Research Article

The Sunny Side of Fairness

Preference for Fairness Activates Reward Circuitry (and Disregarding Unfairness Activates Self-Control Circuitry)

Golnaz Tabibnia, Ajay B. Satpute, and Matthew D. Lieberman

University of California, Los Angeles

ABSTRACT—***Little is known about the positive emotional impact of fairness or the process of resolving conflict between fairness and financial interests. In past research, fairness has covaried with monetary payoff, such that the mental processes underlying preference for fairness and those underlying preference for greater monetary outcome could not be distinguished. We examined self-reported happiness and neural responses to fair and unfair offers while controlling for monetary payoff. Compared with unfair offers of equal monetary value, fair offers led to higher happiness ratings and activation in several reward regions of the brain. Furthermore, the tendency to accept unfair proposals was associated with increased activity in right ventrolateral prefrontal cortex, a region involved in emotion regulation, and with decreased activity in the anterior insula, which has been implicated in negative affect. This work provides evidence that fairness is hedonically valued and that tolerating unfair treatment for material gain involves a pattern of activation resembling suppression of negative affect.***

Anyone who has watched children negotiate how to share a piece of cake knows that humans are exquisitely sensitive to fairness. Although economic models of decision making have traditionally assumed that individuals are motivated solely by material utility (e.g., financial payouts) and are not directly affected by social factors such as fairness (Camerer, Loewenstein, & Prelec, 2005; Kahneman, Knetsch, & Thaler, 1986), there is increasing empirical evidence that fairness does play a role in economic decision making (Fehr & Schmidt, 1999; Sears & Funk, 1991). Fairness in economic-exchange tasks is typically defined as the equitable distribution of an initial stake of money between two people. Because fair outcomes tend to be more materially desirable for the recipient than unfair outcomes in everyday life, it is difficult to distinguish the desire for fairness from the desire for material gain. Bilateral bargaining games, such as the ultimatum game, allow these two potential motives to be examined separately. The results of studies using the ultimatum game indicate that people are sensitive to fairness over and above its consequences for material gain (Güth, Schmittberger, & Schwarze, 1982). Although there is evidence that receiving an unfair proposal is associated with negative emotional responses (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), no study on economic decision making has examined whether a fair proposal produces positive emotional responses beyond those associated with the material gain itself.

To examine the emotional response associated with fair treatment, we conducted two ultimatum-game experiments. In this game, one player proposes how to split a given sum of money, the stake, and another player responds. If the responder accepts, each player keeps the amount allocated by the proposer. If the responder rejects the offer, neither player receives any money. Numerous studies using the ultimatum game have shown that responders do not maximize material utility by accepting every offer, but rather tend to reject offers below 20% of the stake (Camerer & Thaler, 1995), even when there will be no future interactions with the partner (Güth et al., 1982).

In a neuroimaging study of the ultimatum game, Sanfey et al. (2003) observed that being treated unfairly is associated with a negative emotional response, inferred from anterior insula activation. They did not report what regions were more active during fair than during unfair offers. Furthermore, because fair offers (i.e., \$5 out of \$10) were always associated with higher monetary payoff than unfair offers (e.g., \$2 out of \$10), it is difficult to dissociate emotional response to fairness from emotional response to monetary payoff in their study. Hence, it is unclear from these data whether fair treatment is rewarding, in addition to unfair treatment being aversive.

Research on social justice suggests that seeking justice is a basic human impulse (i.e., the justice motive; Tyler, 1991), possibly rooted in a basic social motivation to be accepted (Bau-

Address correspondence to Golnaz Tabibnia, The Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, 760 Westwood Plaza, C8-532, Los Angeles, CA 90095-1759, e-mail: golnaz@ucla.edu.

meister & Leary, 1995). Perceived fair treatment from public institutions (e.g., court, police) has been associated with satisfaction beyond the effects of the material outcomes, such as sentencing (Tyler, 1984). Critically, studies examining the impact of fairness on positive and negative emotions separately, controlling for material outcomes, have found substantial increases in self-rated positive emotions associated with fair treatment (De Cremer & Alberts, 2004; Hegtvedt & Killian, 1999).

If being treated fairly is experienced as rewarding, then people should be happier with a fair offer than with an unfair offer of the same monetary value. Similarly, brain regions associated with reward should be more active during fair than during unfair treatment, after controlling for material utility. These reward regions include the ventral striatum, the amygdala, ventromedial prefrontal cortex (VMPFC), orbitofrontal cortex (OFC), and midbrain dopamine regions (Cardinal, Parkinson, Hall, & Everitt, 2002; Trepel, Fox, & Poldrack, 2005). Although the amygdala has commonly been associated with fear processes, activity in this structure, particularly on the left, has also been associated with reward processes (Hommer et al., 2003; Zalla et al., 2000).

In order to control for material utility, we varied both the offer amount and the stake size across trials (see Fig. 1). On different trials, the same offer amount could represent a large percentage of the total stake (e.g., \$7 out of \$15), and therefore seem fair, or a small percentage of the total stake (e.g., \$7 out of \$23), and therefore seem unfair. Differences in ratings of happiness or reward activations observed in the comparison of such trials cannot be attributed to the magnitude of the monetary reward and thus are reasonably attributed to fairness.

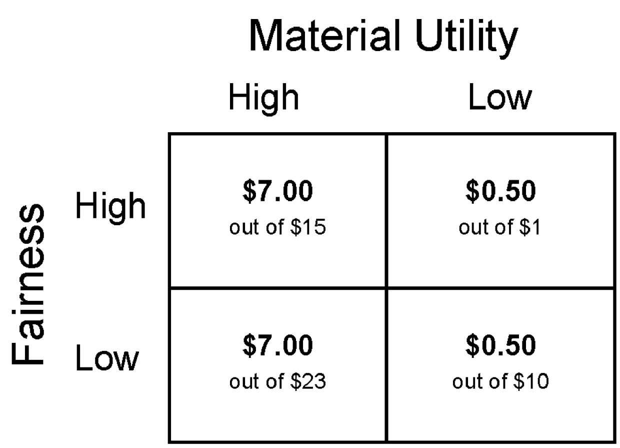

Fig. 1. Illustration of the manipulation of material utility and fairness. In the analysis of fairness preference, trials of equal material utility were divided according to fairness (i.e., the ratio of the offer to the stake). In this example, the offers in the top row are high-fairness offers, and those in the bottom row are low-fairness offers, and each of two monetary outcomes is presented in both a high-fairness offer and a low-fairness offer. Across trials, high- and low-fairness offers had the same average material utility.

We also examined neural response during trials in which fairness and material outcome were at odds—that is, trials on which the offers were unfair but financially desirable (e.g., \$8 out of \$23). Thus, we examined the neural correlates of the tendency to accept unfair offers. Two possibilities were investigated. First, accepted unfair offers may activate reward circuitry to a greater extent than rejected unfair offers; such a pattern would reflect enhanced desire to accept the offers. Second, emotion regulation may be engaged when unfair offers are accepted, which would diminish the anterior insula's response and decrease the desire to reject the offer. In this case, one would expect decreased activity in the anterior insula and increased activity in a prefrontal region that has been associated with emotion regulation, such as the right ventrolateral prefrontal cortex (right VLPFC; Hariri, Bookheimer, & Mazziotta, 2000; Lieberman et al., 2007).

Participants in our studies played the role of responder. In Experiment 1, we measured emotional responses to each offer by obtaining self-ratings of happiness and contempt. In Experiment 2, we measured neural responses to fair and unfair offers using functional magnetic resonance imaging (fMRI). Participants were told that their decisions regarding four randomly selected offers in the experiment would actually be implemented, such that they and the proposers of those offers would be paid or not, according to the responses.

EXPERIMENT 1

Method

Participants and Task

Twenty-nine undergraduates (average age = 20.1 years; 18 females, 11 males) participated after replying to a flyer indicating that they could earn up to \$52 for participation. They were told that the proposers had submitted their offers already and would not be present. Actually, there were no real proposers. During the experiment, each offer was presented as follows (see Fig. 2): First, participants were shown the purported proposer. Then, the stake was indicated, followed by the offer. While the offer was displayed, participants could accept or reject it. After the experiment, all participants were debriefed, paid a total of \$27, and entered in a lottery in which 4 participants were selected to receive an additional \$25. Thus, all participants had a chance of winning "up to \$52," as advertised. Offers ranged from 5% to 50% of the total stake size, and stakes ranged from \$1 to \$30. We selected particular offer values and then matched each with two stake sizes in order to obtain one low and one high ratio of offer to stake size.

Measures

After playing the game, participants were asked to rate (1–7) how much happiness and contempt they felt in response to each of a preselected subset of 28 offers. This subset consisted of 14

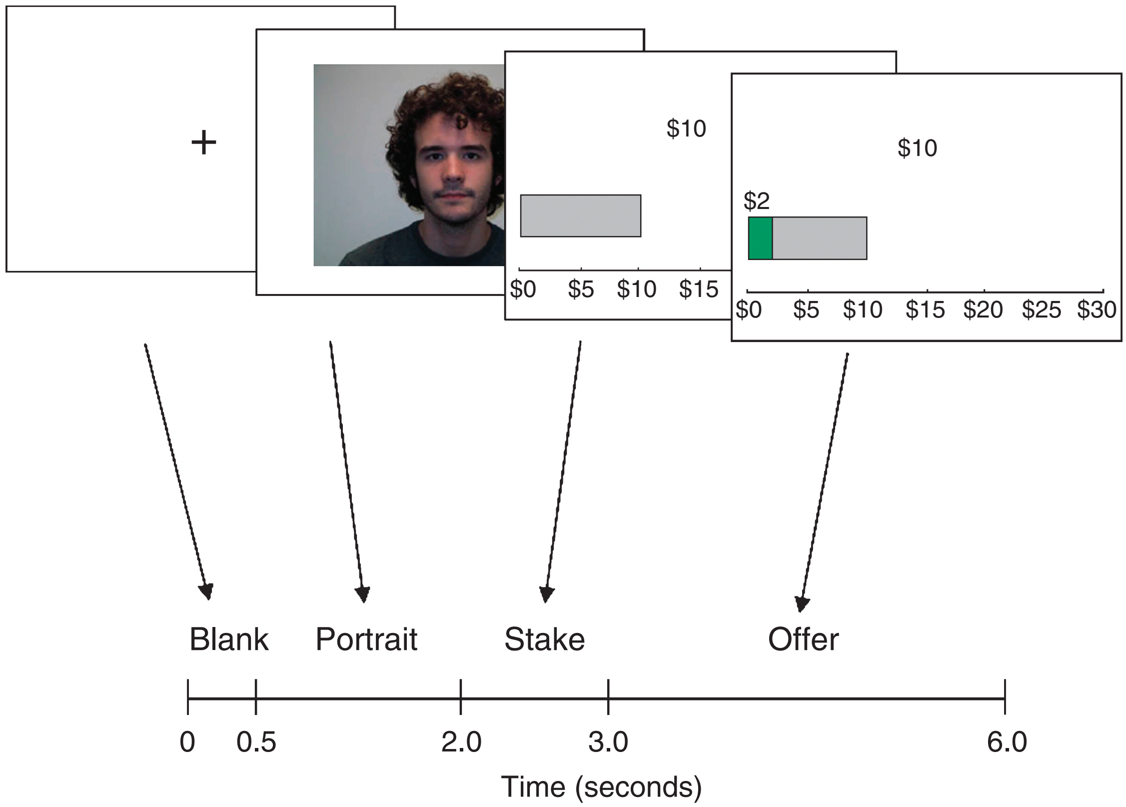

Fig. 2. Diagram illustrating the structure of each 6-s trial: The participant saw a fixation cross for 0.5 s, a picture of the purported proposer for 1.5 s, a display indicating the size of the stake for 1 s, and finally the offer for 3 s. The participant was given the final 3 s of each trial to respond, by pressing one button to ''accept'' and another to ''reject'' the offer. Then the next blank screen appeared for 0.5 s, and so forth.

fair offers ($\geq$ 40% of the stake) and 14 unfair offers ($\leq$ 20% of the stake) that were matched in material utility (e.g., \$2 out of \$4 matched with \$2 out of \$10). In this subset of offers, the ratio of offer to stake size ranged from 5 to 50% (average = 28%); the stakes ranged from \$1 to \$30 (average = \$12.18).

Results

Fairness Predicts Happiness, Independently of Contempt

Happiness ratings were strongly associated with fairness (i.e., percentage of the stake size offered). Participants reported greater happiness for fair offers ($\geq$ 40%) than unfair offers ($\leq$ 20%) of equal monetary value, $t(13) = 7.73$, $p_{rep} > .99$, $d = 4.29$. Similarly, there was a strong correlation between level of fairness and happiness ratings ($r = .89$, $p_{rep} > .99$). A complementary pattern was observed for contempt ratings. Participants reported greater contempt for unfair offers than for fair offers of equal monetary value, $t(13) = 5.51$, $p_{rep} > .99$, $d = 3.06$, and there was a strong correlation between level of fairness and contempt ratings ($r = -.82$, $p_{rep} > .99$).

Given that happiness and contempt were correlated ($r = -.82$, $p_{rep} > .99$), we examined the effect of fairness on happiness after partialing out contempt. Happiness controlling for contempt ($H_{\bar{C}}$) was still associated with fairness (r = .38, $p_{rep} > .95$); however, contempt controlling for happiness ($C_{\bar{H}}$), was not ($r = -.17$, $p_{rep} = .80$).

Fairness Predicts Happiness, Independently of Material Outcome

Offer amount and emotion ratings were not strongly associated. Happiness ratings of high-value offers (> \$2) did not differ from those of low-value offers ($\leq$ \$2) of equal fairness, $t(13) = 1.16$, $p_{rep} = .87$, $d = 0.64$. Although the correlation between offer amount and happiness was marginally significant ($r = .32$, $p_{rep} > .95$), offer amount did not predict $H_{\bar{C}}$ ($r = .28$, $p_{rep} > .92$). Similarly, contempt ratings of high-value offers did not differ from those of low-value offers of equal fairness, $t(13) = 0.39$, $p_{rep} = .60$, $d = 0.15$. The correlations between offer amount and contempt ($r = -.20$, $p_{rep} = .85$) and between offer amount and $C_{\bar{H}}$($r = .11$, $p_{rep} = .71$) were not significant.

After controlling for the effect of offer amount, fairness still predicted happiness ($\beta = .87$, $p_{rep} > .99$) and $H_{\bar{C}}$($\beta = .35$, $p_{rep} > .95$), and fairness still predicted contempt ($\beta = -.81$, $p_{rep} > .99$), but not $C_{\bar{H}}$ ($\beta = -.18$, $p_{rep} = .82$). Together, these results indicate that fairness, independently of offer amount, predicts happiness, independently of contempt.

EXPERIMENT 2

Method

Participants and Task

Twelve undergraduates (average age = 21.8 years; 9 females, 3 males) participated. The task was similar to that in Experiment 1, except that the stakes ranged from \$1 to \$23 (average = \$9.60), and participants underwent fMRI scanning while they considered the offers. After the scanning session, participants indicated what they considered a fair offer for each stake size.

fMRI Acquisition and Analysis

Data were acquired on a GE 3-T full-body scanner. Scanning parameters were identical to those used in our previous studies (see Lieberman, Jarcho, & Satpute, 2004). Each of four functional scans consisted of thirty-one 6-s trials, as well as five 6-s jitter trials. The MR data were analyzed using SPM99 (Wellcome Department of Cognitive Neurology, London). Images for each participant were realigned, slice-timed, normalized to Montreal Neurological Institute (MNI) space, and smoothed with an 8-mm Gaussian kernel, full width at half maximum.

We included in the fairness analysis only matched pairs of trials in which the same amount was accepted in one case and rejected in the other. Thus, we analyzed neural activation in response to offers that possessed matched financial rewards and therefore differed primarily in their perceived fairness.

Events were modeled with a canonical hemodynamic response function time-locked with the onset of the offer.[1] Linear contrasts were employed to assess comparisons of interest within individual participants. Random-effects analyses of the group were computed using the contrast images generated for each participant. For regions of interest, significance was set using an uncorrected p value of .005 (5-voxel threshold). Post hoc analyses were carried out at a p value of .001 (20-voxel threshold). Peristimulus hemodynamic time courses were computed by identifying clusters of activations from the random-effects analyses and then applying to these clusters a selective averaging procedure on a participant-by-participant basis (Poldrack, 2004). Regression analysis of the tendency to accept unfair offers was conducted by performing a group analysis in which each participant's rate of accepting unfair offers was entered as a regressor to identify which activations correlated with the rate.

Results

Behavioral Results

On average, participants accepted 56.3% (SD = 12.3%) of all the offers in the experiment, a result indicating that they were not solely motivated by monetary reward, in which case they would have accepted all offers. This acceptance rate decreased significantly ($p_{\mathrm{rep}} > .95$) as the proportion of the offer relative to the stake size decreased (see Table 1). Across all trials in the study, multiple participants rejected offers as financially desirable as \$8 out of \$23; on average, participants rejected at least one offer as high as \$4.88.

Self-Report Results

Average self-reported estimates of what constituted a fair offer ranged from 45.2% to 48.3% across stakes. For each participant, we calculated the percentage of unfair offers (as identified by the participant's own responses after the scanning session) that were accepted. On average, participants accepted 49.0% of the offers that were below their self-reported fairness thresholds. These results suggest that although participants were influenced by fairness, they were sometimes able to overcome or disregard fairness considerations and make the economically normative decision when fairness and material considerations were at odds.

fMRI Results: Fairness Preference

Several brain regions associated with reward processes were more active during high-fairness offers than during low-fairness offers, after controlling for material utility. Brain regions that showed greater activity for high- than for low-fairness offers were the ventral striatum, the amygdala, VMPFC, OFC, and a midbrain region near the substantia nigra (see Table 2 and Fig. 3). A post hoc whole-brain analysis also revealed sensitivity to increased fairness in lateral temporal cortex ($x = -44$, $y = 18$, $z = -24$), $t(11) = 4.74$, $p_{\mathrm{rep}} > .99$, $d = 2.86$.

In the high- versus low-fairness contrast, participants producing greater activity in the ventral striatum, compared with other participants, tended to produce greater activity in the amygdala ($r = .81$, $p_{\mathrm{rep}} > .95$) and VMPFC ($r = .65$, $p_{\mathrm{rep}} > .95$) as well. This finding is consistent with the hypothesis that the ventral striatum, the amygdala, and VMPFC function together as a motivational circuit related to reward (Trepel et al., 2005).

We examined whether the reward activations could be attributed to a continuation of effects occurring during the first half of each trial, when the face and stake size were presented, prior to when the offer was presented and a choice was made. We

TABLE 1

Likelihood of an Offer Being Accepted as a Function of the Ratio of the Offer to the Stake Size

Ratio of offer to stake	Acceptance rate
50%	97.9%
40–49%	92.3%
30–39%	75.8%
20–29%	44.7%
10–19%	30.8%
<10%	1.4%

[1]These imaging techniques allowed us to determine the blood-oxygenation-level-dependent signal, an index of neuronal activity, associated with specific types of offers.

TABLE 2

A Priori Regions Showing Greater Activation in High-Fairness Than in Low-Fairness Trials

Region	Hemisphere	Coordinates x	y	z	No. of voxels	t
Ventral striatum	Left	−6	4	0	7	4.03
VMPFC	Left	−14	32	−10	10	4.85
	Left	−16	16	−16	6	3.83
	Right	10	60	−4	5	3.82
Orbitofrontal cortex	Right	36	36	−20	24	4.32
Amygdala	Left	−12	−4	−24	10	3.81
	Right	20	−12	−12	32	4.32
Midbrain, SN	Left	−14	−20	−6	9	4.75

Note. No activation was greater in low-fairness than in high-fairness trials. The coordinates are from the Montreal Neurological Institute (MNI) atlas. Significance was based on an uncorrected p value of .005, with a 5-voxel threshold. VMPFC = ventromedial prefrontal cortex; SN = substantia nigra.

performed a new fairness analysis targeting the 3-s period prior to offer onset. No motivational areas active in the fairness analysis were active in this new analysis.

fMRI Results: Accepting Unfair Offers

Our findings for the anterior insula are consistent with those of Sanfey et al. (2003). Specifically, we observed increased anterior insula activity (relative to a resting baseline) during idiographically defined unfair trials that were rejected (x = 36, y = 18, z = −8), $t(11)$ = 4.39, p_{rep} > .99, d = 2.95. We also examined the neural structures activated when responders overcame fairness concerns and accepted offers they considered unfair. The hypothesis that accepted unfair offers activate reward circuitry to a greater extent than rejected unfair offers was not supported by the data. There was no activity in the ventral striatum, the amygdala, or VMPFC (at a liberal statistical threshold of p < .01 uncorrected) when accepted unfair offers were compared with resting baseline. However, a large cluster (k = 834) in right VLPFC was active in this comparison (x = 50, y = 44, z = 8), $t(11)$ = 5.39, p_{rep} > .99,

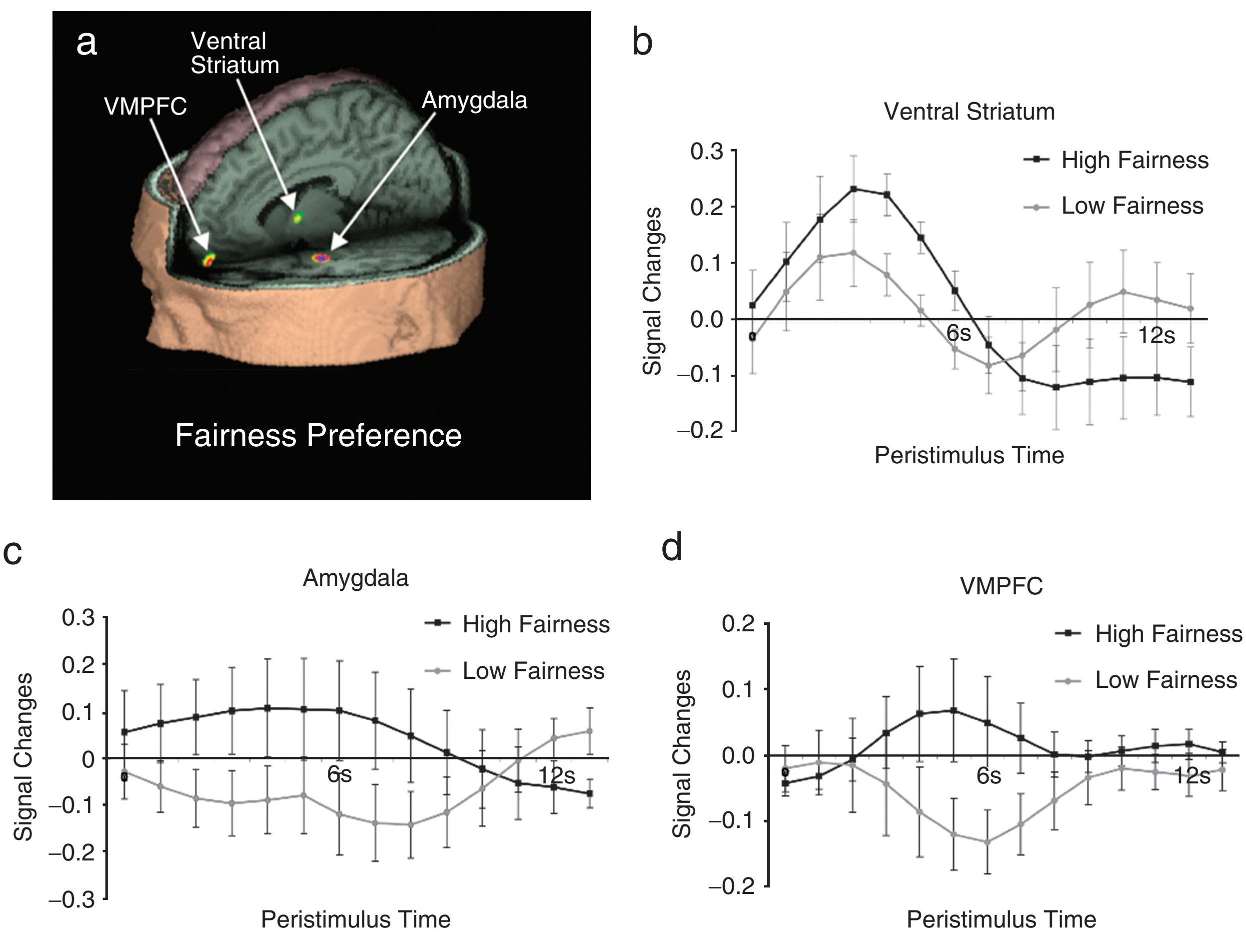

Fig. 3. Ventromedial prefrontal cortex (VMPFC), ventral striatum, and amygdala activation associated with fairness preference. The illustration (a) shows the location of clusters with significantly greater activation in response to fair compared with unfair offers. The graphs present the time course of activity for fair and unfair trials, relative to a resting baseline, in (b) the ventral striatum, (c) the amygdala, and (d) VMPFC. Error bars indicate ± 1 SE. Along the abscissa, 0 s indicates the onset of the offer (which was 3 s after the trial began).

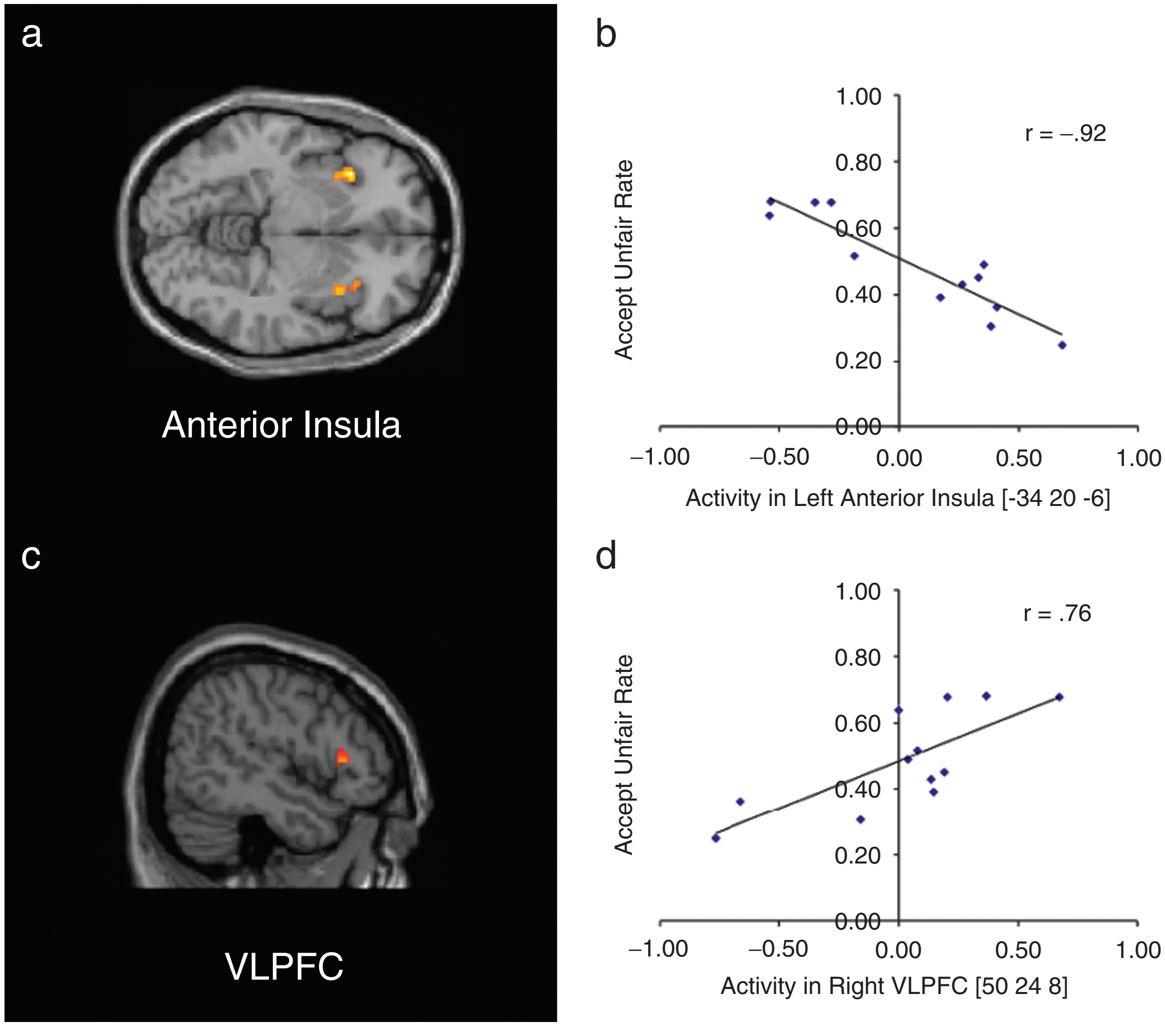

Fig. 4. Brain activation associated with the tendency to accept unfair offers. The illustrations show the location of areas in (a) left anterior insula and (c) right ventrolateral prefrontal cortex (right VLPFC) whose activation predicted this tendency. The corresponding scatter plots (b and d) depict the correlation between signal change in these areas during accepted relative to rejected offers and the rate at which participants accepted offers they later identified as unfair.

$d = 3.25$, a finding consistent with the hypothesis that accepting unfair offers may involve emotion regulation.

To further explore the involvement of right VLPFC and the anterior insula in decisions regarding unfair offers, we specifically compared trials in which unfair offers were accepted with trials in which unfair offers were rejected. As expected, the left anterior insula was less active when unfair offers were accepted than when they were rejected ($x = -28$, $y = 8$, $z = -6$), $t(11) = 4.06$, $p_{rep} > .99$, $d = 2.45$. Activity in this insula region was inversely correlated with right VLPFC activity ($x = 58$, $y = 34$, $z = 10$) during trials in which unfair offers were accepted ($r = -.77$, $p_{rep} > .99$), a finding consistent with the hypothesis that right VLPFC reduces insula activity in such cases.

We also examined the rate at which participants accepted offers they rated unfair in the postscanning questionnaire, regressing this index onto brain activations in the contrast of accepted versus rejected offers. As Figure 4 shows, participants who accepted a higher proportion of idiographically defined unfair offers showed a greater increase in activity in right VLPFC ($x = 50$, $y = 24$, $z = 8$), $t(11) = 3.68$, $p_{rep} > .99$, $d = 2.22$, and a greater decrease in activity in the anterior insula on the left ($x = -34$, $y = 20$, $z = -6$), $t(11) = -7.59$, $p_{rep} > .99$, $d = 4.58$, and right ($x = 32$, $y = 22$, $z = -10$), $t(11) = -6.01$, $p_{rep} > .99$, $d = 3.62$, during trials in which offers were accepted relative to trials in which offers were rejected. Furthermore, activity in right VLPFC was again inversely correlated with activity in the left ($r = -.68$, $p_{rep} > .99$) and right ($r = -.86$, $p_{rep} > .99$) anterior insula. No activity in the ventral striatum, the amygdala, or VMPFC was positively correlated with the tendency to accept unfair offers. These findings suggest a prefrontal down-regulation of negative emotional responses during the process of accepting unfair offers.

If right VLPFC is involved in the acceptance of unfair offers by reducing negative affect associated with the anterior insula, then the relationship between right VLPFC activity and the rate of accepting unfair offers should be mediated by activity in the anterior insula. Indeed, the direct path between right VLPFC activity and the rate of accepting unfair offers ($\beta = .76$, $p_{\mathrm{rep}} >$ .99) was significantly mediated by activity in the left anterior insula (Sobel test: $Z = 2.51$, $p_{\mathrm{rep}} > .99$). After we controlled for activity in this insula region, the remaining path between VLPFC activity and the rate of accepting unfair offers was no longer significant ($\beta = .25$, $p_{\mathrm{rep}} = .94$).

DISCUSSION

Experiment 1 demonstrated that fairness, but not monetary value, predicted self-reported happiness independently of contempt ratings. In Experiment 2, nearly all commonly identified reward areas, including the ventral striatum, the amygdala, VMPFC, and OFC, were associated with fairness preference. Together, these results suggest that individuals react to fairness with a positive hedonic response, rather than that fairness produces a neutral state and only unfairness produces an affective response. These results are consistent with previous reports that reward regions such as the striatum, VMPFC, and the amygdala are responsive to cooperative partners and behavior (King-Casas et al., 2005; Rilling et al., 2002; Singer, Kiebel, Winston, Dolan, & Frith, 2004). Further, these results suggest that fairness processing is relatively automatic and intuitive, as the ventral striatum, the amygdala, and VMPFC have all been associated with automatic and intuitive processes (Lieberman, 2007).

In previous studies of the ultimatum game, fairness and material utility have covaried. Typical offers in these studies have been either completely fair and of highest material utility (e.g., \$5 out of \$10), and therefore easily accepted, or very unfair and of lowest material utility (e.g., \$1 out of \$10), and therefore usually rejected. However, in real life there are situations in which the choice is less straightforward. In the current study, we manipulated conflict-ridden choices by presenting offers that were of considerable material utility but relatively unfair (e.g., \$8 out of \$23).

To the extent that a participant produced increased activity in right VLPFC and decreased activity in the anterior insula, the participant was more likely to accept unfair but financially desirable offers. One interpretation of this result is that participants who accepted unfair offers were better able than others to down-regulate the negative emotional response to unfair treatment. Alternatively, increased activity in anterior insula could reflect the "pain of paying" associated with rejecting an offer (Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007), rather than the social pain of unfair treatment. Thus, participants who were more likely to accept unfair offers may have had lower insula activity because they experienced little pain of paying. However, the overall pattern of results involving right VLPFC is more consistent with the former interpretation. Activity in the insula mediated the relationship between right VLPFC activity and tendency to accept unfair offers, and this finding supports the hypothesis that right VLPFC promotes more normative decision making by down-regulating activity in the anterior insula when unfair offers are considered. Thus, the ability to swallow one's pride, overcome the insult, and take an unfair offer may involve active down-regulation of emotional responses to unfair treatment.

Consistent with the idea that right VLPFC activity drove participants' decisions to accept unfair offers, studies using repetitive transcranial magnetic stimulation (rTMS) have identified the right lateral prefrontal cortex as playing a causal role in rational decision making in the ultimatum game (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; van't Wout, Kahn, Sanfey, & Aleman, 2005). Interestingly, in the current study, accepted unfair offers were not associated with increased activity in reward-related regions, which supports the interpretation that logical rather than hedonic processes guided these particular decisions. In the two rTMS studies, transient disruption of function in right dorsolateral prefrontal cortex (right DLPFC) interfered with rejection of unfair offers. Although the dorsolateral region identified in these studies is structurally and functionally distinct from the VLPFC region identified in the current study, these findings together are consistent with the notion (van't Wout et al., 2005) that the default response in the ultimatum game is to reject an unfair offer; DLPFC may be needed to maintain this default goal, whereas VLPFC may be needed to override it. An alternative interpretation is that DLPFC may be needed to override material self-interest (Knoch et al., 2006), as the role of VLPFC may be to override fairness concerns.

In previous studies, right VLPFC has consistently been associated with the down-regulation of activity in regions supporting negative affect (Lieberman, in press). Specifically, it has been associated with the down-regulation of the amygdala's response to pictures of disturbing scenes (Hariri, Mattay, Tessitore, Fera, & Weinberger, 2003) and pictures of evocative faces (Hariri et al., 2000; Lieberman et al., 2007; Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005), as well as the down-regulation of the anterior cingulate's response to physical (Lieberman, Jarcho, Berman, et al., 2004; Wager et al., 2004) and social (Eisenberger, Lieberman, & Williams, 2003) pain. Right VLPFC has also been linked to reduced susceptibility to amygdala-mediated framing effects and thus to rational decision making (De Martino, Kumaran, Seymour, & Dolan, 2006). Thus, it is plausible that in the current study, right VLPFC down-regulated affect-related activity in the anterior insula to enable the rational decision of accepting an unfair offer.

The fact that we did not find any region exhibiting greater activity for low-fairness offers than for high-fairness offers seems at odds with the results of Sanfey et al. (2003), who reported greater activity in anterior insula and prefrontal areas during unfair than during fair offers. It is possible that the differences in results are due to differences in the experimental designs, such as the use of rapid versus slow event-related designs or the types of fair and unfair trials used. Nonetheless, in both investigations, increased insula activity correlated with the tendency to reject unfair offers, so the studies converge on similar functional interpretations of insula activity.

One may wonder whether the reward activations observed in the fairness-preference analysis were in response to fairness per se or were actually due to the perceived higher probability of receiving the monetary offer in fair trials (which were accepted) than in unfair trials (which were rejected). Although our data cannot fully rule out the latter interpretation, it is unlikely. If greater expectation of monetary payoff was driving reward activity during high-fairness trials in Experiment 2, we should have found reward activity during acceptance of unfair offers, which had the same perceived probability of payoff as the accepted fair offers. However, activation during accepted unfair offers was not greater than baseline activation in any reward-related regions, which suggests that the increased activity in these regions during high-fairness trials was not driven by the perceived probability of monetary payoff.

Finally, the relatively early onset of activity in the amygdala (see Fig. 3c) suggests that this response may have been both related to the offer and a continuation of response from the first half of the trial. However, we did not find significant amygdala activation in a whole-brain analysis of the first half of the trial, even at a lenient threshold ($p_{rep} > .99$, uncorrected). The discrepancy between this null result and the apparent time course of amygdala activation in Figure 3c reflects the fact that whole-brain analyses examine the correlation between the blood-oxygenation-level-dependent response and the canonical hemodynamic response, and do not show effects at any one time point. Despite the absence of significant amygdala activation in this analysis, however, it is difficult to infer that the amygdala activity reported here reflects reward processing per se, as such activity has also been observed in numerous studies of negatively valenced information processing. However, when amygdala activity co-occurs with activity in VMPFC and ventral striatum, it is commonly in the context of reward processing (Cardinal et al., 2002; Petrovich, Holland, & Gallagher, 2005; Trepel et al., 2005). Our confidence in the inference that the amygdala response is related to reward processing is increased because of (a) the convergence of evidence from self-report and (b) the fact that this response was observed in the context of the activation of multiple regions that work together in a network underlying reward processing (Poldrack, 2006). Future studies are needed to elucidate the role of the amygdala in social reward.

In conclusion, these findings suggest that people may prefer fair outcomes at the cost of material utility in part because they hedonically value fairness itself; this preference may not be motivated solely by negative emotional responses to unfairness or by the impersonal application of culture-driven rules. Moreover, when material utility outweighs social utility, people may down-regulate their affect-related neural response to unfair treatment in order to choose the economically desirable option. These results support the notion that the automatic or default reaction in economic decision making is to prefer the fair and refuse the unfair (van't Wout et al., 2005), not just because fair options also tend to be materially advantageous or because unfairness is jarring, but also because fair treatment can be rewarding in itself.

Acknowledgments—We thank the UCLA Brain Mapping Medical Organization, the Ahmanson Foundation, the Pierson-Lovelace Foundation, and the Tamkin Foundation for support and John Monterosso and Daniel Gilbert for helpful comments on earlier drafts.

REFERENCES

Baumeister, R.F., & Leary, M.R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529.

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*, 9–64.

Camerer, C., & Thaler, R.H. (1995). Anomalies: Ultimatums, dictators, and manner. *Journal of Economic Perspectives*, *9*, 209–219.

Cardinal, R.N., Parkinson, J.A., Hall, J., & Everitt, B.J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, *26*, 321–352.

De Cremer, D., & Alberts, H.J.E.M. (2004). When procedural fairness does not influence how positive I feel: The effects of voice and leader selection as a function of belongingness need. *European Journal of Social Psychology*, *34*, 333–344.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R.J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*, 684–687.

Eisenberger, N.I., Lieberman, M.D., & Williams, K.D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, *302*, 290–292.

Fehr, E., & Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*, 367–388.

Hariri, A.R., Bookheimer, S.Y., & Mazziotta, J.C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *NeuroReport*, *11*, 43–48.

Hariri, A.R., Mattay, V.S., Tessitore, A., Fera, F., & Weinberger, D.R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biological Psychiatry*, *53*, 494–501.

Hegtvedt, K.A., & Killian, C. (1999). Fairness and emotions: Reactions to the process and outcomes of negotiations. *Social Forces*, *78*, 269–303.

Hommer, D.W., Knutson, B., Fong, G.W., Bennett, S., Adams, C.M., & Varnera, J.L. (2003). Amygdalar recruitment during anticipation of monetary rewards: An event-related fMRI study. In A. Pitkanen, A. Shekhar, & L. Cahill (Eds.), *Annals of the New York Academy of Sciences: Vol. 985. The amygdala in brain function:*

Basic and clinical approaches (pp. 476–478). New York: New York Academy of Sciences.

Kahneman, D., Knetsch, J.L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, *76*, 728–741.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., & Montague, P.R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.

Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, *53*, 147–156.

Lieberman, M.D. (2007). The X- and C-systems: The neural basis of automatic and controlled social cognition. In E. Harmon-Jones & P. Winkielman (Eds.), *Fundamentals of social neuroscience* (pp. 290–315). New York: Guilford.

Lieberman, M.D. (in press). Why symbolic processing of affect can disrupt negative affect: Social cognitive and affective neuroscience investigations. In A. Todorov, S.T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind*. New York: Oxford University Press.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H., & Way, B.M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity to affective stimuli. *Psychological Science*, *18*, 421–428.

Lieberman, M.D., Hariri, A., Jarcho, J.M., Eisenberger, N.I., & Bookheimer, S.Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, *8*, 720–722.

Lieberman, M.D., Jarcho, J.M., Berman, S., Naliboff, B., Suyenobu, B.Y., Mandelkern, M., & Mayer, E.A. (2004). The neural correlates of placebo effects: A disruption account. *NeuroImage*, *22*, 447–455.

Lieberman, M.D., Jarcho, J.M., & Satpute, A.B. (2004). Evidence-based and intuition-based self knowledge: An fMRI study. *Journal of Personality and Social Psychology*, *87*, 421–435.

Petrovich, G.D., Holland, P.C., & Gallagher, M. (2005). Amygdalar and prefrontal pathways to the lateral hypothalamus are activated by a learned cue that stimulates eating. *Journal of Neuroscience*, *36*, 8295–8302.

Poldrack, R.A. (2004). *SPM_ROI_graph*. Retrieved August 26, 2004, from http://spm-toolbox.sourceforge.net

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., & Kilts, C.D. (2002). Neural basis of social cooperation. *Neuron*, *35*, 395–405.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., & Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*, 1755–1758.

Sears, D.O., & Funk, C.L. (1991). The role of self-interest in social and political attitudes. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 2–91). New York: Academic Press.

Singer, T., Kiebel, S.J., Winston, J.S., Dolan, R.J., & Frith, C.D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *19*, 653–662.

Trepel, C., Fox, C.R., & Poldrack, R.A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Brain Research & Cognitive Brain Research*, *23*, 34–50.

Tyler, T.R. (1984). The role of perceived injustice in defendants' evaluations of their courtroom experience. *Law & Society Review*, *18*, 51–74.

Tyler, T.R. (1991). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology*, *67*, 850–863.

van't Wout, M., Kahn, R.S., Sanfey, A.G., & Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *NeuroReport*, *16*, 1849–1852.

Wager, T.D., Rilling, J.K., Smith, E.E., Sokolik, A., Casey, K.L., Davidson, R.J., et al. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science*, *303*, 1162–1167.

Zalla, T., Koechlin, E., Pietrini, P., Basso, G., Aquino, P., Sirigu, A., & Grafman, J. (2000). Differential amygdala responses to winning and losing: A functional magnetic resonance imaging study in humans. *European Journal of Neuroscience*, *12*, 1764–1770.

(Received 3/21/07; Revision accepted 9/18/07)