, for simplicity the rest of the analyses utilize the simpler version.

vmPFC Responses Encode an Integrated Value Signal at Decision

The model suggests that an integrated value signal is used to make choices. We provide neural evidence of such a signal by estimating a general linear model to identify regions in which BOLD responses correlate positively with the value assigned to proposals at the time of decision, measured by the four-point response scale (1 = Strong-No to 4 = Strong-Yes). Several regions satisfy this property, including a region of vmPFC (p < 0.05, whole-brain corrected [WBC]; Figure 3A; Table S3) that encodes stimulus values at the time of decision in a wide range of tasks (Clithero and Rangel, 2014; Kable and Glimcher, 2007).

Neural Representations of $S_{\text{Self}}$ and $S_{\text{Other}}$

Our model assumes that the overall value assigned to the proposal (and used by the DDM comparator algorithm to generate a choice) is constructed from information about the independent attributes $S_{\text{Self}}$ and $S_{\text{Other}}$. We show that there are neural
signals consistent with representation of these two quantities, using a second model to look for areas in which BOLD responses correlate positively with either $Self or $Other separately (i.e., inputs to the integrated value signal), as well as regions that reflect both (i.e., overall values).

$Self correlates with BOLD responses in a distributed set of regions (Figure 3B; Table S4), including vmPFC ($p < 0.05$, WBC) and the ventral striatum ($p < 0.05$, WBC). $Other correlates with BOLD responses in a distinct and more circumscribed set of regions (Figure 3C; Table S4), including right temporoparietal junction (rTPJ), precuneus (both $p < 0.05$, WBC), and vmPFC ($p < 0.004$ SVC). To determine the specificity of these responses, we looked for regions in which the effect for $Self is stronger than for $Other, and vice versa. Regions responding more strongly to $Self include the ventral striatum ($p < 0.05$, SVC), vmPFC, and areas of visual and somatosensory cortex ($p < 0.05$, WBC). No regions respond more strongly to $Other at our omnibus threshold, although we observe such specificity in the right TPJ at a more liberal threshold ($p < 0.005$, uncorrected). A conjunction analysis (Table S4) shows a region of vmPFC (Figure 3D) responding significantly to both $Self and $Other ($p < 0.05$, SVC). This area also overlaps fully the vmPFC area correlating with overall preference, supporting the idea that it may represent an area where separate attributes are combined into an integrated value signal.

Together, the behavioral and neural results are consistent with the hypothesis that both $Self and $Other, quantities required by the computational model, are independently represented in the brain. The results also support the idea that the vmPFC combines information about $Self and $Other into an overall value and that choices are made by integrating the proposal values using an algorithm that is well captured by the DDM. These results motivate the second part of the paper, in which we use the best-fitting computational model to derive and test several implications of the theory.

**Implication 1: RTs Are Longer for Generous Choices, Particularly for More Selfish Individuals**

Our computational model has the advantage that it provides a theory of the relationship between choices and RTs. This is of particular interest because differences in RT when choosing to act selfishly or generously have been used in several studies to make inferences about the relative automaticity of pro-social behavior (Piovesan and Wengstrom, 2009; Rand et al., 2012). Simulations from the individual models reveal two interesting predictions. First, in the domain of best-fitting parameters for our subjects, the model predicts that on average RTs are longer for trials that result in a generous (G) choice compared to trials resulting in a selfish (S) choice (predicted RT$_G = 2,269$, predicted RT$_S = 2,074$, paired-t$_{50} = 9.37$, $p < 0.0001$; Figure 4A). Second, it predicts that this RT difference is bigger for more selfish subjects (correlation between predicted generosity and difference in G versus S RTs $r_{49} = 0.89$, $p < 0.0001$; Figure 4B). The observed data displayed both patterns. On average, G choices were significantly slower than S choices (RT$_G = 2,300$ ms ± 310, RT$_S = 2,131$ ms ± 280, paired-t$_{43} = 4.97$, $p < 0.0001$; Figure 4C), and the more generous the individual, the smaller this difference ($r_{42} = -0.60$, $p < 0.001$, Figure 4D).

**Implication 2: Neural Response in Valuation and Comparison Regions Is Higher for Generous Choices**

Our computational model suggests a neural corollary of differences in RT: regions whose activity scales with computation in the comparison process should have higher responses during G compared to S choices (predicted comparator response, arbitrary units: Comp$_G = 69.68 \pm 27.97$, S choice = 65.76 ± 27.62, paired-t$_{50} = 6.43$, $p < 0.0001$, see Experimental Procedures for details).
To see why, note that the predicted area-under-the-curve of the accumulator process is larger on longer trials and that inputs into this process must also be sustained until the process terminates at a decision barrier. This prediction is important, since many studies of altruism have observed differential response during pro-social choices in regions like the vmPFC and TPJ and interpreted it as evidence that such choices are rewarding (Zaki and Mitchell, 2011) or that they require the inhibition of selfish impulses by the TPJ (Strombach et al., 2015). In contrast, our model suggests that such differences could be a straightforward by-product of the integration and comparison process.

To test this prediction, we first defined two independent ROIs shown in previous research to have differential response during G choices: (1) a value-modulated vmPFC region (Figure 5A) based on the set of voxels that correlated significantly with stated preference at the time of choice (p < 0.001, uncorrected); and (2) a generosity-related TPJ region (Figure 5B) based on an 8-mm sphere around the peak coordinates of a recent study reporting greater activation in the TPJ when subjects chose generously (Strombach et al., 2015). In both regions, we replicate the pattern of higher response on G versus S choices (both p < 0.02). Critically, however, we also find that differential BOLD on G versus S choices correlates positively with predicted differences in accumulator response in both regions (both rG, < 0.46, p < 0.001). Moreover, accounting for predicted accumulator differences reduces non-significance the differential generosity-related response in both vmPFC (p = 0.92) and TPJ (p = 0.91). In contrast, response in occipital and motor cortices (which show value modulation but are unlikely to perform value integration and comparison) bear little resemblance to predictions of the model (Figure S2).

Implication 3: Relationships between Model Parameters and Generosity
In order to understand the impact on generosity of variation in the different parameters, we simulate model predictions for our task in the regression, all parameters are normalized by their mean and standard deviation in order to assess their influence on a common scale. As illustrated in Figures 6A–6B, we find the expected association between average generosity and both wSelf (β = 4.86, p < 0.0001) and wOther (β = 9.26, p < 0.0001). Intriguingly, the simulations also reveal that a lower starting threshold (β = −0.141, p = 0.0001), and a faster collapse rate (β = 0.21, p < 0.0001) increase generosity. That is, individuals with less stringent decision criteria, leading to less accurate (and in this case, generosity-prone) behavior. The model predicts no relationship with NDTs (p = 0.79).

We use a similar regression to see whether the same relation is evident in the observed data. As with the simulated data, the fitted parameters were z scored to assess their influence on a common scale. Consistent with model predictions, we find that average observed generosity correlates negatively with wSelf (β = −4.86, p < 0.0001) and positively with wOther (β = 9.26, p < 0.0001). Also as predicted, observed generosity correlates negatively with the height of the decision threshold (β = −4.86, p < 0.0001) and positively with the rate at which the threshold collapses toward zero (β = 0.21, p < 0.0001). The NDT parameter is non-significant, as expected.

Implication 4: Errors Are More Likely to Involve Generous Choices
Because choices are stochastic, the model suggests that some may be errors (i.e., a decision in which the option with the higher relative value is not chosen; Bernheim and Rangel, 2005). We use the simulated data to investigate how decision mistakes change with variation in the model parameters, and how this affects generosity. Multivariate regression analyses on the theoretical data, where mistakes can be identified precisely on every
trial, show that error rates decrease with $w_{Self}$ ($\beta = -0.05, p < 0.0001$) and $w_{Other}$ ($\beta = -0.014, p < 0.0001$) and increase with more liberal barrier parameters for $b$ ($\beta = -0.04, p < 0.0001$) and $d$ ($\beta = 0.05, p < 0.0001$). We also assess the relationship between model parameters and the relative percentage of trials that result in generous errors (mistakenly choosing to give to the other) versus selfish errors (mistakenly choosing to keep more money). This assesses whether different parameters increase generosity by increasing errors. As shown in Figure 6C, $w_{Self}$ ($\beta = 0.026, p < 0.0001$) and $w_{Other}$ ($\beta = -0.047, p < 0.0001$) influence the relative balance toward generous errors in opposite ways. Increasing the height of the barrier decreases the bias toward generous errors ($b$: $\beta = -0.008, p < 0.0001$; $d$: $\beta = +0.01, p < 0.0001$; Figure 6D), while NDT has no effect ($\beta = -0.0001, p = 0.89$).

We next use the individually fitted weights $w_{Self}$ and $w_{Other}$ to define the “true” relative value of each proposal, which allows us to estimate the proportion of observed G and S choices for each subject that might reasonably be assumed to be errors. This analysis suggests that G choices were significantly more likely to be errors ($M = 49\% \pm 38\%$) than S choices ($M = 10\% \pm 21\%$, paired-$t_{30} = 5.45, p < 0.0001$).

We carry out a further test of this prediction, using outcome period BOLD responses, based on the following logic. The model suggests that a proposal’s true value should become increasingly clear to the decision circuitry as the amount of accumulated evidence increases over time, because random fluctuations in the signal will tend to cancel out. If subjects continue to accumulate evidence about the proposal even after making a choice (i.e., “double checking” whether they have made an error), then these signals should be quite clean by the time the subject sees the outcome of his choice. If this increased clarity leads a subject to realize at some point after making his choice that it was a mistake, having that mistake overturned during the outcome period (yielding the unchosen but ultimately preferred option) should be perceived as “good news” (i.e., relief), whereas having it implemented should be experienced as “bad news” (i.e., disappointment). If the original choice is actually correct, then reversal of this choice should be perceived negatively. The model thus predicts that reversal of S choices (which simulations suggest are generally likely to be correct) should be associated with negative affect and lower response in brain regions coding for the utility of an outcome, relative to non-reversal. In contrast, because G choices more likely reflect choice errors, reversing them should be more likely to evoke positive affect and greater neural response compared to implementation. Finally, the response in utility-coding areas to reversing a G choice should increase, across subjects, with the model-estimated likelihood that G choices are mistakes.

We tested these predictions by computing the difference between response in the vmPFC to reversal versus implementation of G or S choices, controlling both for the strength of preference at the time of decision, and for actual outcomes received (i.e., the amounts $\$Self$ and $\$Other$ resulting from choice combined with the random implementation, see GLM 1 in Experimental Procedures for details). Consistent with predictions, vmPFC response to reversal versus implementation was significantly higher after G compared to S choices ($p = 0.02, SVC, Figure 7A$). Also as predicted by the model, the difference in response in this region correlated positively with the estimated excess rate of mistakes for G over S choices ($r_{39} = 0.43, p = 0.004, Figure 7B$).

**DISCUSSION**

We have proposed a neurocomputational model of altruistic choice that builds on behavioral, neural, and computational work in non-social domains (Basten et al., 2010; Bogacz et al., 2010; Hare et al., 2011; Heekeren et al., 2008; Ratcliff and McKoon, 2008). In the model, decisions result from the stochastic accumulation of a relative value signal that linearly weights information about payoffs for self and other. Despite its simplicity, the model has considerable explanatory power. It accounts for differences in average levels of altruism and RTs within and across subjects, as well as for neural signals encoded in vmPFC, TPJ, and striatum at the time of choice. Our results provide insight into the common processes at work in altruistic choice and simple non-social decisions, shed light on some of the neural mechanisms specifically involved in the computation of social value, and provide novel insights into the nature of altruistic behavior.
help to incorporate others’ reward into the value signal, they
than lower amounts of $
This representation would appear selectively for higher rather
another person would value the proposal, it is not clear why
computations about belief might be used to determine how
fully explain the pattern of TPJ response in our task. Although
458
empathic and altruistic decision making (Decety and Jackson,
2006; Hare et al., 2010; Morishima et al., 2012; Saxe and Powell,
2006), its precise computational nature remains poorly under-
stood. Our findings show that, at least during altruistic choice, signals in this area may reflect computations related to others’ interests. This signal differs from two popular alternative ac-
counts of TPJ function: that it represents the beliefs of others (Saxe and Powell, 2006) or that it allows attention to be shifted away from the self (Scholz et al., 2009). Neither theory appears to fully explain the pattern of TPJ response in our task. Although computations about belief might be used to determine how another person would value the proposal, it is not clear why this representation would appear selectively for higher rather than lower amounts of $Other. Similarly, if attentional shifts help to incorporate others’ reward into the value signal, they

Simple versus Social Decision Making
A growing body of work suggests that in simple non-social choices the vmPFC receives information from regions computing information about different stimulus attributes (Basten et al., 2010; Hare et al., 2009; Kable and Glimcher, 2007; Lim et al., 2013) and combines it into a relative value signal (Hare et al., 2009; Kable and Glimcher, 2007). This signal is then dynamically integrated in comparator regions using algorithms with properties similar to the DDM (Basten et al., 2010; Hare et al., 2011; Hunt et al., 2012). BOLD responses in our study suggest a similar neural ar-
chitecture for social choice: we observed attribute-coding regions like the striatum and TPJ (which correlated with $Self and $Other, respectively), as well as a vmPFC region that represented $Self and $Other simultaneously and encoded the overall value of a choice.

Neural and Psychological Bases of Pro-social Decision Making
Although several studies have shown that the TPJ plays a role in empathic and altruistic decision making (Decety and Jackson, 2006; Hare et al., 2010; Morishima et al., 2012; Saxe and Powell, 2006), its precise computational nature remains poorly under-
stood. Our findings show that, at least during altruistic choice, signals in this area may reflect computations related to others’ interests. This signal differs from two popular alternative ac-
counts of TPJ function: that it represents the beliefs of others (Saxe and Powell, 2006) or that it allows attention to be shifted away from the self (Scholz et al., 2009). Neither theory appears to fully explain the pattern of TPJ response in our task. Although computations about belief might be used to determine how another person would value the proposal, it is not clear why this representation would appear selectively for higher rather than lower amounts of $Other. Similarly, if attentional shifts help to incorporate others’ reward into the value signal, they

Model Implications
Several implications of the model showcase the value of comput-
tional approaches and provide novel insights into the nature of pro-social behavior. First, consistent with the data, model simu-
lations predict that generous decisions are made more slowly but that this slowdown is less pronounced in more generous subjects. This observation has direct relevance for a literature that has made the case for dual-process models of social decision making based on RT differences between generous and selfish choices (Piovesan and Wengstrom, 2009; Rand et al., 2012; Tinghög et al., 2013). Our results suggest caution in inter-
preting these RT differences, by showing how they can arise without requiring competition between “fast and automatic” and “slow and deliberative” systems. In our model, generous choices are made more slowly if the relative weight placed on the self is higher, but more quickly if weights on others’ payoffs are higher.

Figure 6. Model Implications for Relation between Different Parameters of the Model and Behavior
(A and B) Variation in generosity as weights for self and other vary (A) and as threshold starting height and collapse rate vary (B).
(C and D) Variation in likelihood that a generous choice is a mistake as a function of weights for self and other (C) and threshold parameters (D). Dots represent the estimated parameter values for the 51 subjects who completed the fMRI study, jittered randomly by a small amount to allow visualization of subjects with overlapping values.
Figure 7. Model Implications for the Likelihood that Selfish or Generous Choices Are Errors
(A) A vmPFC region implicated in coding outcome value responded more positively to reversal versus receipt of generous (G) choices compared to selfish (S) choices (p < 0.05, SVC). Differential BOLD response in this region (mean ± SEM) is shown for illustrative purposes only.
(B) vmPFC response to reversal versus receipt of G versus S choices correlated with the DDM-predicted likelihood that a subject’s G choices were more likely to be errors than S choices (i.e., indexing the relief they should feel if those choices are overturned).

This could help to reconcile some of the apparently contradictory results in this literature. Different contexts can evoke dramatically different levels of altruistic or pro-social behavior (Engel, 2011). Studies observing faster RTs for more generous or cooperative choices (Rand et al., 2012) may establish contexts in which, for a variety of reasons, the needs of others are weighted more highly, while studies observing slower RTs (Piovesan and Wengstrom, 2009) may prime subjects toward reduced consideration of others. Note, however, that our results do not undermine the general validity of dual-process frameworks. Indeed, in some respects, our model can be interpreted as involving dual processes with respect to valuing self- and other-interests but suggests that RT data should be used carefully and in conjunction with more formal computational models to derive and test predictions.

Second, the model has similar implications for the interpretation of neural response. It predicts that for subjects with a bias toward the self (i.e., almost everyone), brain areas whose activity scales with computations in the accumulation and comparison process will have greater response on trials resulting in generous choice. We find this pattern in both the TPJ and vmPFC and show that it can be accounted for by the neurocomputational model. These results urge caution in interpreting generosity-specific activation in TPJ as inhibition of selfish impulses (Strombach et al., 2015) or in concluding from activation differences in vmPFC that choosing generously is rewarding (Strombach et al., 2015; Zaki and Mitchell, 2011). A simple neurocomputational model with identical parameters on every trial reproduces these differences without requiring that choosing generously specifically involve either self-control or a special reward value.

A third implication of the model concerns the relationship between individual differences in generosity and specific model parameters. Not surprisingly, generosity increases with the weight to other and decreases with the weight to self. More surprisingly, generosity also increases with less stringent barriers (i.e., lower starting threshold and a faster collapse rate). Thus, systematic differences in altruistic behavior may not reflect different underlying preferences (i.e., weights on self and other), but simply alterations in the amount of noise in the decision process. This observation has important implications for the large body of social decision making literature that has used manipulations that might influence barrier height and response caution, such as time pressure (Rand et al., 2012), cognitive load (Cornelissen et al., 2011), or even electrical brain stimulation (Ruff et al., 2013). The results of these studies are often assumed to support a role of self-control in increasing (or decreasing) consideration of others’ welfare. Our results point to an alternative interpretation and suggest that greater attention should be paid to the precise mechanism of action through which different manipulations influence generosity.

The observation that noise can induce systematic shifts in choice without systematic shifts in preferences leads to the final implication of our model: that a significant fraction of generous choices may be decision mistakes. Results from the outcome period in our study suggest that people track these errors and may feel relieved when the consequences of such errors are avoided due to external contingencies. This insight has profound implications for our understanding of both basic decision making and pro-sociality. It adds to other work on impure altruism (Andreoni, 1990; Andreoni and Bernheim, 2009), suggesting that any single generous act can result from many processes that have little to do with the true value we assign to others’ welfare.

EXPERIMENTAL PROCEDURES
Participants
Male volunteers (n = 122) were recruited in pairs from the Caltech community. Half were active participants who completed the scanning task. The other subjects participated passively as described below. All were right-handed, healthy, had normal/corrected-to-normal vision, were free of psychiatric/neurological conditions, and did not report taking any medications that might interfere with fMRI. All participants received a show-up fee of $30 as well as $0–$100 in additional earnings, depending on the outcome of a randomly chosen experimental trial. We excluded data from ten scanning subjects due to excessive head motion or technical difficulties during scanning (remaining 51 subjects: 18–35 years of age, mean 22.3). Caltech’s Internal Review Board approved all procedures. Subjects provided informed consent prior to participation.

Task
Each participant in a pair arrived separately to the lab and waited in a private area where he received instructions. We randomly designated one participant as the active participant (AP), who completed the tasks described below. We designated the other as the passive partner (PP) who, after receiving instructions, waited in a separate room for the study duration. The PP’s presence created a real and non-deceptive social context for the AP.

The AP made 180 real decisions in a modified Dictator Game. On each trial, he chose between a proposed pair of monetary prizes to himself and his partner and a constant default prize-pair of $50 to both (Figure 1A). Proposed prizes varied from $10 to $100 and were drawn from one of the nine pairs shown in Figure 1B. Each pair appeared 20 times, randomly intermixed across trials, and divided evenly across four scanner runs (five instances/run). To minimize habituation and repetition effects, we randomly jittered proposal amounts by $1–$4, with the exception that amounts above $100 were always jittered.
downward. The side of the screen on which $Self$ appeared was counterbalanced across subjects but was constant throughout the task.

All prize-pairs included one payment below and one payment above the default and thus involved a choice between generous behavior (benefitting the other at a cost to oneself) and selfish behavior (benefitting oneself at a cost to the other). After presentation of the proposal, subjects had up to 4 s to indicate their choice using a four-point scale (Strong No, No, Yes, Strong Yes). This allowed us to simultaneously measure both their decision and the relative value of the proposed payment at the time of choice. The direction of increasing preference (right-to-left or left-to-right) varied on each scan. If the subject failed to respond within 4 s, both individuals received $0 for that trial. Although this time limit could be considered a form of time pressure, pilot testing with free response times suggested that a relative minority of choices (14%) took longer than 4 s and that other basic properties of choice and RT were similar to the current study.

The task also included a second component designed both to increase the anonymity of choices and allow us to test a prediction made by the DDM about the possibility of decision mistakes. After a random delay of 2–4 s following response, the subject’s choice was implemented probabilistically: in 60% of trials he received his chosen option, while in 40% his choice was reversed and he received the alternative non-chosen option. This reversal meant that while it was always in his best interest to choose according to his true preferences, his partner could never be sure about the actual choice made. APs were informed that the PP’s were aware of the probabilistic implementation. The 40% reversal rate was necessary to test key predictions of the model but raises the concern that it alters decision computations. Pilot testing with only 10% choice reversals yielded nearly identical behavioral results, suggesting this is not likely an issue.

Behavioral Definition of Generosity

We label specific decisions as Generous (G) if the AP gave up money to help the PP (i.e., accepting $Self < $50 or rejecting $Self > $50) and as Selfish (S) otherwise. Subject-level generosity was measured by the average amount of money per trial that a subject gave to the PP by choosing generously. Alternative measures of generosity (such as money sacrificed) led to similar results.

Model Estimation

We use maximum likelihood to estimate the value of the free parameters that provide the best fit to the observed choice and RT data, separately for each AP. For assessing the goodness-of-fit of the model, we estimate these parameters separately for half of the trials and test the accuracy of predictions in the other half of the data. For testing model implications, we use the full set of trials for each AP to fit the parameters. Fitting was done in several steps.

First, we ran 1,000 simulations of the DDM to compute the likelihood function over observed choices (Yes/No) and RT bins separately for each parameter pair used in the experiment and each possible combination of parameters. RT bins were specified in 250-ms increments from 0 to 4 s and included one additional bin for non-responses (simulations in which the RDV failed to cross a barrier within the 4 s time limit). The combination of parameters used covers a grid determined by the cross-product of the following sets: $w_{Self} = w_{Other} = \{-0.045, -0.003, -0.0015, 0, 0.0015, 0.003, 0.0045, 0.006, 0.0075, 0.009, 0.0105, 0.012, 0.0135\}$, $NDT = \{0, 0.5, 0.7, 0.9, 1.1, 1.3\}$, $B = \{0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3, 0.32\}$, and $b = \{0, 0.00005, 0.0001, 0.00025, 0.0005, 0.00075, 0.001, 0.005\}$. The range of the grid was chosen by trial and error so that no more than 10% of subjects fell on a boundary edge for any parameter, while keeping the total number of parameter combinations low to minimize exploding computational costs.

Second, for each subject we identified the parameter combination that minimized the negative log-likelihood (NLL) of the selected trials observed for that subject, based on likelihoods generated from the simulated data. If more than one parameter combination resulted in the same minimal NLL, one was randomly selected as the solution.

Model Simulations

To assess model fits to behavior, we used half of each subject’s responses (randomly selected) to find the best-fitting parameters and simulated 1,000 runs of the other half of trials seen by that individual. We then compared observed and simulated values for the average amount given to the partner, average RT on G and S choice trials, and average choices and RTs for particular proposals.

We also simulated data from the best-fitting parameters for all trials in each subject, to explore other model implications. First, we predicted overall response in the accumulator (Implication 2 and Figure 6). We speculated based on prior research (Basten et al., 2010) that several brain regions may contribute to this computation and explored the implications of this architecture for understanding behavioral and neural correlates of generosity. We follow Basten et al. (2010) in defining accumulator response for each trial as $S_i \times R_{DVT}$ and estimate it separately for G and S choices (see GLM 3 below).

Second, we explored how individual variation in model parameters affects generosity (Implication 3). Finally, we used simulations to understand the role of choice errors in producing altruistic behavior (Implication 4 and Figure 7).

fMRI

fMRI data were acquired and preprocessed using standard procedures (see Supplemental Experimental Procedures for details). Using these data, we estimated three different general linear models (GLMs) of BOLD response.

GLM 1

The first GLM served two purposes: (1) to identify regions associated with the overall decision value of the proposal behaviorally expressed at the time of choice, and (2) to test the hypothesis that many generous choices should be considered errors and thus be perceived as good news if they are reversed.

For each subject, we estimated a GLM with AR(1) and the following regressors of interest: (R1) a boxcar function for the choice period on all trials; (R2) R1 modulated by the behaviorally expressed preference, ranging from 1 = Strong No to 4 = Strong Yes; (R3) a boxcar function of 3 s duration for the outcome period; (R4) R3 modulated by the outcome for self on each trial; (R5) R3 modulated by the outcome for other on each trial; (R6) R3 modulated by a variable consisting of a 1 for every trial in which the subject chose generously but the choice was vetoed, a −1 for every trial in which the subject chose generously and the choice was implemented, and 0 otherwise (i.e., after a selfish choice); and (R7) R3 modulated by a variable similar to R6, but which was 1 for veto of selfish choices, −1 for implementation of selfish choices, and 0 otherwise. No orthogonalization was used, allowing regressors to compete fully for explained variance. All regressors of interest were convolved with the canonical form of the hemodynamic response. The model also included motion parameters and session constants as regressors of no interest. Missed response trials were excluded from analysis.

We then computed second-level random effects contrasts with one-sample t tests, using the single-subject parameter estimates to construct several contrasts. We used R2 to determine areas correlated with behaviorally expressed preference at the time of choice. We used R6 and R7, and their difference, to explore activation related to choice reversal. Because outcomes for self and other are entered as modulators, R6 and R7 reflect differences in response over and above those associated purely with the amounts received when the outcome is revealed.

For inference purposes, we imposed a family-wise error cluster-corrected threshold of $p < 0.05$ (based on Gaussian random field theory as implemented in SPM5). We also report results surviving small-volume correction within regions for which we had strong a priori hypotheses (see ROI definition below), including vmPFC and TPJ.

GLM 2

This GLM identified regions in which activity correlates with proposed payments at the time of choice. It included the following regressors: (R1) a boxcar function for the choice period on all trials; (R2) R1 modulated by $Self$ on each trial; (R3) R1 modulated by $Other$ on each trial; (R4) a 3 s boxcar function for the outcome period; (R5) R4 modulated by the outcome for self on each trial; and (R6) R4 modulated by the outcome for other on each trial. All other details are as in GLM 1. Using GLM 2, we calculated three single-subject parametric contrasts: R2 versus zero, R3 versus zero, and R2 versus R3.

GLM 3

We used GLM 3 to test predictions about comparator differences on G versus S choice trials. It included the following regressors: (R1) a boxcar function for the choice period on trials when the subject chose selfishly; (R2) R1 modulated...
by the behaviorally expressed preference at the time of choice; (R3) a boxcar function for the choice period on trials when the subject chose generously; and (R4) R3 modulated by behavioral preference. R5–R9 were identical to R3–R7 from GLM 1. The contrast R3 versus R1 identified regions with differential response for G versus S choices.

Within two independently defined ROIs in vmPFC and TPJ (see Figure 5), we calculated the average value of the contrast R3 – R1 for each subject and regressed it on the model-predicted difference in accumulator activity ($BOLD_{G,S} = \beta_0 + \beta_1 DPM_{G,S} + \theta$). This allowed us to determine whether predicted accumulator differences were associated with BOLD differences on G versus S choice trials, and whether differential generosity-related BOLD response (i.e., $\beta_0$) remained significant after controlling for predicted accumulator differences.

ROI Definition
For use in small-volume corrections, as implemented in SPM5, we defined three a priori regions of interest: a vmPFC region associated with decision value, a vmPFC region associated with outcome value, and bilateral TPJ. We defined decision value-related vmPFC using the conjunction of two recent meta-analyses on decision-related reward valuation (Bartra et al., 2013; Clithero and Rangel, 2014). Outcome value-related vmPFC was defined in a similar way but based on meta-analysis results for value representations at outcome, which may preferentially activate more anterior vmPFC regions (Clithero and Rangel, 2014). The TPJ mask was defined anatomically using the WFU PickAtlas (http://fmri.wfubmc.edu/software/PickAtlas), with a dilation of 3 mm to ensure full coverage of the area. It included bilateral angular and superior temporal gyrus, posterior to $y = -40$ (1,975 voxels), a region encompassing activation peaks from several studies of Theory-of-Mind (Decety and Jackson, 2006; Saxe and Powell, 2006). All masks can be obtained from http://www.ml.caltech.edu/resources/index.html.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, two figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2015.06.031.

AUTHOR CONTRIBUTIONS
C.A.H., B.B., and A.R. designed the experiment. C.A.H. and B.B. collected the data. C.A.H. developed the model and its predictions. C.A.H. analyzed the data. C.A.H., B.B., and A.R. wrote the paper.

ACKNOWLEDGMENTS
Matthew Rabin was an earlier collaborator on this project, which benefited greatly from his insight. This research was supported by NSF-IGERT (B.B.), NSF-Economics (A.R.), NSF-DRMS (A.R.), the Gordon and Betty Moore Foundation (A.R., C.A.H.), and the Lipper Foundation (A.R.). We thank John Clithero and Anita Tuszche for helpful comments.

Received: September 2, 2014
Revised: March 17, 2015
Accepted: June 23, 2015
Published: July 15, 2015

REFERENCES


A Neurocomputational Model of Altruistic Choice and Its Implications

Cendri A. Hutcherson, Benjamin Bushong, and Antonio Rangel
Supplemental Information:
A neurocomputational model of altruistic choice and its implications
Cendri A. Hutcherson, Benjamin Bushong, & Antonio Rangel

Supplementary Figure S1, Related to Figure 2. Model fits to behavior, allowing parameters to vary by whether $Self > Other. Within-subject observed behavior (grey bars) and predicted behavior (red circles) of acceptance likelihood (right) and average response time (left) for each of the 9 proposal-types. Bars indicate mean ± SEM.
Supplementary Figure S2, Related to Figure 5. Poor fit to model predictions for BOLD responses during generous vs. selfish choices in value-modulated regions that are unlikely to be involved in the integration and comparison process. Top: occipital (A) and motor (B) regions correlate with value at the time of choice (P < .0001, uncorrected, masks shown in red). Middle: In contrast to vmPFC and TPJ, both regions show lower response on trials when a subject chose generously. Bars indicate mean BOLD response ± SEM. Bottom: In neither region do individual differences in the response to generous vs. selfish choices correlate with model-predicted comparator differences. *P=.02; **P=.0002.
Supplementary Table S1. Parameter values estimated separately in the realm of advantageous and disadvantageous inequality, Related to Table 1. $w_{Self}$ and $w_{Other}$ represent weights applied to the relative value of $Self$ and $Other$ on each trial compared to the default. NDT: non-decision time. $b$ and $d$: starting value and collapse rate of the decision threshold.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$SS &gt; SO$</th>
<th>$SS &lt; SO$</th>
<th>T-statistic</th>
<th>P-value</th>
<th>R</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{Self}$</td>
<td>.006±.002</td>
<td>.009±.004</td>
<td>5.34</td>
<td>&lt;.001</td>
<td>.16</td>
<td>.25</td>
</tr>
<tr>
<td>$w_{Other}$</td>
<td>.002±.003</td>
<td>.001±.004</td>
<td>2.56</td>
<td>.01</td>
<td>.64</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>NDT</td>
<td>896±210ms</td>
<td>888±231ms</td>
<td>.25</td>
<td>n.s.</td>
<td>.48</td>
<td>.0004</td>
</tr>
<tr>
<td>$b$</td>
<td>.23±.06</td>
<td>.27±.06</td>
<td>3.97</td>
<td>.0002</td>
<td>.42</td>
<td>.002</td>
</tr>
<tr>
<td>$d$</td>
<td>.0005±.0002</td>
<td>.0005±.0002</td>
<td>.28</td>
<td>n.s.</td>
<td>.42</td>
<td>.002</td>
</tr>
</tbody>
</table>
Supplementary Table S2. Association between parameter values, generosity and RT, related to Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Assoc. w/ generosity: $S &gt; $O</th>
<th>Assoc. w/ generosity $S &lt; $O</th>
<th>Assoc. w/ G vs. S RT: $S &gt; $O</th>
<th>Assoc. w/ G vs. S RT: $S &lt; $O</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_{Self}$</td>
<td>-2.82**</td>
<td>-1.82**</td>
<td>+.112*</td>
<td>+.137</td>
</tr>
<tr>
<td>$w_{Other}$</td>
<td>+5.69**</td>
<td>+9.56**</td>
<td>+.004</td>
<td>-.312**</td>
</tr>
<tr>
<td>NDT</td>
<td>-.28</td>
<td>+.21</td>
<td>+.005</td>
<td>+.09</td>
</tr>
<tr>
<td>$b$</td>
<td>-.94*</td>
<td>-.43</td>
<td>+.075</td>
<td>+.07</td>
</tr>
<tr>
<td>$d$</td>
<td>+.46</td>
<td>+.11</td>
<td>-.007</td>
<td>+.01</td>
</tr>
</tbody>
</table>

** $P < .001$ * $P < .05$
Supplementary Table S3. Regions correlating with stated preferences at the time of choice (GLM 1), related to Figure 3A.

<table>
<thead>
<tr>
<th>Region</th>
<th>BA</th>
<th>Cluster Size</th>
<th>Z score</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior frontal gyrus</td>
<td>10</td>
<td>75</td>
<td>4.9</td>
<td>-12</td>
<td>60</td>
<td>27</td>
</tr>
<tr>
<td>Anterior cingulate cortex</td>
<td>24/32</td>
<td>680</td>
<td>5.37</td>
<td>3</td>
<td>39</td>
<td>18</td>
</tr>
<tr>
<td>L Ventromedial prefrontal cortex</td>
<td>11/32</td>
<td>a</td>
<td>5.01</td>
<td>-6</td>
<td>33</td>
<td>-12</td>
</tr>
<tr>
<td>Ventral striatum</td>
<td>a</td>
<td>5.29</td>
<td>9</td>
<td>12</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td>Ventral striatum</td>
<td>a</td>
<td>5.13</td>
<td>-9</td>
<td>12</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td>Middle frontal gyrus</td>
<td>6/8</td>
<td>49</td>
<td>4.36</td>
<td>-21</td>
<td>24</td>
<td>54</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>6</td>
<td>32</td>
<td>4.33</td>
<td>63</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Mid-cingulate cortex</td>
<td>24</td>
<td>53</td>
<td>4.27</td>
<td>-3</td>
<td>-6</td>
<td>39</td>
</tr>
<tr>
<td>Supplementary Motor Area</td>
<td>6</td>
<td>16</td>
<td>4.28</td>
<td>6</td>
<td>-12</td>
<td>72</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>4</td>
<td>90</td>
<td>4.86</td>
<td>-39</td>
<td>-15</td>
<td>57</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>4</td>
<td>216</td>
<td>5.06</td>
<td>-21</td>
<td>-27</td>
<td>72</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>21/22</td>
<td>38</td>
<td>4.63</td>
<td>60</td>
<td>-30</td>
<td>6</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>22/41</td>
<td>179</td>
<td>4.91</td>
<td>-63</td>
<td>-36</td>
<td>9</td>
</tr>
<tr>
<td>Posterior cingulate cortex</td>
<td>31</td>
<td>186</td>
<td>5.64</td>
<td>-6</td>
<td>-42</td>
<td>42</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>37</td>
<td>35</td>
<td>5.42</td>
<td>54</td>
<td>-42</td>
<td>-21</td>
</tr>
<tr>
<td>Inferior parietal cortex</td>
<td>7</td>
<td>72</td>
<td>4.48</td>
<td>-36</td>
<td>-75</td>
<td>42</td>
</tr>
<tr>
<td>Occipital cortex</td>
<td>18/19</td>
<td>3430</td>
<td>6.23</td>
<td>-6</td>
<td>-102</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>18/19</td>
<td>a</td>
<td>5.3</td>
<td>18</td>
<td>-96</td>
<td>15</td>
</tr>
</tbody>
</table>

Note:
Regions are reported if they passed two thresholds: \( P < .0001 \) uncorrected and \( P < .05 \) cluster corrected. A higher threshold was used for reporting because a lower threshold resulted in a single undifferentiated cluster.

a. Distinct peak in larger cluster of activation, reported separately for completeness.
Supplementary Table S4. Neural correlates of $Self and $Other (GLM 2), related to Figure 3B, C.

<table>
<thead>
<tr>
<th>Region</th>
<th>BA</th>
<th>Cluster Size</th>
<th>Z score</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regions associated with $Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R   Anterior cingulate cortex</td>
<td>24</td>
<td>45</td>
<td>3.82*</td>
<td>9</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td>L   Inferior parietal lobule/temporoparietal junction</td>
<td>7/39</td>
<td>217</td>
<td>4.71</td>
<td>-24</td>
<td>-48</td>
<td>24</td>
</tr>
<tr>
<td>L   Precuneus</td>
<td>7/31</td>
<td>494</td>
<td>4.89</td>
<td>-9</td>
<td>-60</td>
<td>45</td>
</tr>
<tr>
<td>R   Temporoparietal junction</td>
<td>39</td>
<td>63</td>
<td>3.92</td>
<td>39</td>
<td>-63</td>
<td>21</td>
</tr>
<tr>
<td>L   Occipital cortex</td>
<td>30</td>
<td>61</td>
<td>4.51</td>
<td>-30</td>
<td>-63</td>
<td>12</td>
</tr>
<tr>
<td>R   Occipital cortex</td>
<td>18</td>
<td>76</td>
<td>4.22</td>
<td>21</td>
<td>-90</td>
<td>-9</td>
</tr>
<tr>
<td>L   Cerebellum</td>
<td>152</td>
<td></td>
<td>4.24</td>
<td>-24</td>
<td>-93</td>
<td>-27</td>
</tr>
<tr>
<td>Occipital cortex</td>
<td>18</td>
<td>a</td>
<td>4.17</td>
<td>-18</td>
<td>-96</td>
<td>-12</td>
</tr>
<tr>
<td><strong>Regions associated with $Self</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R   Anterior cingulate cortex</td>
<td>24</td>
<td>94</td>
<td>5.04</td>
<td>9</td>
<td>36</td>
<td>6</td>
</tr>
<tr>
<td>R   Ventromedial prefrontal cortex</td>
<td>32</td>
<td>20</td>
<td>4.13</td>
<td>6</td>
<td>36</td>
<td>-9</td>
</tr>
<tr>
<td>R   Inferior frontal gyrus</td>
<td>44</td>
<td>61</td>
<td>4.73</td>
<td>45</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>L   Insula</td>
<td>13</td>
<td>20</td>
<td>4.46</td>
<td>-33</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>R   Supplementary motor area</td>
<td>6</td>
<td>212</td>
<td>5.1</td>
<td>9</td>
<td>-9</td>
<td>63</td>
</tr>
<tr>
<td>L   Postcentral gyrus</td>
<td>3</td>
<td>230</td>
<td>5.33</td>
<td>-48</td>
<td>-18</td>
<td>51</td>
</tr>
<tr>
<td>R   Supramarginal gyrus</td>
<td>40</td>
<td>20</td>
<td>4.13</td>
<td>57</td>
<td>-27</td>
<td>27</td>
</tr>
<tr>
<td>L   Postcentral gyrus</td>
<td>2</td>
<td>48</td>
<td>4.5</td>
<td>-15</td>
<td>-30</td>
<td>78</td>
</tr>
<tr>
<td>L   Mid-cingulate cortex</td>
<td>31</td>
<td>84</td>
<td>4.7</td>
<td>-3</td>
<td>-36</td>
<td>42</td>
</tr>
<tr>
<td>L   Superior temporal cortex</td>
<td>41</td>
<td>54</td>
<td>4.76</td>
<td>-51</td>
<td>-36</td>
<td>15</td>
</tr>
<tr>
<td>L   Postcentral gyrus</td>
<td>2</td>
<td>47</td>
<td>4.4</td>
<td>-33</td>
<td>-42</td>
<td>66</td>
</tr>
<tr>
<td>L   Posterior cingulate</td>
<td>23</td>
<td>13</td>
<td>4.23</td>
<td>-3</td>
<td>-42</td>
<td>21</td>
</tr>
<tr>
<td>R   Occipital cortex</td>
<td>18/19</td>
<td>3544</td>
<td>6.73</td>
<td>21</td>
<td>-90</td>
<td>-9</td>
</tr>
<tr>
<td>R   Ventral striatum</td>
<td></td>
<td>43</td>
<td>4.79</td>
<td>9</td>
<td>12</td>
<td>-9</td>
</tr>
<tr>
<td><strong>Regions where association with $Other &gt; $Self</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R   Temporoparietal junction</td>
<td>39/40</td>
<td></td>
<td>3.93†</td>
<td>51</td>
<td>-51</td>
<td>27</td>
</tr>
<tr>
<td><strong>Regions where association with $Self &gt; $Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R   Anterior cingulate cortex</td>
<td>24</td>
<td>304</td>
<td>4.4</td>
<td>0</td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td>R   Inferior frontal gyrus</td>
<td>44</td>
<td>110</td>
<td>4.48</td>
<td>48</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>L   Middle frontal gyrus</td>
<td>6</td>
<td>55</td>
<td>4.32</td>
<td>-48</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>R   Supplementary motor area</td>
<td>6</td>
<td>638</td>
<td>5.34</td>
<td>3</td>
<td>-9</td>
<td>54</td>
</tr>
<tr>
<td>R   Supramarginal gyrus</td>
<td>40</td>
<td>295</td>
<td>4.53</td>
<td>57</td>
<td>-21</td>
<td>30</td>
</tr>
<tr>
<td>L   Postcentral gyrus</td>
<td>3</td>
<td>7849</td>
<td>6.39</td>
<td>-39</td>
<td>-24</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Occipital cortex</td>
<td>18/19</td>
<td>*</td>
<td>5.96</td>
<td>30</td>
<td>-75</td>
</tr>
<tr>
<td>L</td>
<td>Ventral striatum</td>
<td>32</td>
<td>4.01</td>
<td>-3</td>
<td>-6</td>
<td>6</td>
</tr>
<tr>
<td>R</td>
<td>Ventral striatum</td>
<td>208</td>
<td>4.72</td>
<td>9</td>
<td>9</td>
<td>-12</td>
</tr>
</tbody>
</table>

**Conjunction of regions associated with **Self** and **Other**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>Occipital cortex</td>
<td>18</td>
<td>244</td>
<td>-</td>
<td>-25</td>
</tr>
<tr>
<td>R</td>
<td>Occipital cortex</td>
<td>18</td>
<td>92</td>
<td>-</td>
<td>24</td>
</tr>
<tr>
<td>L</td>
<td>Precuneus</td>
<td>7</td>
<td>62</td>
<td>-</td>
<td>-19</td>
</tr>
<tr>
<td>L</td>
<td>Precuneus</td>
<td>7</td>
<td>52</td>
<td>-</td>
<td>-3</td>
</tr>
<tr>
<td>R</td>
<td>Mid cingulate cortex</td>
<td>24</td>
<td>50</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>L</td>
<td>Superior frontal gyrus</td>
<td>6</td>
<td>33</td>
<td>-</td>
<td>-22</td>
</tr>
<tr>
<td>L</td>
<td>Frontopolar cortex</td>
<td>10</td>
<td>28</td>
<td>-</td>
<td>-15</td>
</tr>
<tr>
<td>R</td>
<td>Cerebellum</td>
<td>27</td>
<td>29</td>
<td>-</td>
<td>-72</td>
</tr>
<tr>
<td>R</td>
<td>Ventromedial prefrontal cortex</td>
<td>24/32</td>
<td>22</td>
<td>-</td>
<td>6</td>
</tr>
</tbody>
</table>

Note:
Regions are reported if they passed two thresholds: $P < .001$ uncorrected and $P < .05$ cluster corrected. * $P < .05$, small-volume corrected within a-priori ROI. † $P < .005$, uncorrected. a. Distinct peak in larger cluster of activation, reported separately for completeness; b. Maps thresholded separately at $P < .05$, corrected, with minimum overlap of 20 voxels.
Supplemental Experimental Procedures.

Alternative specification of the DDM. In the main text, we describe the results of computational model-fitting to the observed data. We observed that a simple model in which choices were determined by five parameters (weights on $Self and $Other, height and collapse rate of a decision threshold, and a non-decision time) was capable of closely reproducing within-subject patterns of variation in choice and reaction times (see section titled “The model accurately predicts out-of-sample choice and RT”). However, we also observed that the fit to these simulations was better for trials on which the proposed amount $Self was higher than $Other, and worse for the trials in which it was lower. We speculated that this difference does not arise from a problem with the DDM per se, but rather may reflect a change in one or more of the parameters of the model when in the domains of advantageous inequality ($Self > $Other) and disadvantageous inequality ($Self < $Other). This distinction has been shown to have a considerable impact on both behaviour and neural response in previous work (Charness and Rabin, 2002; Fehr and Schmidt, 1999; Tricomi et al., 2010) so we sought to address this issue here.

We fit the five parameters of the DDM individually to each subject, using the same method as described in the main text and Online Methods, with the difference that these parameters were fit separately for those trials on which $Self > $Other and trials on which $Self < $Other. We then compared these parameters across the two types of trials to determine both whether there was a systematic difference in one or more parameters between the two trial types, whether there was a correlation across the two models in the parameters fit from the two trial types, and, finally, whether this difference might change any of the conclusions drawn from the simpler model fits.
We quantified the improvement in model fit by using a log-likelihood ratio test (LLRT), where the log-likelihood of each response and RT was estimated using the simulated probability distributions under the best-fitting parameters for each of the two models, and summed over all responses to create the total log-likelihood values $L_{Simple}$ and $L_{Alternative}$. The alternative model has ten degrees of freedom, since it estimates each of the five parameters separately for the two halves of the data, while the simple model, with five degrees of freedom, can be considered a nested version of the alternative in which the parameters for the two trial types are constrained to be equal. Because the simpler model is nested within the more complex one, the distribution of the test-statistic $D = -2 * (L_{Simple} - L_{Alternative})$ is distributed approximately as a $\chi^2$ with degrees of freedom equal to the difference in the degrees of freedom of the two models. We observed a log-likelihood value of -21759 for the simple model and -20873 for the alternative model, yielding a $\chi^2(5) = 885.6, P < .001$. Supplemental Figure 1 shows that the model allowing parameters to vary as a function of the relationship between $Self$ and $Other$ indeed produced a better fit to the data, although this improvement was more pronounced for choice data than reaction times.

We next examined which parameters of the model changed significantly between the two trial types. These analyses indicated significant differences in three parameters: the weights given to $Self$ and $Other$, as well as the threshold. Trials with disadvantageous inequality showed a higher weight on $Self$, a lower weight on $Other$, and an increase in the threshold for making a choice (see Supplementary Table S1 for mean and standard deviation for the estimated parameters in the two models, and the results of paired t-tests). Despite this difference, most parameters were correlated across the two halves of the data (see Table S1), with the exception of $w_{Self}$ suggesting that they likely derived from common processes that persisted across the different trial types.

Multiple regression analyses suggested that individual differences in these parameters correlated in a similar way to observed generosity and RT in the full dataset, although
The significance of these parameter values were generally somewhat lower, likely due to the increased noise based on the more limited number of trials. The results of these correlations can be seen in Supplementary Table S2.

**fMRI data acquisition.** BOLD responses were acquired using a Siemens 3.0 Tesla Trio MRI scanner (Erlangen, Germany) to acquire gradient echo T2*-weighted echo-planar (EPI) images. To optimize functional sensitivity in the orbitofrontal cortex (OFC), a key region of interest, we used a tilted acquisition in an oblique orientation of 30° to the anterior commissure–posterior commissure line. In addition, we used a standard eight-channel phased array coil. Each volume comprised 45 axial slices. A total of 960 volumes were collected over four sessions during the experiment in an interleaved ascending manner. The first two volumes of each session were discarded to allow for scanner equilibration. The imaging parameters were as follows: echo time, 30 ms; field of view, 192 mm; in-plane resolution and slice thickness, 3 mm; repetition time, 2.75 s. Whole-brain high resolution T1-weighted structural scans (1 x 1 x 1 mm) were acquired for the 51 subjects and co-registered with their mean EPI images and averaged together to permit anatomical localization of the functional activations at the group level.

**fMRI data pre-processing.** Image analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Images were corrected for slice acquisition time within each volume, motion corrected with realignment to the last volume, spatially normalized to the standard Montreal Neurological Institute EPI template using affine transformation, and spatially smoothed using an isotropic Gaussian kernel with a full width at half maximum of 8 mm. Intensity normalization and high-pass temporal filtering (using a filter width of 128 s) were also applied to the data.
Instructions given to subjects

Welcome! This is an experiment about decision making. We are interested in understanding how people make decisions about outcomes that affect not only themselves, but also other people. It should take about an hour and 45 minutes, including instructions and a few brief questionnaires. For participating, you will receive at least $30. Depending on your choices during the task, you will have the opportunity to earn from $0 up to $100 more. You will be paid in cash for your time and your earnings at the end of the experiment.

In the experiment room next door, another person is participating, who you should consider your partner. Depending on your choices and the outcomes of some random events, you might end up causing this person to end up with $0 up to $100 more than their pay for participating.

The other person is a real person, and the decision you make can have a large impact on their payoffs. Like you, they signed up for this experiment in response to an email. Since your decisions can have a big impact on their payoffs, think carefully about this other person throughout the experiment.

You will not be told who the person you are matched with is, and this other person will not be told that they were matched with you. They will never be given your name or any information about you.

This means that all choices you make in the experiment should be considered anonymous. Your name will never be connected with the choices you make, and neither your partner nor the experimenter will know what you have chosen.

While your decisions are anonymous, remember: they do have a large impact on the other person’s payoffs!

The choices you make involve real money, usually quite large amounts, so please think carefully about each decision.

In the next sections, we will describe precisely the instructions for the task you will be doing in the scanner. Pay attention to these instructions. It is critical that you understand the instructions, since they affect your ability to make good decisions – and potentially more money!

Here is how the experiment will work. Every trial will begin with a ‘+’ in the middle of the screen. Please keep your eyes on this center cross when it appears.

After the central cross appears, you will see an offer. On one side of the screen you will see the amount of money you could win if you decide to accept the offer. On the other side of the screen, you will see the amount of money your partner could win if you accept the offer. We’ll call this the proposed allocation.

The amounts of money will always range between $0 and $100, for both you and the other person.

If you decide to reject the offer, both you and your partner will each receive $50.
We will call this the default allocation, which will be the same for all trials. In every trial, therefore, you are choosing between the proposed allocation and the default allocation.

For instance, if we offer a proposed allocation of $75 for you, $25 for the other person, and you accepted this, this would indicate that you prefer $75 for you ($25 above the default), and $25 total for the other person ($25 below the default).

Although we are asking you to make choice between accepting and rejecting the proposed allocation, we would also like to get a sense of how strongly you feel about this choice. So you should indicate your choice on the following four-point scale:

**Strong Yes:** Indicates you strongly prefer the proposed allocation to the default.

**Yes:** Indicates you weakly prefer the proposed allocation to the default.

**No:** Indicates you weakly prefer the default to the proposed allocation.

**Strong No:** Indicates you strongly prefer the default to the proposed allocation.

You should respond using the keyboard as follows: [picture of applicable keys]

It is important to note: Either “Strong No” or “No” are counted as choosing the default allocation. Either “Strong Yes” or “Yes” are counted as choosing the proposed allocation. You are still just choosing whether to accept or reject the proposal, but you are also indicating how strongly you prefer the proposed or default options.

One other note: You will be required to make your decision within 4 seconds of the appearance of the proposal. If you do not make a response within that amount of time, *both you and the other person will receive $0 for that trial.*

It is therefore in your best interest to respond in a timely manner according to your preference.

How do your choices on each trial translate into a payment at the end?

At the end of the experiment, we will select *one* trial randomly from among all the trials you saw in the experiment. The results of this trial will count for real money.

Therefore, you should treat every trial when it appears as if it could be the one and only trial that finally determines how much you and your partner receive at the end of the experiment. Because only one trial is selected, your decisions on other trials should not in any way affect what you decide to do on the current trial.

There is one other important detail in this experiment.

Although you will be choosing whether you would prefer the proposed or default allocations, this choice alone will not determine how much money you and your partner would receive if the trial is chosen to count. Once you make your decision, your choice will be *probabilistically implemented*, meaning that the outcome you receive may not always be the allocation you chose.

For every trial, we will pay you and the other person the amount associated with your chosen allocation *60% of the time.* The other 40% of the time, we will implement the allocation you did not choose.
So for instance, if you chose to reject a proposal win which both you and your partner receive $15 ($25 less than the default option), there is a 60% chance that you will both receive the default, which is $50 each, and a 40% chance that you both will only receive $25.

You will find out after your choice on every trial whether we are implementing your preferred option, or the other one.

On trials where your choice is implemented, you will see a green check mark.

On trials where your choice is not implemented, you will see a red cross.

On all trials, you will see the amounts that you and your partner will receive if this trial is selected to count at the end of the experiment.

Note: Even though we are probabilistically implementing your choice, your best strategy is still just to choose the allocation you prefer. Most of the time, you will get what you chose, and you always make it more likely to get what you want if you choose it.

Don’t let this part of the experiment confuse you: if you prefer the default allocation, reject the proposed offer. If you prefer the proposed allocation, accept the proposed offer instead.

Note also that this probabilistic implementation is NOT based on the choices your partner makes. Rather, it is simply a random lottery that the computer uses to determine whether to implement your choices, or whether to implement the opposite of your choice.

Your partner does not get a say in whether the choice is implemented or not. If that trial is randomly drawn to count for real at the end of the experiment, your partner will simply have to accept whatever the combination of your choice and the probabilistic implementation turns out to be.

You will now have some practice trials. These trials will not count for anything, but are just to give you a sense for the timing and feel of the task. If you have any questions about the task, please ask the experiment now Otherwise, please proceed to the practice trials

[Four practice trials were given here]

PRIVACY: As stated earlier, you will be making decisions about how to allocate money between yourself and another person. Importantly, all the decision you make are secret and anonymous. The other person will never know your choices. All they will find out is what the one trial selected to count for real money is, what the proposed and default offers on that trial were, and what the outcome was. They will know that your choices was implemented with 60% probability, but will never see your actual choice.

To make sure your choices are truly anonymous, a computer program will be used to randomly determine which trial counts for payment. The experimenter will put the payment determined by the computer in two envelopes, one for you and one for the other person, but will not know what choice you made on that trial.

You will need to fill out and sign a receipt for your payment. Your signature does not need to be legible. The people handling these receipts do not know anything about the experiment, or about the decisions that can be made in this experiment.
Your partner will be paid in the same way. They will receive an envelope and sign a receipt. Inside the envelope will be their payment as well as the information about the trial: what the proposed allocation on the trial was, and what the final outcome was, but not what your choice was.

Using this setup, no one involved with the experiment will ever know what choices you make. Your partner will not be told who you are, and does not have enough information to link your choices to you.

All of the data about your choices will be identified by an anonymous code that will have no connection to your personal information. The previous sections explain the careful procedures we use to make sure of this.

This experiment will produce valid results only if you believe that your decisions are anonymous and secret. This is why we take these issues extremely seriously! We are bound ethically and legally to keep the promises we are making to you in this protocol. We are not allowed to use any deception in this experiment. We will do everything we say, and there not be any surprises or tricks.

If you have any questions about the experiment, please ask the experiment now.

Again, it is important that throughout this experiment, you consider both your feelings about the transfer and the potential impact on the other person. Your decisions can have a big impact on the payment they receive for the experiment, so consider both the pros and cons of each of the transfers.

Take your time! There are a number of transfers to consider, and each has both pros and cons. Try to think carefully about what the other person would in your shoes and how you might feel about the outcome.

Before you being, we would like to ask you a few questions to make sure that you have understood the task. We will ask you to determine the final payoff to you and your partner under different conditions.

Press any key to continue to the short quiz.

[3 quiz questions to ascertain comprehension of the instructions given here.]

Following completion of the quiz, and clarification by the experimenter of any questions that were missed, the participant began the scanning session.]
Supplemental References