Behavioral/Cognitive

Serotonin Modulates Striatal Responses to Fairness and Retaliation in Humans

Molly J. Crockett,[1,2,3] Annemieke Apergis-Schoute,[1,2] Benedikt Herrmann,[4] Matt Lieberman,[5] Ulrich Müller,[1,6] Trevor W. Robbins,[1,2] and Luke Clark[1,2]

[1]Behavioral and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom CB2 1TN, [2]Department of Psychology, University of Cambridge, Cambridge, United Kingdom CB2 1TN, [3]Laboratory for Social and Neural Systems Research, University of Zürich, CH-8006 Zurich, Switzerland, [4]Institute for Public Health and Consumer Protection, Joint Research Centre, European Commission, 21027 Ispra (Varese) Italy, [5]Department of Psychology, University of California Los Angeles, Los Angeles, California 90095, [6]Department of Psychiatry, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom CB2 1TN

Humans are willing to incur personal costs to punish others who violate social norms. Such "costly punishment" is an important force for sustaining human cooperation, but the causal neurobiological determinants of punishment decisions remain unclear. Using a combination of behavioral, pharmacological, and neuroimaging techniques, we show that manipulating the serotonin system in humans alters costly punishment decisions by modulating responses to fairness and retaliation in the striatum. Following dietary depletion of the serotonin precursor tryptophan, participants were more likely to punish those who treated them unfairly, and were slower to accept fair exchanges. Neuroimaging data revealed activations in the ventral and dorsal striatum that were associated with fairness and punishment, respectively. Depletion simultaneously reduced ventral striatal responses to fairness and increased dorsal striatal responses during punishment, an effect that predicted its influence on punishment behavior. Finally, we provide behavioral evidence that serotonin modulates specific retaliation, rather than general norm enforcement: depleted participants were more likely to punish unfair behavior directed toward themselves, but not unfair behavior directed toward others. Our findings demonstrate that serotonin modulates social value processing in the striatum, producing context-dependent effects on social behavior.

Introduction

When deciding how to share resources, humans have a preference for fairness, and some are even willing to incur personal costs to ensure fair outcomes (Camerer, 2003). Such "costly punishment" behavior varies dramatically between individuals and across cultures (Henrich et al., 2006), but the biological basis of this variability remains poorly understood. We recently examined how variation in costly punishment behavior is shaped by serotonin, a neurotransmitter long implicated in social behavior (Kiser et al., 2012). Reducing serotonin levels in humans increased costly punishment (Crockett et al., 2008), while enhancing serotonin function decreased costly punishment (Crockett et al., 2010a).

How might serotonin shape costly punishment decisions? One influential economic model posits that costly punishment is driven by preferences for fair outcomes (defined as equitable wealth distributions; Fehr and Schmidt, 1999). In this model, those who care more about fairness are more likely to punish those who violate fairness norms (Fehr and Fischbacher, 2003; 2004). Fair outcomes activate regions associated with valuation, including the ventral striatum (VS) and medial prefrontal cortex (mPFC) (Tabibnia et al., 2008; Tricomi et al., 2010). Reducing serotonin may therefore increase costly punishment by enhancing the subjective value of fairness and its representation in the VS and mPFC.

Alternative models highlight preferences for reciprocity in driving punishment decisions (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). In these models, individuals gain utility by reducing the payoffs of those who have behaved unfairly toward them (i.e., retaliation). Neuroimaging studies report activation in the dorsal striatum (DS) during retaliation against breaches of trust (de Quervain et al., 2004), unfairness (Sanfey et al., 2003; Strobel et al., 2011), and aggression (Krämer et al., 2007). DS activation is positively correlated with the amount spent to punish, greater during "effective" payoff-reducing punishment than "symbolic" nonpecuniary punishment (de Quervain et al., 2004), and greater when payoff reductions are high than low (Strobel et al., 2011). These patterns suggest that the DS computes the subjective value of reducing the payoffs of norm violators. Thus, reducing serotonin may increase costly punishment by enhancing the subjective value of retaliation and its representation in the DS.

Received June 9, 2012; revised Nov. 23, 2012; accepted Nov. 29, 2012.

Author contributions: M.J.C., B.H., M.L., T.W.R., and L.C. designed research; M.J.C., A.A.-S., and U.M. performed research; M.C. analyzed data; M.J.C., A.A.-S., B.H., M.L., U.M., T.W.R., and L.C. wrote the paper.

This work was supported by a JT McDonnell Network Grant, and was completed within the University of Cambridge Behavioral and Clinical Neuroscience Institute, funded by a joint Award from the Medical Research Council and the Wellcome Trust. We thank the staff at the Wellcome Trust Clinical Research Facility and Wolfson Brain Imaging Centre, M. Franklin, N. Wright, S. Fleming, J. Roiser, I. Krajbich, H. Takahashi, J. Chumbley, C. Ruff, Y. Morishima, and E. Fehr for their assistance.

This article is freely available online through the *J Neurosci* Open Choice option.

Correspondence should be addressed to Molly J. Crockett, Behavioral and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK. E-mail: mollycrockett@gmail.com.

DOI:10.1523/JNEUROSCI.2761-12.2013

 0270-6474/13/333505-09$15.00/0

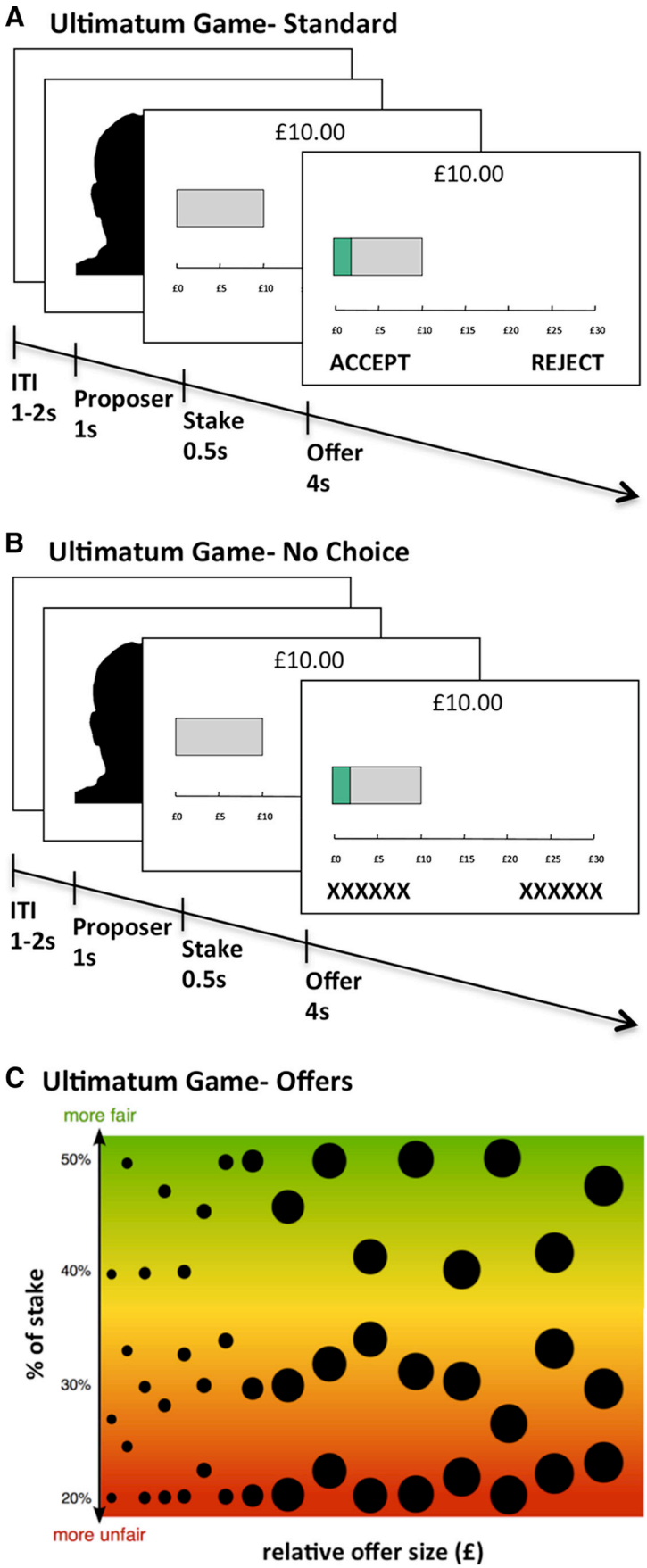

Figure 1. Experimental design. ***A***, In each one-shot ultimatum game, participants viewed a photograph of the Proposer, the amount of the stake, and the offer, and decided whether to accept or reject the offer while the offer was on the screen. ***B***, In the No-Choice condition of the ultimatum game, participants viewed an identical set of offers but their decisions were determined randomly. ***C***, Offers in the ultimatum game. Each bubble represents an offer. The size of the bubble represents its magnitude, and its vertical position corresponds to its fairness. Offer magnitude and offer fairness were not significantly correlated ($r = 0.006$, $p = 0.968$), which permitted us to detect BOLD responses to fairness over and above responses to magnitude.

Note that these models make different predictions about punishment decisions in different social contexts. Whereas retaliation motivates punishment of unfair behavior directed toward oneself (second-party punishment) but not toward others (third-party punishment), fairness preferences motivate punishment of unfair behavior directed both toward oneself and toward others (Fehr and Fischbacher, 2004).

As impaired serotonin function has been linked to reactive aggression (Linnoila et al., 1983; Virkkunen et al., 1994; Higley et al., 1996), we predicted that serotonin regulates retaliatory motives in the context of costly punishment. We tested this hypothesis using a combination of behavioral and neuroimaging methods. If serotonin regulates fairness preferences, serotonin depletion should increase both second- and third-party punishment, and neuroimaging should reveal enhanced fairness-related responses in the mPFC and VS. Conversely, if serotonin regulates retaliatory motives, serotonin depletion should increase only second-party punishment, and neuroimaging should reveal enhanced activity in the DS, specifically during retaliation.

Materials and Methods

Overview. We acquired functional magnetic resonance imaging (fMRI) data while participants decided whether to punish fair and unfair behavior directed toward themselves in a series of one-shot ultimatum games (UGs). In the UG, one player (the proposer) suggests a way to split a sum of money with a second player (the responder). If the responder accepts the offer, both players are paid accordingly. If the responder rejects the offer, neither player is paid. Responders tend to reject offers <20–30% of the total stake, despite the fact that such retaliation is costly (Camerer, 2003). During our UG task, participants decided whether to accept or reject UG offers from human proposers and computer proposers (Fig. 1*A*), and also viewed offers from human proposers in a no-choice condition where subjects were unable to accept or reject (Fig. 1*B*). We included the computer condition as a "nonsocial" comparison condition (Rilling et al., 2002, 2008; Sanfey et al., 2003; Baumgartner et al., 2008),

which enabled us to examine whether acute tryptophan depletion (ATD) affected the neural correlates of social engagement with the UG task. We included the second, no-choice control condition specifically to test the influence of serotonin on neural responses to actively rejecting unfair offers, relative to simply receiving unfair offers.

UG offers ranged from 20–50% of the shared endowment. Importantly, we orthogonalized the material value and the fairness of the offers, which allowed us to parametrically model neural responses to fairness over and above material value (Fig. 1*C*). This was a key aspect of our design, as previous studies have shown that serotonin manipulations can influence behavioral and neural responses to monetary rewards (McCabe et al., 2010; Abler et al., 2012; Seymour et al., 2012). Because fair and unfair offers were matched for material value, this design allowed us to infer that brain regions showing differential responses to fair versus unfair offers were responding to the fairness of the offers and not their material value (Tabibnia et al., 2008).

Outside of the scanner, we assessed participants' willingness to punish fair and unfair behavior directed toward others in a series of one-shot third-party punishment games. In each game, participants had the opportunity to spend a portion of an endowment to reduce the payoff of a proposer who had made a fair or unfair monetary transfer to a "receiver" (see Fig. 5*A*). Proposer transfers in the third-party punishment game ranged from 10 to 50%.

We manipulated serotonin using ATD, a dietary technique that lowers serotonin brain tissue levels (Moja et al., 1989; Stancampiano et al., 1997). Participants completed both tasks twice, once following ATD and once following placebo in a double-blind, counterbalanced crossover design.

Participants. Thirty healthy volunteers (17 females; mean age: 25.1 ± 3.2 years) gave their written informed consent and were financially compensated for participating in this study that was approved by the Cambridgeshire Research Ethics Committee. Exclusion criteria included history of cardiac, hepatic, renal, pulmonary, gastrointestinal, and neurological disorders; medication/recreational drug/tobacco use; and personal/family history of major depression, bipolar affective disorder, or other psychiatric diseases. One participant was excluded from the final analysis due to back-wrapping artifact in the raw imaging data, and another was excluded due to excessive movement in the scanner (>7 mm), leaving 28 participants in the final analysis.

Experimental procedure. Participants attended two experimental sessions, separated by at least 1 week, and were assigned to receive either placebo or ATD on the first session in counterbalanced order. Upon arrival (between 0830 and 1000), participants completed trait questionnaires, gave a blood sample (10 ml), and ingested either the placebo or ATD drink (75 g). After a 5.5 h delay during which participants read or studied in a quiet waiting room, participants gave a second blood sample and completed the UG task in the fMRI scanner.

After exiting the scanner, participants completed the third-party punishment game, and rated on a Likert scale the fairness of six offers representative of those viewed in the scanner. Following this, they completed a reinforced categorization task, the results of which are reported separately (Crockett et al., 2012).

Mood was assessed twice using the Positive and Negative Affect Scale (PANAS; Watson et al., 1988): upon arrival and just before testing. The amino acid drinks used for the ATD procedure were prepared by SHS International, using a standard composition identical to those used in previous studies (Crockett et al., 2008).

At the end of the second session, we asked subjects to guess on which session they received ATD and on which session they received placebo. Group performance was at chance (accuracy mean ± SE: 0.4 ± 0.09). We also asked subjects to rate, on a seven point Likert scale (1 = "not at all," 7 = "completely") the extent to which they believed whether they would be paid for their decisions. These ratings indicated subjects' acceptance of the UG cover story (mean ± SE = 5.39 + 0.35). We note that participants in our study rejected about half of the unfair (20–30%) offers, consistent with the findings of UG experiments that do not use deception (Camerer, 2003).

ATD manipulation check. Blood samples were analyzed for tryptophan as well as the large neutral amino acids (LNAAs) tyrosine, valine, phenylalanine, isoleucine, and leucine. The analysis was performed using HPLC following procedures identical to those described in previous studies from our group (Crockett et al., 2008). ATD resulted in significant reductions in the ratio of tryptophan to other LNAAs (TRP:SigmaLNAA), which is the critical measure for validating the effects of ATD (Booij et al., 2003). A repeated-measures ANOVA revealed a significant two-way interaction between treatment and time ($F_{(1,27)} = 28.605$, $p < 0.0001$), resulting from significant reductions in the TRP:SigmaLNAA ratio 5 h following ATD relative to placebo. Simple effects analyses showed an 85% decrease in the TRP:SigmaLNAA ratio on the ATD session ($t_{(27)} = 12.404$, $p < 0.001$), with no significant change in the TRP:SigmaLNAA ratio on the placebo session ($t_{(27)} = 0.537$, $p = 0.598$).

Consistent with previous studies in healthy volunteers, ATD did not affect subjects' self-reported mood. PANAS scores were analyzed immediately before drink ingestion and immediately before fMRI scanning. A repeated-measures ANOVA with treatment (ATD, placebo) and time point (baseline, +5.5 h) as within-subjects factors found no significant effects of treatment, time point, or their interaction on PANAS-positive affect (all $p > 0.13$) or negative affect (all $p > 0.15$).

Second-party punishment task: ultimatum game. All stimuli were presented using EPrime 1.2. On each trial, participants viewed sequentially a fixation cross (jittered 1–2 s), a photograph of the proposer (1 s), the stake size (0.5 s), and the offer (4 s). While each offer was on the screen, participants pressed a left button to "accept" and a right button to "reject." Offers were divided among three conditions. In the human proposer condition (96 trials), participants responded to offers from human proposers, denoted by a photograph of a person at the start of the trial (Fig. 1*A*). In the computer proposer condition (48 trials), participants responded to offers from computers, denoted by a picture of a computer at the start of the trial. Participants were instructed that in the computer proposer rounds, their decisions would only affect their own payment. In the no-choice condition (48 trials), participants viewed offers from human proposers, denoted by a photograph of the person at the start of the trial, and were presented with the options "xxxxx" and "xxxxx" (Fig. 1*B*). In the no-choice condition, subjects were informed that their decision would be determined by a random device, and were instructed to make a random button press on these trials. Each condition contained an identical set of 48 offers that ranged from 20 of 50% of the stake (Fig. 1*C*). Importantly, we controlled for the material value of the offers such that the same amount could appear as a fair offer (e.g., £5 of 10) or unfair offer (e.g., £5 of 20). Offers were presented in random order across two functional runs (15 min each). Null events (blank screen of duration 6.5–7.5 s) occurred on 30% of trials. Subjects saw the same set of offers in each treatment session. Photographs of proposers were drawn from participants in previous studies and Cambridge residents; subjects understood that the offers were made by participants in previous experiments. Proposer identities were randomly paired with offers across subjects, and subjects saw each photograph only once. Participants were told that one trial would be selected from the experiment, and that they and the proposer would be paid according to their decision on that trial.

Third-party punishment task. In this task, participants observed one-shot dictator games between two other participants (ostensibly volunteers from prior experiments, using the same cover story as in the UG) assigned to the roles of "proposer" and "receiver." On each trial, participants were endowed with £5 and could pay up to 50 p to reduce the proposer's payoff at a 1:3 ratio (i.e., up to £1.50).

According to the third-party punishment instructions, the proposers in this game believed they were playing an ultimatum game; therefore, the proposers knew they could be punished for low offers, and our subjects (the observers) knew this. After the proposers made their offers, we (the experimenters) removed the responder's right to reject the offers; again, our subjects (the observers) knew this. Finally, our subjects (the observers) had the option to pay to reduce the payoffs of the proposers after viewing their offers.

Participants were instructed that one round would be randomly selected and paid out to them, the proposer and the receiver. On each round, participants saw photographs (as in the UG) of the proposer and receiver (2000 ms; Fig. 5*A*) and then saw the proposer's offer and had unlimited time to decide whether to pay to take money away from the

proposer. Participants completed 16 rounds of the third-party punishment game; in these rounds, participants decided whether and how much to spend to reduce the payoffs of proposers that had offered 10, 20, 30, and 50% of the stake (four rounds of each). Compared with the UG, in the third-party punishment game we additionally included extremely inequitable offers (10%) to guard against potential floor effects, because previous studies have shown that people are less willing to engage in third-party than second-party punishment (Fehr and Fischbacher, 2004). The dependent measure was the amount of money paid to reduce the proposer's payoff, as a function of the fairness of the proposer's offer.

Behavioral data analysis. Binary UG choice data (accept/reject) were analyzed using repeated-measures logistic regression implemented with the generalized estimating equations procedure, which generates a χ^2 statistic, 95% confidence interval, and an associated p value. We modeled the within-subjects effects of treatment (ATD, placebo), offer fairness (proportion of stake), and their interaction on rejection decisions. UG reaction time data and third-party punishment choice data (amount spent to punish) were analyzed using repeated-measures ANOVA with treatment (ATD, placebo) and offer fairness as within-subjects factors. For all analyses, gender and treatment order were initially included as between-subjects factors and dropped from subsequent analyses when nonsignificant. Behavioral data analyses were performed using PASW Statistics (v18). On displayed figures, error bars indicate the SE of the difference in means (SED), the appropriate index of variation in within-subject designs.

fMRI data acquisition. A 3 T unit (Tim Trio; Siemens) located at the Wolfson Brain Imaging Centre (Cambridge, UK) was used to collect high-resolution T1-weighted structural images (1 × 1 × 1 mm) for spatial normalization and T2*-weighted echo planar images (32 axial slices, 3 mm thickness; repetition time, 2000 ms; echo time, 30 ms; voxel size, 3 × 3 × 3 mm; field of view, 192 mm).

fMRI preprocessing. All preprocessing and analysis was performed in SPM8 (Wellcome Department of Imaging Neuroscience). Images were realigned to the first scan of the first session and unwarped using field maps; spatially normalized via segmentation of the T1 structural image into gray matter, white matter, and CSF using ICBM tissue probability maps; and spatially smoothed with a Gaussian kernel (8 mm, full-width at half-maximum).

fMRI analysis: fairness. fMRI time series were regressed onto a general linear model containing the following regressors: H1, a stick function denoting a human proposer trial; H2, H1 modulated by offer magnitude; H3, H1 modulated by offer fairness (defined as the proportion of the stake); C1, a stick function denoting a computer proposer trial; C2, C1 modulated by offer magnitude; C3, C1 modulated by offer fairness; N1, a stick function denoting a no-choice trial; N2, N1 modulated by offer magnitude; and N3, N1 modulated by offer fairness. We orthogonalized offer fairness with respect to offer magnitude to identify the independent contribution of fairness to blood oxygenation level-dependent (BOLD) signal after accounting for activity related to offer magnitude. Each regressor was convolved with the canonical hemodynamic response function and its temporal derivative. For all models described, data from ATD and placebo sessions were modeled separately at the first level, and treatment effects were computed at the second level (random-effects analysis) using paired t tests.

fMRI analysis: retaliation. To test the effects of ATD on neural activation associated with the rejection of unfair offers, we created a model with the following regressors: HUA, accepted unfair offer from a human proposer; HUR, rejected unfair offer from a human proposer; HFA, accepted fair offer from a human proposer; CU, unfair offer from a computer proposer; CF, fair offer from a computer proposer; NUL, left button press on an unfair no-choice trial; NUR; right button press on an unfair no-choice trial; and NF, fair offer on a no-choice trial. All regressors were modeled as stick functions and convolved with the canonical hemodynamic response function and its temporal derivative. To maximize the number of trials available for analysis, we defined "unfair" offers as <45% and "fair" offers as 45–50%. We were unable to separately model rejected and accepted unfair offers from computer proposers because a large subset of our participants never rejected offers from computer proposers. We were unable to estimate this model for three subjects due to their choices in the task. For the brain-behavior correlation, we extracted the mean parameter estimate for each subject from a 4 mm radius sphere centered on the peak coordinates of the contrast [HUR_ATD > HUR_PLA] and regressed those values against each subject's change in rejection rate from placebo to ATD.

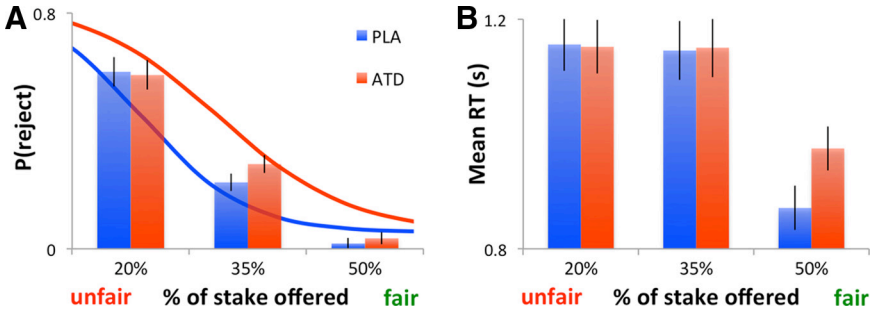

Figure 2. Effects of treatment on ultimatum game behavior (second-party punishment). ***A***, Choice data and estimated logistic model showing that the proportion of offers rejected decreased as offer fairness increased. ATD increased rejection of offers, relative to placebo (PLA), predominantly at intermediate levels of fairness. ***B***, Response times (RTs) in the UG were fastest for fair offers but did not differ significantly between medium and unfair offers. ATD selectively slowed responses to fair offers, relative to PLA. Error bars indicate SED.

fMRI analysis: correction for multiple comparisons. We report as significant only results surviving small-volume correction for multiple comparisons (cluster-level corrected after voxelwise thresholding at $p < 0.005$, $k = 10$). For small-volume correction, anatomical masks based on a priori regions of interest (ROI) were constructed using the Automated Anatomical Labeling anatomical atlas for mPFC (Tzourio-Mazoyer et al., 2002) and an anatomical parcellation of the striatum, which distinguishes ventral and dorsal subdivisions (Martinez et al., 2003). Masking of contrasts was performed using the PickAtlas tool in SPM8 (Maldjian et al., 2003). Small-volume correction was applied based on the number of voxels in the ROI masks. For display purposes, parameter estimates from significant clusters were extracted from 4 mm radius spheres centered on the peak coordinates of the relevant contrast. Some of the results that survived small-volume correction were strong enough to also survive whole-brain correction; we therefore report whole-brain corrected p values for those results.

Results

Behavior: ATD and second-party punishment

First, we examined the behavioral effects of ATD on second-party punishment of human proposers in the UG. In line with previous research, participants were significantly more likely to reject unfair than fair offers (main effect of fairness, $\chi^2_{(1,27)} = 93.539$, $p < 0.001$). The effects of ATD interacted significantly with offer fairness ($\chi^2_{(1,27)} = 6.154$, $p = 0.013$). Consistent with our previous findings (Crockett et al., 2008, 2010a), ATD increased rejection rates relative to placebo, particularly for moderately unfair offers (Fig. 2*A*). The order of treatments (whether subjects received ATD first or placebo first) did not affect the results (all $p > 0.495$).

In addition to increasing rejection rates, ATD altered reaction times. On placebo, participants were fastest to accept equal splits (main effect of fairness, $F_{(3,27)} = 27.563$, $p < 0.001$). ATD slowed reaction times, specifically for the equal splits (treatment-by-fairness interaction, $F_{(3,27)} = 3.461$, $p = 0.039$; Fig. 2*B*). The behavioral effects of ATD did not seem to be driven by changes in

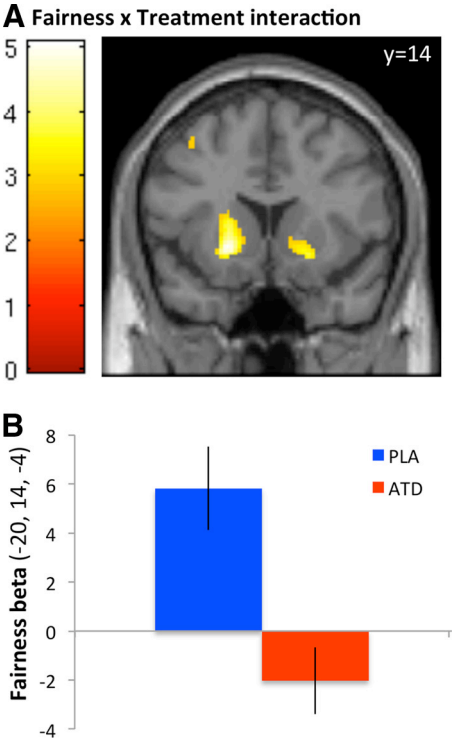

Figure 3. Effects of fairness and treatment in the VS. ***A***, BOLD responses in the bilateral VS showed an interaction between offer fairness and treatment. The image is displayed at $p < 0.005$, uncorrected. ***B***, Fairness parameter estimates for the ATD and placebo (PLA) sessions indicate that the strong positive correlation between VS BOLD response and offer fairness observed on the PLA session was reduced by ATD. Error bars indicate SED.

perceptions of fairness, however; ATD did not significantly influence fairness ratings for a representative sample of offers, collected using a postscanning questionnaire (treatment: $F_{(3,27)} = 2.194$, $p = 0.15$; treatment-by-fairness: $F_{(3,27)} = 0.855$, $p = 0.431$). Thus, the data suggest that ATD influenced preferences about social outcomes, given a set of perceptions about what is fair.

Neuroimaging: ATD and fairness

Note that both retaliatory motives and fairness preferences could drive rejection decisions in the UG. To investigate the motivational processes mediating the effects of ATD on rejection in the UG, we first turned to the fMRI data. To identify brain regions whose response to offer fairness in the UG differed across treatments, offer size and offer fairness on each trial were entered as parametric regressors in a model fitted to the presentation of the offers. From this analysis, we identified regions in which BOLD signal correlated with the fairness of offers from human proposers ($p < 0.05$, cluster level familywise error corrected after voxelwise thresholding at $p < 0.005$, $k = 10$). On the placebo session, fair offers (relative to unfair offers) were associated with activation in the VS ($p = 0.018$, small volume corrected for VS) and mPFC ($p = 0.023$, small volume corrected for mPFC), consistent with previous studies (Tabibnia et al., 2008; Tricomi et al., 2010; Zaki and Mitchell, 2011). Our primary goal was to identify regions that responded differently to offer fairness on ATD versus placebo. We observed a significant interaction between fairness and treatment in the VS ($p = 0.045$, small volume corrected for VS; Fig. 3*A*).

This finding is noteworthy because several previous studies have implicated the VS in representing fairness preferences. In particular, the VS responds to fairness over and above material value (Tabibnia et al., 2008) and shows a pattern of activity consistent with fairness preferences (Tricomi et al., 2010). Thus, if ATD increased costly punishment by enhancing concerns for fairness, we should see stronger VS responses to fairness on ATD, relative to placebo. To address this question, we examined the fairness parameter estimates extracted from the peak activation in the VS, separately for the ATD and placebo sessions. Contrary to our hypothesis, ATD actually reduced VS responses to fairness, relative to placebo (Fig. 3*B*). At a less stringent threshold ($p < 0.001$, uncorrected), the mPFC and midbrain showed a similar pattern to the VS. These findings provide clear evidence against the hypothesis that ATD increased costly punishment by enhancing concerns for fairness.

Previous studies have shown that ATD alters aspects of social perception and appraisal (Williams et al., 2007; Bilderbeck et al., 2011). Thus, ATD may have reduced VS responses to fairness simply by reducing social engagement with the UG task. To address this possibility, we contrasted offers from human proposers with offers from computer proposers. On the placebo session, offers from human proposers (relative to those from computer proposers) were associated with greater activation in several regions associated with motivation, including a cluster encompassing the amygdala and striatum ($p < 0.001$, whole-brain corrected) as well as the mPFC ($p = 0.001$, whole-brain corrected). Importantly, however, ATD did not significantly affect the differential response to human versus computer offers in any of these regions, suggesting that subjects were equally socially engaged with the task on the ATD and placebo sessions.

Neuroimaging: ATD and retaliation

An alternative explanation for the behavioral effects of ATD on costly punishment is that ATD enhanced the subjective value of retaliation. If this is the case, fMRI should reveal stronger responses in reward circuitry during retaliation following ATD, relative to placebo. Specifically, when subjects reject unfair offers in the UG, on ATD we might expect to see enhanced activity in the DS, which has been associated with retaliation in prior studies (de Quervain et al., 2004; Krämer et al., 2007; Strobel et al., 2011).

We tested this hypothesis in a second model that captured the effects of ATD on neural activity during the rejection of unfair offers from human proposers. Our findings supported our prediction: relative to placebo, ATD increased activity in bilateral DS during rejection of unfair offers ($p = 0.003$, whole-brain corrected; Fig. 4*A*).

Next, we tested whether the DS activity enhanced by ATD was associated specifically with the rejection of unfair offers, rather than unfairness per se, by contrasting unfair offers where subjects chose to reject with unfair offers in the no-choice condition. ATD increased DS activity during rejection of unfair offers, relative to unfair offers in the no-choice condition ($p = 0.048$, small volume corrected for DS), demonstrating that the signal in DS enhanced by ATD was specific to costly punishment, rather than unfairness per se.

Finally, we examined whether increases in DS activity during rejection of unfair offers on ATD (relative to placebo) were correlated, across subjects, with increases in rejection behavior on ATD (relative to placebo). Indeed, subjects showing the greatest increases in right DS activity during rejection on ATD were those that also showed the greatest increases in rejection rates on ATD ($r = 0.42$, $p = 0.036$; Fig. 4*B*).

As a robustness check, we conducted an additional analysis to test whether our results support the view that the DS motivates

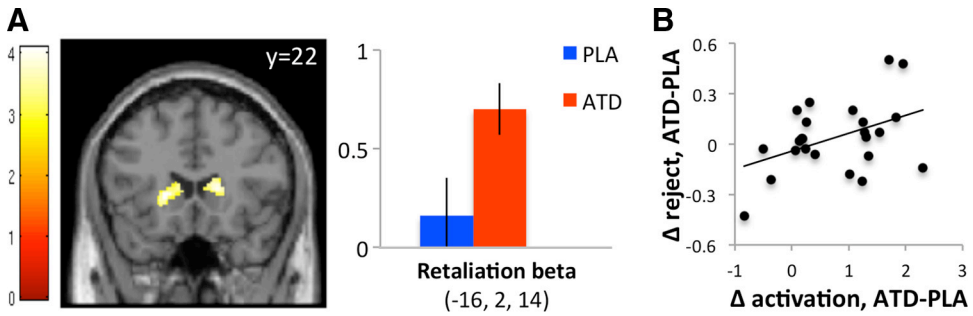

Figure 4. Effects of treatment on DS responses during retaliation. ***A***, In the DS, ATD increased BOLD responses when subjects rejected unfair offers from human proposers. The image is displayed at $p < 0.005$, uncorrected. Error bars indicate SED. ***B***, ATD-induced increases in DS activity during rejection of unfair offers were positively correlated with ATD-induced increases in rejection rates across subjects ($r = 0.42$, $p = 0.036$). PLA, placebo.

costly punishment under baseline (placebo) conditions (de Quervain et al., 2004; Strobel et al., 2011). We regressed participants' rejection rates on the placebo session onto the contrast [Unfair Reject > Unfair No choice] on the placebo session, and observed a cluster in the DS ($p = 0.031$, small volume corrected for DS), in line with our prediction.

Behavior: ATD and third-party punishment

Together, the neuroimaging findings suggest that ATD increased second-party punishment by enhancing retaliatory motives, while at the same time reducing (rather than enhancing) fairness preferences. This set of findings leads to the perhaps counterintuitive hypothesis that serotonin may have different effects on costly punishment in different contexts. Specifically, if impairing serotonin function increases costly punishment by strengthening fairness preferences, following ATD subjects should be more likely to punish unfair behavior directed toward others as well as themselves. Conversely, if impairing serotonin function increases costly punishment by enhancing retaliatory motives, following ATD subjects should be more likely to punish unfair behavior directed only toward themselves. As a final test, we therefore examined the effects of ATD on third-party punishment behavior (Fig. 5*A*).

If ATD increased costly punishment in the UG by enhancing fairness preferences, then we should also observe increased third-party punishment following the ATD treatment, relative to placebo. In fact, we observed a trend in the opposite direction. Participants paid more to punish proposers as their offers became increasingly unfair (main effect of fairness, $F_{(1,27)} = 58.555$, $p < 0.001$), but ATD tended to decrease third-party punishment of unfair, but not fair behavior (fairness-by-treatment interaction, $F_{(1,27)} = 2.709$, $p = 0.050$; Fig. 5*B*). We did not observe any effects of ATD on response times in third-party punishment (treatment: $F_{(1,27)} = 0.450$, $p = 0.508$; treatment × fairness: $F_{(3,27)} = 0.746$, $p = 0.528$).

Discussion

Our findings provide a mechanistic account of how serotonin shapes costly punishment behavior. Supporting our hypothesis, we found neural and behavioral evidence indicating that serotonin regulates retaliatory motives in costly punishment. ATD selectively increased retaliation against unfair behavior directed toward oneself, and enhanced activity in the DS during retaliation. The DS has consistently been implicated in instrumental reward anticipation (O'Doherty, 2004; Tricomi et al., 2004), raising the possibility that ATD may have increased the expected satisfaction resulting from costly punishment. In addition, the DS is involved in avoiding aversive outcomes (Delgado et al., 2008, 2009), which could indicate that ATD enhanced the motivational drive to avoid unfair outcomes. The observed effects of ATD on DS activity during rejection of unfair offers are unlikely to simply reflect changes in reaction time, as ATD did not affect reaction times for unfair offers (Fig. 2*B*).

We observed individual differences in the size of the effect of ATD on costly punishment behavior. Previous studies have shown that the behavioral effects of ATD are moderated by individual differences in genetic polymorphisms (Roiser et al., 2006) or behavioral traits such as aggression (Bjork et al., 2000). In the current study, individual differences in the behavioral effects of ATD were predicted by individual differences in the neural effects of ATD. Participants showing the strongest behavioral effect of ATD on costly punishment also showed the strongest neural effect of ATD on DS activity during punishment. Our data thus dovetail with previous studies implicating the DS in costly punishment (de Quervain et al., 2004; Strobel et al., 2011) and extend them by supporting a causal role for the DS in retaliatory motives. We previously reported data suggesting that enhancing serotonin function reduced costly punishment by increasing aversion to harming others (Crockett et al., 2010a); consistent with this interpretation, the current findings suggest that impairing serotonin function may reduce aversion to harming interaction partners, to the extent that it may even be pleasurable in certain contexts.

Our results also point toward a role for serotonin in enhancing fairness preferences. Reducing central serotonin levels blunted responses in the VS to fairness and slowed response times for accepting fair offers. Previous studies have implicated the VS in processing the subjective value of fair and cooperative social exchanges (Tricomi et al., 2010; Rilling and Sanfey, 2011). We note that if ATD had increased costly punishment by enhancing the salience of fairness preferences, we might have observed increased, rather than decreased VS responses to fairness following ATD. Instead, our results suggest that ATD in fact reduced the subjective value of fairness. This explanation is consistent with our behavioral finding that ATD actually reduced third-party costly punishment, in which fairness preferences play a decisive role (Fehr and Fischbacher, 2004). One limitation of the current study is that the act of punishing is not directly comparable across second-(UG) and third-party conditions, both in terms of how it is accomplished and its cost. However, we note that second- and third-party punishment are rarely equivalent outside of the laboratory, and despite the differences between the tasks, the pattern

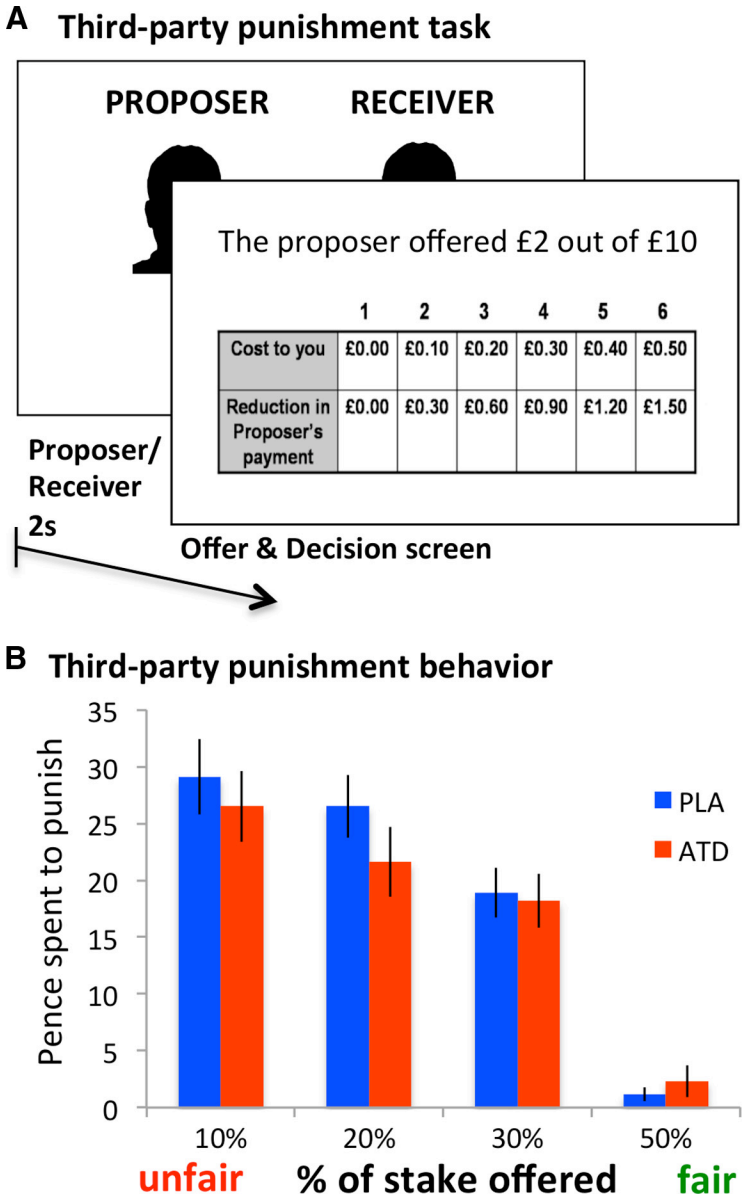

Figure 5. Effects of treatment on third-party punishment behavior. ***A***, In each third-party punishment game, participants viewed a photograph of the Proposer and Receiver, and the Proposer's offer to the Receiver, and decided whether to spend a portion of an endowment to reduce the payoff of the Proposer. ***B***, Choice data from third-party punishment, showing that the amount spent on punishment decreased as offer fairness increased. ATD tended to decrease third-party punishment of unfair, but not fair behavior. Error bars indicate SED. PLA, placebo.

of results we observed argues against a unified punishment motive.

In addition to reducing VS responses to fairness, ATD also influenced decision times for fair offers. On the placebo session, participants were fastest to accept fair offers, similar to a previous study in children (Blake et al., 2011) and recent work showing that cooperative decisions are faster than selfish ones (Rand et al., 2012). In the ATD session, participants were significantly slower to accept fair offers, which could reflect a reduced motivation to engage in cooperative social exchange. This interpretation is consistent with the observation that ATD reduces cooperative decisions in a repeated prisoner's dilemma (Wood et al., 2006), while enhancing serotonin function has the opposite effect (Tse and Bond, 2002). An obvious next step would be to test whether enhancing serotonin function promotes positive reciprocity by boosting VS responses to mutual cooperation (Rilling and Sanfey, 2011).

Previous studies have reported that changes in subjective mood can influence costly punishment behavior (Harlé and Sanfey, 2007; Harlé et al., 2012). To rule out this possibility, we collected self-reported measures of positive and negative affect. Consistent with previous ATD studies, we did not observe significant effects of ATD on subjective ratings of positive or negative affect. Thus, the effects described here are unlikely to be due to ATD-induced changes in subjective mood.

Importantly, our findings suggest some form of motivational opponency between the VS and DS that is modulated by the level of serotonergic activity, with lower levels biasing behavioral control toward the DS. Whether this reflects a more general motivational opponency that also operates in nonsocial circumstances to guide instrumental behavior under conflict is an interesting possibility. Serotonin normally provides inhibitory tone over the striatum (Di Cara et al., 2001), and promotes behavioral inhibition (Crockett et al., 2009, 2012; Boureau and Dayan, 2011). Notably, serotonin activity is diminished when animals can exert control over their environment (Amat et al., 2005). This finely tuned regulatory architecture could adaptively promote instrumental actions in appropriate contexts, and inhibit them when they are likely to be futile. Such contextual sensitivity is especially important for reciprocal social interactions, where a single inappropriate action can have disastrous reputational consequences.

Escalated aggression is one example of inappropriate social behavior, and is associated with impaired serotonin function in humans and primates (Higley et al., 1996; Krakowski, 2003). In primates, low serotonergic activity is implicated specifically in severe, unrestrained aggression that often results in wounding or death, rather than controlled, competitive aggression used to maintain social status (Higley et al., 1996), paralleling studies in humans showing that low serotonin function is associated with impulsive violence, rather than premeditated violence (Linnoila et al., 1983; Virkkunen et al., 1994). The current findings advance our understanding of the role of serotonin in aggressive behavior by shedding light on how impaired serotonin function alters the neural circuitry of aggressive motivation. For individuals with compromised serotonin function, the appetitive drive for retaliation may carry stronger affective weight than the long-term benefits of controlling retaliatory impulses.

Costly punishment behavior is often described as the product of an intentional desire to enforce fairness norms (Knoch et al., 2006; Baumgartner et al., 2011), but the observation that costly

punishment can promote fair behavior in the group (Fehr and Fischbacher, 2003) does not necessarily imply that all punishment is motivated by fairness preferences (Herrmann et al., 2008; Houser and Xiao, 2010; Dreber and Rand, 2012). Indeed, recent findings from public goods games suggest that there are at least two distinct types of costly punishment: moralistic, fairness-based punishment, which is negatively correlated with impulsive choice and competitive, spiteful punishment, which is positively correlated with impulsive choice (Espín et al., 2012). This work is consistent with our previous study showing that impulsive choice and costly punishment in the UG are positively correlated, and increase in tandem with ATD (Crockett et al., 2010b). Collectively, these findings connect serotonin's role in promoting behavioral inhibition with its involvement in regulating retaliation. In social contexts, impaired behavioral inhibition may manifest as a lowered threshold for reactive aggression.

In sum, our findings provide behavioral and neurobiological evidence for multiple motives driving costly punishment: if all punishment were motivated by fairness preferences, then we would have observed similar effects of ATD on second- and third-party punishment. Instead, our neuroimaging data implies that impairing serotonin function enhanced the drive for retaliation while simultaneously reducing fairness preferences. Serotonin may therefore facilitate harmonious social interactions and promote cooperative social exchange by modulating the computation of social value.

References

Abler B, Grön G, Hartmann A, Metzger C, Walter M (2012) Modulation of frontostriatal interaction aligns with reduced primary reward processing under serotonergic drugs. J Neurosci 32:1329–1335. CrossRef Medline

Amat J, Baratta MV, Paul E, Bland ST, Watkins LR, Maier SF (2005) Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. Nat Neurosci 8:365–371. CrossRef Medline

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 58:639–650. CrossRef Medline

Baumgartner T, Knoch D, Hotz P, Eisenegger C, Fehr E (2011) Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nat Neurosci 14:1468–1474. CrossRef Medline

Bilderbeck AC, McCabe C, Wakeley J, McGlone F, Harris T, Cowen PJ, Rogers RD (2011) Serotonergic activity influences the cognitive appraisal of close intimate relationships in healthy adults. Biol Psychiatry 69:720–725. CrossRef Medline

Bjork JM, Dougherty DM, Moeller FG, Swann AC (2000) Differential behavioral effects of plasma tryptophan depletion and loading in aggressive and nonaggressive men. Neuropsychopharmacology 22:357–369. CrossRef Medline

Blake PR, McAuliffe K (2011) "I had so much it didn't seem fair": eight-year-olds reject two forms of inequity. Cognition 120:215–224. CrossRef Medline

Booij L, Van der Does AJ, Riedel WJ (2003) Monoamine depletion in psychiatric and healthy populations: review. Mol Psychiatry 8:951–973. CrossRef Medline

Boureau YL, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74–97. CrossRef Medline

Camerer CF (2003) Behavioral game theory: experiments in strategic interaction. Princeton UP, Princeton, NJ.

Crockett MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. J Neurosci 29:11993–11999. CrossRef Medline

Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW (2008) Serotonin modulates behavioral reactions to unfairness. Science 320:1739. CrossRef Medline

Crockett MJ, Clark L, Hauser MD, Robbins TW (2010a) Serotonin selectively influences moral judgment and behavior through effects on harm aversion. Proc Natl Acad Sci U S A 107:17433–17438. CrossRef Medline

Crockett MJ, Clark L, Lieberman MD, Tabibnia G, Robbins TW (2010b) Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. Emotion 10:855–862. CrossRef Medline

Crockett MJ, Clark L, Apergis-Schoute AM, Morein-Zamir S, Robbins TW (2012) Serotonin modulates the effects of pavlovian aversive predictions on response vigor. Neuropsychopharmacology 37:2244–2252. CrossRef Medline

Delgado MR, Jou RL, Ledoux JE, Phelps EA (2009) Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. Front Behav Neurosci 3:33. CrossRef Medline

Delgado MR, Li J, Schiller D, Phelps EA (2008) The role of the striatum in aversive learning and aversive prediction errors. Philos Trans R Soc Lond B Biol Sci 363:3787–3800. CrossRef Medline

de Quervain DJ, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305:1254–1258. CrossRef Medline

Di Cara B, Dusticier N, Forni C, Lievens JC, Daszuta A (2001) Serotonin depletion produces long lasting increase in striatal glutamatergic transmission. J Neurochem 78:240–248. CrossRef Medline

Dreber A, Rand DG (2012) Retaliation and antisocial punishment are overlooked in many theoretical models as well as behavioral experiments. Behav Brain Sci 35:24. CrossRef Medline

Dufwenberg M, Kirchsteiger G (2004) A theory of sequential reciprocity. Games Econ Behav 47:268–298. CrossRef

Espín AM, Brañas-Garza P, Herrmann B, Gamella JF (2012) Patient and impatient punishers of free-riders. Proc Biol Sci 279:4923–4928. CrossRef Medline

Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425:785–791. CrossRef Medline

Fehr E, Fischbacher U (2004) Third-party punishment and social norms. Evol Hum Behav 25:63–87. CrossRef

Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. Q J Econ 114:817–868. CrossRef

Harlé KM, Sanfey AG (2007) Incidental sadness biases social economic decisions in the Ultimatum Game. Emotion 7:876–881. CrossRef Medline

Harlé KM, Chang LJ, van't Wout M, Sanfey AG (2012) The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. Neuroimage 61:32–40. CrossRef Medline

Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC, Gurven M, Gwako E, Henrich N, Lesorogol C, Marlowe F, Tracer D, Ziker J (2006) Costly punishment across human societies. Science 312:1767–1770. CrossRef Medline

Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. Science 319:1362–1367. CrossRef Medline

Higley JD, King ST Jr, Hasert MF, Champoux M, Suomi SJ, Linnoila M (1996) Stability of interindividual differences in serotonin function and its relationship to severe aggression and competent social behavior in rhesus macaque females. Neuropsychopharmacology 14:67–76. CrossRef Medline

Houser D, Xiao E (2010) Inequality-seeking punishment. Econ Lett 109:20–23. CrossRef

Kiser D, Steemers B, Branchi I, Homberg JR (2012) The reciprocal interaction between serotonin and social behaviour. Neurosci Biobehav Rev 36:786–798. CrossRef Medline

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. CrossRef Medline

Krakowski M (2003) Violence and serotonin: influence of impulse control, affect regulation, and social functioning. J Neuropsychiatry Clin Neurosci 15:294–305. CrossRef Medline

Krämer UM, Jansma H, Tempelmann C, Münte TF (2007) Tit-for-tat: the neural basis of reactive aggression. Neuroimage 38:203–211. CrossRef Medline

Linnoila M, Virkkunen M, Scheinin M, Nuutila A, Rimon R, Goodwin FK (1983) Low cerebrospinal fluid 5-hydroxyindoleacetic acid concentration differentiates impulsive from nonimpulsive violent behavior. Life Sci 33:2609–2614. CrossRef Medline

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239. CrossRef Medline

Martinez D, Slifstein M, Broft A, Mawlawi O, Hwang DR, Huang Y, Cooper

T, Kegeles L, Zarahn E, Abi-Dargham A, Haber SN, Laruelle M (2003) Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: amphetamine-induced dopamine release in the functional subdivisions of the striatum. J Cereb Blood Flow Metab 23: 285–300. Medline

McCabe C, Mishor Z, Cowen PJ, Harmer CJ (2010) Diminished neural processing of aversive and rewarding stimuli during selective serotonin reuptake inhibitor treatment. Biol Psychiatry 67:439–445. CrossRef Medline

Moja EA, Cipolla P, Castoldi D, Tofanetti O (1989) Dose-response decrease in plasma tryptophan and in brain tryptophan and serotonin after tryptophan-free amino acid mixtures in rats. Life Sci 44:971–976. CrossRef Medline

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454. CrossRef Medline

Rabin M (1993) Incorporating fairness into game theory and economics. Am Econ Rev 83:1281–1302.

Rand DG, Greene JD, Nowak MA (2012) Spontaneous giving and calculated greed. Nature 489:427–430. CrossRef Medline

Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. Annu Rev Psychol 62:23–48. CrossRef Medline

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) A neural basis for social cooperation. Neuron 35:395–405. CrossRef Medline

Rilling JK, Goldsmith DR, Glenn AL, Jairam MR, Elfenbein HA, Dagenais JE, Murdock CD, Pagnoni G (2008) The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia 46:1256–1266. CrossRef Medline

Roiser JP, Blackwell AD, Cools R, Clark L, Rubinsztein DC, Robbins TW, Sahakian BJ (2006) Serotonin transporter polymorphism mediates vulnerability to loss of incentive motivation following acute tryptophan depletion. Neuropsychopharmacology 31:2264–2272. Medline

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. CrossRef Medline

Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. J Neurosci 32:5833–5842. CrossRef Medline

Stancampiano R, Melis F, Sarais L, Cocco S, Cugusi C, Fadda F (1997) Acute administration of a tryptophan-free amino acid mixture decreases 5-HT release in rat hippocampus in vivo. Am J Physiol 272:R991–R994. Medline

Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, Kirsch P (2011) Beyond revenge: neural and genetic bases of altruistic punishment. Neuroimage 54:671–680. CrossRef Medline

Tabibnia G, Satpute AB, Lieberman MD (2008) The sunny side of fairness. Psychol Sci 19:339–347. CrossRef Medline

Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. Neuron 41:281–292. CrossRef Medline

Tricomi E, Rangel A, Camerer CF, O'Doherty JP (2010) Neural evidence for inequality-averse social preferences. Nature 463:1089–1091. CrossRef Medline

Tse WS, Bond AJ (2002) Serotonergic intervention affects both social dominance and affiliative behaviour. Psychopharmacology 161:324–330. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289. CrossRef Medline

Virkkunen M, Rawlings R, Tokola R, Poland RE, Guidotti A, Nemeroff C, Bissette G, Kalogeras K, Karonen SL, Linnoila M (1994) CSF biochemistries, glucose metabolism, and diurnal activity rhythms in alcoholic, violent offenders, fire setters, and healthy volunteers. Arch Gen Psychiatry 51:20–27. CrossRef Medline

Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 54:1063–1070. CrossRef Medline

Williams JH, Perrett DI, Waiter GD, Pechey S (2007) Differential effects of tryptophan depletion on emotion processing according to face direction. Soc Cogn Affect Neurosci 2:264–273. CrossRef Medline

Wood RM, Rilling JK, Sanfey AG, Bhagwagar Z, Rogers RD (2006) Effects of tryptophan depletion on the performance of an iterated prisoner's dilemma game in healthy adults. Neuropsychopharmacology 31:1075–1084. CrossRef Medline

Zaki J, Mitchell JP (2011) Equitable decision making is associated with neural markers of intrinsic value. Proc Natl Acad Sci U S A 108:19761–19766. CrossRef Medline