Equitable decision making is associated with neural markers of intrinsic value

Jamil Zaki[1] and Jason P. Mitchell

Department of Psychology, Harvard University, Cambridge, MA 02138

Edited by Edward E. Smith, Columbia University, New York, NY, and approved October 19, 2011 (received for review August 19, 2011)

Standard economic and evolutionary models assume that humans are fundamentally selfish. On this view, any acts of prosociality—such as cooperation, giving, and other forms of altruism—result from covert attempts to avoid social injunctions against selfishness. However, even in the absence of social pressure, individuals routinely forego personal gain to share resources with others. Such anomalous giving cannot be accounted for by standard models of social behavior. Recent observations have suggested that, instead, prosocial behavior may reflect an intrinsic value placed on social ideals such as equity and charity. Here, we show that, consistent with this alternative account, making equitable interpersonal decisions engaged neural structures involved in computing subjective value, even when doing so required foregoing material resources. By contrast, making inequitable decisions produced activity in the anterior insula, a region linked to the experience of subjective disutility. Moreover, inequity-related insula response predicted individuals' unwillingness to make inequitable choices. Together, these data suggest that prosocial behavior is not simply a response to external pressure, but instead represents an intrinsic, and intrinsically social, class of reward.

functional MRI | neuroeconomics | game theory | orbitofrontal cortex

Standard models of decision making assume that humans act to maximize personal gains. Fairness and altruism have long presented a problem for this view, because humans frequently forego personal gains by sharing resources with others (1, 2). The principal attempt to resolve this paradox posits that what passes as prosocial behavior actually reflects selfish attempts to protect one's reputation or avoid retribution. Because injustice is received poorly by others (2), selfishly motivated individuals might act prosocially simply to avoid negative social consequences (3–5). On this view, societies must curtail selfishness actively through sanctions against "cheaters," threats of damage to one's reputation, and the continuous imposition of social norms (6–8). Consistent with this account, magnifying the salience of extrinsic motivators, such as the possibility of punishment or the importance of reputation, drastically increases overt displays of fairness (1, 6). Further, equitable action in the presence of extrinsic pressure is accompanied by engagement of neural structures associated with inhibiting prepotent responses (9, 10), suggesting that adhering to social principles such as equity requires overcoming more basic, self-serving impulses.

A particular class of observations from game theory contradicts this view. In the dictator game, actors divide resources between themselves and others as they see fit, without the possibility of sanctions or reputation costs (11). Surprisingly, even in the absence of such threats, the majority of participants share significant amounts of money with anonymous others (12). Although individuals can be induced to behave even more fairly when extrinsic motivators are present, consistent nonzero rates of giving in the absence of social pressure represent significant anomalies within models of rationally selfish economic choice.

Because such anomalous giving cannot be explained adequately within the standard model, we are forced to look for additional sources of human prosociality. One alternative to the standard model suggests that cooperation and altruism may originate not in rational self-interest but in affective responses to social behavior (5, 13). On this view, social principles—such as fairness, reciprocity, and cooperation—have an intrinsic value of their own that accompanies prosocial behavior (14, 15). That is, although individuals undoubtedly value personal gain, they also prize prosocial outcomes and feel a strong aversion to unfairness, inequity, and selfishness. Increasing evidence in favor of this viewpoint has emerged from recent human neuroimaging studies. A rich and growing body of neuroscientific research has demonstrated reliably that activity in mesolimbic dopaminergic targets—including the ventral striatum and orbitofrontal cortex (OFC)—scales linearly with the subjective value of a wide variety of reward types in both humans and other animals (16–19). Several studies now have demonstrated that these same regions also respond to observing prosocial action such as reciprocity and equitable resource distribution (20, 21), suggesting that prosociality may be imbued with intrinsic value (22).

However, the majority of these studies have focused on passive observation or reception of prosocial actions, leaving unclear whether prosocial action also is experienced as valuable by decision makers. Indeed, given the standard model of prosocial action as a capitulation to social pressure (5, 8, 9), the claim that prosociality instead might be associated with subjective reward remains controversial. Evidence in favor of such an alternative could provide critical insight into the mechanisms underlying many forms of prosociality, especially "anomalous" generosity in the absence of external social pressures. Here, we sought to provide such evidence by examining the neural bases of equitable and inequitable social decision making. We reasoned that if equity is experienced as rewarding by decision makers, then adherence to this principle—irrespective of accompanying personal gain—should engage value-related neural structures. If, on the other hand, equitable action is governed by inhibition of prepotent selfish impulses, it should be accompanied by engagement of neural structures involved in exerting cognitive control, and value-related brain activity should strictly track individuals' personal gains.

Results

Participants ($n = 15$) were scanned using functional MRI while they played a modified dictator game (10) in which they made iterated choices about whether to allocate varying amounts of money to themselves or to another person (hereafter designated the "receiver" see *Methods* for details). Critically, the receiver could not punish unfair choices, and participants understood that they would have no further interactions with the receiver, providing a rationally selfish participant with no incentive for producing any equitable behavior.

Each round of the game began with two monetary offers, one associated with the participant and the other with the receiver

Author contributions: J.Z. and J.P.M. designed research; J.Z. performed research; J.Z. analyzed data; and J.Z. and J.P.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: zaki@wjh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112324108/-/DCSupplemental.

(Fig. 1*A*). Participants decided which of the two offers to enact. The trial structure of the game was such that, on some rounds, participants could allocate more money to themselves than they could to the receiver (e.g., $2.00 for self vs. $1.25 for the receiver). On other rounds, the opposite was true (e.g., $1.50 for the self vs. $3.00 for the receiver). Thus, participants' decisions could vary both in their generosity (giving to another person at a cost to one's self) and their equity (impartially allocating resources to the person who stood to gain the most) (Fig. 1*B*). For example, allocating $2.00 to oneself as opposed to $1.25 to a receiver would be both self-serving and equitable, whereas allocating $3.00 to the receiver rather than $1.50 to oneself would be generous and equitable (Fig. 1*B*). During trials in which participants stood to gain more than the receiver, they allocated money to themselves 83.2% of the time (self-serving/equitable choices). During trials on which the receiver stood to gain more than the participant, participants were equally likely to allocate money to themselves (self-serving/inequitable choices) or to the receiver (generous/equitable choices): 44.6% vs. 55.2%, respectively. Reaction times were marginally longer for self-serving/inequitable trials (mean = 669 ms, SD = 214) than for self-serving/equitable [mean = 580 ms, SD = 151, $t(14) = 1.64$, $P = 0.12$, difference between the means (Cohen's d) = 0.44] and for self-serving/inequitable trials than for generous/equitable trials [mean = 556 ms, SD = 128, $t(14) = 1.84$, $P = 0.09$, $d = 0.49$]. Generous/inequitable choices, in which participants stood to gain more than the receiver but nonetheless allocated money to the receiver, were too rare to model in subsequent analyses. On average, participants chose to allocate 22.2% (SD = 11.9) of the total available resources to the receiver even in the absence of external pressures to act generously; this proportion of anomolous giving resembles proportions identified in previous studies of the dictator game (12).

On the basis of participants' choices, we used a whole-brain contrast to isolate neural activity associated with making equitable as opposed to inequitable decisions, irrespective of whether equitable decisions were self-serving or generous. Critically, all analyses controlled for the amount of money participants stood to gain on each trial, ensuring that observed patterns of neural response did not reflect the magnitude of possible gains but were related specifically to participants' choice-type. This analysis revealed engagement of the OFC (Fig. 2*A* and Table S1) related to making equitable as opposed to inequitable choices. Condition-specific parameter estimates revealed that this region responded equivalently during both self-serving and generous choices provided that these choices were equitable (Fig. 2*B*). Thus, OFC activity here cannot have reflected (*i*) the presence or absence of personal gain or (*ii*) the "warm glow" of acting generously per se (11), because generous/equitable and self-serving/equitable trials differed along both of these dimensions but engaged this region similarly. Remarkably, the lowest OFC response was observed when participants chose to allocate money inequitably to themselves, even though these choices resulted in real financial gain for the participant, suggesting that a motivation to maximize social outcomes may "crowd out" the value associated with personal gains.

Did equity-related activity occur in an area of OFC more generally responsive to value? To address this question, we localized brain activity that was responsive to a separate set of "pure" gain trials. This condition mimicked the trial structure of the dictator game but provided participants with monetary gains of varying amounts at no cost to the receiver and with no option to act prosocially. A parametric, whole-brain analysis revealed that OFC activity scaled with the amount of money participants gained in these trials (*x*/*y*/*z* coordinates: 8, 50, −6; 363 voxels) (Fig. 2*C*), consistent with this region's role in computing goal value (18, 23). We then interrogated this independently defined region for differences across choice-types. Consistent with the primary analysis, this OFC region-of-interest was more engaged by making equitable choices than by making inequitable choices [self-serving/equitable vs. self-serving/inequitable: $t(14) = 3.56$, $P < 0.01$, $d = 0.95$; generous/equitable vs. self-serving/inequitable: $t(14) = 2.66$, $P < 0.05$, $d = 0.71$] but did not differentiate between self-serving and generous forms of equitable choices [$t(14) = 0.11$, $P > 0.80$, $d = 0.03$] (Fig. 2*D*). These data further bolster the conclusion that decision makers place intrinsic value on equity.

If making equitable decisions is associated with subjective value, it follows that making inequitable choices should produce a sense of subjective disutility. To test this possibility, we contrasted brain activity accompanying inequitable as opposed to equitable (self-serving and generous) choice-types. Inequitable choices preferentially engaged the anterior insula (AI), a region commonly associated with aversive emotional states such as disgust and pain (24) (Fig. 3 *A* and *B* and Table S1). Further, participants who more strongly engaged the AI during inequitable decision making also made fewer inequitable choices

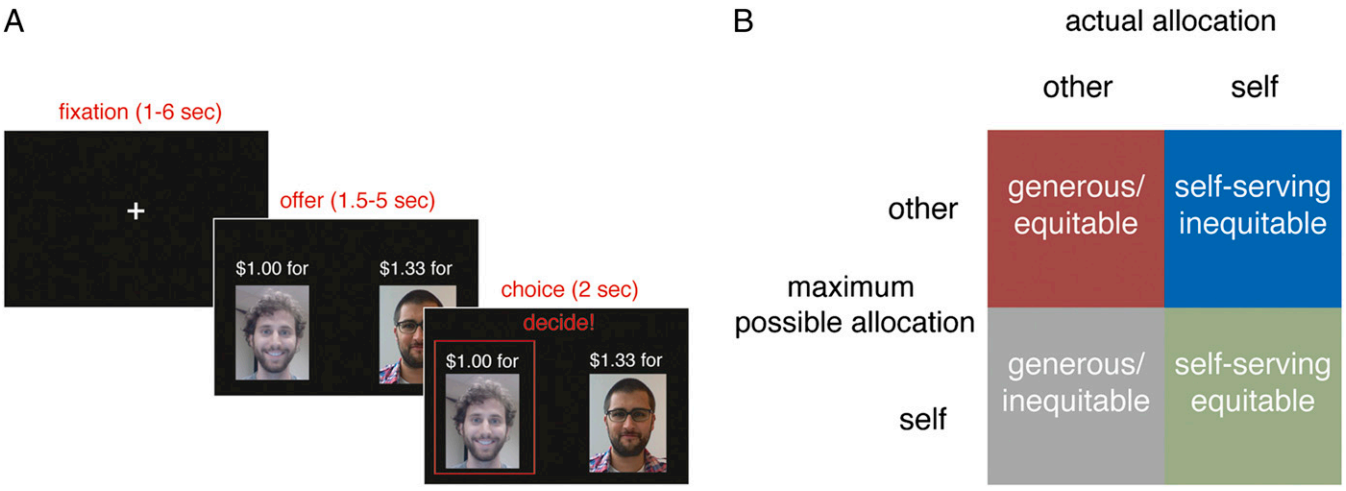

Fig. 1. (*A*) On each round of the task, participants were presented with a choice between allocating separate amounts of money either to themselves or to the receiver. (*B*) Individuals' decisions to act equitably (impartially maximizing the dyad's earnings) and generously (giving money to the receiver at a cost to themselves) varied independently as a function of (*i*) which person (self or other) stood to gain the most on a given round and (*ii*) participants' actual allocation decisions on that round.

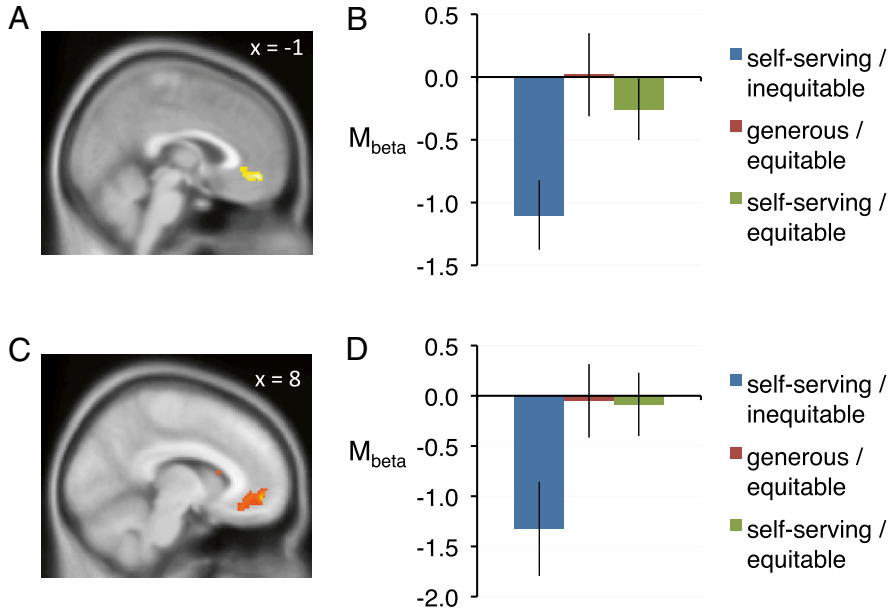

Fig. 2. (*A*) A random effects, whole-brain contrast of equitable > inequitable decisions revealed activity in the OFC. (*B*) The response of this region was reduced significantly during inequitable choices, even though such decisions led to the receipt of money by the participant. (*C*) A random effects, whole-brain parametric analysis revealed that response in a similar region of OFC correlated with the magnitude of monetary gains when nothing was at stake for the recipient. (*D*) Responses in this independently defined region were reduced significantly during inequitable choices and did not differentiate between different types of equitable choice. Error bars represent SEM.

($r = -0.64$, $P < 0.01$) (Fig. 3*C*). That is, participants who demonstrated the highest AI response to acting unfairly were the least likely to act unfairly. [The parallel correlation in the OFC—relating activity from the equitable > inequitable contrast to individuals' frequency of equitable decision making—was positive but nonsignificant ($r = 0.24$, $P = 0.22$).] To rule out the possibility that this finding merely reflects an artifact of the different number of inequitable choices across participants, we conducted an additional analysis in which we modeled inequity-related AI activity using an identical number of inequitable trials for each participant (*Methods*). Inequity-related AI activity modeled in this way negatively correlated with participants' overall tendency to act inequitably ($r = -0.60$, $P = 0.02$), consistent with the original analysis.

Discussion

The current findings complement and extend key insights from the growing field of social neuroeconomics. Research in this field has demonstrated that brain regions involved in representing subjective value—including the OFC—respond to a number of purely social outcomes, including watching others receive rewards (25, 26), act cooperatively (20), and distribute money equitably (21). The current study adds to this work by demonstrating that not only the observation of socially appropriate choices but also the decision to act prosocially engages the OFC. Although OFC activity is an indirect measure of value, equitable decisions in this study engaged a specific region also associated with receiving personal rewards, suggesting that this region indeed is involved in computing subjective value. Thus this study extends recent prior demonstrations that charitable donation engages reward-related brain regions (27, 28). Here, we document that the intrinsic value of giving does not merely reflect idiosyncratic responses to individual charities but generalizes broadly to the principle of upholding equity.

By contrast, inequitable decision making was accompanied by engagement of the AI, a region previously associated with subjective disutility (29). AI activity in this context cannot be ascribed to being self-serving (which also occurred during self-serving/equitable trials) or to being offered less money than the receiver (which also occurred during generous/equitable trials) but likely represented a response to inequity itself. Further, the AI cluster identified here is quite similar (Euclidian distance = 12.15 mm) to one previously associated with being the recipient of inequitable offers in an ultimatum game and with rejecting such offers (30). This parallel across studies suggests that inequity may be aversive not only to those affected by it but also to those responsible for producing it. Finally, participants who engaged the AI most strongly while acting inequitably also acted inequitably least often, suggesting that their own affective responses to inequity were sufficient to reduce inequitable behavior, even in the absence of external threats of punishment and despite the monetary benefits of unfair behavior. Of course, AI activity is associated with a number of other subjective states in addition to disutility per se (24, 31), and the current study does not demonstrate a causal role of brain activity in preventing unfair behavior. Nonetheless, the current findings offer suggestive evidence that acting inequitably—even when it is profitable—may be experienced as aversive.

Interestingly, the current findings contrast with recent work suggesting that prosocial behavior emerges primarily as a response to threats of punishment (9). In this earlier study, researchers compared neural responses during a dictator game (similar to the one used here) with those during a second game in which receivers had the option of punishing unfair decision makers. Threats of punishment engaged brain regions involved in exerting control over prepotent responses, suggesting that participants effortfully inhibited their impulse to act selfishly only when prompted by extrinsic motivators. However, the undermining effects of extrinsic reinforcement on intrinsic motivation have been well known to social psychologists for decades: After receiving external inducements to engage in an enjoyable behavior, the frequency of that behavior decreases in the absence of the inducement (32, 33). Recent data suggest that such "undermining" of intrinsic value is reflected in decreased en-

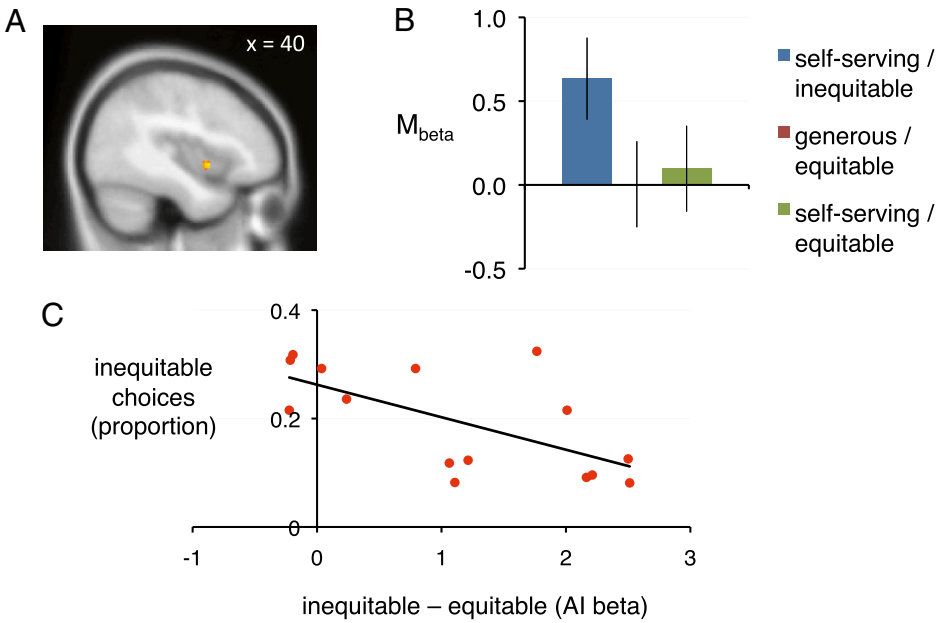

Fig. 3. (*A*) A random-effects, whole-brain contrast of inequitable > equitable decisions revealed activity in the AI. (*B*) This region was engaged preferentially for inequitable as opposed to equitable decisions. (*C*) AI β-values for the contrast of inequitable > equitable decisions were negatively correlated with the proportion of all trials on which participants made inequitable choices.

gagement of reward-related neural structures (34). We suspect that the presence of external motivation (punishment threats) may diminish the intrinsic value otherwise associated with equitable choices; indeed, in some cases, sanctions reduce, rather than increase, prosocial behavior (35). However, because this earlier study did not analyze brain activity following equitable vs. inequitable decisions within the dictator game, it could not index the intrinsic value associated with fairness or how its value may change under external inducements to act prosocially.

It is important to note that although the current study minimized external social pressure—most importantly, the threat of retribution for unfair actions—it is possible that participants nonetheless felt pressure to act prosocially because their decisions were observed by the experimenters or because they did not fully trust that their choices would in fact be obscured from the receiver. A number of studies have dealt with experimenter effects through clever double-blind procedures in which experimenters remain ignorant about participant choices. These studies have produced mixed results: in some cases drastically (but not completely) reducing generosity in dictator games (36), in others leaving generosity unaffected (37), and in yet others demonstrating an effect mediated by other experimental factors such as receivers' deservingness (38). Other types of subtle external pressures (e.g., participant mistrust of task instructions) are more difficult to rule out. Overall, demand characteristics are a vexing and unruly problem in game theoretic paradigms and render true isolation of intrinsic motives toward prosociality methodologically difficult.

That said, the current experimental situation represents a case in which neuroimaging data may help clarify otherwise ambiguous sources of behavior. Specifically, if participants in this study indeed suppressed their selfish impulses in the face of external pressure to act equitably, equitable decision making likely would have engaged neural structures associated with exertion of such regulation, including lateral prefrontal or anterior cingulate cortex (cf. ref. 9). However, the contrast of equitable vs. inequitable choices did not reveal engagement in these regions, even at a lenient threshold of $P < 0.05$, uncorrected for multiple comparisons, suggesting that the exertion of control did not feature prominently into participants' prosocial decisions. Instead, equity was accompanied by a wholly different pattern of brain activity consistent with the experience of subjective reward. Of course, neuroimaging data alone do not provide conclusive evidence of particular subjective states. Nonetheless, such converging data bolster the claim that participants likely were not merely responding to pressure from the experimental setting.

Sources of prosocial behavior vary broadly and in many cases likely include a combination of external pressures and intrinsic value that remain incompletely characterized. For example, although the current study documents the likely experience of subjective value accompanying equitable behavior, it does not demonstrate whether this value plays a causal role in producing such behavior. Future work should manipulate more independently the external social pressures as well as internal motivation to act prosocially, to produce a richer understanding of these factors' contributions to prosociality. Especially important to this endeavor will be examining how contextual variance may mediate the relationship between external and internal motivations, on the one hand, and generous or equitable behavior on the other. For example, individuals may require external pressures to act equitably during competitive interactions or when interacting with outgroup members. By contrast, in a cooperative interaction or when interacting with ingroup members, intrinsic motives may play a larger role in promoting equitability and cooperation.

Fairness simultaneously stands out among humans' most notable and most puzzling behaviors. Although prosociality often can be explained as a response to external pressures such as threats of punishment or to reputation, it cannot always be dismissed in this way. Many behaviors (e.g., anonymous helping) suggest the workings of a deeper, more intrinsic source of prosociality, one aimed at maximizing social—not personal—outcomes. Here, we document such an instance by demonstrating that, in some cases, fairness can be its own reward.

Methods

Participants. Nineteen right-handed participants (12 male, mean age = 23.2 y) with no history of psychiatric or neurological disorders completed the study in exchange for monetary compensation. Informed consent was obtained in

accordance with the regulations of the Committee on the Use of Human Subjects at Harvard University. Four participants produced no generous/equitable choices; their data could not be modeled and were excluded, resulting in a final dataset of 15 participants (9 male, mean age = 21.8 y).

Protocol. Before scanning, participants were introduced to a confederate whom they believed was a second participant. An experimenter informed both the actual participant and the confederate that the study had been designed to examine how individuals form impressions about real other people they had actually met, as opposed to imaginary or fictional others. As such, one of the participants would enter the scanner and make decisions about the other participant while the second participant completed unrelated tasks outside the scanner. The experimenter did not mention an economic decision-making task. Through an ostensibly random (but actually fixed) assignment, the participant always was chosen to enter the scanner, whereas the confederate (hereafter, the "receiver") supposedly was assigned to complete unrelated behavioral tasks outside the scanner.

After entering the scanner, participants completed a short social judgment task described elsewhere (39, 40). They then were introduced to the modified dictator game (see *SI Methods* for the full instructions given to participants). Participants were told that they would make repeated decisions about whether to allocate money to themselves or to the receiver. They were told that five of their decisions, chosen at random, would be enacted (i.e., the money they allocated to themselves or the receiver on those trials would actually be paid out to that person). Importantly, participants also were told that the receiver would not know that the participant had completed the dictator game; instead, extra compensation would simply be included in the payment later mailed to the receiver (thus further minimizing any influence of reputation motives on participants' decisions).

Following this instruction period, participants completed 210 rounds of the dictator game, segregated across three functional MRI runs. After a jittered interstimulus interval (1–6 s), participants were presented with two options. On each side of the screen, a photograph of the participant or of the receiver appeared (each taken immediately before the start of the experiment). The text "$X.XX for" appeared above each photograph, where "X.XX" corresponded to a monetary amount. Thus each trial contained two choices for a participant. For example, the left side of the screen might read "$1.00 for" above a photograph of the participant, and the right side of the screen might read "$1.50 for" above a photograph of the receiver. The side of the screen on which each potential target (the participant and the receiver) appeared varied across trials. These options remained on the screen for a jittered interval (1.5–5 s) before participants were asked to choose between them. To indicate the period during which participants should indicate their choice, the word "Decide" appeared on the screen, after which participants were given 2 s to choose between the two options. Their choice then appeared as a red box surrounding the option they chose, which was displayed on the screen for the remainder of the choice period.

The amounts that each person stood to gain varied parametrically across trials but always adhered to one of a set of six ratios specifying the relationship between the two monetary amounts: 3:1, 2:1, 3:2, 4:3, 5:4, and 1:1. For each trial, a random value between $0.00 and $3.00 was chosen, and a second value was determined by transforming the first value according to the ratio that applied during that trial. These two amounts then were paired pseudorandomly with the two targets (the participant and the receiver). For example, if—in a given trial—the amount of $1.50 was selected, and the ratio was 2:1, then the other choice presented would be $0.75. Note that each ratio thus could produce two relationships between the amounts that the participant and the receiver stood to gain. Assigning $0.75 to the participant, and $1.50 to the receiver would produce a 1:2 ratio between possible self and other gains. If the opposite assignment were made, the self: other ratio would be 2:1. As such, 11 total ratios could relate potential self and other gains: 3:1, 2:1, 3:2, 4:3, 5:4, 1:1, 4:5, 3:4, 2:3, 1:2, and 1:3. Note also that if the ratio on a given trial were larger than 1:1 (e.g., 3:1), it would be equitable for participants to act in a self-serving way (allocating money to themselves), whereas if this ratio were smaller than 1:1, it would be equitable for participants to act generously (allocating money to the receiver). The maximum amount that either the participant or receiver stood to gain on a given trial was $9.00, and trials were organized so that the total amounts of money available to the participant and the receiver over the course of the entire study were comparable ($P > 0.40$).

The choice paradigm included 15 trials adhering to each of the 11 ratios listed above. In addition, we included 15 "pure-self" and 15 "pure-other" reward trials. These trials presented rewards of varying amounts to either the participant or to the receiver while maintaining procedural elements that allowed for parallelism with the main dictator game. Specifically, during pure-self trials the participant was presented with offers of a nonzero amount of money (e.g., $1.00) for herself or $0.00 for the receiver. During pure-other trials, the participant were presented with offers of $0.00 for herself and a nonzero amount of money for the receiver. As such, these trials maintained the same visual and response features of dictator game trials but actually provided rewards only for participants themselves or for the receiver. In other words, these choices represented "costless" rewards for the participant and the receiver, respectively. Finally, we included 15 nonreward trials in which participants chose between $0.00 for themselves and $0.00 for the receiver. The choices included in the neuroimaging analyses reported here, as well as the proportions of inequitable trials plotted in Fig. 3*C*, are based on the 150 trials in which the participant and receiver stood to gain unequal, nonzero amounts of money.

Neuroimaging Acquisition and Analysis. Imaging data were collected on a 3-Tesla Siemens Trio scanner using a gradient-echo echo-planar pulse sequence (31 axial slices, 5-mm thick; 1-mm skip; repetition time = 2 s; echo time = 35 ms; 3.75 × 3.75 in-plane resolution). A high-resolution T1-weighted structural scan (magnetization-prepared rapid gradient echo) was collected before the functional runs. Stimuli were presented on a screen at the end of the magnet bore using the Psychophysics Toolbox for MATLAB (41). Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement.

MRI data were preprocessed and analyzed using SPM statistical parametric mapping image analysis software (Wellcome Department of Cognitive Neurology, London). First, functional data were time-corrected for differences in acquisition time between slices for each whole-brain volume and realigned to correct for head movement. Functional data then were transformed into a standard anatomical space (3-mm isotropic voxels) based on the International Consortium of Brain Mapping ICBM 152 brain template (Montreal Neurological Institute). Normalized data then were spatially smoothed (6 mm full width at one-half maximum) using a Gaussian kernel. Statistical analyses were performed using the general linear model in which the event-related design was modeled using a canonical hemodynamic response function, its temporal derivative, and additional covariates of no interest (a session mean and a linear trend). This analysis was performed individually for each participant, and contrast images for each participant subsequently were entered into a second-level analysis treating participants as a random effect. Brain regions that differentiated between conditions were identified using a statistical criterion of 55 or more contiguous voxels at a voxel-wise threshold of $P < 0.005$. These height and extent thresholds were selected on the basis of a Monte Carlo simulation implemented in MATLAB, to correspond with an overall false-positive rate of less than 5%, corrected for multiple comparisons (42).

Our main analysis consisted of a general linear model that examined the "decision" phase (when participants made a choice about which monetary allocation to enact) of the choice task. This model included regressors for each decision condition of interest: self-serving/inequitable, self-serving/equitable, and generous/equitable. Generous/inequitable decisions (in which the participant stood more to gain than the receiver but nonetheless gave to the other person) were too rare to model and are not included here. The primary analysis in each model consisted of whole-brain, random-effects contrasts between [self-serving/equitable + generous/equitable] > self-serving/inequitable, as well as the opposite contrast, self-serving/inequitable > [self-serving/equitable + generous/equitable]. Thus, this analysis isolated clusters of brain activity that differentiated between equitable choices of either type vs. inequitable choices. Critically, this model included the amount of money participants stood to gain on each trial as covariate of no interest. This analytic approach ensured that resulting brain activity did not reflect the magnitude of potential personal gain involved in each trial. Therefore, resulting patterns of neural response represent the effects of particular decision-types (i.e., equitable or inequitable), irrespective of this potentially confounding intrapersonal variable. All significant activations identified by this contrast are listed in Table S1.

Additionally, to isolate brain regions contributing to the receipt and anticipation of reward to the self, we examined blood oxygen level-dependent (BOLD) differences for the pure-self condition. We constructed a separate generalized linear model (GLM) that specifically modeled pure-self trials (in which only the participant could gain money) and included a parametric regressor representing the amount of money the participant gained on that particular trial. This analysis allowed us to localize brain regions in which response scaled with the magnitude of pure, self-oriented reward using data independent of equitable or inequitable choice trials. As expected, this contrast produced activation of the OFC. We then further interrogated this independently defined region-of-interest for differences in

BOLD response as a function of choice type (equitable vs. inequitable) in the set of trials during which participants were obliged to allocate nonzero amounts of money to themselves or the receiver. We first defined a sphere (radius = 6 mm) about the activation peak related to pure self gains and then extracted parameter estimates (β-weights) from this region of interest related to each choice condition (self-serving/inequitable, self-serving/equitable, and generous/equitable). We then compared these β-weights across conditions using paired-sample *t* tests.

Finally, after identifying activity in the AI related to making inequitable, as opposed to equitable, choices, we explored the relationship between this activity and individuals' decision-making patterns. Specifically, we defined a sphere (radius = 6 mm) about the insula peak defined by the group contrast of self-serving/inequitable > self-serving/equitable and generous/equitable (coordinates: 40, 4, −4) and then extracted β-weights from this region (related to the same contrast) for each participant. These participant-specific β-weights then were correlated with the proportion of trials in which each participant produced self-serving/inequitable trials.

This correlation analysis, however, included a potential confound: Participants who made less inequitable choices also produced a smaller number of trials through which to model inequity-related AI activity. To eliminate this potential concern, we generated a separate GLM analysis that modeled a subset of 10 randomly selected self-serving/inequitable trials per participant, thus standardizing the number of trials in this regressor across all participants. We then recomputed a contrast of inequitable > equitable trials and extracted β-values from a 6-mm sphere surrounding the inequity-related group AI peak described above (coordinates: 40, 4, −4). Finally, we extracted β-values from each participant from this peak and correlated these values with participants' proportion of inequitable choices, as described above. This analysis reproduced the correlation, thus obviating any potential confounds related to variance in participants' number of inequitable choices.

ACKNOWLEDGMENTS. We thank T. J. Eisenstein and Adam Waytz for advice and assistance. This work was supported by a grant from the Templeton Foundation for Positive Neuroscience.

1. Camerer C (2003) *Behavioral Game Theory* (Princeton University Press, Princeton).
2. Henrich J, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav Brain Sci* 28:795–815.
3. Nowak MA, Page KM, Sigmund K (2000) Fairness versus reason in the ultimatum game. *Science* 289(5485):1773–1775.
4. Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868.
5. Trivers R (1971) The evolution of reciprocal altruism. *Q Rev Biol* 46:35–57.
6. Milinski M, Semmann D, Krambeck HJ (2002) Reputation helps solve the 'tragedy of the commons'. *Nature* 415:424–426.
7. Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328:617–620.
8. Gintis H (2008) Behavior. Punishment and cooperation. *Science* 319:1345–1346.
9. Spitzer M, Fischbacher U, Herrnberger B, Grön G, Fehr E (2007) The neural signature of social norm compliance. *Neuron* 56(1):185–196.
10. Forsythe R, Horowitz J, Savin N, Sefton M (1994) Fairness in simple bargaining experiments. *Games Econ Behav* 6:347–369.
11. Andreoni J (1990) Impure altruism and donations to public goods: A theory of warm-glow giving. *Econ J* 100:464–477.
12. Andreoni J, Miller J (2002) Giving according to GARP: An experimental study of rationality and altruism. *Econometrica* 70:737–753.
13. de Waal FB (2008) Putting the altruism back into altruism: The evolution of empathy. *Annu Rev Psychol* 59:279–300.
14. Dawes CT, Fowler JH, Johnson T, McElreath R, Smirnov O (2007) Egalitarian motives in humans. *Nature* 446:794–796.
15. Bolton G, Ockenfels A (2000) ERC: A theory of equity, reciprocity, and competition. *Am Econ Rev* 90(1):166–193.
16. Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* 9:545–556.
17. Rolls ET (2004) The functions of the orbitofrontal cortex. *Brain Cogn* 55(1):11–29.
18. Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. *Science* 315:515–518.
19. Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441(7090):223–226.
20. Rilling J, et al. (2002) A neural basis for social cooperation. *Neuron* 35:395–405.
21. Tricomi E, Rangel A, Camerer CF, O'Doherty JP (2010) Neural evidence for inequality-averse social preferences. *Nature* 463:1089–1091.
22. Fehr E, Camerer CF (2007) Social neuroeconomics: The neural circuitry of social preferences. *Trends Cogn Sci* 11:419–427.
23. Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. *J Neurosci* 25:4806–4812.
24. Wager TD, Feldman Barrett L (2004) From affect to control: Functional specialization of the insula in motivation and regulation. *PsycExtra,* available at http://wagerlab.colorado.edu.
25. Harbaugh WT, Mayr U, Burghart DR (2007) Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316:1622–1625.
26. Mobbs D, et al. (2009) A key role for similarity in vicarious reward. *Science* 324:900.
27. Hare TA, Camerer CF, Knoepfle DT, 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.
28. Moll J, et al. (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci USA* 103:15623–15628.
29. Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53(1):147–156.
30. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the Ultimatum Game. *Science* 300:1755–1758.
31. Craig AD (2002) How do you feel? Interoception: The sense of the physiological condition of the body. *Nat Rev Neurosci* 3:655–666.
32. Lepper M, Greene D, Nisbett R (1973) Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *J Pers Soc Psychol* 28(1):129–137.
33. Deci E (1975) *Intrinsic Motivation* (Plenum, New York).
34. Murayama K, Matsumoto M, Izuma K, Matsumoto K (2010) Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proc Natl Acad Sci USA* 107:20911–20916.
35. Fehr E, Rockenbach B (2003) Detrimental effects of sanctions on human altruism. *Nature* 422(6928):137–140.
36. Hoffman E, McCabe K, Shachat K, Smith V (1994) Preferences, property rights and anonymity in bargaining games. *Games Econ Behav* 7:346–380.
37. Bolton G, Katok E, Zwick R (1998) Dictator game giving: Rules of fairness versus acts of kindness. *Int J Game Theory* 27(2):269–299.
38. Eckel C, Grossman P (1996) Altruism in anonymous dictator games. *Games Econ Behav* 16(2):181–191.
39. Jenkins AC, Macrae CN, Mitchell JP (2008) Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci USA* 105:4507–4512.
40. Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50(4):655–663.
41. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
42. Slotnick SD, Moo LR, Segal JB, Hart J, Jr. (2003) Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res Cogn Brain Res* 17(1):75–82.