UC Riverside
UC Riverside Previously Published Works

Title
Trusting outgroup, but not ingroup members, requires control: neural and behavioral evidence

Permalink
https://escholarship.org/uc/item/1fh6r2f7

Journal
Social Cognitive and Affective Neuroscience, 12(3)

ISSN
1749-5016

Authors
Hughes, Brent L
Ambady, Nalini
Zaki, Jamil

Publication Date
2017-03-01

DOI
10.1093/scan/nsw139

Peer reviewed

Trusting outgroup, but not ingroup members, requires control: neural and behavioral evidence

Brent L. Hughes[1,2], Nalini Ambady[1,†], and Jamil Zaki[1]

[1]Stanford University, Stanford, CA, USA and [2]University of California, Riverside, CA, USA

[†]In memoriam.
Correspondence should be addressed to Brent L. Hughes, Department of Psychology, University of California, 900 University Ave, Riverside, CA 92521, USA.
E-mail: brent.hughes@ucr.edu

Abstract

Trust and cooperation often break down across group boundaries, contributing to pernicious consequences, from polarized political structures to intractable conflict. As such, addressing such conflicts require first understanding why trust is reduced in intergroup settings. Here, we clarify the structure of intergroup trust using neuroscientific and behavioral methods. We found that trusting ingroup members produced activity in brain areas associated with reward, whereas trusting outgroup members produced activity in areas associated with top-down control. Behaviorally, time pressure— which reduces people's ability to exert control—reduced individuals' trust in outgroup, but not ingroup members. These data suggest that the exertion of control can help recover trust in intergroup settings, offering potential avenues for reducing intergroup failures in trust and the consequences of these failures.

Key words: intergroup dynamics, prosociality, reward, top-down control, neuroeconomics

Introduction

Global communication, diplomacy and trade increasingly require individuals to cooperate with members of other ethnic, social and cultural groups (Friedman, 2006), but this need often clashes with longstanding parochial preferences (Choi and Bowles, 2007; Balliet *et al.*, 2014). Individuals tend to favor members of their own groups over outsiders and cooperate less with outgroup, when compared with ingroup members (Levine *et al.*, 2005), propensities that likely reflect evolutionary pressures of social living (Kurzban and Leary, 2001). Such ingroup favoritism, although in some ways evolutionarily adaptive, often limits people's ability to meet modern demands and form coalitions across groups. Failures in intergroup relations, in turn, produce a number of pernicious consequences, from increasingly polarized political structures to intractable conflict (Ross and Ward, 1995).

Ingroup favoritism expresses itself through a number of psychological channels, including favorable beliefs about ingroup, when compared with outgroup members (Tajfel, 1982),

dehumanization of outgroup members (Haslam and Loughnan, 2014) and distrust of outgroup members (Brewer, 1999; Stanley et al., 2011; Balliet *et al.*, 2014). Here we focus on the last of these factors. Interpersonal trust is key to successful cooperation and—at a larger scale—economic growth (Zak and Knack, 2001), and intergroup breakdowns in trust foment conflict (Ross and Ward, 1995). As such, understanding intergroup distrust is critical to minimizing such conflicts.

Here we adjudicate between two predictions about intergroup trust and the psychological mechanisms of trust more broadly. On one hand, trust might be driven by a *single mechanism*. One candidate mechanism is subjective value. Individuals could find trust valuable for at least two reasons: (i) it constitutes an affiliative act that helps them identify with a group (Baumeister and Leary, 1995) and (ii) it often produces cooperative outcomes that benefit all individuals (Zaki and Mitchell, 2013). If trust is driven by value calculations, then intergroup failures in trust might reflect a lack of subjective value associated with trusting outgroup members. Under this model,

intergroup effects on trust would reflect a difference in the *degree* to which trust is valued across group boundaries, which in turn produces behavioral discrepancies in trust.

On the other hand, trust might be driven by *multiple mechanisms*. For instance, even when trust is not experienced as subjectively valuable, individuals might exert *top-down control* to trust others when in their best interest. Under such a dual-process model, intergroup trust might reflect fundamentally different *kinds* of processes. In particular, people might trust ingroup members to the extent that it produces subjective value, but uniquely require top-down control to trust outgroup members (Hughes and Zaki, 2015). If this is the case, intergroup failures in trust could reflect individuals' failure to engage such deliberative control.

Research across social and biological science provides useful neural and behavioral 'markers' of subjective value and top-down control. First, dissociable brain systems reliably track the experience of value and control. A large body of work demonstrates that activity in the mesolimbic dopaminergic system—including the striatum and ventromedial prefrontal cortex (vMPFC)—scales linearly with the subjective value assigned to a wide variety of stimuli (Bartra *et al.*, 2013; Ruff and Fehr, 2014), including decisions to trust others (Rilling et al., 2002; Fareri *et al.*, 2015). In contrast, a prefrontal cortical network—including the dorsal anterior cingulate (dACC) and lateral prefrontal cortex (LPFC)—commonly responds to tasks that require top-down control, such as conflict monitoring, reappraisal and error correction (Badre, 2008; Shackman *et al.*, 2011; Shenhav *et al.*, 2013). Activity in dACC and LPFC further tracks cognitive effort and choice difficulty (Shenhav *et al.*, 2014).

The dissociation between value and control also characterizes value-based decision-making. As an example, consider intertemporal choice. Single mechanism accounts of this phenomenon suggest that the value individuals assign to immediate versus delayed rewards is represented through a single, integrated valuation mechanism, marked by activity in the striatum (Kable and Glimcher, 2007). In contrast, dual-process accounts suggests that whereas immediate or 'hot' rewards are represented through value computations in striatum, activation of regions involved in executive control, including dACC and LPFC, are engaged to promote delayed or 'cool' rewards (McClure *et al.*, 2004; Bhanji and Beer, 2012), including through the modulation of value-responsive regions (Hare *et al.*, 2009).

Here, we leveraged these neural markers of subjective value and control to adjudicate between models of intergroup trust. If trust reflects a single value-computation process (e.g. subjective value, Kable and Glimcher, 2007), then trusting outgroup, when compared with ingroup members, should produce less activity in mesolimbic structures, and this discrepancy should track individuals' unwillingness to trust across group boundaries. Alternatively, under a dual-process model (e.g. McClure *et al.*, 2004; Hare *et al.*, 2009), trusting outgroup, when compared with ingroup members should uniquely produce activity in structures related to control, including dACC and LPFC. Further, this control-related activity should (i) track individuals' ability to overcome affective biases and trust outgroup members and (ii) modulate value signals in the striatum during outgroup trust decisions.

A second marker of control often used to test dual-process models is reaction time. Classic theoretical and behavioral work demonstrates that responses that require top-down control become disrupted when processing-time or cognitive resources are limited (Kahneman, 2003). This generates a second clear prediction about intergroup trust. In particular, if outgroup, but not ingroup, trust requires control, then limiting participants' time to deliberate about their decisions should reduce their outgroup trust, but leave ingroup trust unaffected.

Here, we employed an experimental economics approach in conjunction with neuroimaging to test these predictions about single and dual-process models of intergroup trust. Participants completed modified trust games in which they made single-shot decisions about whether to trust ingroup and outgroup members with real money. Participants were assigned to be *investors* and decided how much money to send to in- and outgroup *trustees*. The amount of money entrusted is quadrupled, and the trustee can either cooperate and return half the money (thereby doubling the initial investment) or defect and keep the entire sum. The best payoff for both players is to behave unselfishly and cooperate. However, the best payoff for the trustee is to behave selfishly and defect. Therefore, sharing money is in the investor's best interest, but only if the trustee unselfishly cooperates. As such, investment in this task serves as a behavioral proxy of trust.

In the first experiment, we used functional magnetic resonance imaging (fMRI) to investigate whether (i) intergroup trust processes differ only in the extent to which they draw on value-related neural activity (consistent with a difference of *degree*) or (ii) outgroup trust uniquely engages regions involved in top-down control (consistent with a difference of *kind*). In the second experiment, we investigated whether limiting participants' response time differentially affects competitive outgroup, as compared to ingroup trust.

Methods

Study 1

Participants. Twenty-six participants were recruited in compliance with the human subjects regulations of Stanford University and compensated with $15/h or course credit (15 females, mean age = 19.1 years, s.d. = 1.1). This sample size was determined *a priori* to provide power of 0.80 to detect differences in intergroup trust (ingroup *vs* outgroup) based on an estimated effect size of $d = 0.88$, derived from recent studies that also employ a within-subject, repeated-measures trust game to investigate cooperation (Delgado *et al.*, 2005). Five participants were excluded from analyses (three participants were excluded due to excessive head motion >2.0 mm from one volume of acquisition to the next along any dimension or across the duration of the scan, one participant was excluded due to the scanner malfunction, one participant was excluded for failing to respond to over 50% of trials). The remaining 21 participants (12 females, mean age = 18.8 years, s.d. = 0.75) were all right-handed, native English speakers, free from medications and psychological and neurological conditions, and had normal or corrected-to-normal vision. Because the present study examines intergroup trust, all participants were prescreened to ensure that they were members of the ingroup condition (i.e. Stanford University students). Finally, participants completed the Collective Self-Esteem Scale (CSES: Luhtanen and Crocker, 1992) to ensure that they would experience a motivation to favor ingroup members. Participants all reported positive associations with their social identity as Stanford students ($M = 5.5$, s.d. = 0.4).

Procedure. Participants completed a modified version of the trust game in which they made iterated choices about how much money ($0 to $4, in increments of $1) to invest with trustees. Critically, trustees were ostensibly either ingroup members (i.e. Stanford University students) or outgroup members (i.e. UC-

Berkeley students). In addition, participants completed a series of trust decisions with trustees devoid of group membership information to serve as control trials. The control trials established a behavioral baseline of trust and made the group membership manipulation salient. Participants were instructed that the amount they sent to the ostensible trustees would be quadrupled and then allocated based on the trustee's decision. Specifically, the trustee could either equally share the money and double the participant's investment, or keep all of the money for themselves. Participants were told that the outcomes of three randomly selected interactions would be chosen to count, and that they could earn significant amounts of money or lose the investments in these interactions. In actuality, all trustees were fake and participants were all paid a fixed $5 bonus for their participation.

During the fMRI session, participants made single-shot trust decisions to invest with 150 trustees (50 ingroup, 50 outgroup and 50 control). For each trial, participants saw a picture of the trustee and group membership information (or no information) and had 3 s to decide how much money ($0 to $4 in increments of $1) to entrust. Underneath the photograph of the trustee were dollar amounts ($0 to $4 in increments of $1) ascending from left to right. Trial order was randomized and each trial was followed by a random jittered inter-trial interval (1, 3 or 5 s). Visual stimuli were presented using E-prime and projected onto a large-screen flat-panel display monitor that participants viewed in a mirror mounted on the scanner. Following the completion of the task, participants were shown the outcomes of three randomly selected interactions and paid their bonus money based on these investments.

Stimuli. The interaction partners were represented by facial photographs along with their group membership: ingroup was depicted by a Stanford University logo, outgroup was depicted by a Cal logo, or control (no group membership displayed). Photographs were drawn from the first author's photo database and consisted of color pictures of forward-looking male faces with neutral expressions. Photographs were randomly-distributed to belong to the ingroup, outgroup or control conditions and were balanced across race.

MRI data acquisition. All images were collected on a 3.0T GE Discovery MR750 scanner at the Center for Cognitive and Neurobiological Imaging at Stanford University. Functional images were acquired with a T2*-weighted gradient echo pulse sequence (TR = 2 s, TE = 24 ms, flip angle = 77°) with each volume consisting of 46 axial slices (2.9-mm-thick slices, in-plane resolution 2.9 mm isotropic, no gap, interleaved acquisition). Functional images were collected in one run (consisting of 150 trials). High-resolution structural scans were acquired with a T1-weighted pulse sequence (TR = 7.2 ms, TW = 2.8 ms, flip angle = 12°) after functional scans, to facilitate their localization and co-registration.

MRI data analysis. All statistical analyses were conducted using SPM8 (Wellcome Department of Cognitive Neurology). Functional images were reconstructed from *k*-space using a linear time interpolation algorithm to double the effective sampling rate. Image volumes were corrected for slice-timing skew using temporal sinc interpolation and for movement using rigid-body transformation parameters. Functional data and structural data were co-registered and normalized into a standard anatomical space (2-mm isotropic voxels) based on the echo planar imaging and T1 templates (Montreal Neurological Institute), respectively. Images were smoothed with a 5-mm full-width at half-maximum Gaussian kernel. To remove drifts within sessions, a high-pass filter with a cut-off period of 128 s

was applied. Visual inspection of motion correction estimates confirmed that no subject's head motion exceeded 2.0 mm in any dimension from one volume acquisition to the next.

Functional images were analyzed to identify neural activity that was parametrically modulated by trust amounts for ingroup and outgroup targets, and neural activity associated with individual differences in intergroup trust bias. Two analytic approaches were used.

The first analytic approach sought to identify neural activity associated with parametric increases in trust decisions based on group membership. The parametric analytic approach provided estimates of trust at a trial-by-trial, within-subject level, by capitalizing on the repeated-measures nature of our experimental design. The GLM consisted of three regressors of interest: the trust decision periods for ingroup, outgroup, and control conditions. These regressors were modeled as stick functions at the onset of each trial and convolved with a canonical (double-gamma) hemodynamic response function. In addition, each onset was weighted by the dollar amount sent to the trustee during each trial. To do so, parametric modulators of the dollar amount sent to the trustee on each trial were included for each of the three regressors of interest (ingroup, outgroup and control). Finally, six regressors of non-interest modeled participant head movement during the scan.

First, contrasts identified neural activity related to parametric increases in trust regardless of group membership (ingroup and outgroup trust *vs* baseline). Next, we sought to examine whether intergroup trust is supported by dissociable neural systems. To do so, we sought to identify brain regions that (i) parametrically tracked increases in trust decisions for ingroup targets and (ii) did so to a greater extent for ingroup *vs* outgroup targets. Similarly, we sought to identify brain regions that (i) parametrically tracked increases in outgroup trust and (ii) did so to a greater extent for outgroup versus ingroup targets. Therefore, the analysis strategy proceeded in two steps. First, we examined neural activity that parametrically tracked trust amounts during the ingroup condition significantly above baseline (ingroup parametric condition > baseline). Second, we masked the results by brain regions that tracked trust amount for the ingroup condition significantly more strongly than the outgroup condition (ingroup > outgroup parametric contrast), using the minimum statistic approach (Nichols *et al.*, 2005). We similarly isolated neural systems that parametrically tracked increases in outgroup trust significantly above baseline (outgroup parametric condition > baseline), and masked the results by brain regions that significantly tracked trust amount for the outgroup over and above the ingroup condition (outgroup > ingroup parametric contrast).

The aim of the second analytic approach was to identify whether brain regions that parametrically tracked trust for in- and outgroup members at a within-subject level were also associated with intergroup breakdowns in trust at a between-subject level, that is, unevenly favoring ingroup over outgroup trust. To test this question, we first computed a trust bias score (ingroup trust amount—outgroup trust amount) for each participant. Next, we used multiple regression to test whether individual differences in trust bias significantly correlated with neural activation in the ingroup > outgroup main effect contrast. Given that the small samples that are common in fMRI designs provide limited power to test individual differences, we limited our search to brain regions that were parametrically associated with trust within-subjects. To do so, we masked the individual difference analysis by brain regions that were parametrically modulated by ingroup and outgroup trust.
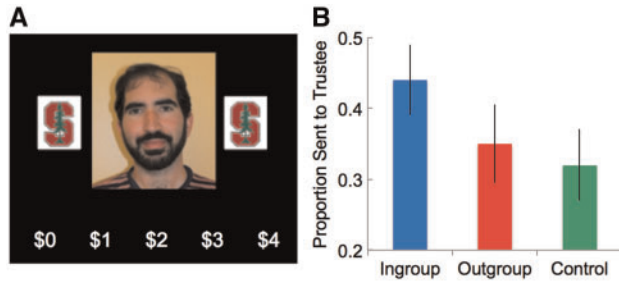
Fig. 1. Study 1 task and behavioral responses. (A) Example ingroup trust trial. On each trial, participants had 3 s to decide how much money to invest with trustees. Trustees were either ingroup members (Stanford students) or outgroup members (Cal students), which were represented by the school logo appearing next to the trustee photographs. (B) Behavioral responses varied significantly by group: participants trusted ingroup members with a significantly greater proportion of money than outgroup members or control. Error bars represent SEM.

Table 1. Brain regions that parametrically track increases in trust overall, irrespective of group

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
Cerebellum		18, −55, −18	6.27	1301
Postcentral gyrus	4	−63, −14, 28	5.99	796
dMPFC	10	6, 48, 16	5.43	956
dACC	9/24	18, 38, 24	5.27	
vMPFC	11	0, 50, −8	4.57	
Right hippocampus		32, −12, −20	5.07	135
Left hippocampus		−32, −24, −18	4.56	109
PCC	31	−8, −44, 32	4.52	418
Inferior temporal	37	−44, −52, −16	4.50	132
LPFC	9/46	50, 20, 28	4.50	148
Right striatum		10, 20, 0	3.70	66
Left striatum		−12, 16, 6	3.61	65

dMPFC, dorsomedial prefrontal cortex; dACC, dorsal anterior cingulate cortex; vMPFC, ventromedial prefrontal cortex; PCC, posterior cingulate cortex; LPFC, lateral prefrontal cortex.

Main effect maps were thresholded at $P < 0.005$, with a spatial extent threshold of $k = 23$, corresponding to a threshold of $P < 0.05$ corrected for multiple comparison (derived with 15 000 Monte Carlo simulations using the current release of the AFNI program 3dClustSim). To compute the thresholds for maps of the two-way conjunctions, we used Fisher's method (Fisher, 1925), which combines probabilities of multiple hypothesis tests using the following formula:

$$x^2 = -2 \sum_{i=1}^{k} \log e(P_i),$$

where P_i is the P-value for the ith test being combined, k is the number of tests being combined and the resulting statistic has a χ^2 distribution with $2k$ degrees of freedom. Thus, thresholding each test of a two-way conjunction at P-values of 0.024 corresponded to a combined voxel-wise threshold P-value of 0.005, with an extent threshold of $k = 23$, corresponding to a corrected threshold of $P < 0.05$ (derived with Monte Carlo simulations using AFNI program 3dClustSim).

Psychophysiological interaction (PPI) analysis. Finally, we conducted PPI analyses to identify brain regions (e.g. dACC, LPFC and other control-related areas) exhibiting an increase in correlation with the striatum during trust decisions. In particular, we were interested whether the strength of such correlations would be greater for outgroup *vs* ingroup trust decisions.

First, we defined the seed volume of interest (VOI) as the striatum region that tracks parametric trial-by-trial increases in trust regardless of group membership. For each participant, we extracted time-series of activation from the striatum mask VOI described above. Variance associated with the six motion regressors was removed from the extracted time-series. Next, we computed a first-level whole-brain GLM for each participant that included the following regressors: (i) an interaction between neural activity in striatum and the onset of trust decisions for ingroup, outgroup and control targets convolved with the canonical HRF, (ii) a regressor specifying the onset of all trials convolved with the canonical HRF and (iii) the original BOLD eigenvariate from the striatum seed VOI (i.e. average time-series from the striatum mask). Finally, the GLM included six nuisance regressors that modeled head motion during the scan across the whole brain volume, as well as additional nuisance regressors for single time-points where motion spikes occurred. We used two different metrics to assess motion spikes: (i) frame-to-frame head motion that exceeded 1.0 mm in rotation/

translation parameters and (2) signal intensity differences for each volume greater than 2.5 s.d. from the global mean signal intensity. All motion spikes were removed via regression (Satterthwaite et al., 2013; Power et al., 2015). This procedure removed an average of 0.17% of volumes (range 0–1.2%). These additional procedures were implemented in order to reduce false positives in connectivity analyses, which are especially sensitive to head motion.

Single subject contrasts were then calculated (Trust irrespective of group > baseline; Outgroup trust > Ingroup trust; Ingroup trust > Outgroup trust). Second-level group contrasts were calculated based on the single subject contrasts. Results report areas exhibiting positive correlations with striatum, as captured by the first regressor in the GLM. Contrasts were thresholded using the 3dClustSim thresholding procedure described above.

Study 2

Participants. A total of 616 US participants (248 female, mean age = 30.69 years, s.d. = 10.08) were recruited from Amazon's Mechanical Turk. Unlike Study 1, which employed a repeated-measures design, Study 2 employed a between-subject and single-shot design. The sample size was pre-determined to provide power of 0.80 to detect differences in intergroup trust based on an estimated effect size of $d = 0.35$, derived from a recent quantitative meta-analysis on intergroup cooperation (Balliet et al., 2014). Data collected through Mechanical Turk are as reliable as those collected through traditional methods (Buhrmester et al., 2011). Participants were excluded ($n = 79$) from analyses if they failed attention checks (catch scenarios: Oppenheimer et al., 2009).

Procedure. Participants were introduced to the trust game, in which they were endowed with $0.20 and would have the option to invest their money with another MTurk worker from around the world. We instructed participants that they stood to gain substantial bonus payments based on their investments, but also risked losing their endowment altogether. Participants first entered their name and nationality (we limited participation to American MTurk workers from the USA).

Table 2. Brain regions that parametrically track ingroup trust

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
Cerebellum		18, −46, −22	6.81	604
Postcentral gyrus	4	−54, −16, 34	5.96	260
Left hippocampus		−26, −12, −20	5.16	117
Left striatum		−24, 22, 10	5.08	120
Inferior parietal	49	−49, −46, 64	5.04	375
vMPFC	10/11	0, 38, 8	4.06	280
Right striatum		10, 20, 2	3.86	145
Right hippocampus		24, −8, −20	4.01	96
Inferior temporal	20	−52, 44, −14	3.58	66
Precuneus	7	−8, −58, 42	3.26	66

vMPFC, ventromedial prefrontal cortex.

Table 3. Brain regions that parametrically track outgroup trust

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
dACC	9/24	18, 38, 24	6.43	511
dMPFC	10	6, 50, 16		
vMPFC	11	8, 40, −10		
Cerebellum		24, −64, −26	4.24	449
LPFC	46	40, 30, 22	4.00	52
PCC	23	8, −54, 24	3.48	42
Inferior parietal	2	−44, −28, 48	3.42	44

dACC, dorsal anterior cingulate cortex; dMPFC, dorsomedial prefrontal cortex; vMPFC, ventromedial prefrontal cortex; LPFC, lateral prefrontal cortex; PCC, posterior cingulate cortex.

Next, participants were introduced to the Trustee for the trust game. Participants were randomly paired with either 'David from the United States' and were shown an American flag (*Ingroup* condition), or with 'Diego from Spain' and were shown a Spanish flag (*Outgroup* condition). Participants then completed a single round from the trust task. Participants used a slider in increments of $0.01 and decided how much of their $0.20 endowment to invest with the Trustee. Critically, participants were randomly assigned to one of the two versions of the Trust Game used in prior research (Rand *et al.*, 2014): participants were either forced to make their decisions in less than 10 s (intuition condition), or were told to deliberate for at least 10 s before making their decision. Specifically, in the *Intuition* condition, participants were told to make their decisions 'as quickly as possible' and that they had 10 s to make their decision (a timer counted down from 10 to highlight the speeded nature of the task). In the *Deliberation* condition, participants were told to 'stop and think before making their decision' and waited for 10 s before they could enter a response (a timer counted up to 10 to highlight that they must wait until the timer has elapsed to enter their response).

Results

Study 1

Behavioral results. Overall, participants trusted ingroup members more than outgroup members [ingroup: $M = \$1.72$, s.d. $= 0.46$; outgroup: $M = \$1.45$, s.d. $= 0.58$; $t(20) = 2.80$, $P = 0.010$, $d = 0.51$]. Participants also trusted ingroup members more than controls [ingroup: $M = \$1.72$, s.d. $= 0.44$; control: $M = \$1.28$, s.d. $= 0.58$;

$t(20) = 3.98$, $P < 0.001$, $d = 0.84$], but did not distrust outgroup members less than controls [outgroup: $M = \$1.45$, s.d. $= 0.58$; control: $M = \$1.28$, s.d. $= 0.58$; $t(20) = 1.85$, $P = 0.08$, $d = 0.29$] (Figure 1B). Finally, there were no significant differences in reaction time between conditions [ingroup RT: $M = 1.71$s, s.d. $= 0.25$s; outgroup RT: $M = 1.69$s, s.d. $= 0.29$s; control RT: $M = 1.66$, s.d. $= 0.26$s; $t(20) = 0.78$, $P > 0.25$, $d = 0.07$].

Neuroimaging results. Next, we sought to address whether intergroup trust is supported by dissociable underlying mechanisms. Because we were interested in the correspondence between brain activity and trust behavior, we employed trial-by-trial parametric analyses, within subjects, to separately isolate brain regions in which engagement significantly tracked (i) the amount of money that investors trusted ingroup trustees above baseline and (ii) outgroup trustees above baseline (for trust-related brain activity irrespective of group; Table 1). We found that activity in bilateral striatum and vMPFC (among other regions) parametrically tracked increases in ingroup trust (Table 2). Conversely, activity in dACC and LPFC (among other regions) parametrically tracked increases in outgroup trust (Table 3).

Next, we isolated brain regions involved in preferentially trusting ingroup, when compared with outgroup, members (and *vice versa*). We find that activation in striatum tracks ingroup trust to a significantly greater degree than outgroup trust (ingroup parametric condition > outgroup parametric condition; Figure 2A). In contrast, we find that activation in dACC tracks outgroup trust to a significantly greater degree than ingroup trust (outgroup parametric condition > ingroup parametric condition; Figure 2B).

The data from the within-subject parametric contrasts suggest that ingroup and outgroup trust decisions are supported by dissociable underlying mechanisms. Whereas striatum activation significantly tracks the degree of ingroup (but not outgroup) trust, dACC activation significantly tracks outgroup (but not ingroup) trust. If this dissociation relates functionally to intergroup failures in trust, then it should also track individuals' intergroup trust *bias*, or their tendency to favor trusting ingroup over outgroup members. To address this question, we operationalized trust bias by computing difference scores between each participant's ingroup trust rating and outgroup trust rating. Next, we conducted a regression analysis that examined whether individual differences in trust bias correlate with greater neural activation during ingroup over outgroup trust decisions at a between-subject level (ingroup trust > outgroup trust contrast).
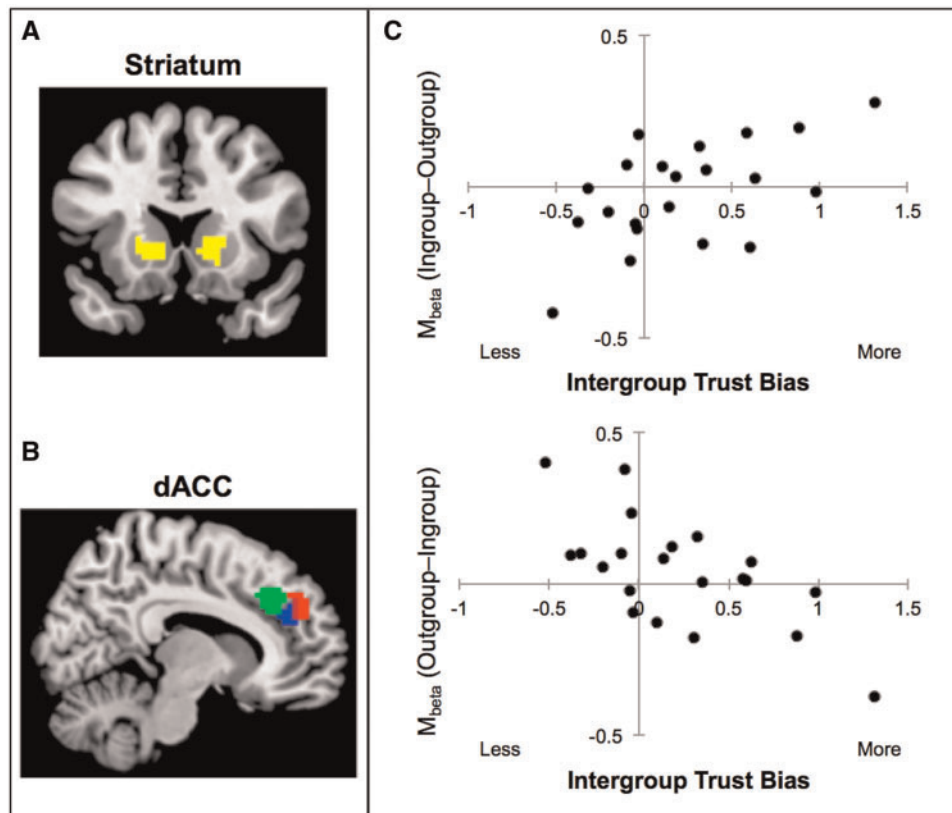
Fig. 2. Neural activation from within-subject parametric analyses, between-subject regression, and psychophysiological interaction (PPI) analyses. (A) Parametric analyses revealed that increases in ingroup (but not outgroup) trust were significantly modulated by increases in striatum activation (left: x, y, $z = -18$, 10, 8; $t = 3.19$, $k = 288$; right: x, y, $z = 10$, 20, 2; $t = 3.17$, $k = 141$). Bilateral striatum depicted in yellow was used as the seed VOI in subsequent PPI analyses. (B) Conversely, increases in outgroup (but not ingroup) trust were significantly modulated by increases in dACC activation (x, y, $z = 14$, 42, 26; $t = 2.95$, $k = 67$). Red cluster was modulated by parametric trial-by-trial outgroup trust; green cluster was significantly correlated with the reduction of trust bias across participants; blue cluster was functionally coupled with striatum during outgroup trust. (C) Whole-brain regression analyses revealed that activation in striatum—in clusters that parametrically tracked ingroup trust within participants—also significantly correlated with trust bias across participants (right striatum: x, y, $z = 16$, 14, -2; left striatum: x, y, $z = 16$, 14, -2; $t = 3.37$, $k = 250$). Second, dACC activation—in clusters that parametrically tracked outgroup trust within participants—significantly correlated with trust bias correction across participants (x, y, $z = 12$, 36, 36; $t = 3.92$, $k = 40$).

This analysis revealed that striatal activity—in clusters that also parametrically tracked ingroup trust within participants—significantly correlated with trust bias across participants (Figure 2C and Table 3). That is, participants who engaged striatum relatively strongly during ingroup trust decisions were also *more* behaviorally biased, trusting ingroup over outgroup members. Second, dACC and LPFC activity—in clusters that also tracked outgroup trust within participants—significantly correlated with trust bias *correction* across participants (Figure 2B, C and Table 4). That is, participants that engaged greater dACC and LPFC activation during outgroup trust decisions exhibited *less* intergroup trust bias.

Finally, we conducted PPI analyses to identify brain regions (e.g. dACC, LPFC and other control-related areas) exhibiting an increase in correlation with the striatum during trust decisions. In particular, we were interested whether the strength of such correlations would be greater for outgroup versus ingroup trust decisions. If the exertion of control is needed to increase outgroup trust, then functional connectivity between striatum and control-related regions (e.g. dACC, LPFC) might be stronger for outgroup *vs* ingroup targets. To address this question, we conducted a PPI analysis that identifies brain regions that are functionally coupled with the striatum VOI that tracks trial-by-trial trust amounts (see the Methods section for more detail). Across group membership conditions, we find that striatum activation

is significantly correlated with activation in control-related brain regions, including the dACC, LPFC and TPJ (Table 5). Crucially, the strength of connectivity between activation in striatum and clusters in dACC and LPFC that also tracked increases in outgroup trust was greater for outgroup than ingroup targets (outgroup PPI > ingroup PPI; Figure 2C and Table 6). These findings raise the possibility that control-related activation in dACC and LPFC may modulate subjective value-related activity in striatum to increase outgroup trust. This possibility is consistent with extant work showing increased connectivity between striatum and vMPFC with control-related activation in dACC and LPFC in order to pursue long-term goals (e.g. Hare *et al.*, 2009).

Study 2

The results of Study 1 suggest that ingroup and outgroup trust are supported by psychological mechanisms that differ in *kind*, not *degree*. Specifically, whereas neural circuitry associated with subjective value calculation (e.g. striatum) tracked ingroup (but not outgroup) trust, cognitive control systems (e.g. dACC) tracked outgroup (but not ingroup) trust. Taken together, these neural findings led us to a novel behavioral prediction that further tests a dual system model of intergroup trust. If outgroup, but not ingroup trust requires the exertion of cognitive control,

Table 4. Brain regions that significantly correlate with individual differences in intergroup trust bias

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
Striatum		−18, 16, −2	3.37	250
		16, 14, −2		
dACC	9/32	12, 36, 36	3.92	40
LPFC	46	46, 26, 32	3.99	99

dACC, dorsal anterior cingulate cortex; LPFC, lateral prefrontal cortex.

Table 5. Brain regions functionally connected with bilateral striatum during intergroup trust decisions, irrespective of group

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
dACC	9/32	3, 26, 31	7.22	30
LPFC	46	33, 44, 28	8.46	166
		−30, 44, 31	9.88	180
TPJ		−60, −37, 37	7.10	35

dACC, dorsal anterior cingulate cortex; LPFC, lateral prefrontal cortex; TPJ, temporoparietal junction.

Fig 3. Behavioral results (Study 2) revealed a significant interaction. In the intuitive condition, participants trusted ingroup members with a significantly greater proportion of money than outgroup members. In the deliberative condition, there was no significant difference between ingroup and outgroup trust. Error bars represent SEM.

Table 6. Brain regions more strongly functionally coupled with bilateral striatum during outgroup (versus ingroup) trust decisions

Region of activation	BA	Coordinates (x, y, z, in mm)	T-stat	Cluster size
dACC	9/32	−9, 38, 31	3.09	129
LPFC	46	42, 18, 31	2.46	29

dACC, dorsal anterior cingulate cortex; LPFC, lateral prefrontal cortex.

then limiting the processing time needed to exert control should reduce outgroup trust, but leave ingroup trust unaffected. To test this prediction, a separate sample of participants completed a single-shot trust game with an outgroup member or ingroup member in one of the two conditions: an intuitive (speeded) condition, or a deliberative (non-speeded) condition (Study 2).

We submitted participants' trust decisions to a 2 (group: ingroup *vs* outgroup) × 2 (speed: intuition *vs* deliberation) ANOVA. This analysis revealed a significant group × speed interaction [$F(3,533)=4.57$, $P = 0.033$, $\eta\ P = 0.01$] (Figure 3). Participants trusted ingroup members equally across the intuitive ($M = ¢14.06$, s.d. $= 6.78$) and deliberative ($M = ¢14.53$, s.d. $= 6.09$) conditions [$t(258)= 0.58$, $P > 0.250$, $d = 0.07$]. However, participants trusted outgroup members significantly less in the intuitive ($M = ¢11.93$, s.d. $= 6.87$) *vs* deliberative ($M = ¢14.78$, s.d. $= 6.08$) conditions [$t(275)= −3.65$, $P < 0.001$, $d = 0.42$]. These findings suggest that people intuitively and automatically trust ingroup members, but require cognitive resources to trust outgroup members even in non-competitive intergroup settings.

Study 2 demonstrated relatively small effects of deliberation on intergroup cooperation ($d = 0.42$). Nonetheless, there are at least two reasons to consider these effects noteworthy. First, effects in these ranges are thought to inform us about the cognitive basis of cooperation in general (Rand *et al.*, 2012, 2014), and intergroup cooperation in particular (Balliet *et al.*, 2014). Second, small effect sizes associated with automatic or intuitive processes, such as implicit bias measures (Greenwald *et al.*, 2015), predict consequential forms of real world discrimination.

Discussion

Together, these experiments support a dual-process model of intergroup trust, through multiple converging methods. We capitalized on behavioral and neural markers of reward and control to examine whet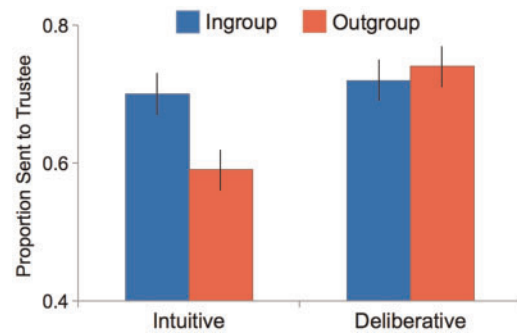her these processes make dissociable contributions to intergroup trust. Across two studies, we found that outgroup trust tracked dACC activation (Study 1) and increased when participants had time to deliberate about their decisions (Study 2). Conversely, ingroup trust tracked striatum activation (Study 1) and was unaffected by deliberation time (Study 2).

The pattern of behavioral results further suggests that bounded rationality in intergroup trust does not produce irrationally high ingroup trust (as ingroup trust did not decrease with deliberation), but rather irrationally low outgroup trust. People trusted outgroup members just as much as ingroup members when they had time to deliberate, but intuitively distrust outgroup members when processing time was limited (Study 2). Moreover, these behaviors with small-stakes economic games should translate to more consequential settings, because game theoretic behavior is relatively scale invariant (Johansson-Stenman *et al.*, 2005; Kocher et al., 2008; Amir *et al.*, 2012), even when stakes are raised to three times participants' monthly expenditure (Cameron, 1999; Henrich *et al.*, 2001). Taken together, this work demonstrates the translational value of 'behavioral phenotypes' from game theory in informing us about larger decisions, such as whether or not to cooperate across group lines.

Finally, the dissociable neural mechanisms underlying intergroup trust also relate to individual differences in intergroup *breakdowns in trust*. Specifically, participants that engaged greater striatum activation during ingroup trust decisions were *more* biased toward trusting ingroup over outgroup members. Participants that engaged greater dACC and LPFC activation during outgroup trust decisions were *less* biased toward trusting ingroup over outgroup members. Moreover, participants that engaged greater dACC and LPFC activation during outgroup trust decisions also engaged greater striatum activation during

outgroup (*vs* ingroup) interactions. These findings are consistent with a dual-system model of decision-making, and suggest that dACC and LPFC activation may increase outgroup trust by modulating value-related activation in striatum. However, as highlighted in Study 1 Methods above, whereas repeated-measure fMRI designs provide stable estimates of within-subject variability, the small samples employed in fMRI research provide low power to robustly detect between-subjects effects. Future work should employ larger samples to further examine the relationship between activation in these systems and individual differences in intergroup trust bias. Taken together, the findings across two studies suggest that people find it intuitive and subjectively valuable to cooperate with ingroup members, but uniquely exert control to override intuitive distrust of outgroup members.

These findings extend key insights across a number of research domains. First, these data add nuance to evidence about the psychological structure of prosociality. A wellspring of evidence collected over the last decade suggests that prosocial acts such as altruistic giving and cooperation are reward-driven and intuitive. First, brain regions involved in representing subjective value—including the striatum and vMPFC—respond when individuals act in ways that are cooperative (Rilling et al., 2002), fair (Zaki and Mitchell, 2011) or charitable (Hare et al., 2010). Second, individuals pressured to make decisions quickly become more, not less cooperative (Rand et al., 2014), suggesting that prosociality is intuitive (Zaki and Mitchell, 2013). The current study conditionalizes these insights, by demonstrating that intuitive, reward-based features of trust are likely bounded by group membership. These data are consistent with the Social Heuristics Hypothesis (Rand et al., 2014), which suggests that expectations about social behaviors become 'hard coded' into intuitive responses, such that individuals learn to intuitively cooperate or defect in social settings based on past experience (Rand and Nowak, 2013). In this context, intergroup encounters might generate a heuristic, intuitive *distrust*, which individuals must override through deliberative efforts. These efforts may lead people to realize that trust—even of outgroup members—may be payoff maximizing, 'nudging' people toward outgroup trust.

Our findings also contribute to a scientific understanding of intergroup cognition. The majority of neuroscientific examinations of intergroup cognition focus on racial group boundaries. However, not all intergroup contexts are alike (Cikara and Van Bavel, 2014). Race differs in crucial ways from other intergroup contexts: race can induce motivations not common to other intergroup settings. For instance, many people do not wish to appear racially biased, and exert deliberative control to suppress their implicit stereotypes (Devine, 1989). As such, individuals are more likely to express racial biases through implicit channels, which are largely outside deliberative control (Dovidio et al., 1997; Payne, 2001). Other intergroup contexts differ fundamentally from race along this dimension. For instance, in competitive contexts ranging from sports rivalries to war, it is often socially acceptable to express out-group antipathy (Cikara and Van Bavel, 2014). People in these non-race contexts might not be motivated to suppress outgroup bias, and often eagerly air outgroup antipathy.

Here, we show that non-racial intergroup contexts are similar to interracial contexts in some—but not all—of their component processes. On one hand, neuroscientific work on race shows that overriding implicit racial biases engages regions involved in top-down control, such as dACC and LPFC (Cunningham et al., 2004; Amodio et al., 2008). Here, we demonstrate that these features of control extend to other intergroup

contexts, namely, overriding intuitive outgroup *distrust*. On the other hand, neuroscientific work on race shows that decisions to trust outgroup members engage regions associated with subjective value, such as the striatum (Stanley et al., 2012). This could signal the value or goal relevance of successfully suppressing unwanted racial biases. The current study demonstrates that in competitive intergroup contexts, striatum activation tracks ingroup, not outgroup trust. In these non-racial contexts, striatum activation may signal the value or goal relevance of affiliation with ingroup members, as compared to the value of avoiding the appearance of bias. Together, these findings deepen our understanding of the shared and distinct features that characterize trust in racial versus non-racial intergroup contexts.

Finally, our findings contribute to a growing body of research on intergroup conflict resolution. Most interventions designed to reduce intergroup conflict trade in the idea of increasing positive and decreasing negative reactions toward outgroup members. These types of interventions include increasing social contact with outgroup members and promoting a sense of common identity (Pettigrew and Tropp, 2006), taking the perspective of outgroup members (Galinsky and Moskowitz, 2000) and regulating negative emotions about outgroup members (Halperin et al., 2013). Although these interventions often foster positive intergroup relations, they also backfire (Dixon et al., 2010). For example, encouraging positive intergroup affect can delegitimize negative affect experienced by low power groups in response to inequality, which may promote rather than reduce injustice.

The current findings suggest that—at least in the domain of trust—encouraging individuals to experience positive affect toward outgroup members is not the only useful strategy to address intergroup trust failures. This is because fundamentally different mechanisms—including not only intuitive, affective experience, but also deliberative control—can support intergroup trust. As such, interventions that encourage 'cool' or 'System 2' processing during intergroup interactions (Metcalfe and Mischel, 1999; Kahneman, 2003) might increase outgroup trust even without changing individuals' affective experiences. In the long-term, cognitive strategies that increase intergroup cooperation may reduce conflict and promote social change.

References

Amir, O., Rand, D.G., Gal, Y.K. (2012). Economic games on the internet: the effect of $1 stakes. *PLOS: One*. doi: 0.1371/journal.pone.0031461

Amodio, D.M., Devine, P.G., Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: the role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*, 94, 60–74.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, **12**(5), 193–200.

Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, **76**, 412–27.

Balliet, D., Wu, J., De Dreu, C.K.W. (2014). Ingroup favoritism in cooperation: a meta- analysis. *Psychological Bulletin*, **140**(6), 1556.

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.

Bhanji, J.P., Beer, J.S. (2012). Taking a different perspective: mindset influences neural regions that represent value and choice. *Social Cognitive and Affective Neuroscience*, **7**(7), 782–93.

Brewer, M.B. (1999). The psychology of prejudice: ingroup love and outgroup hate? *Journal of Social Issues*, **55**, 429–44.

Buhrmester, M., Kwang, T., Gosling, S.D. (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, **6**, 3–5.

Cameron, L.A. (1999). Raising the stakes in the ultimatum game: experimental evidence from Indonesia. *Economic Inquiry*, **37**(1), 47–59.

Choi, J.K., Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, **318**, 636–40.

Cikara, M., Van Bavel, J.J. (2014). The neuroscience of intergroup relations: an integrative review. *Perspectives on Psychological Science*, **9**(3), 245–74.

Cunningham, W.A., Johnson, M.K., Raye, C.L., Gatenby, J.C., Gore, J.C., Banaji, M.R. (2004). Separable neural components in the processing of Black and White Faces. *Psychological Science*, **15**, 806–13.

Delgado, M.R., Frank, R.H., Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, **8**, 1611–8.

Devine, P.G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, **56**, 5–18.

Dixon, J., Tropp, L.R., Durrheim, K., Tredoux, C. (2010). "Let them eat harmony": prejudice-reduction strategies and attitudes of historically disadvantaged groups. *Current Directions in Psychological Science*, **19**, 76–80.

Dovidio, J.F., Kawakami, K., Johnson, C., Johnson, B., Howard, A. (1997). On the nature of prejudice: automatic and controlled processes. *Journal of Experimental Social Psychology*, **33**, 510–40.

Fareri, D.S., Chang, L.J., Delgado, M.R. (2015). Computational substrates of social value in interpersonal collaboration. *The Journal of Neuroscience*, **35**(21), 8170–80.

Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Friedman, T.L. (2006). *The world is flat: The globalized world in the twenty-first century*. London: Penguin.

Galinsky, A.D., Moskowitz, G.B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, **78**, 708–24.

Greenwald, A.G., Banaji, M.R., Nosek, B.A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology*, **108**, 553–61.

Halperin, E., Porat, R., Tamir, M., Gross, J.J. (2013). Can emotion regulation change political attitudes in intractable conflicts? From the laboratory to the field. *Psychological Science*, **24**, 106–11.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience*, **30**, 583–90.

Hare, T.A., Camerer, C.F., Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, **324**(5927), 646–8.

Haslam, N., Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, **65**, 399–423.

Henrich, J., Boyd, R., Bowles, S., *et al.* (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 73–8.

Hughes, B.L., Zaki, J. (2015). The neuroscience of motivated cognition. *Trends in Cognitive Sciences*, **19**, 62–4.

Johansson-Stenman, O., Mahmud, M., Martinsson, P. (2005). Does stake size matter in trust games? *Economics Letters*, **88**(3), 365–9.

Kable, J.W., Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, **10**(12), 1625–33.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, **58**, 697–720.

Kocher, M.G., Martinsson, P., Visser, M. (2008). Does stake size matter for cooperation and punishment? *Economics Letters*, **99**(3), 508–11.

Kurzban, R., Leary, M.R. (2001). Evolutionary origins of stigmatization: the functions of social exclusion. *Psychological Bulletin*, **127**, 187–208.

Levine, M., Prosser, A., Evans, D., Reicher, S. (2005). Identity and emergency intervention: how social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, **31**, 443–53.

Luhtanen, R., Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, **18**(3), 302–18.

Metcalfe, J., Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, **106**(1), 3–19.

McClure, S.M., Laibson, D.I., Loewenstein, G., Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, **306**(5695), 503–7.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, **25**, 653–60.

Oppenheimer, D.M., Meyvis, T., Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, **45**(4), 867–72.

Payne, K.B. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, **81**, 181–92.

Pettigrew, T.F., Tropp, L.R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, **90**, 751–83.

Power, J.D., Schlaggar, B.L., Petersen, S.E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage*, **105**, 536–51.

Rand, D.G., Greene, J.D., Nowak, M.A. (2012). Spontaneous giving and calculated greed. *Nature*, **489**(7416), 427–30.

Rand, D.G., Nowak, M.A. (2013). Human cooperation. *Trends in Cognitive Sciences*, **17**, 413–25.

Rand, D.G., Peysakhovich, A., Kraft-Todd, G.T., *et al.* (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, **5**, 1–12.

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

Ross, L., Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology*, **27**, 255–304.

Ruff, C.C., Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, **15**, 549–62.

Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., *et al.* (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, **64**, 240–56.

Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, **12**(3), 154–67.

Shenhav, A., Straccia, M.A., Cohen, J.D., Botvinick, M.M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, **17**, 1249–54.

Shenhav, A., Botvinick, M.M., Cohen, J.D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, **79**(2), 217–40.

Stanley, D.A., Sokol-Hessner, P., Banaji, M.R., Phelps, E.A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, **108**, 7710–5.

Stanley, D.A., Sokol-Hessner, P., Fareri, D.S., *et al.* (2012). Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 744–53.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, **33**, 1–39.

Zak, P.J., Knack, S. (2001). Trust and growth. *The Economic Journal*, **111**, 295–321.

Zaki, J., Mitchell, J.P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences*, **108**, 19761–6.

Zaki, J., Mitchell, J.P. (2013). Intuitive Prosociality. *Current Directions in Psychological Science*, **22**, 466–70.